A Comparison of Orthologous Proteins in Homotypic Vacuole Fusion

_____

A Thesis

Presented to

The Division of Mathematics and Natural Sciences

Reed College

_____

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

_____

Aden J. O'Brien

May 2025

Approved for the Division
(Biology)

_____

Anna Ritz

Acknowledgements

I want to thank a few people.

Preface

This is an example of a thesis setup to use the reed thesis document class (for LaTeX) and the R bookdown package, in general.

List of Abbreviations

PPI | Protein Protein Interactions

Table of Contents

List of Tables

List of Figures

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here.

Dedication

You can have a dedication here if you wish.

Introduction

At a cursory glance, vacuoles may seem to be relatively simple organelles, storing water, nutrients, and waste products while lacking intra-organelle machinery in comparison to many of the more 'glamorous' organelles. However, despite their simplicity, vacuoles play crucial roles in various cellular processes across all eukaryotes. These functions include, but are not limited to, waste disposal, maintaining internal cell pressure, structural support, storage for nutrients, ions, pigments, and water, endocytosis, exocytosis, osmoregulation, autophagy, and defense against pathogens.

One such role of vacuoles is the regulation of stomatal opening and closing in the vast majority of terrestrial plants. The intake of carbon dioxide and other atmospheric gasses is essential for photosynthesis in all plants. However, to absorb atmospheric gasses plants must expose their internal structures to the environment which causes water stored within the plant to evaporate. Therefore, plants must carefully balance their CO2 absorption with water availability. Stomata are small mouth shaped pores present on the epidermis of nearly all land plants and are responsible for regulating gas exchange. The term stomata refers to a two part structure consisting of the stomatal aperture and two guard cells which surround the stomatal aperture creating a mouth like appearance. In the stomata's open state, both guard cells are deflated and the stomatal aperture is exposed allowing for the free exchange of atmospheric gasses and water between the plant and its environment. In the stomata's closed state, the guard cells inflate, covering the stomatal aperture and preventing gasses from entering or leaving the plant. In the deflated state, guard cells contain many small vacuoles filled primarily with water. In the inflated state, the vacuoles within guard cells fuse together to form much larger vacuoles, causing the guard cells to swell and cover the stomatal aperture. (write something about how fusion is critical to the functioning of vacuoles)

In contrast to their simplistic internal structure, the process of vacuole fusion is remarkably complex. Consequently, researchers have turned to the study of model organisms in an attempt to better understand the components and signaling mech-

anisms involved in vacuole fusion. *S. Cerevisiae*, also known as brewers yeast, has proven to be an ideal candidate for studying vacuole fusion (specifically homotypic vacuole fusion) due to its extensively studied genome, easy to visualize vacuoles, and our ability to isolate, purify, and store its vacuoles in high quantities. Furthermore, nearly 50 years of extensive research has been devoted to the study of yeast vacuole fusion providing researchers with a strong foundation of research to draw from.

Chapter 1

Yeast Homotypic Vacuole Fusion

The process of homotypic vacuole fusion in *S. Cerevisiae* is commonly divided into four general stages: Priming, Tethering, Docking, and Fusion.

1.1 Priming

1.1.1 Components

- SNARE Complex

    - SNARE proteins have two separate classification systems based on functionality and structural features respectively. Functional classification splits SNAREs into two groups: v-SNAREs which are localized to the vesicle, and t-SNAREs which are localized to the target membrane. Likewise, structural classification also splits SNAREs into two groups: Q-SNAREs which contain the central amino acid residue glutamine, and R-SNAREs which contain the central amino acid residue arginine. While it is true that R-SNAREs are often also v-SNAREs and Q-SNAREs are often also t-SNAREs, this is not always the case. There are four distinct SNARE proteins present on the yeast vacuole membrane, three Q-SNAREs: Qa (Vam3), Qb (Vti1), Qc (Vam7), and one R-SNARE: R (Nyv1). In their inactive state, the four SNARE proteins form tightly bound bundles known as cis-SNARE complexes along the vacuole membrane.

- Vps1

    - Vps1 is responsible for sequestering individual Qa SNAREs, preventing them from forming stable cis-SNARE complexes along with the R SNARE

and the other two Q SNAREs. It is thought that Vps1 polymerizes around the Qa SNARE domain, sequestering a population of Qa SNAREs which are able to undergo HOPS tethering but are unable to bind to other SNARE proteins. Vps1 must then be released prior to trans-SNARE assembly in a process involving Sec18 GTP hydrolysis independent from Sec17.

- LMA1

  - While Sec18 is bound to the cis-SNARE complex, LMA1 is bound to Sec18. Upon Sec18 mediated ATP hydrolysis, LMA1 is released from Sec18 and is bound to the Qa SNARE, stabilizing the protein.

- Sec17

  - Sec17 is responsible for the recruitment of Sec18 to cis-SNARE complexes. It functions as an intermediary by binding to both Sec18 and the cis-SNARE complex.

- Sec18

  - Sec18 is responsible for ATP hydrolysis leading to the disassembly of the cis-SNARE complex and the transfer of LMA1 to the Qa SNARE.

- Phosphoinositides

  - The primary role of phosphoinositides during the priming stage is the release of monomeric Sec18 from the vacuole membrane. Phosphatidic acid (PA) is responsible for binding and sequestering Sec18 along the vacuole membrane. The release of Sec18 from the membrane is catalyzed by the phosphatidic acid phosphatase Pah1p which converts PA to diacylglycerol (DAG). Once released from the membrane the Sec18 monomers assemble into the Sec18 hexamer which is the active form responsible for cis-SNARE disassembly.

- Pah1

  - A phosphatidic acid phosphatase responsible for the conversion of PA to DAG

1.1.2   Process

It is important to note that the protein complexes present on the vacuole membrane prior to fusion are often the result of prior fusion reactions. For example, the cis-SNARE complex is assembled during the final step of the fusion process and remains bound in its cis conformation until the next fusion event. Likewise Sec17 is bound to the cis-SNARE complex in the same fashion. Prior to the initation of priming, Sec18 is bound to the membrane by Phosphatidic acid (PA) and acts as a membrane receptor for the protein LMA1. To initiate priming, PA is converted to Diacylglycerol (DAG) by the phosphatidic acid phosphatase Pah1. This conversion releases monomeric Sec18 from the membrane. The Sec18 monomers then assemble into hexameric complexes and are recruited to cis-SNARE complexes by Sec17. Sec18 binds to Sec17 and hydrolyzes ATP, facilitating the disassembly of the cis-SNARE complex into individual SNARE proteins. During this reaction, LMA1 dissociates from Sec18 and binds to the Qa SNARE (Vam3), providing stabilization. While three of the four SNARE proteins are membrane-anchored, the Qc SNARE (Vam7) is bound only to the other SNARE proteins within the cis-SNARE complex. Upon Sec18 mediated disassembly, the Qc SNARE and Sec17 are released into the cytosol. There is also evidence that bundles of Qa SNAREs exist along the membrane during the priming process. These Qa SNARE groups are sequestered by polymerized Vps1 proteins which prevent Qa snares from associating with other SNARE proteins. Additionally, several phosphoinositides such as Ergesterol and PI(4,5)P2 have been implicated in the priming stage of homotypic vacuole fusion although their functions have yet to be defined.

1.2   Tethering & Docking

1.2.1   Components

- Ypt7

    – A Ras-like GTPase which, in conjunction with the HOPS complex, acts as a tether holding the two opposing membranes in close proximity. Ypt7 is present in both its GTP and GDP forms along the vacuole membrane. While Ypt7 is able to associate with the non-phosphorylated HOPS complex in both its GTP and GDP forms, it can only associate with the phosphorylated HOPS complex in its GTP form.

- HOPS Complex

  - A hexameric protein complex consisting of Vps39,Vps41,Vps33, Vps11, Vps16, and Vps18. The HOPS complex plays two main roles in yeast homotypic vacuole fusion: Vacuole tethering via Ypt7 binding, and organization of SNARE proteins.

- Ccz1–Mon1 Complex

  - A protein complex responsible for the conversion of Ypt7-GDP to Ypt7-GTP. Mon1 is released from the vacuole in an ATP-dependant fashion at which point it dimerizes with Ccz1 throug a currently uncharacterized mechanism.

- PI3P

  - A regulatory lipid which recruits the Ccz1–Mon1 Complex to Ypt7 in its GDP form. PI3P is also responsible for binding the Qc SNARE Vam7 to the membrane and recruiting the HOPS complex.

- PI

  - Phosphatidylinositol is a glycerophospholipid present in the cytosolic leaflet of the vacuole membrane. While it serves no direct function in vacuole fusion, it serves as a precursor lipid for a variety of Phosphoinositides such as PI3P and PI(4,5)P.

- Vps34/Vps15 Complex

  - Vps34 is a PI3 kinase responsible for the conversion of membrane bound Phosphatidylinositol (PI) to PI3P. It forms complexes with Vps15 which acts as a membrane anchor. Once bound to the membrane, Vps34 is activated by its effector protein Gpa1 and converts PI into PI3P.

- Vam7

  - Origially part of the cis-SNARE complex, Vam7 is expelled into the cytosol following cis-SNARE dissassembly. Vam7 is then recruited to the vacuole membrane by PI3P and is partially responsible for HOPS complex tethering.

- Yck3

  - A vacuolar kinase responsible for the phosphorylation of the HOPS subunit Vps41.

- Gpa1

  - The alpha subunit of a heterotrimeric guanine nucleotide-binding protein responsible for the activation of Vps34.

- Rho-type GTPases & Bem1

  - Two Rho-type GTPases (Rho1 and Cdc42) are implicated in the vacuole fusion process and have been shown to activate following Ypt7 tethering but before trans-SNARE complex formation. Bem1 is a scaffolding protein which serves both a structural role in Cdc42 signaling complexes and a regulatory role by stabilizing polarized localization of activated Cdc42. Rho-type GTPases play a critical role in actin cytoskeleton architecture, and while their exact function in homotypic vacuole fusion is currently undefined, it has been theorized that actin cytoskeleton dynamics may play a regulatory role in the process.

1.2.2 Process

In the initial stage of the tethering process, both Vps41 (a HOPS complex subunit) and Ypt7 must be phosphorylated by Yck3 and the Ccz1-Mon1 complex respectively. during this step, Vps34 is recruited to the membrane by the membrane bound Vps15. This Vps34/15 complex is then activated by Gpa1 to synthesize PI3P from PI. Following these phosphorylation events, the HOPS complex binds to Ypt7 proteins on the two opposing membranes tethering the two vesicles in close proximity. The HOPS complex shares a binding affinity for both Vam7 and the PI3P which promotes the localization of both components along the vertex ring at which point HOPS-Vam7 binding initiates trans-SNARE assembly. It is important to note that this process not only tethers the two membranes together but also correctly positions the SNARE proteins on the opposing membranes as antiparallel bound SNAREs often lead to fusion inactive complexes. In addition to the numerous regulatory mechanisms above, tethering is also mediated by the dynamin-like Vps1 protein. As discussed in the priming stage, Vps1 sequesters a population of individual Qa SNARE proteins which remain separate from cis-SNARE complexes. Qa SNARE bundles associated with Vps1 can undergo tethering upon contact with the HOPS complex; however, to form

trans-SNARE complexes, Vps1 must be depolymerized by Sec18 to release the individual Qa-SNAREs. While tethering and docking are generally defined as separate processes in yeast homotypic vacuole fusion, the distinction between when tethering ends and docking begins is not clearly delinated.

Docking is said to be completed once trans-SNARE complexes have been established, bridging the two opposing membranes. This process is primarily mediated by the HOPS complex. During the docking process, the HOPS complex protects trans SNARE complexes from Sec18 mediated dissassembly, initiates trans SNARE complex assembly by binding to the Qa SNARE and R SNARE on opposing membranes, proofreads SNARE complexes allowing only correctly assembled SNARES to undergo trans-complex formation, and recruits the remaining Q-SNAREs to its binding site to complete trans SNARE complex assembly. While it has been shown that Qa and R SNARE HOPS association precedes the recruitment of the remaining Q-SNAREs, the mechanism behind the recruitment of the remaining Q-SNAREs and the order of their assembly is currently unknown. (add section about luminal acidification/ Ca+)

1.3   Fusion

1.3.1   Components

- V-ATPase Complex

    - A 13 member protein complex split into two sectors. The V0 sector is comprised of VPH1, VMA3, VMA11, VMA16, and VMA6. The V1 sector is comprised of VMA1, VMA2, VMA4, VMA5, VMA7, VMA8, VMA10, AND VMA13. While the V-ATPase complex is commonly implicated in the active transport of protons across the vacuolar membrane, evidence points to a separate role of the V-ATPase complex in homotypic vacuole fusion. Analysis of V-ATPase activity during vacuole fusion has shown that the V0 sectors of V-ATPase complexes on the opposing membranes form inter-membrane V0 complexes. It has been proposed that these inter-membrane V0 complexes act as the initial fusion pores during the process of vertex ring hemifusion.

- Cadmodulin

    - A regulatory protein which responds to Ca+ concentration and is bound to the membrane in a Ca+ dependant fashion. Cadmodulin binds to the

V0 sector of the V-ATPase complex and, in response to Ca+, induces the formation of a V0 protien pore.

- VTC Complex

    - A 4 member protein complex comprised of VTC1, VTC2, VTC3, and VTC4. Coimmunoprecipitation studies have shown that the VTC complex binds to the VPH1 protein present in the V0 sector of V-ATPase. This binding activity is theorized to induce conformational changes in the V0 sector and induce the opening of the fusion pore.

1.3.2 Process

The final fusion event is perhaps the least understood stage of *S. Cerevisiae* vacuole fusion. Although a number of components have been linked to the final fusion process through genetic deletion screens and co-immunoprecipitation, the mechanisms by which many components regulate the fusion event have yet to be revealed. The final fusion event begins once trans-SNARE complexes have been established between two vacuoles. The assembly of trans-SNARE complexes induces the formation of the vertex ring, a region in which the two vacuole membranes are pressed tightly together forming a flattened boundary. Bilayer and Lumenal Content Mixing is thought to occur via vertex ring hemifusion. Whereas the classical model of hemifusion involves the zippering of membranes proceeding outward from a single fusion pore, the vertex ring hemifusion model proceeds via two separate fusion pores resulting in a fused vacuole containing smaller intralumenal membrane vesicles. These intralumenal membrane vesicles can be observed in *S. Cerevisiae* post fusion which provides strong evidence for the vertex ring hemifusion model as the classical model of hemifusion cannot account for the formation of such vesicles. The activity of the V-ATP complex is one of the several proposed mechanisms by which hemifusion occurs. In response to luminal acidification, the V1 subunit is released from the membrane bound V0 subunit. Concomitantly, in response to Ca+ efflux triggered by trans-SNARE formation, cadmodulin binds to and triggers conformational change in the V0 subunit. The VTC complex is thought to bind to, and stabilize these primed V0 subunits. Following priming, two V0 subunits on the opposing membranes form trans-V0 complexes which are thought to act as fusion pores.

Chapter 2

Guard Cell Homotypic Vacuole Fusion

Chapter 3

Computational Methods

3.1   Data Collection

All protein-protein interaction data for both *A. Thaliana* and *S. Cerevisiae* were downloaded from the STRING database (`https://string-db.org/`). For each organism, two files were downloaded: protein-info and protein-links. The protein-info files contain the STRING ID of each protein, the proteins common name, the proteins size, and any annotations attached to the protein. The protein-links files contain protein-protein pairs representing individual protein interactions, a number of scoring parameters, and a combined score value which represents the certainty of the interaction. The *A. Thaliana* dataset consisted of 27,463 proteins and 12,557,495 protein-protein interactions while the *S. Cerevisiae* dataset contained 6601 proteins and 2,824,843 protein-protein interactions.

3.2   Thresholding and Pruning

The STRING database uses a number of distinct evidence souces to calculate the confidence score of each protein protein interaction. Several of these sources use homologous protein interactions as evidence to support the possiblily of identical protein protein interactions occuring in the organism of interest. Due to the nature of this study, interaction evidence obtained through the comparison of homologs must be excluded to prevent the possiblilty of data contamination. Fortunatly, the STRING database provides their edge scoring algorithm as a python script which allowed for evidence channels obtained through homology to be removed and for the combined scores to be re-calculated. The homolog-free edgelist was then thresholded, allowing

only PPI's with a combined confidence score greater than 900 to be included in the dataset. This step reduced the *S. Cerevisiae* edge count to 69,784 and the *A. Thaliana* edge count to 49,604. To further isolate proteins which are likely to implicated in *S. Cerevisiae* vacuole fusion, a literature review was conducted with the goal of creating a list of core homotypic vacuole fusion proteins. A total of 45 proteins were identified in the literature review. 43 of the 45 proteins names were successfully converted to STRING IDs, however, PBI2 has no recorded PPI's with a confidence score greater than 900. Proteins without recorded PPI's cause errors when integrating the identified proteins of interest into the thresholded network and PBI2 has thus been removed from the core protein list. It is impractical to expect every PPI invovled in homotypic vacuole fusion to be identified through literature review, therfore, a K-hop algorithm was employed to create a subgraph which likely includes the majority of PPIs involved in the fusion process. The K-hop algorithm requires a set of initial nodes (in this case the 42 core proteins) and a k value (k=2 for the *S. Cerevisiae* dataset) which dictates the number of 'hops' the algorithm performs. This algorithm produces a subgraph which contains the initial nodes and all initial node neighbors that are within k edges from the initial nodes. The *A. Thaliana* dataset was pruned through a similar mechanism. However, due to the lack of literature focusing on *A. Thaliana* guard cell homotypic vacuole fusion, a literature review was insufficient for the purpose of idenitfying a set of core proteins. Therefore, a set of *S. Cerevisiae* homologous proteins was used as the basis for constructing an *A. Thaliana* core protein set. Using the STRING API's 'homology_best?' method, 37 of the 42 *S. Cerevisiae* core proteins were matched with homologous proteins from the *A. Thaliana* dataset. STRING's 'homology_best?' method employs the Smith–Waterman algorithm to perform local sequence alignment and identify likely homologous proteins. This method produces a score which describes the similarity of the two proteins where a higher score denotes a greater similarity between the two proteins. The list of *A. Thaliana* core homologous proteins was thresholded to include only proteins with SW scores greater than 550 and pruned with the K-hop algorithm using parameters identical to those used in *S. Cerevisiae* pruning.
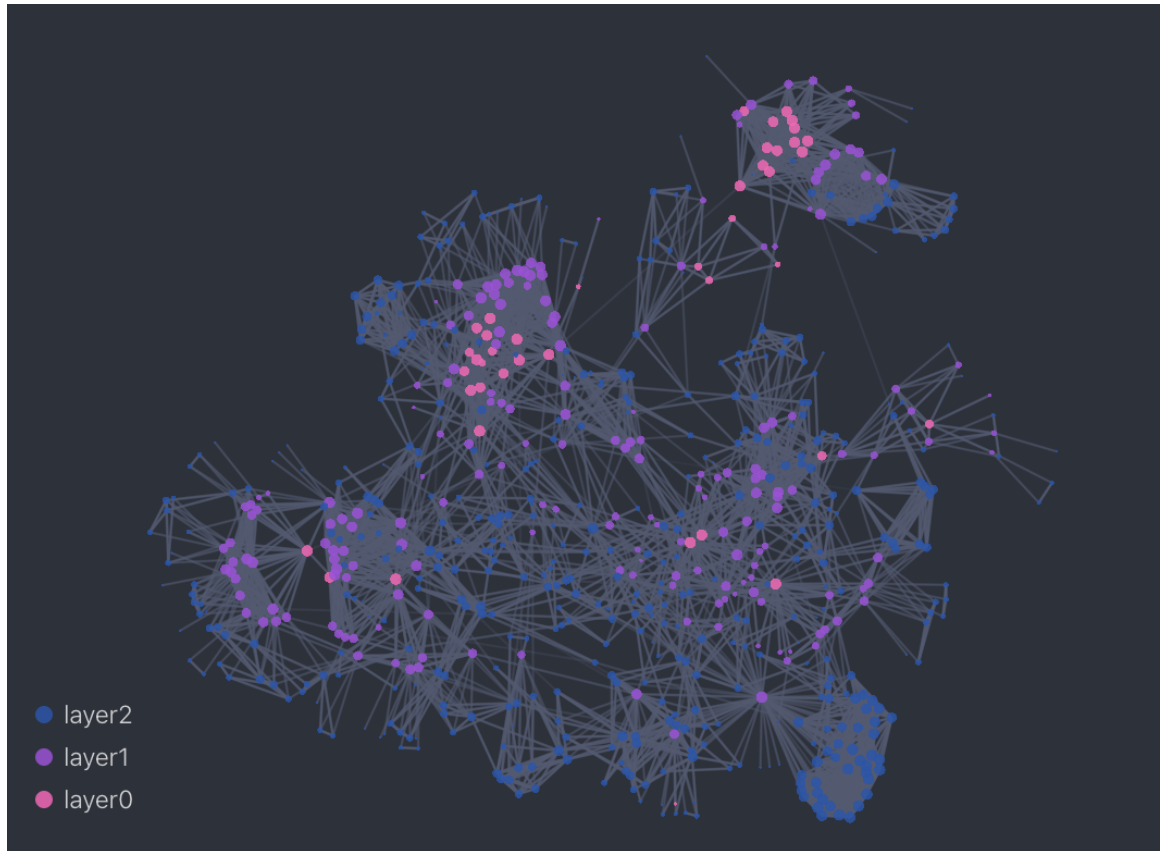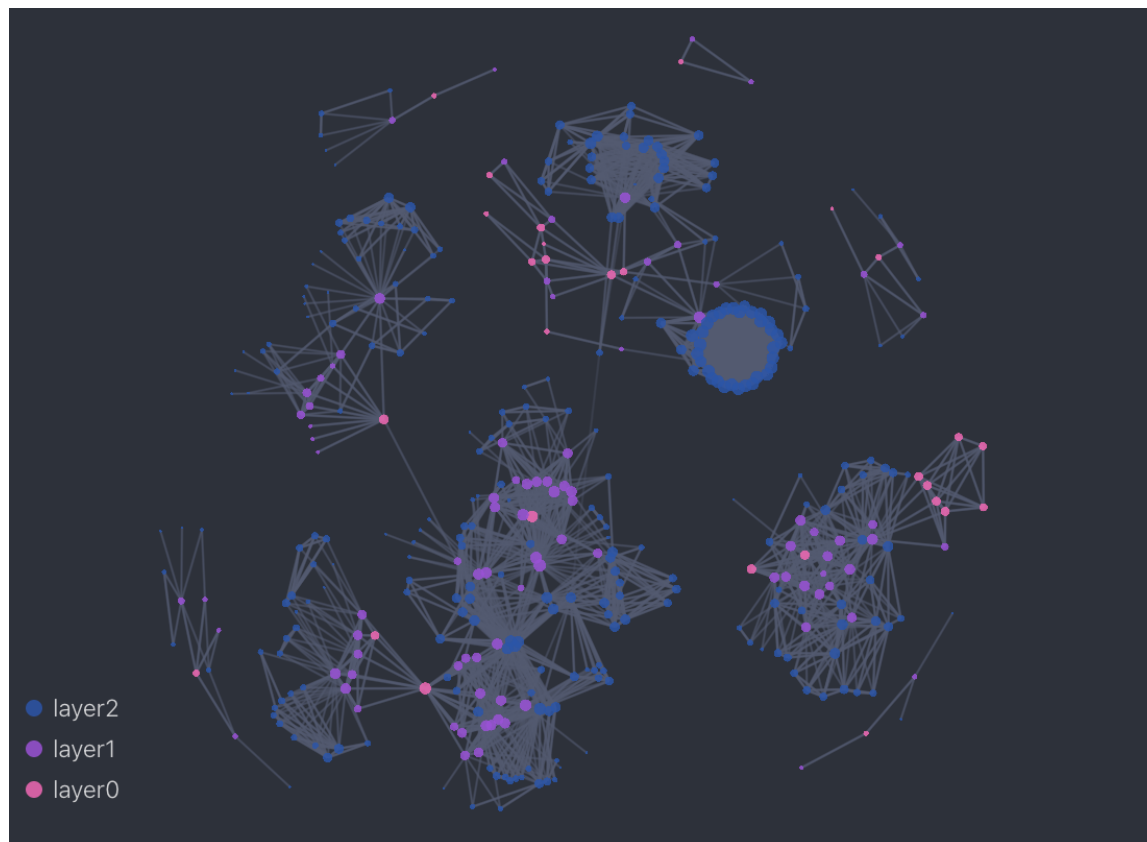
Figure 3.1: S. Cerevisiae subgraph

Figure 3.2: A. Thaliana subgraph

3.3   Network Importance Analysis

The steps above resulted in two datasets, one for each organism, each containing three 'layers' consisting of a set of core proteins, the immidiate neighbors of the core proteins, and the neighbors of the immediate neighbors (henceforth referred to as layer 0, layer 1 and layer 2 respectively). All layers were analyzed to determine the degree centrality and clustering coefficient of their constituent nodes. Degree centrality measures the number of edges connecting the node of interest to other nodes within the graph (ie. if a node is connected to 5 neighbors in the graph that node will have a degree centrality of 5). Nodes with higher degree are often thought to be more 'important' to the process modeled by the network and the distribution of node degrees in a network often occurs in a scale free fashion where a small number of nodes have very high degree centrailty while the majority of the nodes have low degree centrality. This is thought to be somewhat of an evolutionary defense mechanism against genetic mutation as loss of function mutations in proteins with low degree centrality are both more likely to happen and are less likely to

cause major disruptions in signaling pathways. Clustering coefficinet measures the interconnectedness of nodes in a network. Individual node clustering coefficients are calculated by counting the number of edges between a nodes neighbors and dividing that number by the number of possible connnections between those neighbors. These individual clustering coefficients can be used as a metric for node importance similar to degree or averaged to determine the overal clustering coefficent of the network. Layer 0 of each organism was also analyzed to determine the proportion of homologous edges (ie. edges in which both proteins are homologous and the edge between them appears in both organisms).

3.4 Network Visualization

Two python packages were used to create visual representations of the *S. Cerevisiae* and *A. Thaliana* PPI networks. Cosmograph was used to create two network visualizations for each organism, one which displays a subgraph containing the initial core proteins and their k-neighbors, and one which displays the aforementioned subgraph within the context of the full STRING PPI network. Pymnet was used to construct a multi-layer network visualization in which each of the two layers represents the PPI's between the core proteins of each organism, and edges between the layers represent homologous proteins present in each organim.
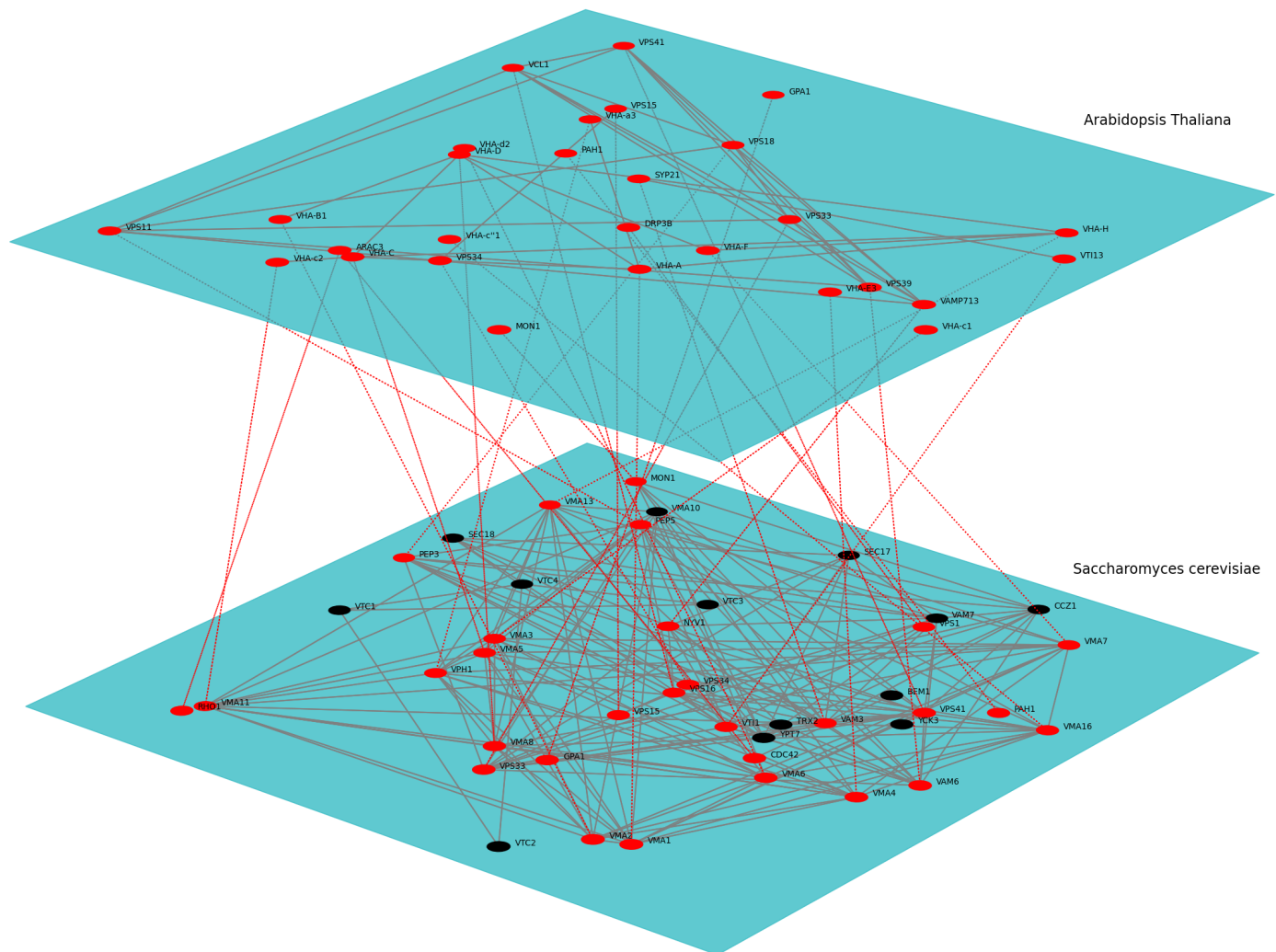
Figure 3.3: Homologous edges between A. Thaliana and S. Cervisiae

Chapter 4

Ortholog Comparisons

4.1   Layer 0 Homologs and Properties

For the *A. Thaliana* dataset, layer 0 represented 0.1% of the total nodes and 0.1% of the total edges and had an average clustering coefficient of 0.49. For the *S. Cerevisiae* dataset, layer 0 represented 0.6% of the total nodes and 0.4% of the total edges. There were a total of 66 homologous edges between layer 0 of the two organisms. These represented 94% (66/70) of the total edges in the initial *A. Thaliana* layer and 25% (66/256) of the total edges in the initial *S. Cerevisiae* layer.

4.2   Layer 1 Homologs and Properties

For the *A. Thaliana* dataset, layer 1 represented 0.3% of the total nodes and 1% of the total edges and had an average clustering coefficient of 0.59. For the *S. Cerevisiae* dataset, layer 1 represented 2% of the total nodes and 3% of the total edges. There were a total of 202 homologous edges between layer 1 of the two organisms. These represented 23% (202/844) of the total edges in layer 1 of the *A. Thaliana* dataset and 15% (202/1336) of the total edges in layer 1 of the *S. Cerevisiae* dataset.

4.3   Layer 2 Homologs and Properties

For the *A. Thaliana* dataset, layer 2 represented 1% of the total nodes and 6% of the total edges and had an average clustering coefficient of 0.59. For the *S. Cerevisiae* dataset, layer 2 represented 7% of the total nodes and 8% of the total edges. There were a total of 294 homologous edges between layer 2 of the two organisms. These

represented 8% (294/3492) of the total edges in layer 2 of the *A. Thaliana* dataset and 11% (294/2633) of the total edges in layer 2 of the *S. Cerevisiae* dataset.
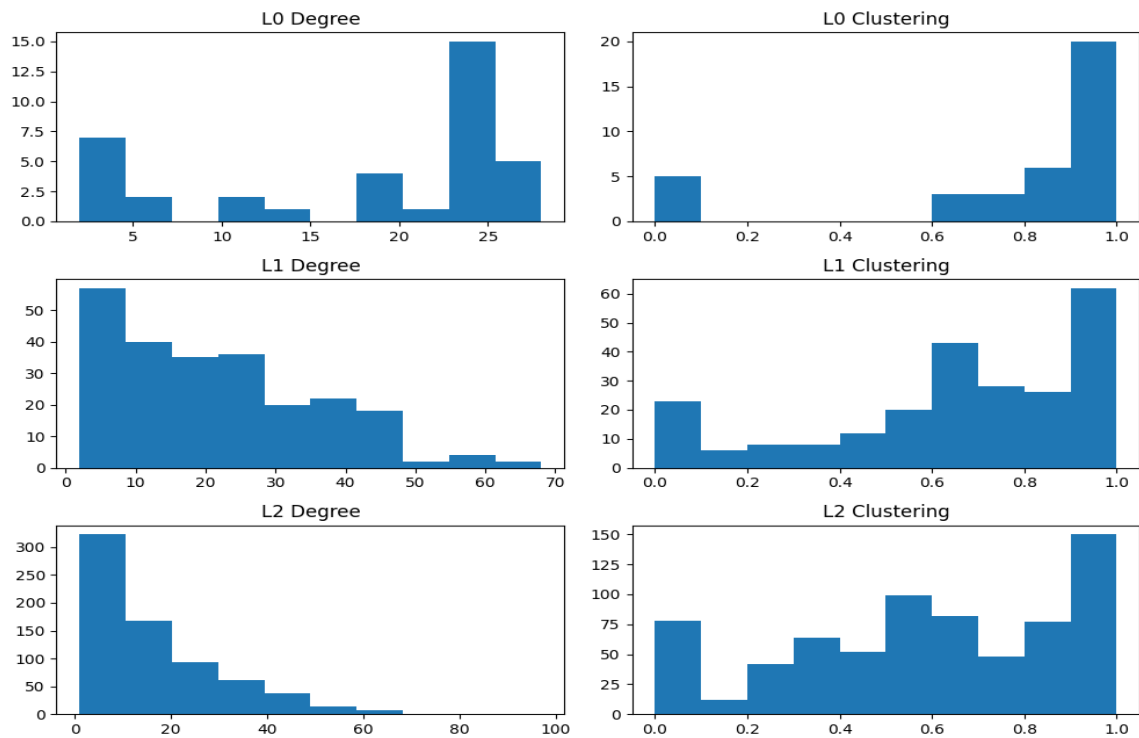
Figure 4.1: S. Cerevisiae layer statistics
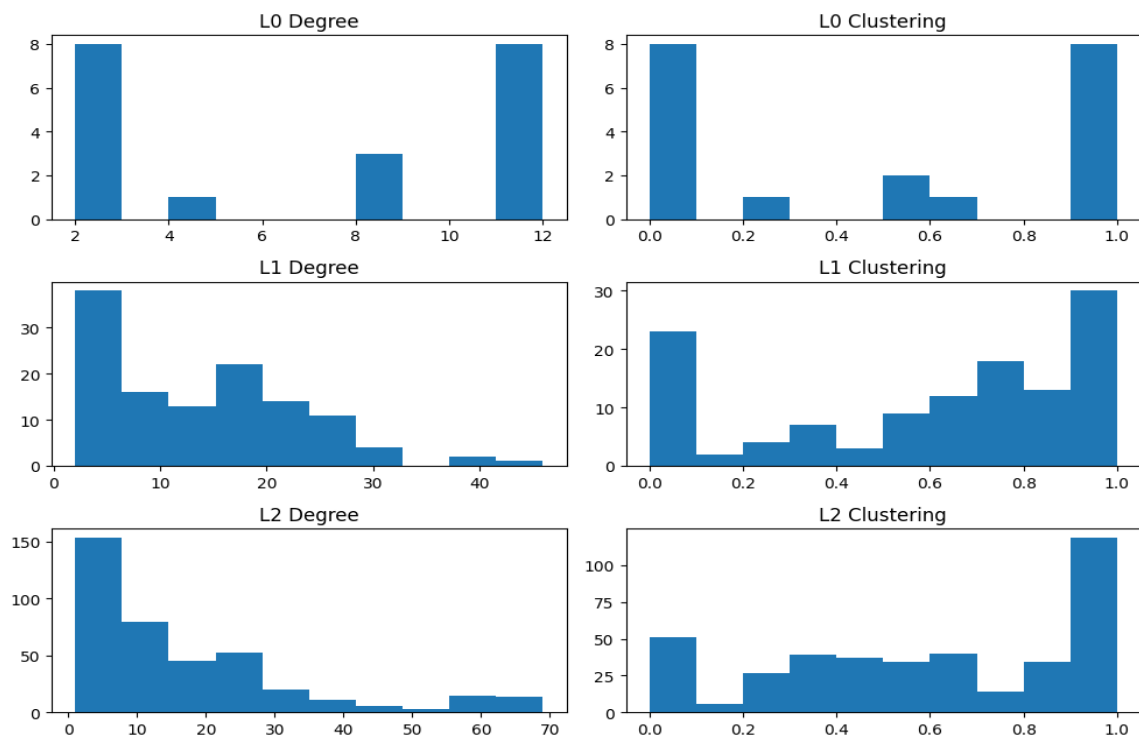
Figure 4.2: A. Thaliana layer statistics

Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Appendix

Protein Tables

S. Cerevisiae

A. Thaliana

References

Seals, Darren F., Gary Eitzen, Nathan Margolis, William T. Wickner, and Albert Price. 2000. "A Ypt/Rab Effector Complex Containing the Sec1 Homolog Vps33p Is Required for Homotypic Vacuole Fusion." Proceedings of the NationalAcademy of Sciences of the United States of America 97(17):9402–7. `doi:10.1073/pnas.97.17.9402`.