

Introduction

Based on current understanding, mitochondrion had a bacterial ancestor and was obtained by proto-eukaryotic cell via endosymbiosis, which happened once in eukaryotic history. Over the course of evolution mitochondrial genome has seen a significant reduction in size, due to loss of its genes or their one-way transfer to the nucleus. These processes led to great variety of mitogenomes across different groups, with lots of rare genes and their functions remaining unknown.

Aim: a search and phylogenomic analysis of rarely occurring protein-coding genes in the mitochondrial genomes of various groups of photosynthetic organisms.

Materials and methods

The overall workflow is shown in **Fig. 1**. MitoCOGs, which are clusters of orthologous groups (COGs [1]) constructed from mitochondrial genomes that describe typical mitochondrial genes, have been described previously [2]. The HMMER v3.4 software was used to compare sequences to 140 MitoCOGs HMM profiles. iTOL v.7 was used for tree visualization. MMSEQS2 tool was used for clustering.

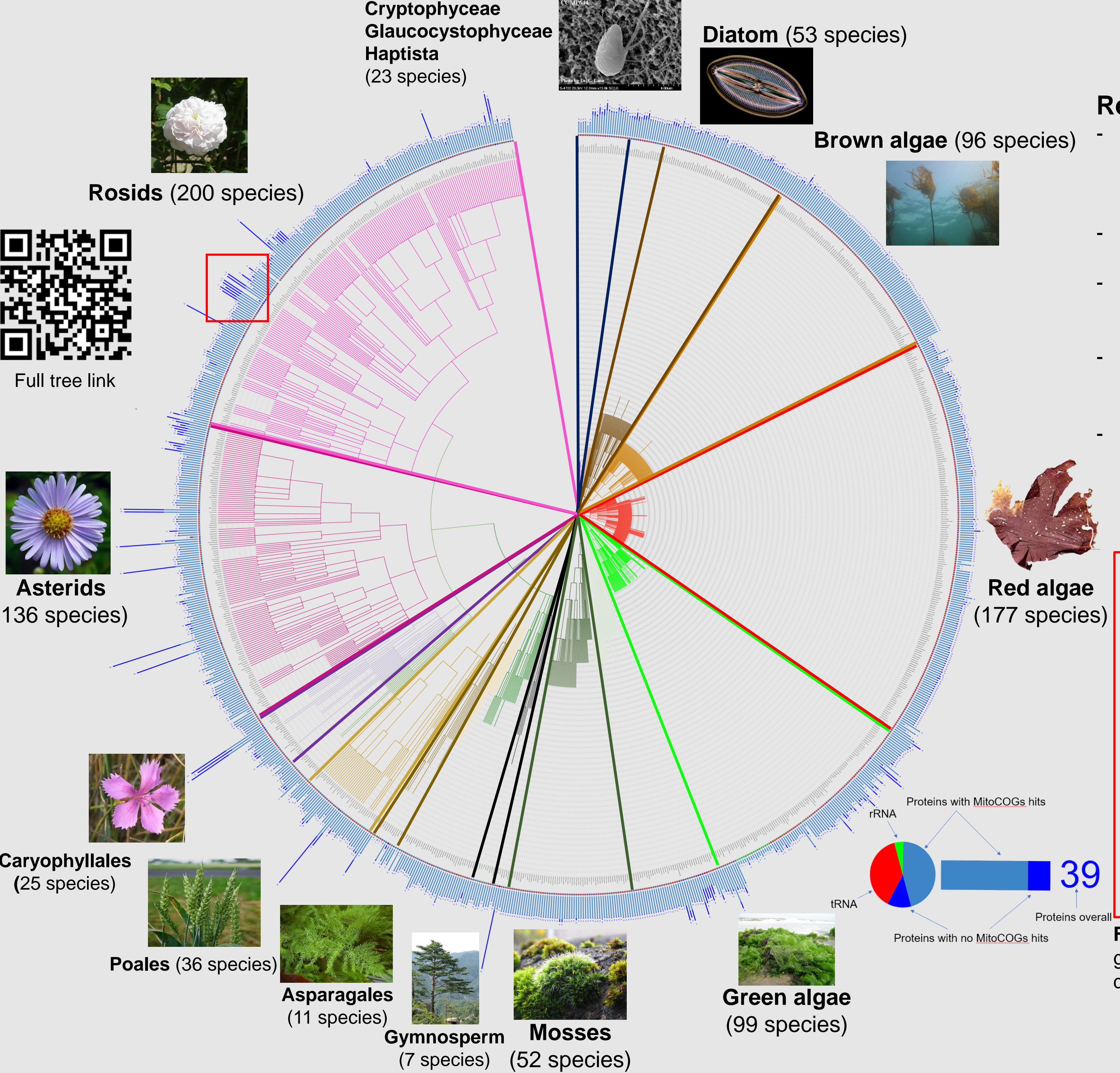


Fig. 2. Taxonomy tree of 1 049 photosynthetic organisms, built upon NCBI Taxonomy database. Each sector is colored according to its corresponding taxonomic group. Red frame indicates Brassiceae tribe. Diagrams near species names how many proteins (with and without MitoCogs hits), tRNA and rRNA are encoded in their mitogenome (example diagram in Fig. 3). *Images source: Wikipedia.com*

Table 1. Average genome composition of different organism groups presented in Figure 2

Taxonomic group	tRNA	rRNA	Proteins							
			Total	No MitoCOGs	with MitoCOGs					
					Total	ETC proteins	Polymerases	Ribosomal proteins	Housekeeping proteins	Unknown
CGH	26	2	34	4	30	18	0	11	1	0
Bacillariophyceae	24	2	38	3	35	17	1	16	1	0
Phaeophyceae	25	3	37	0	37	17	0	17	1	2
Rhodophyta	21	2	24	2	22	17	0	4	1	0
Chlorophyta	22	5	26	3	23	15	0	8	0	0
Bryophyta	24	3	40	1	39	19	0	15	5	0
Acrogymnospermae	18	5	36	1	35	18	0	11	6	0
Asparagales	7	1	13	1	12	7	0	3	2	0
Poales	24	5	39	5	34	17	0	9	6	2
Caryophyllales	25	4	46	10	36	19	3	7	6	1
Asterids	22	4	42	7	35	18	0	10	6	1
Rosids	22	3	38	5	33	18	0	9	6	0

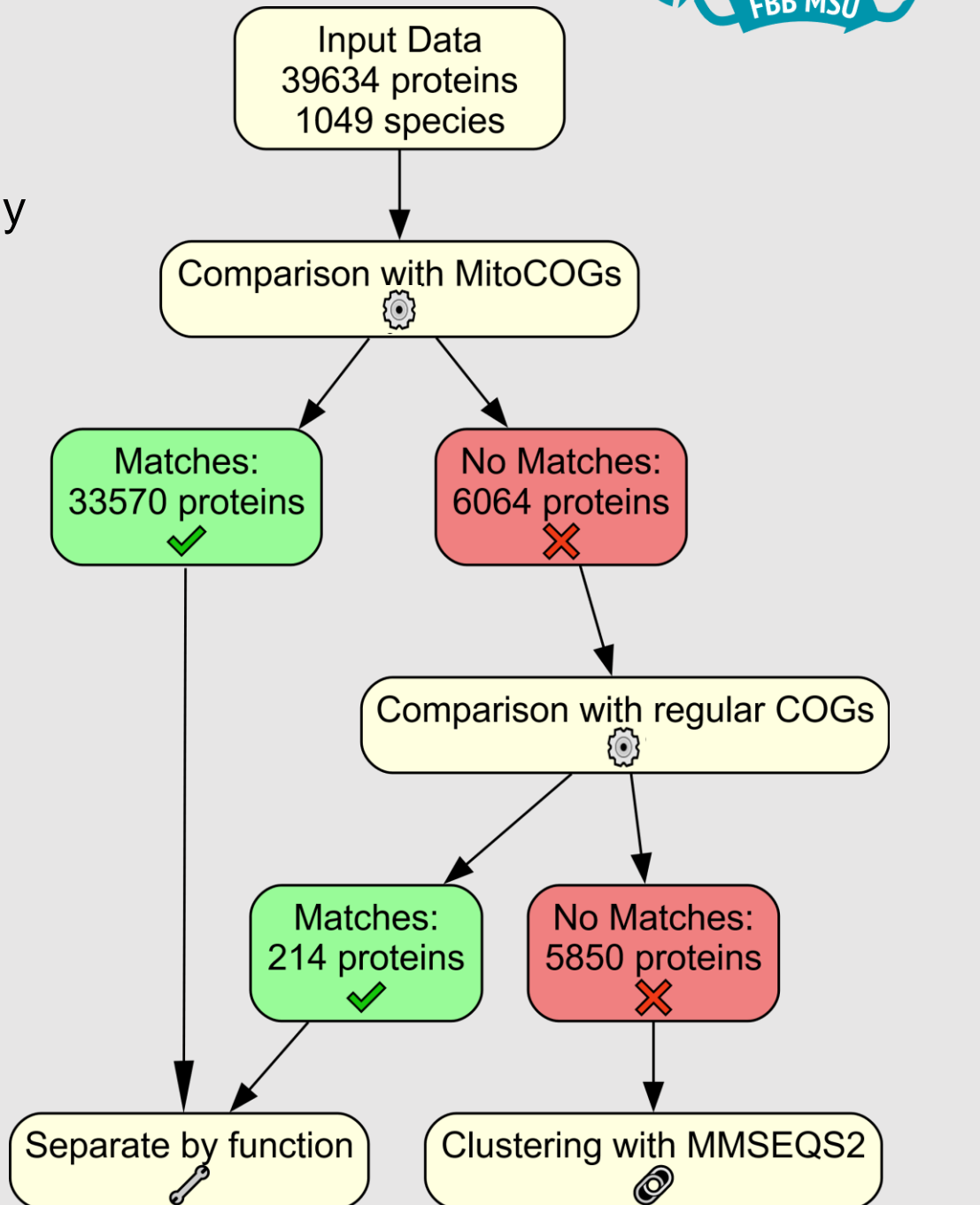


Fig. 1. Overall workflow of the study

Results

- 39 634 proteins were obtained from genome annotation and manual translation: 28 322 belong to green plants, 11 312 – to red algae and other photosynthetic organisms.
- 33 570 of them had at least one hit from MitoCOGs profiles.
- Of 6 064 proteins that had no hits – 214 had at least one hit from COGs profiles, some even containing proteins specific to chloroplasts.
- Other 5 850 were clustered, 743 of them had a domain within the F_O component of ATP synthase domain.
- For individual taxonomic groups, the average compositions of mitogenomes were calculated and the categories of MitoCOGs to which the found proteins belong were determined (**Table 1**).

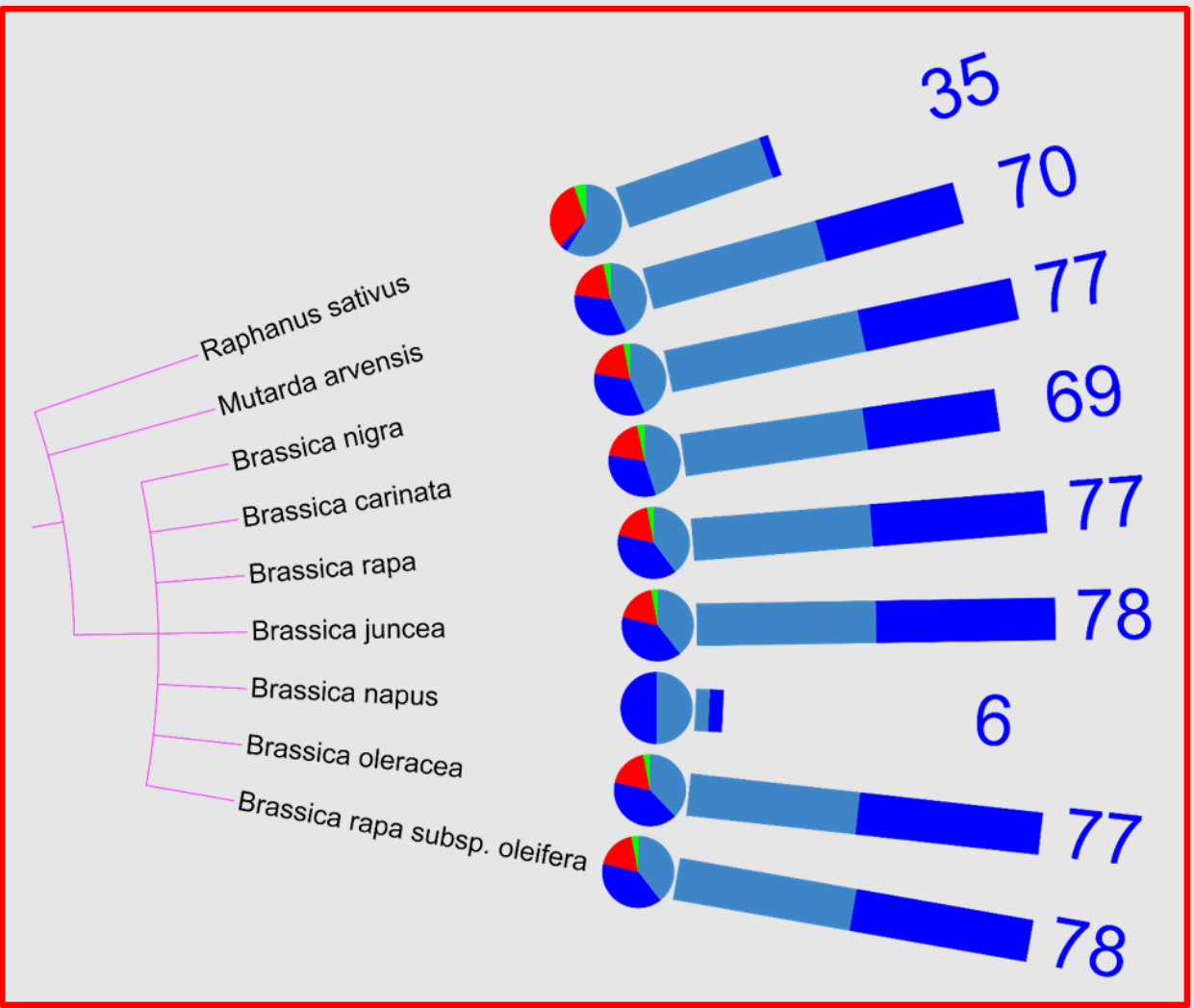


Fig. 3. The close-up of tribe Brassiceae with great rare gene content (right) and an example diagram (left)

Conclusions

- The composition of the mitogenomes in the studied photosynthetic organisms is relatively conserved.
- In all taxonomic groups considered, there is a large number of proteins (5 850 in total) that cannot be assigned to at least one protein family frequently occurring in mitogenomes and, therefore, require more detailed analysis and comparison with prokaryotic proteins.

References:

[1] Galperin, M. Y., Makarova K. S., Wolf Y. I., Koonin E. V. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res*, 43(Database issue), D261-269.

[2] Kannan S, Rogozin IB, Koonin EV. (2014) MitoCOGs: clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes, *BMC Evolutionary Biology* 14:237.