

Network analysis of complex trait evolution: C4 photosynthesis

Elli Cryan^{*+x}, Daniel Kliebenstein^{*}, Jeffrey Ross-Ibarra⁺

University of California, Davis: Departments of ^{*}Plant Sciences ⁺Evolution and Ecology; ^xPlant Biology Graduate Group

Genes evolve as a network

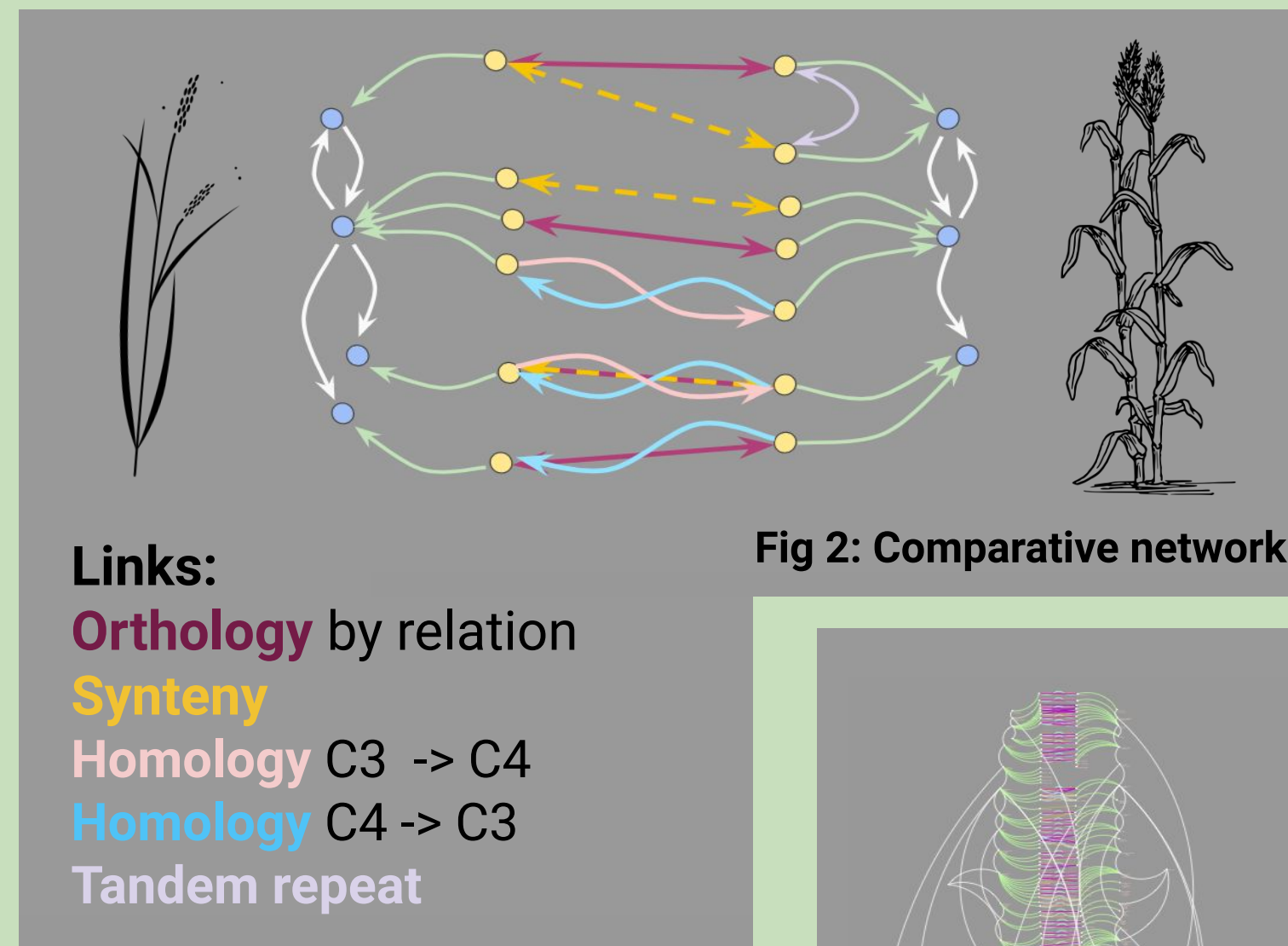
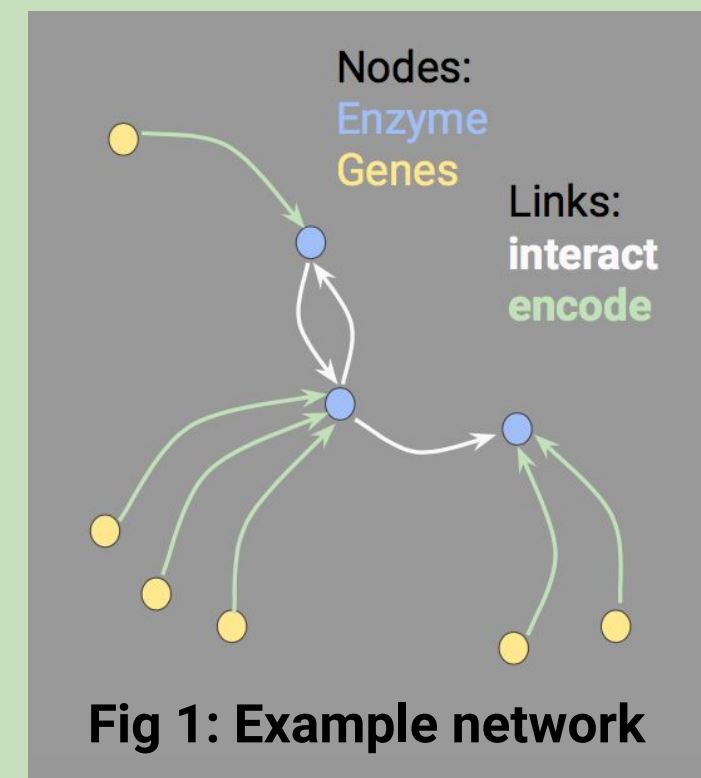
A complex trait can be represented by a network of interactions between metabolites and proteins underpinned by a **network of genes that encode the proteins**. Studying these genes in a network context allows complex trait evolution to be represented by both changes in network architecture and gene sequences across evolutionary time. Network architecture changes include the rewiring of links between nodes as genes are exapted to new functions and the creation of new nodes in cases of gene duplications. Additionally, network architecture properties can give insight into the 'choice' of which gene in a family will be exapted. Established comparative measures of gene sequence evolution like **Dn/Ds** can be mapped onto the network to provide a more holistic view of complex trait evolution.

C4 photosynthesis

An important complex trait in plants is C4 photosynthesis, which is an adaptation that reduces the incidence of photorespiration by partitioning gas exchange and carbon capture to different cells. C4 photosynthesis has evolved from a C3 photosynthesis ancestral state in more than sixty documented independent events. By **comparing the gene networks of pairs of species on either side of independent C4 evolution events**, we hope to better understand the process of complex trait evolution. Each event is considered a single biological replicate. We first compare two grass crops with abundant genomic resources: C4 sorghum (*Sorghum bicolor*) and C3 rice (*Oryza sativa japonica*) (Fig 3). All enzymes involved in the C4 network predate the evolution of the trait and are present in C3-type plants.

Model structure

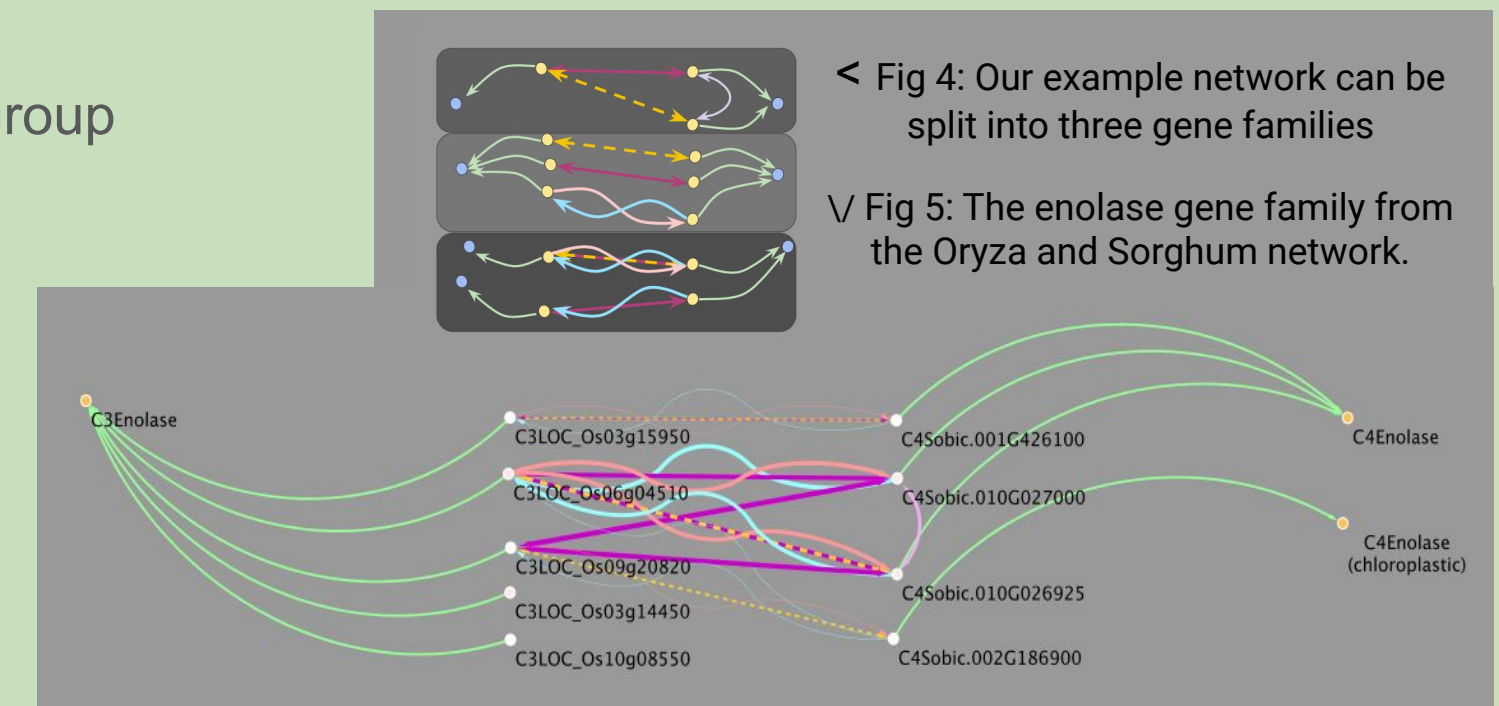
Metabolic pathways are typically depicted as metabolite nodes linked by enzymes. In contrast, we use enzymes as nodes linked by interaction (direct and indirect) between enzymes through the passing of metabolites (Fig 1). We add functional gene nodes linked to enzyme nodes by transcription.



Comparing networks

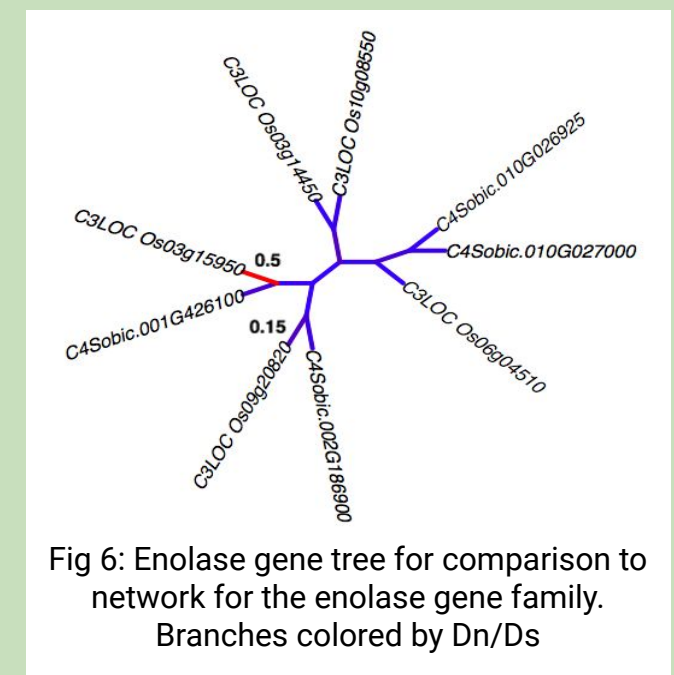
Because we include genes in our network, **we can align two functional networks, which represent two sides of an evolutionary event**, by orthology, synteny, homology, and tandem duplication links (Fig 2). These links can be weighted by Dn/Ds values to provide information about selection on genes.

Fig 3: Full network *Oryza* & *Sorghum*



Gene families

Gene nodes and the links between them form connected components. These components are subnetworks that represent gene families (Fig 4). **Subnetworks provide details about the properties of genes that are exapted or rewired into the C4 network.**



Bioinformatics Methods

Metabolic pathway (protein nodes and interaction links) from literature review. Genes (gene nodes and encoding links) identified by EC number, filtered by functionality. Synteny links pulled from CoGe, orthology and homology links pulled from Phytozome, tandem repeat links from genomic location. Codon-aware alignment with muscle via TranslatorX with predicted transit peptides (TransitP) removed. Dn/Ds values calculated with PAML, input trees from RAXML.

Conclusions

Genes do not evolve independently. Instead genes evolve as networks that contribute to phenotypes that impact fitness. **Network modeling of proteins, genes, and their interactions on both sides of an evolutionary event uncovers details of gene exaptation**, or the network rewiring that comprises the new complex trait. After framing other C4 events in this network context, network measures like centrality and betweenness will be compared to the incidence of exaptation in gene family subnetworks and in the larger network.

Funding

UC Davis: Plant Sciences GSR, Jastro Award, Graduate Studies Computational GSR