Multi-enzyme Phylogenies of Anaplerotic CO_2 Fixation and Glyoxylate-cycle Enzymes Illuminate Mixed Evolutionary Origins and Complex Carbon Assimilation Capacity in *Euglena gracilis*

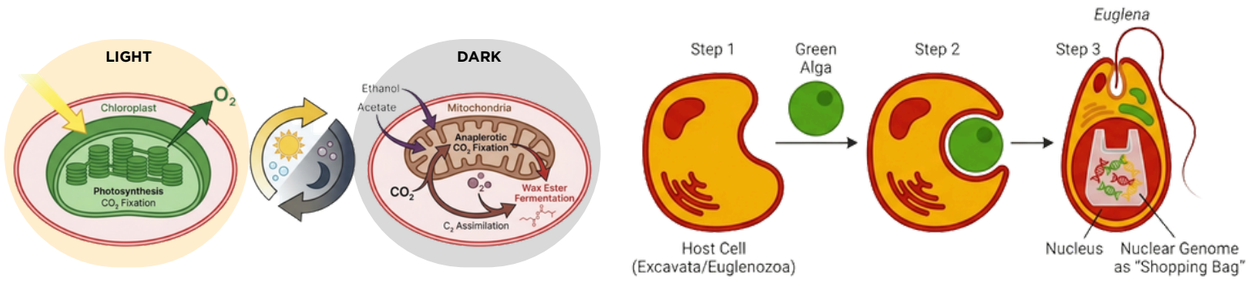
Nattharat Suchatpong [4], Patcharaporn Wongkham [4], Pimmada Thongprasri [4], Klatawan Srisawang [4], Chinnapat Buntan [4], Sahutchai Inwongwan [1,2,3]
[1] Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
[2] Algal and Cyanobacterial Research Laboratory, Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
[3] Center of Excellence in Microbial Diversity and Sustainable Utilization, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
[4] Data Science Research Center, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

ABSTRACT

Euglena gracilis exhibits exceptional carbon metabolic flexibility, including inorganic carbon assimilation under heterotrophic and oxygen limited conditions and efficient utilization of C2 substrates. To investigate how these capabilities may have been evolutionarily assembled, we performed coordinated phylogenetic analyses of enzymes central to anaplerotic CO_2 fixation and C2 metabolism: phosphoenolpyruvate carboxylase (PEPC), phosphoenolpyruvate carboxykinase (PEPCK), pyruvate carboxylase (PC), malic enzyme (ME), and glyoxylate cycle components (isocitrate lyase, ICL; malate synthase, MS), alongside the characteristic *Euglena* malate synthase isocitrate lyase fusion (MS ICL; EgGCE). Candidate sequences were retrieved from de novo transcriptome annotation and retained using similarity screening (at least 50 percent query coverage and at least 50 percent identity) with domain architecture verification, then analysed with curated reference homologs spanning major eukaryotic lineages. Nucleotide sequences were aligned with MAFFT, curated to retain phylogenetically informative regions, and used to infer maximum likelihood trees within the NGPhylogeny.fr framework; trees were annotated and visualized in iTOL using consistent group colouring (*Euglena* in red, plants in green, protists in yellow, animals in purple, fungi in brown). Across enzyme families, *E. gracilis* candidates exhibited distinct evolutionary topologies. PC showed the most compact pattern, comprising a closely related pair (Clusters 3571.40347 and 3571.40844) plus an additional candidate (Cluster 3571.39909). In contrast, PEPCK and ME were more heterogeneous. In the PEPCK phylogeny, one candidate (Cluster 3571.40400) grouped with an amoebozoan reference, whereas three candidates (Clusters 3571.37627, 3571.39711, and 3571.42343) formed a *Euglena* clade positioned with kinetoplastid sequences, indicating at least two divergent PEPCK lineages. PEPC candidates separated into two placements, one among algal and protist references and another within a stramenopile enriched grouping. ME exhibited the greatest diversification, with seven *Euglena* candidates distributed across multiple regions of the tree, including a compact clade shared with an additional euglenid reference and several distinct placements near protist dominated branches. For glyoxylate cycle functions, the ICL candidate (Cluster 3571.39868) showed a distinct placement, while MS candidates split into plant or algal associated versus protist associated affinities. Together, these results provide a reproducible comparative phylogenetic resource for central carbon assimilation enzymes in *E. gracilis* and generate testable hypotheses linking enzyme family diversification and mixed evolutionary histories to heterotrophic CO_2 fixation and C2 metabolic capacity.

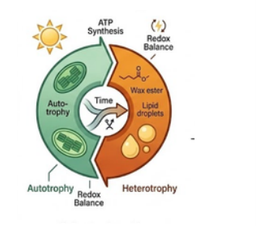
INTRODUCTION

- **A "Metabolic Shapeshifter":** *Euglena* occupies a unique evolutionary position, easily transitioning between phototrophy and heterotrophy.
- **The "Shopping Bag" Model:** The nuclear genome of *E. gracilis* has a complex history, receiving gene contributions from multiple photosynthetic lineages via secondary endosymbiosis, suggesting mixed ancestry for its metabolic functions.

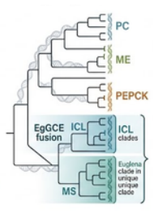


- **The Dark Carbon Challenge:** Under dark and oxygen-limited conditions, *Euglena* utilizes non-Rubisco, anaplerotic CO_2 assimilation routes to feed into central carbon intermediates, a crucial step for anaerobic survival like wax ester fermentation.
- **Core Question:** Do the enzyme families enabling these complex carbon capture and C2 assimilation processes share a single, coherent origin, or do they exhibit mosaic evolutionary signatures?

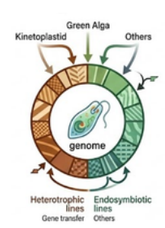
OBJECTIVES



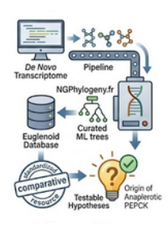
Investigate the evolutionary assembly of metabolic flexibility



Map the phylogenies of key carbon assimilation enzymes



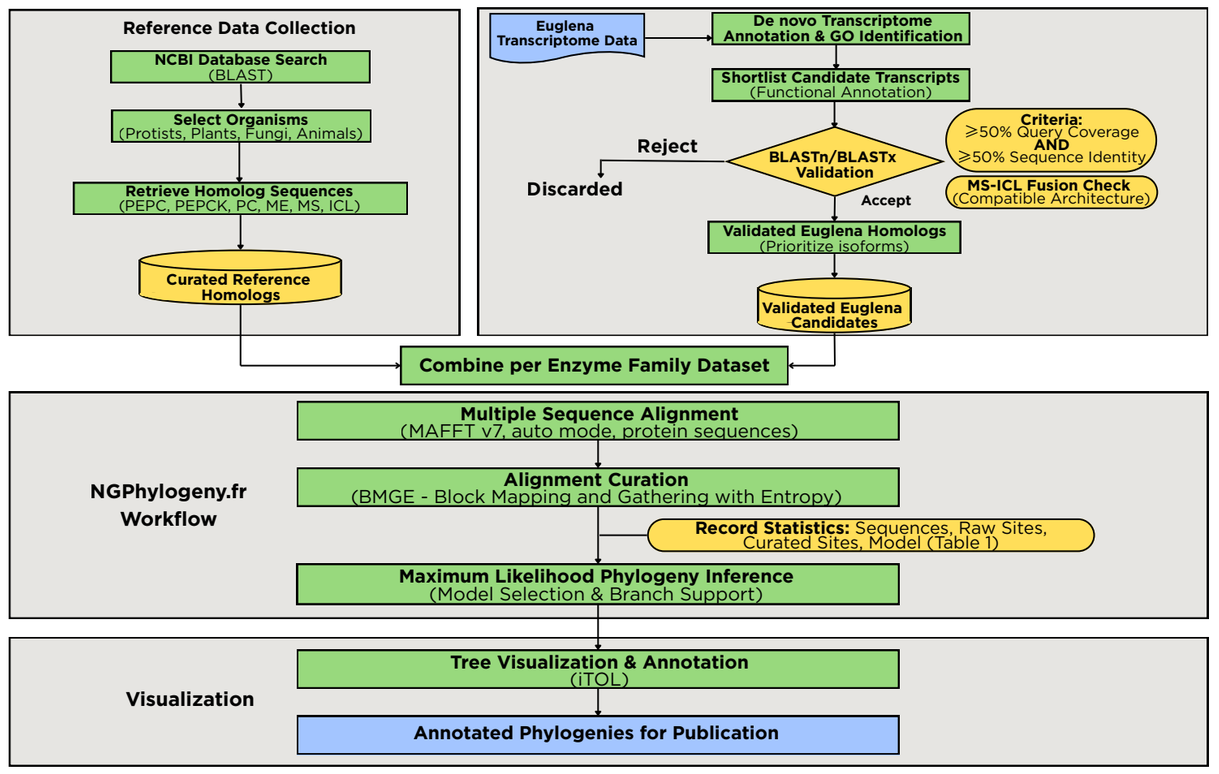
Test for **"Mosaic"** evolutionary signatures



Generate a reproducible resource and testable hypotheses

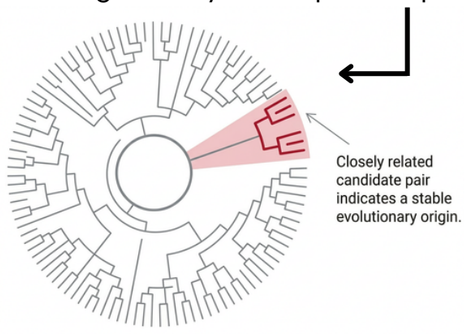
TOOLS

METHODOLOGY

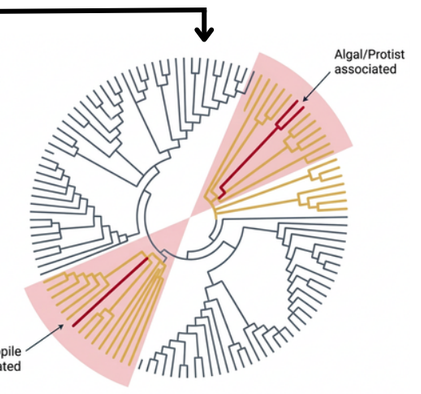


RESULTS

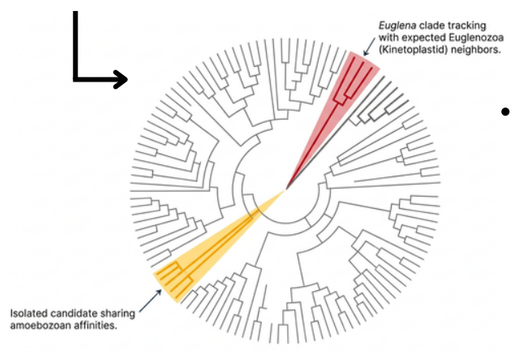
- **Pyruvate Carboxylase (PC) – The Conserved Core:** Displayed the most coherent *E. gracilis* signal, forming a closely related pair that points to a relatively conserved evolutionary anchor.



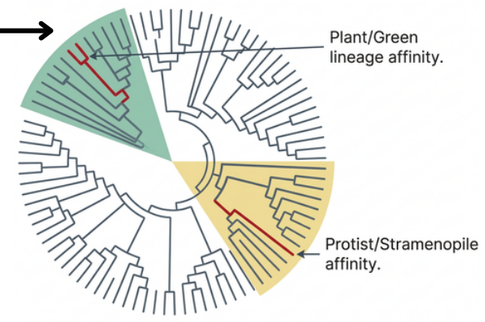
- **Phosphoenolpyruvate Carboxylase (PEPC) – Distinct Placements:** Separated into two different phylogenetic neighborhoods: an algal/protist region and a stramenopile-rich region.



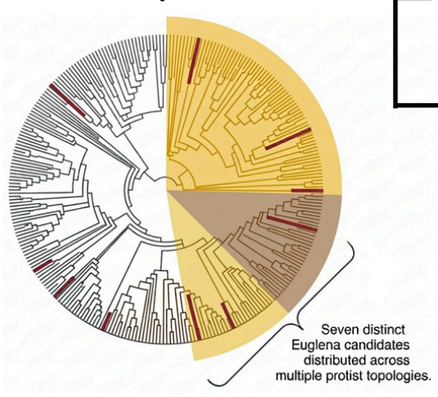
- **Phosphoenolpyruvate Carboxykinase (PEPCK) – Divergent Lineages:** Separated into at least two distinct lineages; one formed a clade adjacent to kinetoplastids, while another placed entirely separately near an amoebozoa reference.



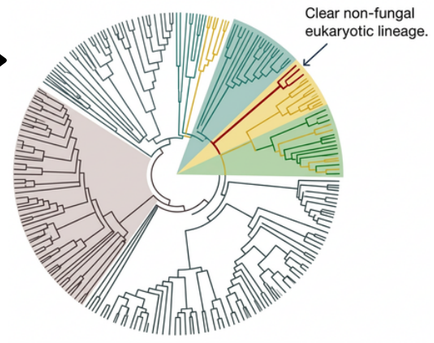
- **Malate Synthase (MS) –** Candidates split, showing contrasting affinities (plant/algal vs. protist-associated).



- **Malic Enzyme (ME) – Highest Diversification:** Exhibited extreme heterogeneity. Multiple Euglena candidates were widely dispersed across several distinct protist-dominated branches, consistent with an expanded and complex isoenzyme repertoire.



- **Isocitrate lyase (ICL) –** Recovered a clear, distinct eukaryotic ICL candidate



DISCUSSIONS & CONCLUSIONS

- **The "Shopping Bag" Genome:** *E. gracilis* contains genes from multiple photosynthetic lineages, resulting in extensive gene family expansion.
- **Enzyme Evolutionary Patterns:** Anaplerotic nodes display a structural spectrum—ranging from highly conserved core functions (PC) to widely dispersed multi-lineage systems (ME).
- **Metabolic Plasticity:** This mosaic architecture provides the genomic foundation to rewire metabolic fluxes, supporting survival in dark anaerobiosis and hypoxia.
- **Future Perspectives:** Further studies require sequence clustering to resolve transcriptome redundancy and proteomic integration for functional validation.

Enzyme topologies show highly variable evolutionary origins in *E. gracilis*. These results support a mosaic evolutionary framework in which conserved core functions coexist with diversified enzyme repertoires, offering testable hypotheses for how evolutionary remodelling of key pathway nodes contributes to the exceptional metabolic plasticity of *E. gracilis* under heterotrophic and oxygen-limited conditions.

REFERENCES

UNDERGRADUATE RESEARCH, INNOVATION, AND COOPERATIVE EDUCATION SYMPOSIUM 2025
Faculty of Science, Chiang Mai University