THESIS

GENOME-WIDE SIGNATURES OF COEVOLUTION IN PARASITIC AND AUTOTROPHIC PLANTS

Submitted by

Chris deRoux

Department of Biology

In partial fulfilment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2025

Master's Committee:

Advisor: Dan Sloan

Evan Forsythe

Marek Borowiec

ABSTRACT

GENOME-WIDE SIGNATURES OF COEVOLUTION IN PARASITIC AND AUTOTROPHIC PLANTS

Parasitism has repeatedly evolved in flowering plants, resulting in convergent genomic, metabolic, and morphological adaptations. Evolutionary Rate Covariation (ERC) analysis is a phylogenomic method used to uncover gene pairs whose functional interactions are under shared selection pressures. In this study, we applied the ERCnet pipeline to a paired sample of parasites and autotrophic relatives to investigate patterns of gene coevolution common to trophic mode transitions. Few networks of genes sharing ERC exhibited trophic-mode-driven differences in evolutionary rates, suggesting divergent selection pressures on shared interactions. Gene Ontology (GO) analysis yielded underrepresentation of plastid-related genes in coevolving networks. Meanwhile, two smaller networks highlighted the slow evolution of developmental and fatty acid metabolism genes in parasites. Covarying genes involved in the WUSCHEL/CLAVATA3 (WUS/CLV3) signaling pathway, key to shoot development, evolved slowly in parasites, implicating them in body plan reorganization. Additionally, genes involved in fatty acid metabolism were overrepresented in a slow-evolving parasite cluster, hinting at parasite dependence on conserving this pathway while losing photosynthetic plastid functions. These results indicate conservation of gene coevolution patterns between the two trophic modes, with a few notable exceptions. Expanding ERC analysis of parasitic plants to

include more species could provide greater insight into divergent and conserved networks of gene interactions.

TABLE OF CONTENTS

1. INTRODUCTION

The ancestral state of plants is photosynthetic autotrophy, with plants capturing light energy from their abiotic environment. However, heterotrophy has evolved independently and convergently 12-13 times in angiosperms [1]. Parasitic plants obtain some or all of their organic carbon from other plants, whether directly by host invasion [2] or indirectly by exploiting mycorrhizal networks [3]. Although parasitic plants are deleterious to the health of host plants, parasites in general and parasitic plants specifically are drivers of ecological diversity and organismal adaptation [4]. Dependence on host-derived carbon ranges from facultative parasitism to obligate reliance on host plants. Parasitic plant morphology is equally diverse; parasitic plants can retain stems and leaves on the surface or dwell exclusively within the soil or their hosts, only emerging to flower. Some parasitic plants retain pigment-containing chloroplasts, which allow for continued autotrophy in addition to heterotrophy, while others lose chloroplasts, only retaining plastids.

The molecular adaptations of parasitic plants have been best studied in agricultural pest species, uncovering striking adaptations like RNA sharing between host and parasite as in the dodder *Cuscuta* [5]. Meanwhile, less economically impactful species like those in

[1] Westwood et al., "The Evolution of Parasitism in Plants."
[2] Yoshida et al., "The Haustorium, a Specialized Invasive Organ in Parasitic Plants."
[3] Merckx, *Mycoheterotrophy*.
[4] Betts et al., "High Parasite Diversity Accelerates Host Adaptation and Diversification."
[5] Johnson and Axtell, "Small RNA Warfare"; Shahid et al., "MicroRNAs from the Parasitic Plant Cuscuta Campestris Target Host Messenger RNAs."

the genus *Rafflesia* are morphologically well-characterized; however, the underlying molecular adaptations are in the early stages of characterization [6]. This disparity leaves valuable information to be gleaned from broader phylogenetic approaches to characterizing plant parasitism [7].

Each independent evolution of parasitism presents a unique co-option of ancestral plant structures and molecular systems for new purposes [8]. Though distinct in their evolutionary histories, these independent adaptations provide insights into the convergent evolution of parasites. Similar selection pressures (or relaxation of pressures) acting on genetically isolated taxa inform us about those selection pressures themselves as well as common adaptations associated with the evolution of parasitism. Host-penetrating structures called haustoria have evolved convergently within plants, and parasitic plants share traits common to parasites across the tree of life such as body plan reduction and rapid evolution [9]; [10]

The adoption of parasitism leads to relaxed selection on genes that code for functions offloaded to its host, resulting in gene loss and unconstrained mutation. Parasitic plants demonstrate this through the reduction or outright loss of the chloroplast genome [11].

---

[6] Mursyidah et al., "Dissecting the Biology of Rafflesia Species."
[7] Clarke et al., "Molecular Dialog Between Parasitic Plants and Their Hosts."
[8] Ichihashi et al., "Transcriptomic and Metabolomic Reprogramming from Roots to Haustoria in the Parasitic Plant, Thesium Chinense."
[9] Bromham, Cowman, and Lanfear, "Parasitic Plants Have Increased Rates of Molecular Evolution across All Three Genomes."
[10] Jackson, "The Evolution of Parasite Genomes and the Origins of Parasitism."
[11] Wicke and Naumann, "Molecular Evolution of Plastid Genomes in Parasitic Flowering Plants."

Conversely, parasites benefit from rapid evolution in evolutionary arms races with their hosts [12]. Variation in evolutionary rate between taxa is not always due to drastic differences in lifestyle but is instead widespread and multifactorial. Species may evolve faster or slower than others due to relaxed or constrained selection, generation time, metabolic rate, differences in replication machinery and DNA damage repair, among other factors [13]. As a result, lineage-specific evolutionary rates fluctuate over time. These factors underlying variation in evolutionary rate apply not only to taxa but to individual genes. Natural selection acts differentially on genes depending on their importance to the organism [14]. Even specific regions within genes may be subject to more constrained selection than others [15]. At one extreme end of rapid evolutionary rate, strong positive selection favors mutations in genes such as viral spike proteins[16] whose evolutionary rate must outpace host immune adaptation. Lack of selection in favor of or against mutations occurs pseudogenes, which are non-translated descendants of genes that presumably evolve neutrally and accumulate mutations at an intermediate rate due to the lack of both positive and purifying selection on maintenance of sequence [17]. At the other extreme,

---

[12] Betts et al., "High Parasite Diversity Accelerates Host Adaptation and Diversification."
[13] Voskarides, Dweep, and Chrysostomou, "Evidence That DNA Repair Genes, a Family of Tumor Suppressor Genes, Are Associated with Evolution Rate and Size of Genomes"; Volkova et al., "Mutational Signatures Are Jointly Shaped by DNA Damage and Repair."
[14] Wicke et al., "Mechanistic Model of Evolutionary Rate Variation En Route to a Nonphotosynthetic Lifestyle in Plants"; McNeal et al., "Complete Plastid Genome Sequences Suggest Strong Selection for Retention of Photosynthetic Genes in the Parasitic Plant Genus Cuscuta."
[15] Echave, Spielman, and Wilke, "Causes of Evolutionary Rate Variation among Protein Sites."
[16] Amicone et al., "Mutation Rate of SARS-CoV-2 and Emergence of Mutators during Experimental Evolution."
[17] Cheetham, Faulkner, and Dinger, "Overcoming Challenges and Dogmas to Understand the Functions of Pseudogenes."

ultraconserved elements are noncoding regulatory regions under stringent purifying selection, remaining conserved across disparate taxa, providing information about deep evolutionary divergences [18]. As with evolutionary rate variation between taxa, between-gene evolutionary rate similarities can stem from selection pressure alone, or other causes like gene expression levels, gene functions, and chromatin accessibility. Genomic regions evolving slower than the background rate often indicate purifying selection on that region [19], as seen with ultraconserved elements. Prior knowledge of the causes of rate variation helps predict the evolutionary rates of genes in question. For example, a gene that is highly expressed is examined for mutation more often by transcription-coupled DNA repair, suggesting a slow mutation rate[20]. Evolutionary rate variation and its causes are reciprocally informative, in the sense that genes with conserved functions tend to exhibit sequence conservation, and vice versa.

Variation in the evolutionary rate of genes across branches in phylogenies provides information about the selection pressures acting on those genes [21]. Genes that share similar changes in evolutionary rates across species may be under similar constraints. This similarity in evolutionary rate between genes is called Evolutionary Rate Covariation (ERC), a signal that can be found through correlation analyses of gene substitution rates [22]. ERC

[18] Ryu, Seridi, and Ravasi, "The Evolution of Ultraconserved Elements with Different Phylogenetic Origins."
[19] Wicke et al., "Mechanistic Model of Evolutionary Rate Variation En Route to a Nonphotosynthetic Lifestyle in Plants."
[20] Drummond et al., "Why Highly Expressed Proteins Evolve Slowly."
[21] Echave, Spielman, and Wilke, "Causes of Evolutionary Rate Variation among Protein Sites."
[22] De Juan, Pazos, and Valencia, "Emerging Methods in Protein Co-Evolution."

signal between genes provides information about the functions of those genes within fast-evolving or slow-evolving taxa [23]. Rather than examining gene evolutionary rate in isolation and determining whether autotrophs or parasites evolve faster at that locus, ERC provides insight into the relationships between genes, evidencing which interactions are under shared functional constraints. By organizing pairs of genes exhibiting ERC into networks of coevolving genes, we can visualize clusters of genes that share similar evolutionary rates and investigate these clusters for shared functionality using existing functional datasets. This can both assess the informativeness of ERC findings and direct future research towards evidently coevolving clusters of genes common to the parasitic lifestyle. ERC is often associated with adaptations occurring in a subset of lineages over others.

Here, we aim to detect ERC between genes involved in the evolution of parasitism in angiosperms. Characterizing ERC between genes in the nuclear genomes of parasitic plants and comparing these patterns of covariation to those found in the autotrophic relatives of these parasites, we aim to answer one small part of the larger question of what adaptations drive the parasitic lifestyle to emerge repeatedly in angiosperms. Specifically, which coevolutionary relationships between genes are under differential selection pressure in parasitic plants? ERC between genes indicates gene co-evolution, that is, reciprocal selection pressures resulting from functional interactions between genes and even direct residue contact between the proteins they encode. However, genes may also

[23] Clark, Alani, and Aquadro, "Evolutionary Rate Covariation Reveals Shared Functionality and Coexpression of Genes"; Little, Chikina, and Clark, "Evolutionary Rate Covariation Is a Reliable Predictor of Co-Functional Interactions but Not Necessarily Physical Interactions."

exhibit ERC because they have related functions associated with co-expression or activity in a shared metabolic pathway even if they are not co-evolving in a strict sense [24]. Untangling the source of ERC signal requires biological interpretation of experimentally confirmed or putative functions of the two genes observed to co-vary. Many approaches have been developed using molecular phylogenetic methods to discover ERC between genes; however, the comparison of gene trees based on orthology requires similar enough topology to overlap branches between gene trees. This assumption of one-to-one orthology within compared trees is especially challenged by plant genomes which tend to develop many paralogous copies of an ancestral gene. These paralogous copies become isolated lines of descent, potentially complicating branch length estimations[25]. Paralogs can render one or both gene trees in a pairwise comparison topologically inconsistent, making branch length comparisons unfeasible unless the overall tree topology matches. Detecting and limiting the number of paralogs in ERC analysis alleviates this issue, ensuring that gene tree comparisons are pairwise rather than between families of paralogs. ERCnet is an ERC analysis pipeline designed for use on paralog-heavy plant genomes[26], fitting paralogous genes in gene families into tree topologies comparable 1:1 with those of other gene families despite duplications.

---

[24] Little, Chikina, and Clark, "Evolutionary Rate Covariation Is a Reliable Predictor of Co-Functional Interactions but Not Necessarily Physical Interactions."
[25] Forsythe, Williams, and Sloan, "Genome-Wide Signatures of Plastid-Nuclear Coevolution Point to Repeated Perturbations of Plastid Proteostasis Systems across Angiosperms."
[26] Forsythe et al., "Phylogenomic Prediction of Interaction Networks in the Presence of Gene Duplication."

Observing ERC between genes in species under shared selection pressures, as is the case for independently evolved parasites, could reveal previously undiscovered interactions between genes responsible for the evolution of parasitism. In the context of plants diverging from autotrophic ancestors, parasitic lineages should undergo relaxations in the evolutionary rates shared between genes involved in lost functions such as chloroplast maintenance [27], photosynthesis, and leaf development. Likewise, genes involved in parasitic gains of function such as haustorial development [28] and host manipulation may exhibit ERC due to their involvement in shared pathways with fitness implications for host-dependent parasites. These genes may be under positive selection in favor of mutations supporting these novel adaptations. By identifying genes that exhibit ERC across both autotrophic plants and closely related parasitic species, we may uncover a wealth of information by studying gene coevolution driving the success of both autotrophic and parasitic plants. We may also find genes that are commonly co-opted for adaptation to parasitic lifestyles [29].

[27] Forsythe, Williams, and Sloan, "Genome-Wide Signatures of Plastid-Nuclear Coevolution Point to Repeated Perturbations of Plastid Proteostasis Systems across Angiosperms."
[28] Ichihashi et al., "Transcriptomic and Metabolomic Reprogramming from Roots to Haustoria in the Parasitic Plant, Thesium Chinense."
[29] Yang et al., "Comparative Transcriptome Analyses Reveal Core Parasitism Genes and Suggest Gene Duplication and Repurposing as Sources of Structural Novelty."

2. RESULTS

2.1 Most Orthogroups Showed ERC with One or More Other Orthogroups

We collected protein sequences from the genomes/transcriptomes of four pairs of heterotrophic and autotrophic angiosperm relatives along with the outgroup *Amborella trichopoda* (Figure 1). These sequences were grouped into 63,360 sets of homologous sequences known as "orthogroups" using Orthofinder, which were then filtered based on multiple criteria to exclude orthogroups that were unsuitable for ERC analyses (see Methods).

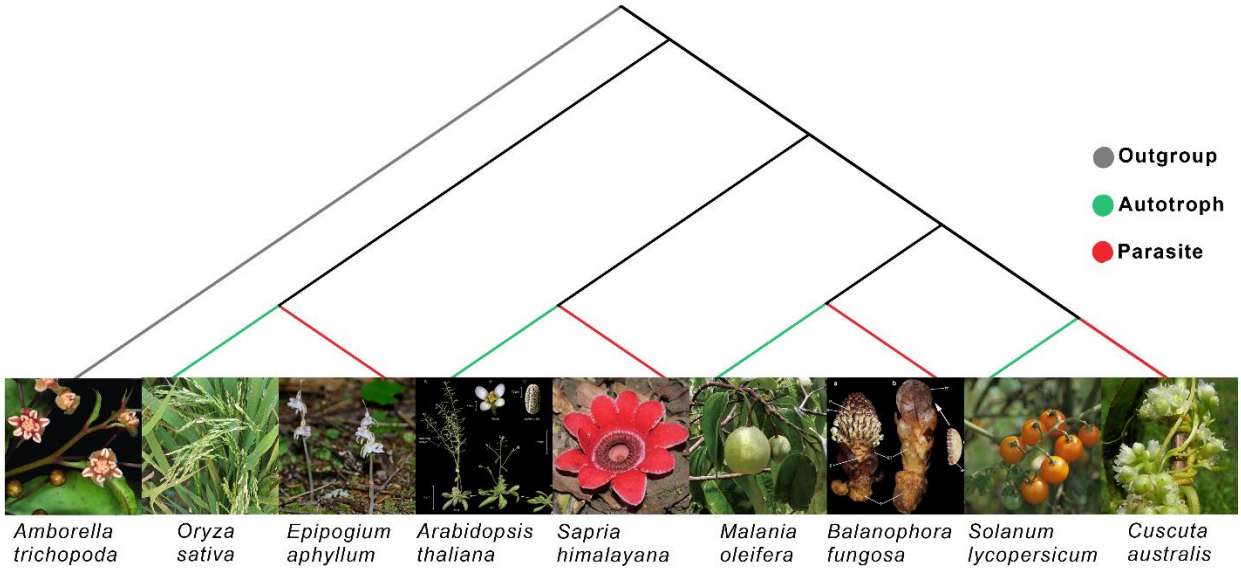

Figure 1

Ultrametric tree of species pairs whose genomes or transcriptomes provided protein-coding gene sequences for ERC analysis. Each pair represents an independent evolution of parasitism. [30]

Out of 5,726 hierarchical orthogroups analyzed for ERC, 5,678 were found in at least one significant covarying pair. These findings resulted from running 14,728,952 pairwise tests. The strength of ERC signal varied for each pair, with uncorrected Pearson p-values for significance in covariance ranging from just below 0.01 (the applied cutoff for this analysis) to 2.63e-9. Pearson R^2 values, i.e. the degree to which the best fit correlation line explained variance in branch length, ranged from 0.69622 to 0.99997.

[30] Cui et al., "Ethylene Signaling Mediates Host Invasion by Parasitic Plants"; Rose, *Cuscuta Australis Flower5*; Duan et al., "Genome-Wide Analysis of the MADS-Box Gene Family in Holoparasitic Plants (Balanophora Subcupularis and Balanophora Fungosa Var. Globosa)"; Nobuyuki TANAKA et al., "Contributions to the Flora of Myanmar IV"; Jasinski et al., "Improving Seed Oil and Protein Content in *Brassicaceae*"; Grosse-Veldmann, "Amborella Trichopoda Cultivation of the Most Ancestral Angiosperm in Botanic Gardens"; Zięba et al., "Przyczynki do flory Tatrzańskiego Parku Narodowego"; 阿橋花譜 HQ Flower Guide, *Rice*.

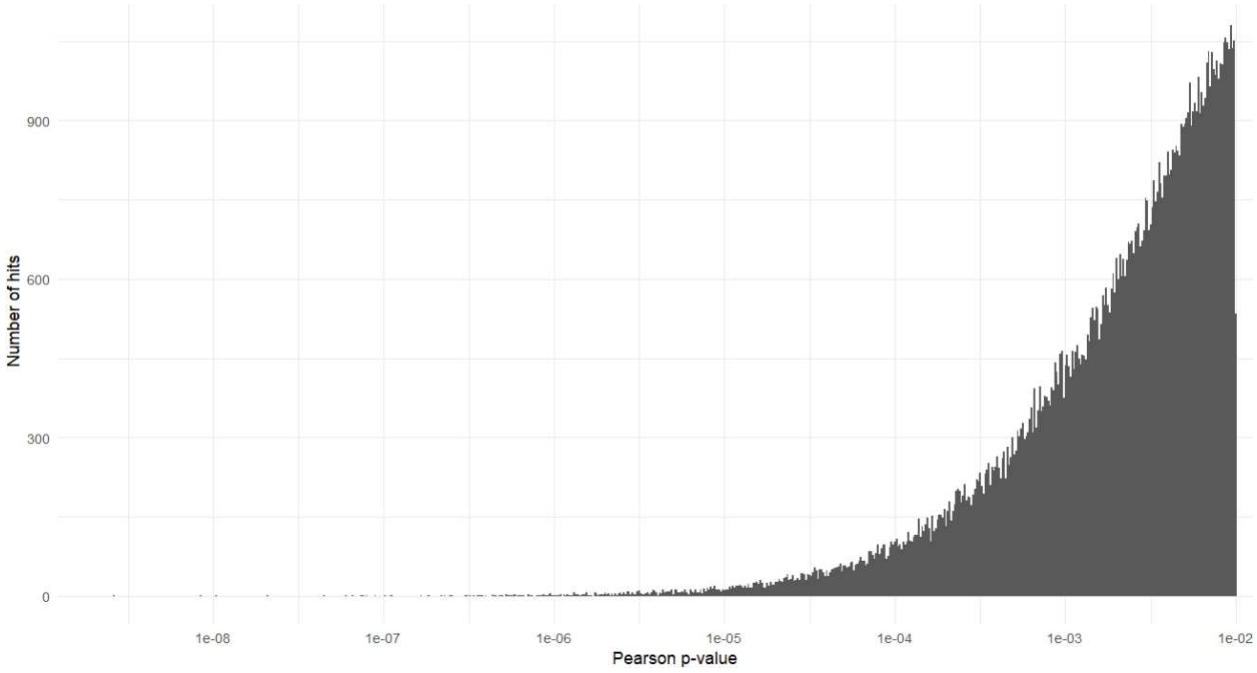

Figure 2

Histogram of ERC hits ordered by Pearson correlation p-value.

2.2 Strongest Hits Included Known Functional Interactions and Unknown Gene Functions

As expected, most hits showed significance closer to the <0.01 cutoff, with a small number at the opposite extreme (Figure 2). The ortholog pairs showcasing the strongest evidence of covariance are summarized in Table 1. Four top hits include covariation between at least one Arabidopsis gene of unknown function. While not immediately informative, strong evidence of covariation within this species sample could indicate shared functionality between experimentally characterized genes and those belonging only to broad protein families. Two pairs share functionality in cell wall synthesis, deposition,

and plasticity. AT2G42300 regulates hypocotyl elongation through cell wall plasticity, and is in strong ERC with AT5G47780, part of the biosynthetic pathway for pectin, a gelatinous cell wall component with alterable plasticity,[31] suggesting shared functions between regulator and regulatory target. The AT1G72670 gene orients the preprophase band during plant cell division, deciding where the new wall will be deposited between daughter cells, while its strongly covarying partner AT5G64470, a O-acetyltransferase which acts on polysaccharides in the cell wall, playing a role in directional plant development. The involvement of these two genes in cell wall deposition and chemistry explains their strong ERC, and may indicate a role for cell wall construction in host-parasite relationships in addition to cell wall destruction.

Table 1

Pairs of genes showing ERC with the lowest p-values, identified by their *Arabidopsis thaliana* homolog and function if known. R^2 quantifies the amount of variance explained by

[31] Cao et al., "Egg-Box Model-Based Gelation of Alginate and Pectin."

the best fit trend line describing correlated branch lengths.

Strongest Hits					
A. tha Ortholog	Function	*A. tha* Ortholog	Function	Pearson R^2	Pearson p-value
AT2G40970	Cold Response	AT4G12540	Unknown	0.9994305	2.63e-09
AT2G44970	Unknown	AT4G24175	Unknown	0.9998496	8.48e-09
AT2G37960	Unknown	AT1G76340	GDP-L-galactose transporter	0.9990099	1.05e-08
AT2G42300	Regulates hypocotyl elongation	AT5G47780	Pectin biosynthesis	0.9986991	2.07e-08
AT1G55546	4-phosphate cytidylyltransferase	AT3G48470	Embryo development	0.9982276	4.49e-08
AT3G51390	Vacuole regulator, palmitoylation	AT1G20580	Core spliceosome protein	0.9979978	6.09e-08
AT1G80210	BRCA1 homolog	AT1G30320	Symbiosis signaling	0.9979384	6.56e-08
AT3G27000	Microtutbule nucleation	AT2G18990	Microtubule organization, viral transport	0.9978278	7.47e-08
AT1G04590	Unknown	AT3G19760	Stress-related initiation factor	0.9978090	7.64e-08
AT1G72670	Preprophase band determination	AT5G64470	Polysaccharide O-acetyltransferase	0.9999673	7.95e-08

2.3 ERC Results Validated by Measuring Covariation Between Co-Localizing Genes

To validate whether the correlations between genes reflected known biological causes of covariation, we hypothesized that genes that are targeted to the same cellular compartment (mitochondrion, plastid, or other) would share co-functionality and thus be likely to produce ERC signal. Regardless of p-value cutoff or branch length measurement method, the subcellular localization of surviving genes had a small but positive association with whether these genes covaried (i.e., assortativity). ERC was calculated using two methods for branch length, Root-to-Tip (R2T) and Branch-by-Branch (BXB). R2T methods produced datasets with larger assortativity effect sizes (assortativity coefficient) and lower p-values (assortativity p-value) than BXB for this species sample, so R2T will be used going forward. Lower p-value thresholds for ERC hits (ERC p-value) constricted the dataset to fewer genes with stronger ERC signal (Table 2). The assortativity p-value was <0.01 for each

of the R2T branch measurement ERC p-value cutoff thresholds, with confidence in assortativity increasing with the number of genes passing each cutoff.

Branch Length Method	ERC p-value	Number of Hits	Assortativity Coefficient	Assortativity p-value
BXB	0.00001	195	0.03195489	0.32604612
BXB	0.0001	1012	0.05330467	0.02774153
BXB	0.001	3663	0.00897061	0.13929992
BXB	0.01	4770	0.00914357	0.001585
R2T	0.00001	609	0.1035344	0.00136301
R2T	0.0001	2259	0.02930936	0.01422059
R2T	0.001	4683	0.02915396	1.1160E-07
R2T	0.01	5678	0.02089091	3.8037E-20

Table 2. Summary of assortativity of proteins with the same subcellular localization in ERC networks

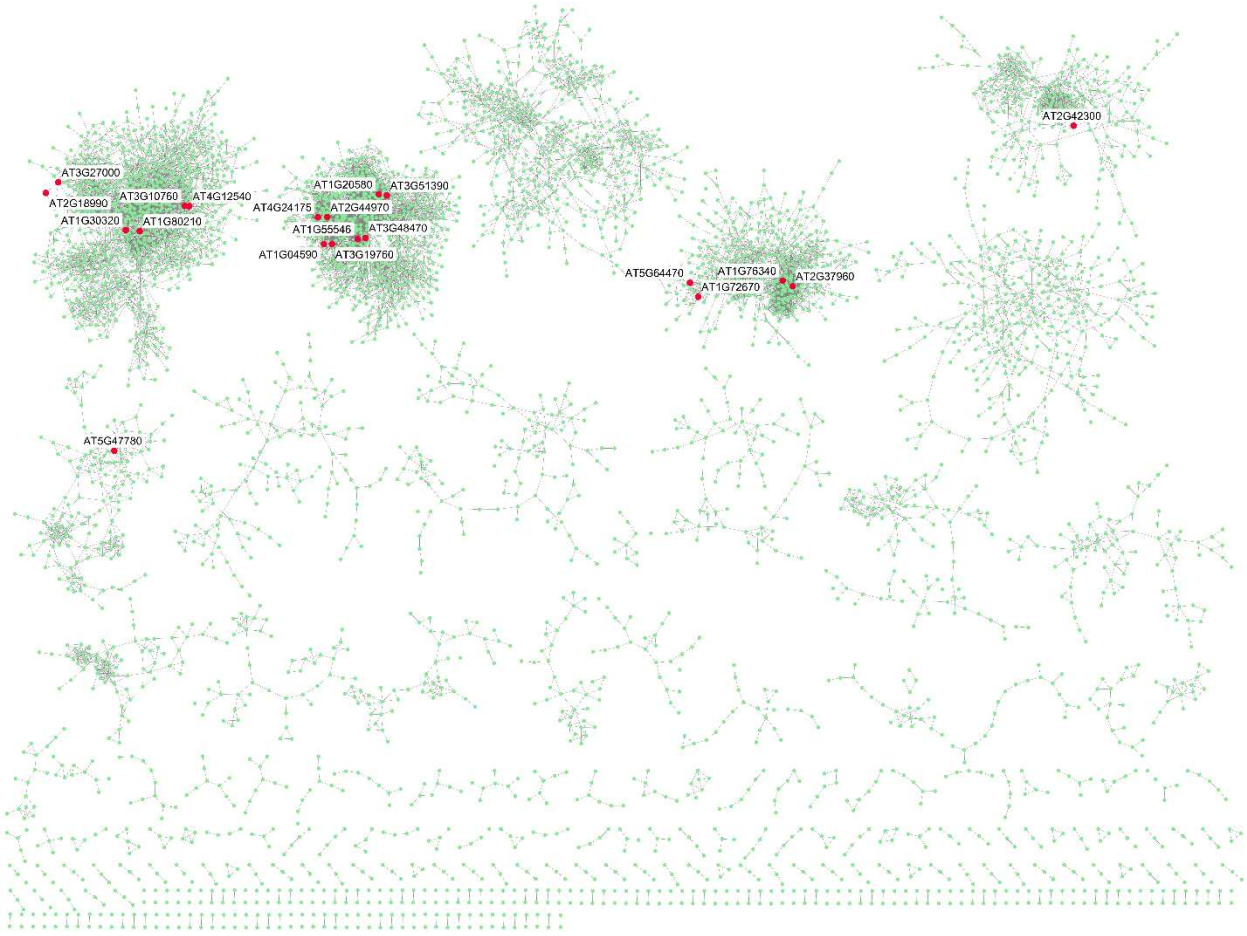

Figure 3

Network communities of covarying orthogroups. Nodes represent orthogroups, and lines indicate significant ERC for Pearson P<0.001 between two orthogroups. The top 10 pairs exhibiting the strongest signature of ERC are highlighted in red. Clusters are numbered arbitrarily for indexing purposes.

2.4 Network Analysis Reveals Orthogroups as Interaction Hubs

Though ERC data is generated using pairwise comparisons, evidence of gene coevolution may point to shared functionality in a broader pathway or process involving multiple genes. Only 113 genes out of 4,683 were found to have significant ERC with only

one other gene, typified by the lone pairs at the bottom of the network visualization (Figure 3). The remaining 4,570 are better understood in communities of coevolving genes. Some genes are peripheral in these networks, only showing ERC with one or two neighbors, while others like *CHR5* (*Chromatin Remodeling Factor 5,* AT2G13370) are hubs showing significant ERC with 245 other genes. *CHR5* is implicated in both plant development and immunity through its nucleosome repositioning in response to stimuli [32], possessing a high number (estimated 21) of MAP-kinase docking sites [33], so its interconnectedness is in line with its highly interactive biological functions in *Arabidopsis*.

The 10 strongest hit pairs are exclusively found in the most populous clusters containing hundreds to thousands of genes. GLay clustering generally kept these strongly correlated pairs together, only once separating a strong hit into two community clusters.

[32] Zou et al., "The Arabidopsis Chromatin-Remodeling Factor CHR5 Regulates Plant Immune Responses and Nucleosome Occupancy."
[33] Rayapuram et al., "Chromatin Phosphoproteomics Unravels a Function for AT-Hook Motif Nuclear Localized Protein AHL13 in PAMP-Triggered Immunity."

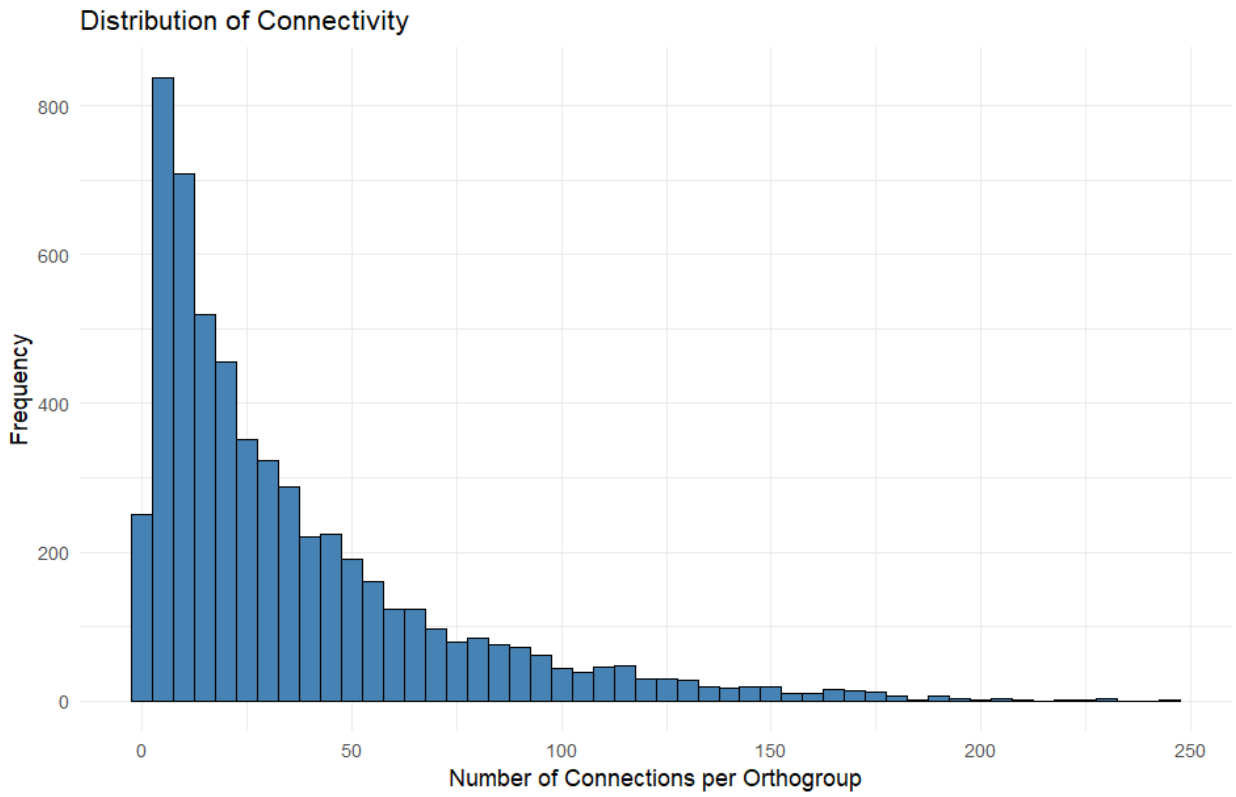

Figure 4

Distribution of the number of ERC hits between each orthogroup and other orthogroups.

Few orthogroups were found in only one pair, with the mode being 5 connections.

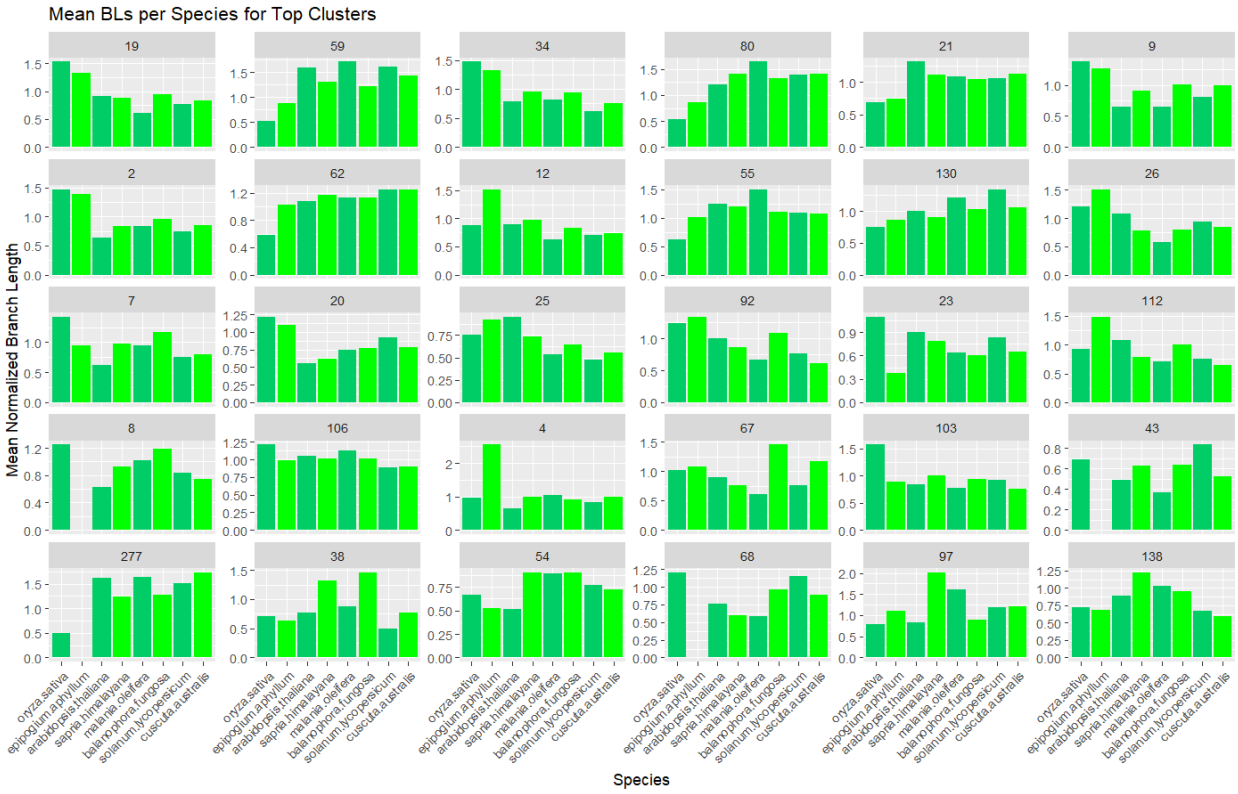

Figure 5

Normalized branch lengths averaged across all genes in the largest 30 network clusters (each panel represents a cluster labeled with their ID numbers). Dark green bars indicate autotrophic species, while light green indicated heterotrophs.

2.5 Little Difference in Branch Length Means between Trophic Groups

Following Glay clustering, we measured the mean branch lengths for each species' orthologous copy of gene families found within clusters showing ERC. These mean branch lengths per cluster are shown in Figure 5 for the largest 30 clusters. Species means were then treated as data points in a test for difference in the mean of means between autotrophs and heterotrophs as trophic categories. While we expected these clusters of coevolving genes to evidence faster or slower evolution in one trophic mode or the other

depending on relaxed or heightened constraint on shared functions, none of the 30 largest orthogroup clusters showed significant differences in mean branch lengths between autotrophs and heterotrophs (Student's t-test $p<0.05$). Only 12 out of 285 clusters showed significant difference in mean between these trophic groups. These clusters were small, numbering 2-5 covarying genes. Due to significance being found only in the smallest samples of branch lengths, this is more likely due to sampling bias than biological explanation. With a basic Student's t-test not correcting for multiple testing, this level of significance can be expected to occur from random chance.

Overall, the two monocot species (*Oryza sativa* and *Epipogium aphyllum*) together diverged in mean branch length compared to the dicots sampled. The differences in branch length attributable to phylogenetic divergence are greater in magnitude than those which can be explained by trophic mode. *Epipogium aphyllum* was absent data for mean branch length comparison due to sharing no ortholog with the other species. This is best explained by the source of its proteome for this analysis being a transcriptome which necessarily only captures genes being expressed by adult plants in sampled tissues, rather than all genes present in a nuclear genome.

When comparing nearest related autotroph-heterotroph pairs, clusters 23 and 106 stand out as having one trophic mode or the other showing faster evolution despite there being no difference when comparing branch lengths for the trophic level groups as a whole. Autotroph lineages in clusters 23 and 106 appear to be evolving faster than their closest heterotrophic relatives for genes in those respective clusters. The 2nd largest cluster 59 shares a similar pattern with autotrophs evolving faster, excepting within the monocot pair.

2.6 Plastid-related GO Terms are Depleted in Large Clusters

The community cluster containing the highest number of orthogroups, cluster 19, shows no difference in mean branch length for its 840 orthologs either in student's t-test or pairwise comparisons. However, the enriched and depleted gene ontology terms for this network of covarying genes are informative about gene coevolution within this sample and may highlight issues with comparative genomics studies involving large numbers of parasitic plants. For statistically significant False-Discovery-Rate (FDR) corrected GO terms, plastid-related genes made up most of the depleted terms. Similar depletion of plastid genes was present in GO analysis of the second largest cluster 59, containing 666 orthogroups.

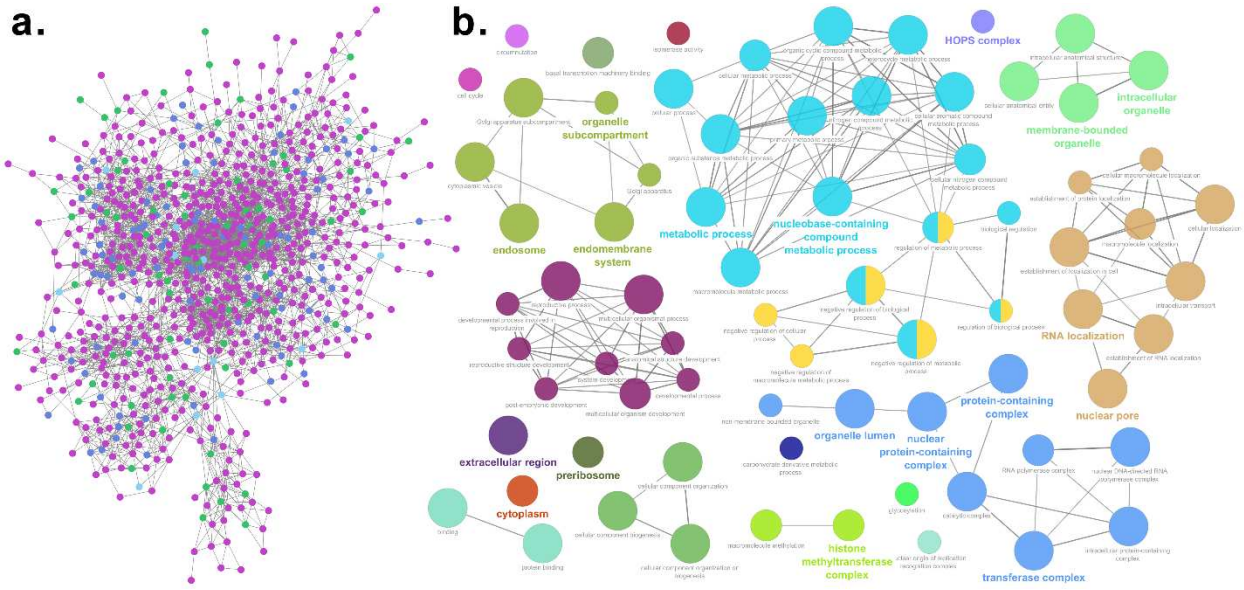

Figure 6

A) Community cluster 19 network visualization of ERC relationships between orthogroups as circular nodes. Node color: purple = not targeted to cellular compartment, green =

targeted to plastid, dark blue = targeted to mitochondrion, light blue = subcellular targeting unknown. B) Community cluster 19 GO enriched terms, grouped by relatedness of terms. Each color indicates terms fitting the same level 1 GO term, while nodes indicate more specific terms of higher level. Labels in color are significantly enriched or depleted after FDR correction. Circle size corresponds to Bonferroni-corrected p-value.

Enriched terms in cluster 19 skewed heavily towards mRNA metabolism, processing, and transport. This is the case for biological process, molecular function, and cellular component GO analyses of cluster 19. While this information has little bearing on comparisons between parasitic and autotrophic plant biology, the clustering of genes showing ERC alongside GO enrichment of terms related to transcription, post-transcriptional processing, mRNA cleavage, nuclear pore components, and mRNA localization is validating for this ERCnet run's ability to return biologically sensible results. Shared constraints on evolutionary rate on genes involved in each stage of mRNA function, from transcription to translation, could lead to evident coevolution between these genes within the 8 genomes studied. RNA exchange has been observed between haustorial parasitic plants and their hosts [34], so it may be worth further examining whether mRNA transport is similarly constrained in both groups, or whether its overrepresentation in covarying genes is driven by the huastorial parasites.

While the 2nd largest cluster 59 shares depletion of plastid genes with cluster 19, it differs in the enriched GO terms. Proteasome and ribosome-related genes were

---

[34] Shahid et al., "MicroRNAs from the Parasitic Plant Cuscuta Campestris Target Host Messenger RNAs"; Johnson and Axtell, "Small RNA Warfare."

overrepresented by 10-60 genes out of 666 total, while other GO terms were overrepresented by smaller numbers of 4-5 genes. Pairwise comparison of per-species branch length means in cluster 59 highlights faster evolving autotrophs in all lineages except the monocots. Proteasome gene coevolution in cluster 59 may be driven by shared evolutionary constraints on proteasome components across all plants, but proteases are involved in both cell wall degradation during haustorial attack and degradation of host defense proteins [35]. Parasite dependence on proteases for host invasion and manipulation may be driving differences in branch length for the eudicot plant species, while the mycoheterotrophic *Epipogium* lacks this constraint due to attacking its host indirectly through mycorrhizal networks.

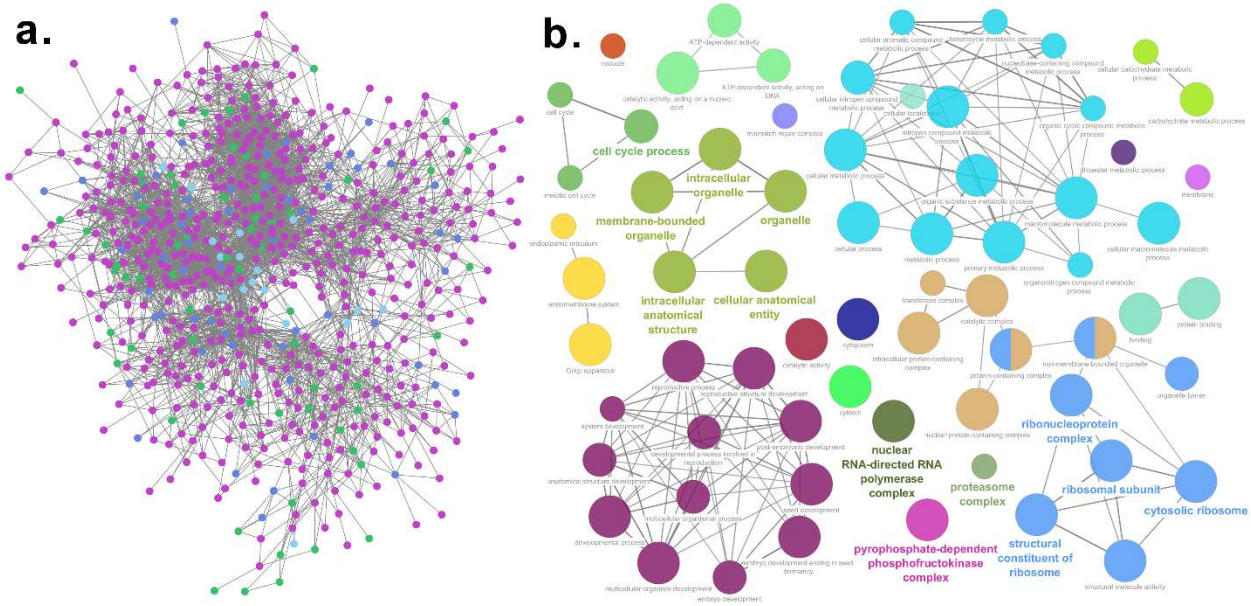

Figure 7

[35] Mitsumasu, Seto, and Yoshida, "Apoplastic Interactions between Plants and Plant Root Intruders."

A) Cluster 59 network visualization of ERC relationships between orthogroups. Node color: purple = not targeted to cellular compartment, green = targeted to plastid, dark blue = targeted to mitochondrion, light blue = subcellular targeting unknown. B) Community cluster 59 GO enriched terms, grouped by relatedness of terms. Each color indicates terms fitting the same level 1 GO term, while nodes indicate more specific terms of higher level. Labels in color are significantly enriched or depleted after FDR correction. Circle size corresponds to Bonferroni-corrected p-value.

2.7 Clusters Involving Faster Evolving Autotrophs were Enriched for Development and Immunity Genes

Cluster 23 stands out among the 30 largest clusters for having all four autotrophs evolving faster than their closest heterotrophic relative. Additionally, the number of GO enriched terms was higher than that observed in cluster of similar size. The preponderance of enriched terms involving plant development, signaling, and immune responses merited investigation, as these covarying genes may be under greater evolutionary constraint in parasitic plants relative to autotrophic relatives.

The most notable genes in this cluster are OBERON (OBE1 and OBE2) and TOPLESS (TPL). In *Arabidopsis*, OBE1 and OBE2 are redundant paralogous regulators of the key meristem genes WUSCHEL (WUS), CLAVATA3 (CLV3), PLETHORA1 (PLT1) and PLETHORA2 (PLT2.) OBERON is a homeobox gene critical for body plan organization, acting as a transcription factor for genes in the WUSCHEL-CLAVATA pathway governing the shoot

apical meristem, the pool of stem cells from which all aerial tissues in plants derive. At least one functional OBERON gene is required for meristem maintenance, and double KO mutants in *Arabidopsis* do not develop beyond the first pair of true leaves [36]. TPL acts as a corepressor in many *Arabidopsis* immune and hormone responses, in addition to its role in the WUS-CLV feedback loop, where it assists WUS in repressing stem cell differentiation in the shoot apical meristem [37]. TPL shares a hormonal response pathway with another gene in cluster 23, MYB DOMAIN PROTEIN 21 (MYB21), a MYB transcription factor involved in jasmonic acid response during stamen development [38].

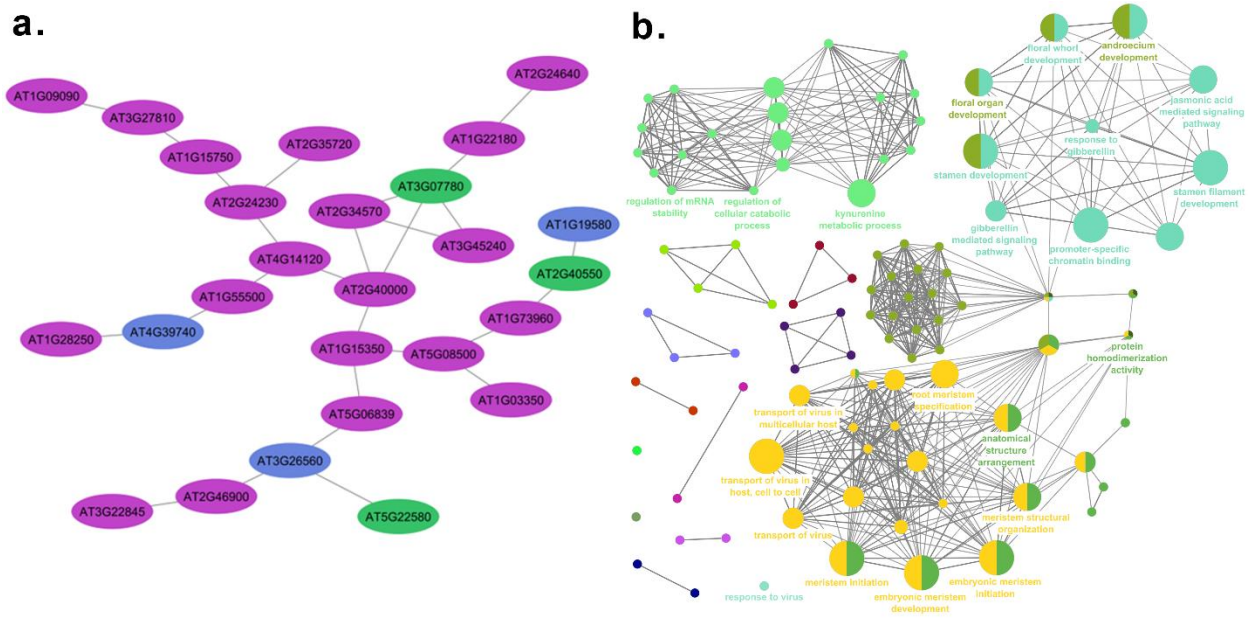

Figure 8

[36] Saiga et al., "The *Arabidopsis OBERON1* and *OBERON2* Genes Encode Plant Homeodomain Finger Proteins and Are Required for Apical Meristem Maintenance."
[37] Causier et al., "The TOPLESS Interactome."
[38] Sachdev et al., "The *Arabidopsis* ARID–HMG DNA-BINDING PROTEIN 15 Modulates Jasmonic Acid Signaling by Regulating MYC2 during Pollen Development."

A) Cluster 23 network visualization. Node color: purple = not targeted to cellular compartment, green = targeted to plastid, dark blue = targeted to mitochondrion. B) Community cluster 23 GO enriched terms, grouped by relatedness of terms. Each color indicates terms fitting the same level 1 GO term, while nodes indicate more specific terms of higher level. Labels in color are significantly enriched or depleted after FDR correction. Circle size corresponds to Bonferroni-corrected p-value.

The ERC found between developmental genes in this cluster, coupled with the slower rate of evolution in parasitic plants compared to their nearest autotrophic relative, may indicate that parasites share increased constraint on the evolution of these correlated development genes. Given that the body plans of parasitic plants differ greatly from those of autotrophs in the lack of leaf development or the absence of aerial structures beyond flowers, the homeobox gene OBERON may play an independently evolved, altered role in maintenance of the shoot apical meristem in plants lacking leaves.

Similarly to cluster 23, cluster 106 showed a difference in branch length between autotrophs and heterotrophs when compared pairwise. However, the GO enriched terms were less numerous, and few genes were characterized beyond their belonging to broad protein families or possessing common domains. The most notable gene in cluster 106 is embryo defective 3003 (EMB3003), which encodes a subunit of the plastid targeted pyruvate dehydrogenase in Arabidopsis [39]. Fatty acid biosynthesis is one of few conserved

---

[39] Yang et al., "Heterodera Avenae GLAND5 Effector Interacts With Pyruvate Dehydrogenase Subunit of Plant to Promote Nematode Parasitism."

functions of the plastid even in non-photosynthetic plants [40]. Within EMB3003's community cluster are several other genes involved in fatty acid metabolism. EMB3003's slower rate of evolution in heterotrophic plants and correlated evolutionary rate with other fatty acid metabolic pathways, may highlight heightened constraints on the few remaining plastid-targeted genes in autotrophic plants. Plastid fatty acids also play a role in plant immunity against viruses and nematodes [41].

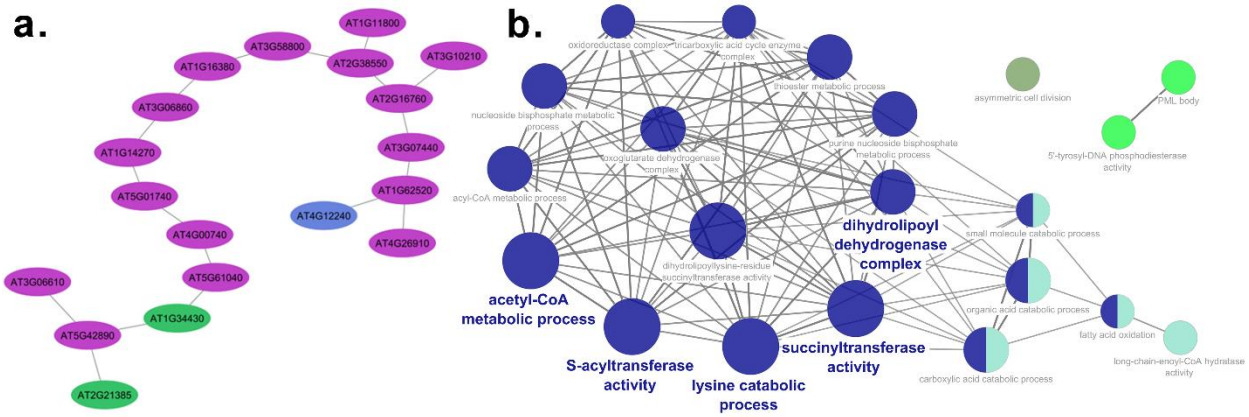

Figure 9

Cluster 106 network visualization. Node color: purple = not targeted to cellular compartment, green = targeted to plastid, dark blue = targeted to mitochondrion. B) Community cluster 106 GO enriched terms, grouped by relatedness of terms. Each color indicates terms fitting the same level 1 GO term, while nodes indicate more specific terms

[40] Bungard, "Photosynthetic Evolution in Parasitic Plants."
[41] Yang et al., "Heterodera Avenae GLAND5 Effector Interacts With Pyruvate Dehydrogenase Subunit of Plant to Promote Nematode Parasitism"; Chandra-Shekara et al., "Plastidial Fatty Acid Levels Regulate Resistance Gene-Dependent Defense Signaling in *Arabidopsis*."

of higher level. Labels in color are significantly enriched or depleted after FDR correction. Circle size corresponds to Bonferroni-corrected p-value.

3. DISCUSSION

3.1 Both Trophic Mode and Lineage Divergence Inform Gene Coevolution in Plants

The most exciting utility of ERC in identifying genes under differential evolutionary constraint in heterotrophic versus autotrophic plants lies in its ability to highlight gene co-option in derived heterotrophic trophic modes. However, our expectations of widespread differences in evolutionary constraint on covarying genes depending on trophic mode were not borne out by the data. Networks of genes showing evidence of covariance in evolutionary rate were not found to have significant differences in rate between autotrophs and heterotrophs. Some gene clusters showed differences when examined using the phylogenetic framework of paired autotroph and heterotroph species which underpinned the study. The clusters large enough for gene ontology analysis showed a mixture of sparse enrichment results (See Supplemental Results), depletion of plastid genes (Figures 7 and 8), and, in a few cases (Figures 9 and 10) networks of genes possibly implicated in the adoption of heterotrophic trophic modes and the loss of autotrophy.

While the original question is largely unanswered by this study, with a few notable exceptions, we did reach unexpected conclusions about gene coevolution in plants, especially the coevolutionary networks emerging from this species sample. Differences in branch length means across network clusters were far more strongly driven by the deep divergence between monocots and eudicots than by differences in trophic mode (Figure 5). A future expanded analysis including more heterotroph and autotroph species on both

sides of the monocot-eudicot divergence could alleviate this. On its own, this finding highlights differences in coevolutionary patterns between orthologous genes found in monocots and eudicots, which was not the subject of this study but emerged as a pattern regardless. This lineage divergence effect on ERC could be explored to further characterize the differences in gene interactions between these two groups and identify co-evolving gene pairs with divergent functions.

3.2 Limitations in the Use of Subcellular Targeting Assortativity to Evaluate ERC Result Quality

Subcellular targeting assortativity was used to evaluate the quality of ERC analysis methods. This assessment strategy led to disuse of ERCnet's Branch-by-branch evolutionary rate measurement feature [42] due to it resulting in smaller assortativity effect sizes with larger p-values. While the assortativity coefficient was significant and positive, its effect size was small (Table 2) even in the selected ERC method and p-value cutoff.

Additionally, the expectation of positive correlations between subcellular targeting and ERC between targeted genes may have been flawed for this species sample, granted the loss of many plastid-targeted genes in four of eight species [43]. Granted its curtailed usefulness in this sample, selecting ERC analysis methods based on effect size and significance may not have improved the quality of downstream analysis. However,

---

[42] Forsythe, Williams, and Sloan, "Genome-Wide Signatures of Plastid-Nuclear Coevolution Point to Repeated Perturbations of Plastid Proteostasis Systems across Angiosperms."

[43] Wicke et al., "Mechanistic Model of Evolutionary Rate Variation En Route to a Nonphotosynthetic Lifestyle in Plants."

including subcellular targeting in the dataset of ERC hits did assist in interpreting covarying networks, indicating the target compartments of subcommunities.

3.3 Strongest Hits Include Pairs with Known Co-Functional Genes and Pairs with No Reasonable Cause for Coevolution

The detection of covariance between two microtubule organization genes, AT3G27000 and AT2G18990, is a validating result [44]. While protein products of these genes likely do not directly interact, their involvement in microtubule nucleation and organization respectively fits the co-functionality interpretation of ERC [45].

Two pairs of cell-wall-related genes exhibited some of the most statistically significant ERC [46], in addition to having nearly all variance in one gene tree's branch length explained by that of the other (Table 1). This supports the reliability of our data in recovering gene pairs involved in known, shared pathways. Additionally, this finding is consistent with our expectations of finding co-functionality between covarying genes. Our ERC results and the existing body of gene functionality research may be reciprocally illuminating.

---

[44] Havelková et al., "Arp2/3 Complex Subunit ARPC2 Binds to Microtubules"; Mathioudakis et al., "A Thioredoxin Domain-Containing Protein Interacts with Pepino Mosaic Virus Triple Gene Block Protein 1."

[45] Little, Chikina, and Clark, "Evolutionary Rate Covariation Is a Reliable Predictor of Co-Functional Interactions but Not Necessarily Physical Interactions."

[46] Yang et al., "Two bHLH Transcription Factors, bHLH48 and bHLH60, Associate with Phytochrome Interacting Factor 7 to Regulate Hypocotyl Elongation in Arabidopsis"; De Godoy et al., "Galacturonosyltransferase 4 Silencing Alters Pectin Composition and Carbon Partitioning in Tomato"; Kumari et al., "IQ67 DOMAIN Proteins Facilitate Preprophase Band Formation and Division-Plane Orientation"; Lunin et al., "Molecular Mechanism of Polysaccharide Acetylation by the Arabidopsis Xylan *O* -Acetyltransferase XOAT1."

Most notably, granted the species sampling for this study, this covariation between cell-wall-related genes could indicate co-option of cell wall maintenance genes in routes of haustorial attack, which involve breakdown of cell walls at the host-parasite interface. [47]

Despite the aforementioned pairs being both validating and potentially relevant to the question at hand, three of the pairs with the strongest evidence of ERC (Table 1) share no reasonable functional association beyond existence in the same organism. These gene trees may share branch length covariance due to reasons other than interaction between gene products, or they may be false positives given the small sample size of species.

Notably absent from the strongest hits were Clp proteases. Prior ERC analysis of autotrophic plant genomes using ERCnet recovered Clp proteases as a network hub, covarying strongly with other genes. In this analysis, Clp was found in significant pairs, but the correlative p-values were not especially low, nor was Clp as interconnected as expected [48].

3.4 Network Hub Connectivity Identifies Key Regulator Genes

Beyond pairs of genes with strong covariation are network hubs consisting of single genes which covaried with a large number of other genes. These network hubs may represent key regulators of genes in their immediate neighborhoods, as was the case with

[47] Yoshida et al., "The Haustorium, a Specialized Invasive Organ in Parasitic Plants."
[48] Forsythe, Williams, and Sloan, "Genome-Wide Signatures of Plastid-Nuclear Coevolution Point to Repeated Perturbations of Plastid Proteostasis Systems across Angiosperms."

the experimentally confirmed regulator of adherens in Drosophila [49]. This key regulatory role seems to be the case with the most interconnected hub, CHR5, which is a chromatin remodeling gene that regulates the expression of numerous target genes by modifying chromatin accessibility to transcription machinery [50]. The CHR5 protein contains the largest number of mitogen-activated protein kinase (MAPK) docking sites found in *Arabidopsis* proteins, indicating its propensity for modification by other proteins in its protein-protein interaction network. A possible next step in this analysis would be examining whether the nodes connected to highly interconnected hubs like CHR5 include known interactors.

3.5 Community Cluster Structure and Size Reflect Patterns of Gene Covariation

At the broadest scale, community clusters represent networks of orthogroups exhibiting similar patterns of evolutionary rate covariation. While network edges represent significant ERC pairs, proximal but unconnected nodes may still share similar branch length patterns due to their mutual connections with a common intermediary node (Su et al. 2010). However, our expectation that community clusters would exhibit differences in branch length driven by trophic mode were not met. While all parasite species in this sample share similar levels of host-dependence, their adaptations to the parasitic lifestyle differ vastly. *Epipogium* differs from the other species in its evolution of mycoheterotrophy, whereas the other species are huastorial feeders. Meanwhile, Cuscuta lost its root system

[49] Raza et al., "Evolutionary Rate Covariation Analysis of E-Cadherin Identifies Raskol as a Regulator of Cell Adhesion and Actin Dynamics in Drosophila."
[50] Zou et al., "The Arabidopsis Chromatin-Remodeling Factor CHR5 Regulates Plant Immune Responses and Nucleosome Occupancy."

while maintaining shoots, whereas the remaining parasites lost their shoots with the exception of flowers. These four parasites may share more in common with their nearest autotrophic relatives than the more distantly related parasites happening to share trophic modes. The expected convergence of gene coevolution adaptations to the parasitic lifestyle may be overwhelmed by the differences in strategy between parasites.

The small species sample size and attendant lack of ortholog representation for each gene family may have led to lack of statistical power at a scale to obscure positive signals across the ERC network. The loss of genes, particularly in the holoparasites making up half of the sample, could have reduced the number of branch lengths available for comparison between trophic modes. This was evident in the case of *Epipogium*, a parasite whose proteome consisted of transcriptome data, its absent genes resulting in no evolutionary rate data to compare with other species (Figure 5). The lack of difference between evolutionary rates per trophic mode in community clusters may have been due to methodological flaws. Community clustering using the Glay algorithm may not have drawn boundaries between communities in a way that reflected biological interaction networks.

3.6 Underrepresentation of Plastid Genes in Large Clusters May Reflect Reduced Covariation or Statistical Power Limits

The largest covarying gene networks (Figure 6 and 8) mainly evidenced a depletion of chloroplast-related genes. Plastid-related genes being notably absent from the largest networks of genes connected by evolutionary rate covariation (ERC) may indicate their exclusion from interacting pathways. Alternatively, these terms may have been depleted in

the largest network clusters due to the heterotrophic plants lacking ERC between genes whose shared function was lost, rather than the genes themselves.

However, since half of the species included in this comparative genomics study were parasitic plants, their loss of plastid-related genes may have led to an underrepresentation of detected ERC between nuclear-encoded plastid-related genes. A simple majority (5/8) in species representation for both orthogroups was required for ERC analysis, to exclude branch length comparisons between even smaller sample sizes than the number of species included in this pilot study. Shared gene losses in the parasitic subset of genomes may have led to reductions in statistical power of correlation, resulting in lack of significant ERC. When all genes in significant hits were analyzed for GO term enrichment, there were no significant results after FDR correction, indicating that plastid-gene depletion mainly affected the two largest clusters.

While this may be evidence of a lack of covariation between nuclear-encoded plastid-related genes or their exclusion from co-functional clusters, it is also possible that gene loss in the sample drives GO depletion results. In this case, inclusion of just 5 orthologs in ERC analysis due to gene loss in heterotrophic plants could lead to deficits in statistical significance.

Future analyses should account for widespread gene loss in a substantial portion of sampled species as a potential source of deviations from expected GO-enrichment terms relative to a photosynthetic model species. Going forward, either demanding higher species representation than a simple majority or including more species in analysis could

curtail this effect of gene loss on ERC significance. Still, in the context of this study, the lack of ERC between plastid-related genes may indicate a relaxation of constraints on their co-functionality in the species sampled.

3.7 Slower Evolution of Developmental and Fatty Acid Metabolic Genes in Parasitic Plants Indicates Functional Divergence and Retention

GO enriched terms in cluster 23 showed the most promise for further investigation. Genes known to interact in shared developmental pathways in *Arabidopsis*, i.e. the *WUS/CLV3* feedback loop [51], exhibited ERC, alongside a collection of plant development and immunity genes (Figure 8). The slower evolutionary rate of these genes in heterotroph species compared to their nearest related autotroph species indicates their heightened evolutionary constraints in heterotrophic branches. This could indicate increased dependence on the pathways shared by these genes in heterotrophs for their survival, or more interestingly, co-option and the acquisition of derived function in heterotrophic plants.

The major body plan reorganizations of parasitic plants, including radical changes such as loss of the root system or aerial tissues aside from flowers, may be governed by altered interactions between GO enriched homeobox genes in this network and the basic organization of the plant [52]. Host developmental manipulation is another possible driver for derived function of plant development genes in heterotrophs. The possibility that these

[51] Somssich et al., "CLAVATA-WUSCHEL Signaling in the Shoot Meristem."
[52] Mukherjee, Brocchieri, and Burglin, "A Comprehensive Classification and Evolutionary Analysis of Plant Homeobox Genes."

genes are more critical for body plan reorganization or host manipulation in heterotrophs than in autotrophs warrants further exploration of possible derived gene product function.

Lastly, in cluster 106, genes involved in fatty acid biosynthesis, one of the few essential plastid functions retained in non-photosynthetic plants (Chandra-Shekara et al. 2007; S. Yang et al. 2019), evolved at a slower rate in heterotrophs (Figure 9). While this cluster may have been too small to overall have enriched plastid-related genes, the ERC between plastid-targeted fatty acid biosynthetic proteins and other proteins involved in fatty acid metabolism contravenes the effects seen in the largest clusters of fewer than expected plastid-related genes. While many plastid-associated functions are lost or unconstrained in heterotrophs, certain metabolic pathways remain under selective pressure, resembling previous findings of surprising gene retention and slow substitution rate in parasitic plants [53].

3.8 Next Steps: Expanding Comparative ERC Analysis of Parasitic Plants and Implementing Methodological Changes

Clusters 23 and 106 tell a more informative story of ERC within easily identifiable pathways due to their size being intermediate between clusters containing thousands of genes and the other extreme end of single pairs. Medium-sized clusters of genes in ERC may be beneficial to favor in expansions on this project, rather than the unwieldy largest clusters which were favored here. Since bounds on network clusters are down to

[53] Wicke et al., "Mechanistic Model of Evolutionary Rate Variation En Route to a Nonphotosynthetic Lifestyle in Plants."

somewhat arbitrary algorithmic decisions, drawing boundaries in favor of smaller interaction networks may yield more readable results.

More species should be included in the expansion of this project, in order to eliminate the problems arising from small sample size. Coupled with the addition of more species should be the inclusion of hemiparasitic plants, which are present in most clades containing holoparasites with the exception of Rafflesiaceae. This would introduce an intermediate state between autotrophy and heterotrophy and open up an avenue of study for genes whose co-evolution is of particular importance to plants retaining the capacity for both trophic modes.

As more plant genomic resources become available, these should be prioritized over transcriptomic datasets. The absence of transcripts in the *Epipogium* proteome compared to genomic proteomes based on gene models severely limited the number of overlapping branches available for ERC analysis. The inclusion of both genomic and transcriptomic datasets may have confounded gene loss and simple absence from the pool of mRNA present in the sampled tissue.

Increasing the species sample size and including hemiparasites could result in finding the expected differences in branch length for co-evolving genes between trophic modes, or alternately this could rule out insufficient statistical power as the factor underlying the absence of expected signal. This, combined with experimentation in analysis parameters and the splitting of network clusters, should more satisfactorily explore differences in ERC between trophic modes in plants.

4. METHODS

Plant genomes with gene models were pulled from NCBI repositories or requested from paper authors. (Table 3) Only gene model data from these genomes was included for ERC analysis, excluding non-protein-coding sequences. Protein-coding genes encoded in plant mitochondrial and chloroplast genomes were also excluded, since this study was aiming to identify nuclear-nuclear genetic interaction. One transcriptome, that of *Epipogium aphyllum*, was included in lieu of gene model sequence data. The early-branching *Amborella trichopoda*, sister species to all other angiosperms, was selected as the outgroup. Its gene models were not included in ERC analysis.

Table 3

Genomic and transcriptomic resources used as proteome datasets.

Species	Genome/Transcripto	Accession Numbers	Source Journals
Amborella	https://www.ncbi.nlm.	PRJNA10719	https://onlinelibrary.
Oryza sativa	https://www.ncbi.nlm.	PRJDB1747	https://pubmed.ncb
Epipogium aphyllum	https://www.ncbi.nlm.	PRJNA330626	https://bmcgenomi
Arabidopsis thaliana	https://www.ncbi.nlm.	PRJNA10719	
Sapria himalayana	https://www.ncbi.nlm.	PRJNA797720	https://bmcbiol.bio
Malania oleifera	https://www.ncbi.nlm.	PRJNA472200	https://www.nature.
Balanophora fungosa	https://db.cngb.org/se	CNA0050684	https://www.nature.

Solanum	https://www.ncbi.nlm.	PRJNA119	https://pubmed.ncb
Cuscuta australis	https://www.ncbi.nlm.	PRJNA394036	n/a

Gene models derived from reference genomes or transcripts derived from transcriptomes were standardized to amino acid sequences rather than nucleotide sequences. Where multiple isoforms of a single gene were found, the longest isoform was chosen as the representative so as to capture as much sequence information as possible, dropping all shorter isoforms from analysis.

Our species sample included genomes with varying levels of gene annotation, so relying on experimental validation of gene homology for all genes was not feasible. Instead, the comparative genomics software OrthoFinder was used to sort genes from all species into putative gene family trees based on sequence similarity [54]. While ideally the species tree generated by running OrthoFinder should be used to inform gene-tree-species-tree reconciliation in latter steps of running ERCnet, the species tree produced from our sample of protein coding genes did not reflect the consensus on deep divergence between these pairs of angiosperms, even after re-rooting the tree. A constraint tree was used instead, constructed to better reflect the consensus on angiosperm relationships [55] rather than relying on our limited dataset with sequences not selected for use in resolving deep phylogeny in plants based on putative orthologs.

[54] Emms and Kelly, "OrthoFinder," December 2019; Emms and Kelly, "OrthoFinder," August 6, 2015.
[55] One Thousand Plant Transcriptomes Initiative, "One Thousand Plant Transcriptomes and the Phylogenomics of Green Plants."

ERCnet is available on github at https://github.com/EvanForsythe/ERCnet. Orthogroups output by Orthofinder were input into ERCnet's phylogenomic analysis. Not all species in the sample had orthologs in every gene tree, while others had multiple paralogs in the same gene family. These branch lengths from the root of gene trees to tips of extant nodes were measured, and those branch length sums normalized by the genome-wide evolutionary rate to distinguish gene evolutionary rate from species-to-species differences in substitution rate.

The maximum allowable number of paralogs for each gene family tree was 3, and the per-species branch lengths were calculated using the mean of these three paralogs' branch lengths. The minimum number of overlapping branches between gene tree undergoing pairwise comparison was 5, a simple majority of ingroup species.

Evidence of covariation between gene pairs was calculated in an all-by-all pairwise manner, with correlation in branch lengths across species being measured for each unique pair of orthologs. Only correlations with Pearson p-values <0.01 were considered hits, and only genes involved in at least one significantly correlated pair were included in downstream analysis. A more stringent Pearson correlation p-value of <0.001 was used to narrow down the dataset used for network analysis, selected to allow a large number of genes into network analysis while preserving high assortativity and manageable network complexity.

Networks of ERC hits produced by ERNnet were visualized using Cytoscape [56]. To disentangle high network connectivity from a "hairball" including nearly all genes involved in ERC hits, Glay clustering [57] was done using the Girvan-Newman fast greedy algorithm. The resulting community clusters (Figure 3) were used to create lists of genes analyzed for GO enrichment. The PANTHER website and knowledge base [58] was used to compare network cluster lists to the background list of genes which went into ERC analysis.

[56] Shannon et al., "Cytoscape."
[57] Su et al., "GLay."
[58] Thomas et al., "<span Style="font-Variant."

5. REFERENCES

Amicone, Massimo, Vítor Borges, Maria João Alves, Joana Isidro, Líbia Zé-Zé, Sílvia Duarte, Luís Vieira, Raquel Guiomar, João Paulo Gomes, and Isabel Gordo. "Mutation Rate of SARS-CoV-2 and Emergence of Mutators during Experimental Evolution." *Evolution, Medicine, and Public Health* 10, no. 1 (January 5, 2022): 142–55. https://doi.org/10.1093/emph/eoac010.

Betts, A., C. Gray, M. Zelek, R. C. MacLean, and K. C. King. "High Parasite Diversity Accelerates Host Adaptation and Diversification." *Science* 360, no. 6391 (May 25, 2018): 907–11. https://doi.org/10.1126/science.aam9974.

Bromham, Lindell, Peter F Cowman, and Robert Lanfear. "Parasitic Plants Have Increased Rates of Molecular Evolution across All Three Genomes." *BMC Evolutionary Biology* 13, no. 1 (2013): 126. https://doi.org/10.1186/1471-2148-13-126.

Bungard, Ralph A. "Photosynthetic Evolution in Parasitic Plants: Insight from the Chloroplast Genome." *BioEssays* 26, no. 3 (March 2004): 235–47. https://doi.org/10.1002/bies.10405.

Cao, Lianqi, Wei Lu, Analucia Mata, Katsuyoshi Nishinari, and Yapeng Fang. "Egg-Box Model-Based Gelation of Alginate and Pectin: A Review." *Carbohydrate Polymers* 242 (August 2020): 116389. https://doi.org/10.1016/j.carbpol.2020.116389.

Causier, Barry, Mary Ashworth, Wenjia Guo, and Brendan Davies. "The TOPLESS Interactome: A Framework for Gene Repression in Arabidopsis." *Plant Physiology* 158, no. 1 (January 2, 2012): 423–38. https://doi.org/10.1104/pp.111.186999.

Chandra-Shekara, A. C., Srivathsa C. Venugopal, Subhankar Roy Barman, Aardra Kachroo, and Pradeep Kachroo. "Plastidial Fatty Acid Levels Regulate Resistance Gene-Dependent Defense Signaling in *Arabidopsis*." *Proceedings of the National Academy of Sciences* 104, no. 17 (April 24, 2007): 7277–82. https://doi.org/10.1073/pnas.0609259104.

Cheetham, Seth W., Geoffrey J. Faulkner, and Marcel E. Dinger. "Overcoming Challenges and Dogmas to Understand the Functions of Pseudogenes." *Nature Reviews Genetics* 21, no. 3 (March 2020): 191–201. https://doi.org/10.1038/s41576-019-0196-1.

Clark, Nathan L., Eric Alani, and Charles F. Aquadro. "Evolutionary Rate Covariation Reveals Shared Functionality and Coexpression of Genes." *Genome Research* 22, no. 4 (April 2012): 714–20. https://doi.org/10.1101/gr.132647.111.

Clarke, Christopher R., Michael P. Timko, John I. Yoder, Michael J. Axtell, and James H. Westwood. "Molecular Dialog Between Parasitic Plants and Their Hosts." *Annual Review of Phytopathology* 57, no. 1 (August 25, 2019): 279–99. https://doi.org/10.1146/annurev-phyto-082718-100043.

Cui, Songkui, Tomoya Kubota, Tomoaki Nishiyama, Juliane K. Ishida, Shuji Shigenobu, Tomoko F. Shibata, Atsushi Toyoda, Mitsuyasu Hasebe, Ken Shirasu, and Satoko Yoshida. "Ethylene Signaling Mediates Host Invasion by Parasitic Plants." *Science Advances* 6, no. 44 (October 30, 2020): eabc2385. https://doi.org/10.1126/sciadv.abc2385.

De Godoy, Fabiana, Luisa Bermúdez, Bruno Silvestre Lira, Amanda Pereira De Souza, Paula Elbl, Diego Demarco, Saleh Alseekh, et al. "Galacturonosyltransferase 4 Silencing Alters Pectin Composition and Carbon Partitioning in Tomato." *Journal of Experimental Botany* 64, no. 8 (May 2013): 2449–66. https://doi.org/10.1093/jxb/ert106.

De Juan, David, Florencio Pazos, and Alfonso Valencia. "Emerging Methods in Protein Co-Evolution." *Nature Reviews Genetics* 14, no. 4 (April 2013): 249–61. https://doi.org/10.1038/nrg3414.

Drummond, D. Allan, Jesse D. Bloom, Christoph Adami, Claus O. Wilke, and Frances H. Arnold. "Why Highly Expressed Proteins Evolve Slowly." *Proceedings of the National Academy of Sciences* 102, no. 40 (October 4, 2005): 14338–43. https://doi.org/10.1073/pnas.0504070102.

Duan, Kunyu, Hui Fu, Dongming Fang, Kaimeng Wang, Wen Zhang, Huan Liu, Sunil Kumar Sahu, and Xiaoli Chen. "Genome-Wide Analysis of the MADS-Box Gene Family in Holoparasitic Plants (Balanophora Subcupularis and Balanophora Fungosa Var. Globosa)." *Frontiers in Plant Science* 13 (May 31, 2022): 846697. https://doi.org/10.3389/fpls.2022.846697.

Echave, Julian, Stephanie J. Spielman, and Claus O. Wilke. "Causes of Evolutionary Rate Variation among Protein Sites." *Nature Reviews Genetics* 17, no. 2 (February 2016): 109–21. https://doi.org/10.1038/nrg.2015.18.

Emms, David M., and Steven Kelly. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20, no. 1 (December 2019): 238. https://doi.org/10.1186/s13059-019-1832-y.

———. "OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy." *Genome Biology* 16, no. 1 (August 6, 2015): 157. https://doi.org/10.1186/s13059-015-0721-2.

Forsythe, Evan S, Tony C Gatts, Linnea E Lane, Chris deRoux, Monica Berggren, Elizabeth A Rehmann, Emily N Zak, Trinity Bartel, Luna L'Argent, and Daniel B Sloan. "Phylogenomic Prediction of Interaction Networks in the Presence of Gene Duplication," August 8, 2024. https://doi.org/10.1101/2024.08.06.606904.

Forsythe, Evan S, Alissa M Williams, and Daniel B Sloan. "Genome-Wide Signatures of Plastid-Nuclear Coevolution Point to Repeated Perturbations of Plastid Proteostasis Systems across Angiosperms." *The Plant Cell* 33, no. 4 (May 31, 2021): 980–97. https://doi.org/10.1093/plcell/koab021.

Grosse-Veldmann, Bernadette. "Amborella Trichopoda Cultivation of the Most Ancestral Angiosperm in Botanic Gardens." *Sibbaldia*, no. 9 (n.d.). https://doi.org/Amborella trichopoda Cultivation of the most ancestral angiosperm in botanic gardens.

Havelková, Lenka, Gitanjali Nanda, Jan Martinek, Erica Bellinvia, Lenka Sikorová, Kateřina Šlajcherová, Daniela Seifertová, et al. "Arp2/3 Complex Subunit ARPC2 Binds to Microtubules." *Plant Science* 241 (December 2015): 96–108. https://doi.org/10.1016/j.plantsci.2015.10.001.

Ichihashi, Yasunori, Miyako Kusano, Makoto Kobayashi, Kenji Suetsugu, Satoko Yoshida, Takanori Wakatake, Kie Kumaishi, Arisa Shibata, Kazuki Saito, and Ken Shirasu. "Transcriptomic and Metabolomic Reprogramming from Roots to Haustoria in the Parasitic Plant, Thesium Chinense." *Plant and Cell Physiology* 59, no. 4 (April 1, 2018): 729–38. https://doi.org/10.1093/pcp/pcx200.

Jackson, Andrew P. "The Evolution of Parasite Genomes and the Origins of Parasitism." *Parasitology* 142, no. S1 (February 2015): S1–5. https://doi.org/10.1017/S0031182014001516.

Jasinski, Sophie, Fabien Chardon, Nathalie Nesi, Alain Lécureuil, and Philippe Guerche. "Improving Seed Oil and Protein Content in *Brassicaceae* : Some New Genetic Insights from *Arabidopsis Thaliana*." *OCL* 25, no. 6 (November 2018): D603. https://doi.org/10.1051/ocl/2018047.

Johnson, Nathan R, and Michael J Axtell. "Small RNA Warfare: Exploring Origins and Function of Trans-Species microRNAs from the Parasitic Plant Cuscuta." *Current Opinion in Plant Biology* 50 (August 2019): 76–81. https://doi.org/10.1016/j.pbi.2019.03.014.

Kumari, Pratibha, Pradeep Dahiya, Pantelis Livanos, Luise Zergiebel, Malte Kölling, Yvonne Poeschl, Gina Stamm, et al. "IQ67 DOMAIN Proteins Facilitate Preprophase Band Formation and

Division-Plane Orientation." *Nature Plants* 7, no. 6 (May 24, 2021): 739–47. https://doi.org/10.1038/s41477-021-00923-z.

Little, Jordan, Maria Chikina, and Nathan L Clark. "Evolutionary Rate Covariation Is a Reliable Predictor of Co-Functional Interactions but Not Necessarily Physical Interactions." Edited by Detlef Weigel. *eLife* 12 (February 28, 2024): RP93333. https://doi.org/10.7554/eLife.93333.

Lunin, Vladimir V., Hsin-Tzu Wang, Vivek S. Bharadwaj, Markus Alahuhta, Maria J. Peña, Jeong-Yeh Yang, Stephanie A. Archer-Hartmann, et al. "Molecular Mechanism of Polysaccharide Acetylation by the Arabidopsis Xylan *O* -Acetyltransferase XOAT1." *The Plant Cell* 32, no. 7 (July 2020): 2367–82. https://doi.org/10.1105/tpc.20.00028.

Mathioudakis, Matthaios M., Souheyla Khechmar, Carolyn A. Owen, Vicente Medina, Karima Ben Mansour, Weronika Tomaszewska, Theodore Spanos, Panagiotis F. Sarris, and Ioannis C. Livieratos. "A Thioredoxin Domain-Containing Protein Interacts with Pepino Mosaic Virus Triple Gene Block Protein 1." *International Journal of Molecular Sciences* 19, no. 12 (November 25, 2018): 3747. https://doi.org/10.3390/ijms19123747.

McNeal, Joel R., Jennifer V. Kuehl, Jeffrey L. Boore, and Claude W. de Pamphilis. "Complete Plastid Genome Sequences Suggest Strong Selection for Retention of Photosynthetic Genes in the Parasitic Plant Genus Cuscuta." *BMC Plant Biology* 7, no. 1 (October 24, 2007): 57. https://doi.org/10.1186/1471-2229-7-57.

Merckx, Vincent, ed. *Mycoheterotrophy: The Biology of Plants Living on Fungi*. New York, NY: Springer New York, 2013. https://doi.org/10.1007/978-1-4614-5209-6.

Mitsumasu, Kanako, Yoshiya Seto, and Satoko Yoshida. "Apoplastic Interactions between Plants and Plant Root Intruders." *Frontiers in Plant Science* 6 (August 14, 2015). https://doi.org/10.3389/fpls.2015.00617.

Mukherjee, K., L. Brocchieri, and T. R. Burglin. "A Comprehensive Classification and Evolutionary Analysis of Plant Homeobox Genes." *Molecular Biology and Evolution* 26, no. 12 (December 1, 2009): 2775–94. https://doi.org/10.1093/molbev/msp201.

Mursyidah, Anwarali-Khan, Mohamad Hafizzudin-Fedeli, Nor Azlan Nor Muhammad, A Latiff, Mohd Firdaus-Raih, and Kiew-Lian Wan. "Dissecting the Biology of Rafflesia Species: Current Progress and Future Directions Made Possible with High-Throughput Sequencing Data." *Plant and Cell Physiology* 64, no. 4 (April 1, 2023): 368–77. https://doi.org/10.1093/pcp/pcad004.

Nobuyuki TANAKA, Hidetoshi NAGAMASU, Shuichiro TAGANE, Mu Mu AUNG, Aung Khaing WIN, and Phyu Phyu HNIN. "Contributions to the Flora of Myanmar IV: A New Species and a Newly Recorded Taxon of the Genus Sapria (Rafflesiaceae)." *TAIWANIA* 64, no. 4 (December 1, 2019). https://doi.org/10.6165/tai.2019.64.357.

One Thousand Plant Transcriptomes Initiative. "One Thousand Plant Transcriptomes and the Phylogenomics of Green Plants." *Nature* 574, no. 7780 (October 31, 2019): 679–85. https://doi.org/10.1038/s41586-019-1693-2.

Rayapuram, Naganand, Mai Jarad, Hanna M. Alhoraibi, Jean Bigeard, Aala A. Abulfaraj, Ronny Völz, Kiruthiga Gayathri Mariappan, et al. "Chromatin Phosphoproteomics Unravels a Function for AT-Hook Motif Nuclear Localized Protein AHL13 in PAMP-Triggered Immunity." *Proceedings of the National Academy of Sciences* 118, no. 3 (January 19, 2021): e2004670118. https://doi.org/10.1073/pnas.2004670118.

Raza, Qanber, Jae Young Choi, Yang Li, Roisin M. O'Dowd, Simon C. Watkins, Maria Chikina, Yang Hong, Nathan L. Clark, and Adam V. Kwiatkowski. "Evolutionary Rate Covariation Analysis of E-Cadherin Identifies Raskol as a Regulator of Cell Adhesion and Actin Dynamics in

Drosophila." *PLOS Genetics* 15, no. 2 (February 14, 2019): e1007720. https://doi.org/10.1371/journal.pgen.1007720.
Rose, Harry. *Cuscuta Australis Flower5*. January 26, 2013. Photograph. https://www.flickr.com/photos/macleaygrassman/8418559318/in/photostream/.
Ryu, Taewoo, Loqmane Seridi, and Timothy Ravasi. "The Evolution of Ultraconserved Elements with Different Phylogenetic Origins." *BMC Evolutionary Biology* 12, no. 1 (December 5, 2012): 236. https://doi.org/10.1186/1471-2148-12-236.
Sachdev, Sonal, Ruby Biswas, Adrita Roy, Ayantika Nandi, Vishal Roy, Sabini Basu, and Shubho Chaudhuri. "The *Arabidopsis* ARID–HMG DNA-BINDING PROTEIN 15 Modulates Jasmonic Acid Signaling by Regulating MYC2 during Pollen Development." *Plant Physiology* 196, no. 2 (October 1, 2024): 996–1013. https://doi.org/10.1093/plphys/kiae355.
Saiga, Shunsuke, Chihiro Furumizu, Ryusuke Yokoyama, Tetsuya Kurata, Shusei Sato, Tomohiko Kato, Satoshi Tabata, Mitsuhiro Suzuki, and Yoshibumi Komeda. "The *Arabidopsis OBERON1* and *OBERON2* Genes Encode Plant Homeodomain Finger Proteins and Are Required for Apical Meristem Maintenance." *Development* 135, no. 10 (May 15, 2008): 1751–59. https://doi.org/10.1242/dev.014993.
Shahid, Saima, Gunjune Kim, Nathan R. Johnson, Eric Wafula, Feng Wang, Ceyda Coruh, Vivian Bernal-Galeano, et al. "MicroRNAs from the Parasitic Plant Cuscuta Campestris Target Host Messenger RNAs." *Nature* 553, no. 7686 (January 2018): 82–85. https://doi.org/10.1038/nature25027.
Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome Research* 13, no. 11 (November 2003): 2498–2504. https://doi.org/10.1101/gr.1239303.
Somssich, Marc, Byoung Il Je, Rüdiger Simon, and David Jackson. "CLAVATA-WUSCHEL Signaling in the Shoot Meristem." *Development* 143, no. 18 (September 15, 2016): 3238–48. https://doi.org/10.1242/dev.133645.
Su, Gang, Allan Kuchinsky, John H. Morris, David J. States, and Fan Meng. "GLay: Community Structure Analysis of Biological Networks." *Bioinformatics* 26, no. 24 (December 15, 2010): 3135–37. https://doi.org/10.1093/bioinformatics/btq596.
Thomas, Paul D., Dustin Ebert, Anushya Muruganujan, Tremayne Mushayahama, Laurent-Philippe Albou, and Huaiyu Mi. "PANTHER : Making Genome-scale Phylogenetics Accessible to All." *Protein Science* 31, no. 1 (January 2022): 8–22. https://doi.org/10.1002/pro.4218.
Volkova, Nadezda V., Bettina Meier, Víctor González-Huici, Simone Bertolini, Santiago Gonzalez, Harald Vöhringer, Federico Abascal, et al. "Mutational Signatures Are Jointly Shaped by DNA Damage and Repair." *Nature Communications* 11, no. 1 (May 1, 2020): 2169. https://doi.org/10.1038/s41467-020-15912-7.
Voskarides, Konstantinos, Harsh Dweep, and Charalambos Chrysostomou. "Evidence That DNA Repair Genes, a Family of Tumor Suppressor Genes, Are Associated with Evolution Rate and Size of Genomes." *Human Genomics* 13, no. 1 (June 7, 2019): 26. https://doi.org/10.1186/s40246-019-0210-x.
Westwood, James H., John I. Yoder, Michael P. Timko, and Claude W. dePamphilis. "The Evolution of Parasitism in Plants." *Trends in Plant Science* 15, no. 4 (April 1, 2010): 227–35. https://doi.org/10.1016/j.tplants.2010.01.004.
Wicke, Susann, Kai F. Müller, Claude W. dePamphilis, Dietmar Quandt, Sidonie Bellot, and Gerald M. Schneeweiss. "Mechanistic Model of Evolutionary Rate Variation En Route to a Nonphotosynthetic Lifestyle in Plants." *Proceedings of the National Academy of Sciences* 113, no. 32 (August 9, 2016): 9045–50. https://doi.org/10.1073/pnas.1607576113.

Wicke, Susann, and Julia Naumann. “Molecular Evolution of Plastid Genomes in Parasitic Flowering Plants.” In *Advances in Botanical Research*, 85:315–47. Elsevier, 2018. https://doi.org/10.1016/bs.abr.2017.11.014.

Yang, Chuanwei, Sha Huang, Yue Zeng, Chang Liu, Qinyi Ma, Jose Pruneda-Paz, Steve A. Kay, and Lin Li. “Two bHLH Transcription Factors, bHLH48 and bHLH60, Associate with Phytochrome Interacting Factor 7 to Regulate Hypocotyl Elongation in Arabidopsis.” *Cell Reports* 35, no. 5 (May 2021): 109054. https://doi.org/10.1016/j.celrep.2021.109054.

Yang, Shanshan, Lingling Pan, Yongpan Chen, Dan Yang, Qian Liu, and Heng Jian. “Heterodera Avenae GLAND5 Effector Interacts With Pyruvate Dehydrogenase Subunit of Plant to Promote Nematode Parasitism.” *Frontiers in Microbiology* 10 (June 4, 2019): 1241. https://doi.org/10.3389/fmicb.2019.01241.

Yang, Zhenzhen, Eric K. Wafula, Loren A. Honaas, Huiting Zhang, Malay Das, Monica Fernandez-Aparicio, Kan Huang, et al. “Comparative Transcriptome Analyses Reveal Core Parasitism Genes and Suggest Gene Duplication and Repurposing as Sources of Structural Novelty.” *Molecular Biology and Evolution* 32, no. 3 (March 1, 2015): 767–90. https://doi.org/10.1093/molbev/msu343.

Yoshida, Satoko, Songkui Cui, Yasunori Ichihashi, and Ken Shirasu. “The Haustorium, a Specialized Invasive Organ in Parasitic Plants.” *Annual Review of Plant Biology* 67, no. 1 (April 29, 2016): 643–67. https://doi.org/10.1146/annurev-arplant-043015-111702.

Zięba, Antoni, Sławomir Wróbel, Paweł Kauzal, Anna Delimat, Anna Maria Ociepa, Maciej Kozak, Katarzyna Kozłowska-Kozak, et al. “Przyczynki do flory Tatrzańskiego Parku Narodowego.” *Fragmenta Floristica et Geobotanica Polonica*, December 2020, 379–94. https://doi.org/10.35535/ffgp-2020-0022.

Zou, Baohong, Qi Sun, Wenli Zhang, Yuan Ding, Dong-Lei Yang, Zhenying Shi, and Jian Hua. “The Arabidopsis Chromatin-Remodeling Factor CHR5 Regulates Plant Immune Responses and Nucleosome Occupancy.” *Plant and Cell Physiology* 58, no. 12 (December 1, 2017): 2202–16. https://doi.org/10.1093/pcp/pcx155.

阿橋花譜 HQ Flower Guide. *稻 Oryza Sativa [香港西貢獅子會自然教育中心 Saikung, Hong Kong]*. July 6, 2013. https://www.flickr.com/photos/nhq9801/9222670472/in/photostream/.