Chair for Bioinformatics
Department of Bioinformatics and Computational Biology
Technical University of Munich

Investigating the Roots of Kleptoplasty

Rushali Chopra, Maximilian Harl, Matthäus Reschka, Lara Rolf, Annika Siebenhaar

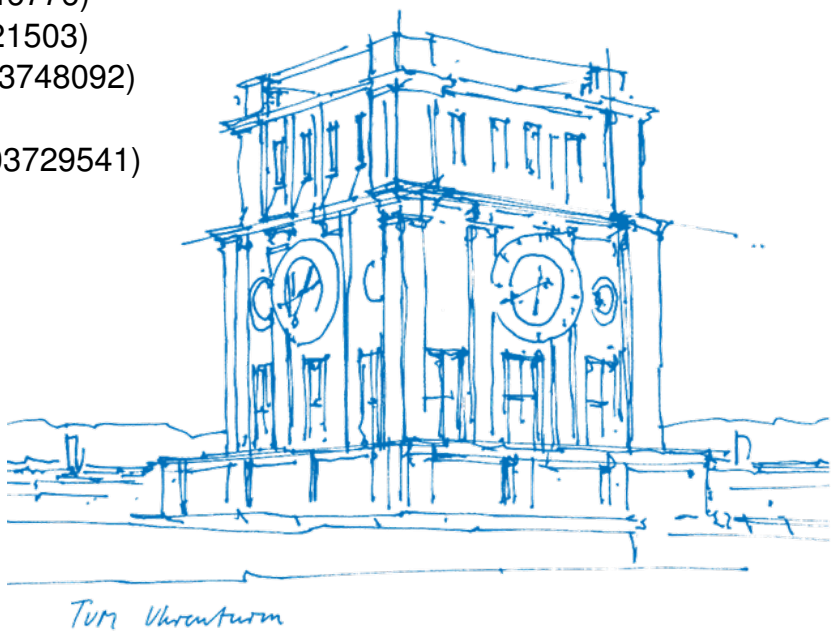
Project Report "Praktikum Genomorientierte Bioinformatik"

Examiner
Prof. Dr. Burkhard Rost
Department of Bioinformatics and Computational Biology

Supervised by
Dr. Ivan Koludarov

Submitted by
Rushali Chopra (M.Nr: 03746776)
Maximilian Harl (M.Nr: 03721503)
Matthäus Reschka (M.Nr: 03748092)
Lara Rolf (M.Nr: 03765909)
Annika Siebenhaar (M.Nr: 03729541)

Submitted on
13.04.2025



Abstract

This study investigates whether structural or evolutionary adaptations in the ftsH gene, a metalloprotease essential for Photosystem II repair, are linked to chloroplast retention in kleptoplastic sea slugs. We analyzed ftsH gene and protein sequences across 10 gastropods (5 kleptoplastic: *Elysia chlorotica*, *Elysia timida*, *Elysia crispata*, *Plakobranchus ocellatus*, *Elysia marginata*; 5 non-kleptoplastic: *Aplysia californica*, *Arion vulgaris*, *Candidula unifasciata*, *Berghia stephaniae*, *Pomacea canaliculata*). Analysis was done using exon homology analysis, structural modeling via AlphaFold, phylogenetic reconstruction, domain prediction and protein space clustering. Despite high variability in exon structure (1–22 exons) and protein length (261–933 AA), all ftsH sequences conserved core functional domains. None of our approaches revealed patterns associated with kleptoplasty, refocusing emphasis on other theories like chloroplast genome autonomy.

Contents

Introduction

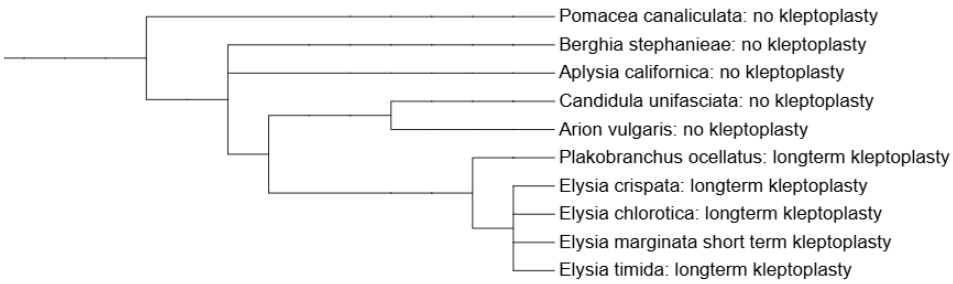
Kleptoplasty, the mechanism by which sacoglossan sea slugs sequester functional chloroplasts from algae that maintain their photosynthetic activity, still represents an unanswered biological question (Cruz et al., 2020; Kawaguti & Yamasu, 1965). Among proposed mechanisms, horizontal gene transfer of algal nuclear genes to slug genomes was initially hypothesized but remains contentious (Bhattacharya et al., 2013). As an alternative, it has been suggested that the retained chloroplasts may be unusually autonomous, containing a larger set of genes essential for photosynthesis than those found in land plant plastids (De Vries et al., 2013). A central gene of interest is ftsH, an ATP-dependent metalloprotease that plays a critical role in degrading photodamaged D1 proteins from photosystem II. This is an essential process for avoiding photoinhibition caused by reactive oxygen species (Lindahl et al., 2000). Whether kleptoplastic slugs have evolved specialized variants of ftsH to support stolen chloroplasts remains unknown.

In this report, we employ comparative genomics and structural analysis to investigate a link between ftsH and kleptoplastic capability. By analyzing ftsH sequences, domains, and evolutionary trajectories across gastropods with divergent plastid-retention capacities (see Figure 1), we disentangle conserved roles of ftsH from potential lineage-specific innovations. This is a step toward understanding one of the most fascinating cross-kingdom interactions in biology—how animals borrow and sustain plant-like functionality.

Methods

The ftsH protein and nucleotide sequences were extracted from NCBI. A custom database was compiled, including sequences from multiple gastropod species, as previously introduced with and without kleptoplastic ability, as well as the algal reference *V.*

Figure 1
Gastropod Species and their kleptoplastic ability.



litorea. The genome and annotation sources for these species are listed in Table 5 in the Supplementary Materials.

BLAST

To find similarities among ftsH sequences, gene regions and exons, we employed BLAST (Altschul et al., 1990). Both the online NCBI BLAST tool and a local installation of NCBI-BLAST v2.16.0+ were used. As our custom database of ftsH sequences was already established, we used blastn for nucleotide comparisons and tblastn to search for protein matches in translated nucleotide datasets. Default parameters for word size and gap penalties were retained, with an e-value threshold of 0.05 applied to filter out non-significant matches (Koludarov et al., 2020).

Alignment

We conducted codon-aware multiple sequence alignments using MACSE (Ranwez et al., 2011). This tool preserves reading frames and accounts for potential frameshifts in protein-coding sequences, providing biologically meaningful alignments. We then visualized it in JalView (Waterhouse et al., 2009). We additionally performed an exon homology screen using a custom exon database of 336 exons from 27 genes in 8 organisms. Homology was evaluated via All-vs-All, Organsim-vs-Exons, and Coding Sequence-vs-Exons BLAST comparisons, following standard comparative genomics approaches (Altschul et al., 1990). To visualize relationships among exon sequences,

we also applied PaSiMap clustering, as implemented in JalView using default settings (Waterhouse et al., 2009).

Phylogenetic Tree

Phylogenetic analysis was conducted using the Phylogeny.fr 'One Click' mode (Dereeper et al., 2008). Input sequences are aligned using MUSCLE with default parameters, and the alignment is curated with Gblocks to remove poorly aligned positions. The phylogenetic tree is then constructed using PhyML, employing the default substitution model and parameters. Branch support is assessed using the approximate likelihood-ratio test (aLRT). The tree was then visualized using the iTOL v7 web tool (Dereeper et al., 2008).

Protein structure simulation

We simulated the 3D-structures using AlphaFold 3 (Abramson et al., 2024). The alignment was done with the Pairwise Structure Alignment tool from PDB (Bittrich et al., 2024) using TM-align and jFATCAT.

Embeddings and Clustering

To analyze the ftsH sequence space, we collected 15,846 sequences from UniProt (annotation score ≥ 4), including our targets. Sequences lacking protein family annotations were excluded. Each sequence was embedded using ProtT5-XL-U50 in half-precision mode (Elnaggar et al., 2020). Percent identity was calculated via MAFFT v7.526 (auto settings) (Katoh et al., 2002) and MEGA v12.09 (Tamura et al., 2021), using Elysia chlorotica's ftsH (A0A3S0ZH14) as the 100% reference. UMAP (McInnes et al., 2020) and ProtSpace (Senoner et al., 2025) (default settings) were applied for dimensionality reduction, clustering and visualization.

Protein Domain Analysis

To better understand the functional components of the ftsH protein and identify evolutionarily conserved domains, we analyzed domains using HMMER (Eddy, 2024). This tool uses Hidden Markov models to detect protein domains based on sequence profiles.

Results

Gene Region Analysis

We extracted the gene regions from our genes, as displayed in Figures 2 and 8. The gene regions of ftsH in the gastropods, as well as chloroplasts, were not conserved enough to locate potentially unannotated ftsH genes. They were mainly determined based on the existing annotations. If a gene was not named, the sequence was blasted, and the tables were amended with the blast results.

Exon Structure and Homology

Exon structure and protein length varied widely across species (e.g., exon count: 1–22; protein length: 261–933 AA), as summarized in Table 1. Gene names were serialized for better readability. For the corresponding official gene identifiers used in the genome annotations (GFF files), see Table 4. No clear pattern emerged linking exon number or protein size to kleptoplasty.

Figure 2
Gene regions of ftsH in gastropods.

	E. chlorotica ftsH 1	E. chlorotica ftsH 2	E. chlorotica ftsH 3	E. Marginata ftsH 1	E. marginata ftsH 2	E. marginata ftsH 3	E. marginata ftsH 4	E. marginata ftsH 5	P. ocellatus ftsH 1	P. ocellatus ftsH 2	P. ocellatus ftsH 3	P. ocellatus ftsH 4	E. timida ftsH 1	E. timida ftsH 2	E. timida ftsH 3	C. unifasciata ftsH 1	C. unifasciata ftsH 2	C. unifasciata ftsH 3	P. canaliculata ftsH 1	P. canaliculata ftsH 2	A. californica ftsH 1	A. california ftsH 2	A. california ftsH 3
5	YERD	HEBP2		SIRT5			GYRA		no blast match		NWD1		CEBPA/B	SLC52A3	K02A2.6-like	no blast match			ZNFX1	LOC112576475	LOC101861385	TAF4	
4	ACCC	PNKP		PDP1			GYRB		OSBP		no blast match		PPARG	SLC22A15	K02A2.6-like	no blast match			ZNFX1	LOC112576474	SLC6A19,	KCTD15	
3	BCCP	PNKP	scaffold end	EIF3			IF-2		ACNAT1	scaffold end	no blast match	scaffold end	PSMG2	no blast match	NPFFR2	RRBP1	scaffold end	scaffold end	ZNFX1	LOC112576465	SLC6A19	SMB1	scaffold end
2	RSMA	FKH1	NWD1	SETMAR			RNR		EIF2AK2	no blast match	no blast match	NPY	LOC131939765	SLC25A13	UBE2Q2	ANKRD12	EMC4	LOC112576482	LOC112576482	ALPK1	TSNAXIP1	TMEM64	LOC101861846
1	SYR	EIF2AK1	no blast match	UBE2Q2	scaffold end	scaffold end	adenylosuccinate synthetase	scaffold end	LOC106011851	no blast match	no blast match	no blast match	ANKAR	LOC106074731	TIGD4	no blast match	BOP1	neo-calmodulin-like	LOC112577282	ALPK1	UBC-25	NUDCD1	LOC106013230
0	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	ftsH	AFG3-like protein 2	paraplegin-like	Paraplegin	AFG3-like protein	YME1L
1	scaffold end	WD40_zf-RING_UBOX	NOX2	scaffold end	scaffold end	scaffold end	no blast match	UROD	no blast match	EIF2AK2	MPD1	RPS28	no blast match	LOC135113735	EIF2AK2	OSGIN2	neo-calmodulin-like	COPZ1	NOP9	ANKRD11	ANKRD11	TMEM170	shematrin-like protein 1
2		UNC50	no blast match				RRM1	HMBS	no blast match	UGT2A1	TTLL5	scaffold end	LOC118478016	SMARCB1	PPIL6	neo-calmodulin-like	scaffold end	PFDN5	LOC112577107	CYPB	FOXC1	LOC101859492	LOC101860604
3		ITFG1	no blast match				FBA1	Fe-S oxidoreductase-like protein	endonuclease-reverse transcriptase	PPIL6	CYT1		DDO-like	PSME3	PNKP	scaffold end		GGACT	LOC112577108	CYPB	scaffold end	LOC106012095	ATAD2
4		CFAP20	no blast match				ALDOB	ACAT1	scaffold end	PNKP	KDELC1		TLR4	LOC131929795	TRAF7			ABHD2	LOC112553154	SNRK		scaffold end	BRD4A
5		no blast match	no blast match				scaffold end	COX1		TRAF7	PO21_19		DUS2-like	MRC1	TIP			BMPR1A	RMDN1	LOC112553348			scaffold end

Table 1
Exon count and protein length (amino acids) of ftsH genes in gastropods.

Gene	Exon Count	Protein Length (AA)
A. californica ftsH 1	20	816
A. californica ftsH 2	20	746
A. californica ftsH 3	20	765
C. unifasciata ftsH 1	15	738
C. unifasciata ftsH 2	15	652
C. unifasciata ftsH 3	14	585
E. chlorotica ftsH 1	20	778
E. crisparta ftsH 1	21	744
E. crisparta ftsH 2	20	771
E. marginata ftsH 1	22	650
E. marginata ftsH 2	7	513
E. marginata ftsH 3	10	933
E. marginata ftsH 4	1	261
E. marginata ftsH 5	1	374
E. timida ftsH 1	20	776
E. timida ftsH 2	19	761
E. timida ftsH 3	18	767
P. canaliculata ftsH 1	19	748
P. canaliculata ftsH 2	19	779
P. ocellatus ftsH 1	5	773
P. ocellatus ftsH 2	19	421
P. ocellatus ftsH 3	20	784
P. ocellatus ftsH 4	12	280

Our BLAST-based homology search between exons across species revealed low overall exon homology, with only a few exons (e.g., Exon 8 from *E. marginata ftsH 1* and Exon 14 from *E. crisparta ftsH 1*) shared across multiple species (see Table 2). We then applied PaSiMap using Jalview (see Figure 3) to cluster our exons. No distinct cluster corresponding to kleptoplasty could be identified. We also assessed sequence conservation with alignments using MACSE. A single highly conserved region was identified across all species. However, the majority of exon sequences displayed high variability and poor alignment outside this core region, as can be seen in Figure 4.

Table 2
Homologous exons across our database of ftsH genes. Each row lists exons identified as homologous.

E. marginata (ftsH 1)	*P. ocellatus* (ftsH 2)	*E. timida* (ftsH 3)	*E. crisparta* (ftsH 1)
Exon 4	Exon 4	Exon 4	
Exon 8			Exon 14
Exon 9			Exon 13
		Exon 6	Exon 16
Exon 12	Exon 12		
	Exon 13		Exon 9
	Exon 14		Exon 8

Figure 3
PaSiMap of exon database.

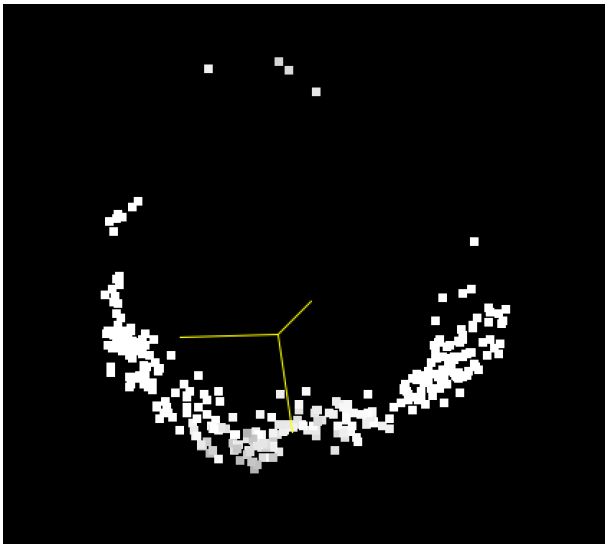
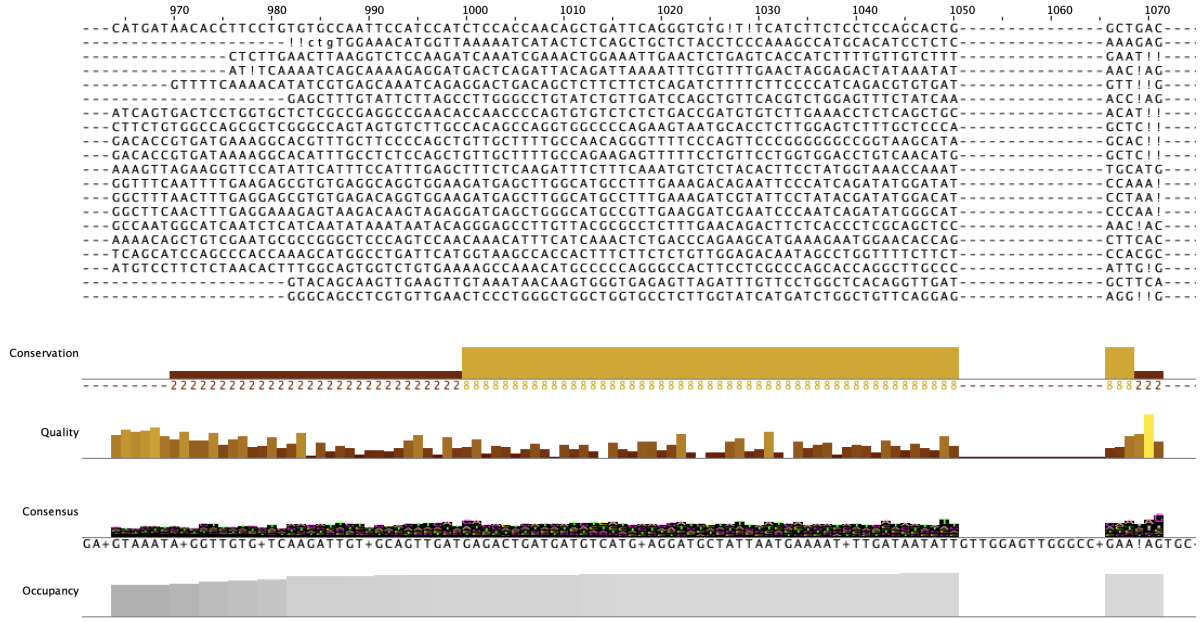


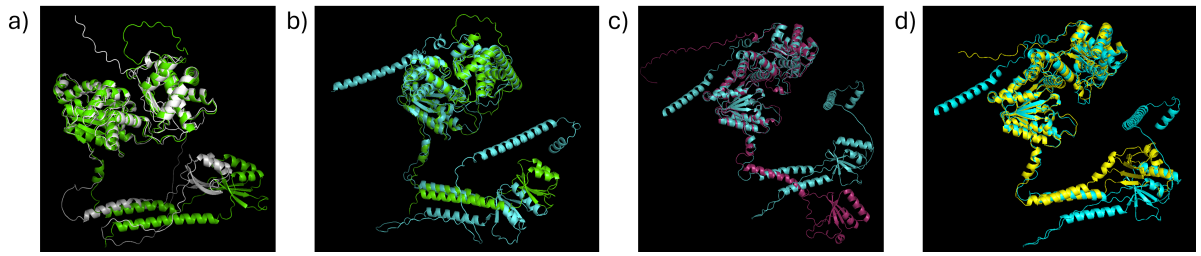
Figure 4
Conserved Region of exon alignment. Only randomly selected exon nucleotides are shown. Gap is created by one exon of P. Ocellatus and one of E. Timida.



Protein Structure

Since we speculated that the ftsH genes in the kleptoplastic slugs have the same function as the ftsH in the plants, we compared the 3D-structures from all the gastropod ftsH proteins to the ftsH protein from the *V. litorea* chloroplast. Visual inspection of the protein alignments showed that proteins with similar lengths had a quite conserved C-terminal domain. The results can be seen in the Supplementary Material Figure 9, 10 and Table 6. Additionally, we aligned *V. litorea* chloroplast ftsH with ftsH 8 from the land plant *A. thaliana*. To investigate if there is a difference between longterm retention and shortterm retention slugs, we compared the ftsH from *E. chlorotica* with *E. marginata*. We also aligned a non kleptoplastic slug ftsH (from *C. unifasciata*) with an ftsH from *E. chlorotica*. For this we took the ftsHs from the gastropods, that had the best match with the *V. litorea* chloroplast ftsH. These alignments can be seen in Figure 5. Neither the TM-score from the alignments nor the visual inspections suggest a meaningful difference in the ftsH proteins and therefore their kleptoplastic ability.

Figure 5
3D structure alignment of a) V. litorea (green) and A. thaliana (white), b) V. litorea (green) and E. chlorotica (blue), c) E. chlorotica (blue) and E. marginata (pink) and d) E. chlorotica (blue) and C. unifasciata (yellow).



Manual Gene Annotataion

We attempted manual re-annotation of some genes to improve the accuracy of the automatic gene annotations we used. We focused on genes with discrepancies in the annotations such as low exon counts. The low exon homology as well as short scaffolds

of the genome assemblies hindered the manual re-annotation of ftsH genes. Attempts to identify missing exons by blasting our curated database of 336 exons from 27 ftsH genes against the gene regions of ftsH genes with low exon counts were unsuccessful. Aligning known ftsH coding sequences to the regions also failed to reveal misannotated exon structures. We also used Geneious (Kearse et al., 2012) to identify open reading frames in the surrounding regions and assessed their homology to known ftsH proteins using BLAST and InterProScan (Jones et al., 2014) domain analysis. However, none of these approaches yielded sufficient evidence to support re-annotation.

Phylogenetic Tree

Figure 6 illustrates the phylogenetic tree of our analyzed ftsH protein sequences and is divided into two main clusters. The upper cluster includes bacterial and algal ftsH proteins and a distinct group of gastropod sequences. The lower cluster is composed entirely of gastropods. It is further split into two subclusters, including land and sea gastropods. Since almost all clusters contain snails with kleptoplasty, we can identify no clear correlation between these groupings and kleptoplastic ability.

Protein clustering

The protein cluster analysis of kleptoplasty using Protspace shows that nearly all snails, independent of their ability of kleptoplasty, are clustered together, as shown in Figure 7. In this cluster, we can see snails with short-term, with long-term, and without kleptoplasty, as well as organisms without known kleptoplastic ability in grey. As shown in Figure 8, where the data points are colored by their percent identity, varying sequence identities among functionally similar proteins suggest that functional conservation is independent of the sequence similarity. This leads to the conclusion that there is no clear correlation between the ftsH protein and kleptoplastic ability.

Figure 6
Phylogenetic tree of the analyzed ftsH proteins.

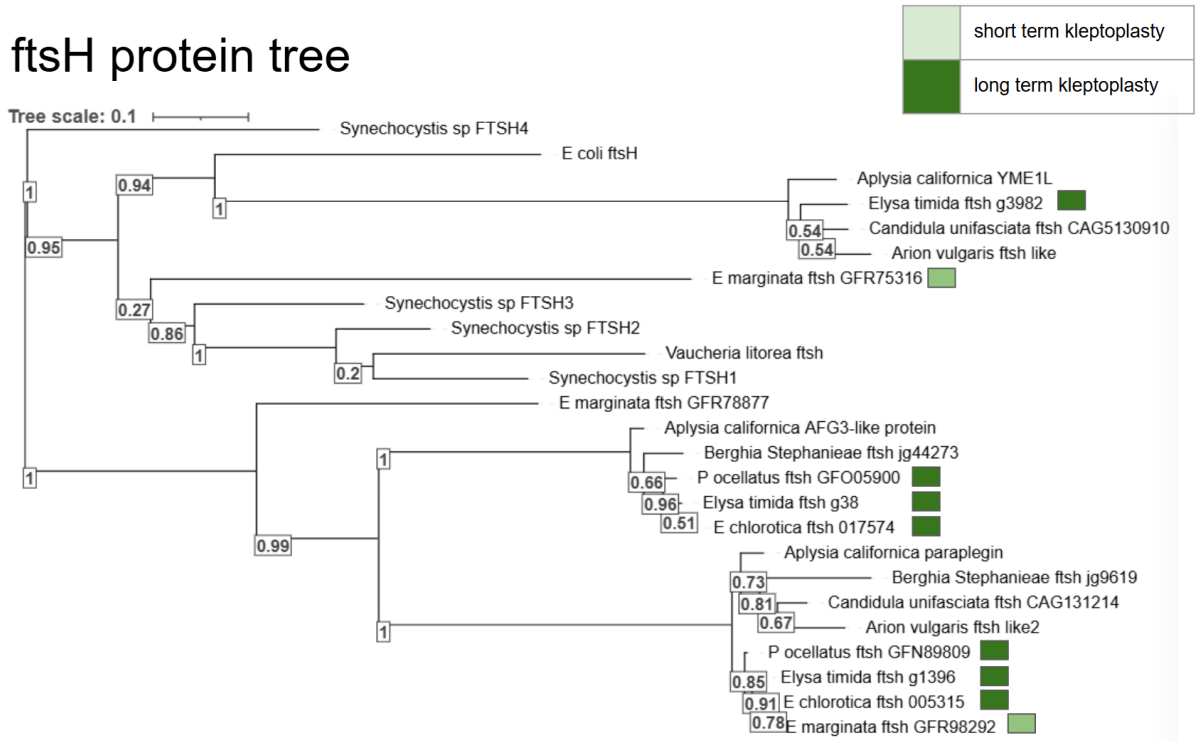


Figure 7
ProtSapce clustering, Feature: Kleptoplasty ability.
Blue circle: main cluster gastropoda, Yellow circle: V. litorea.

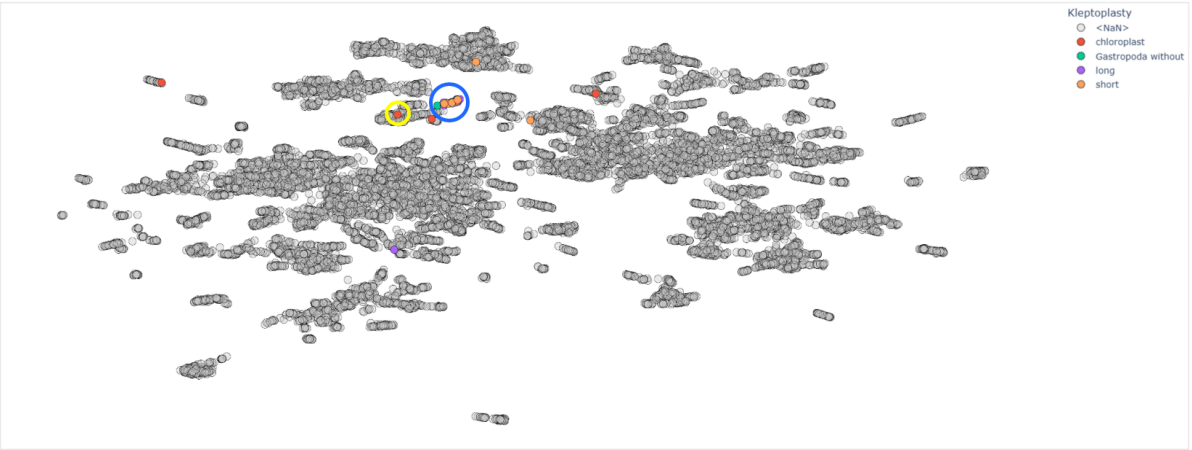
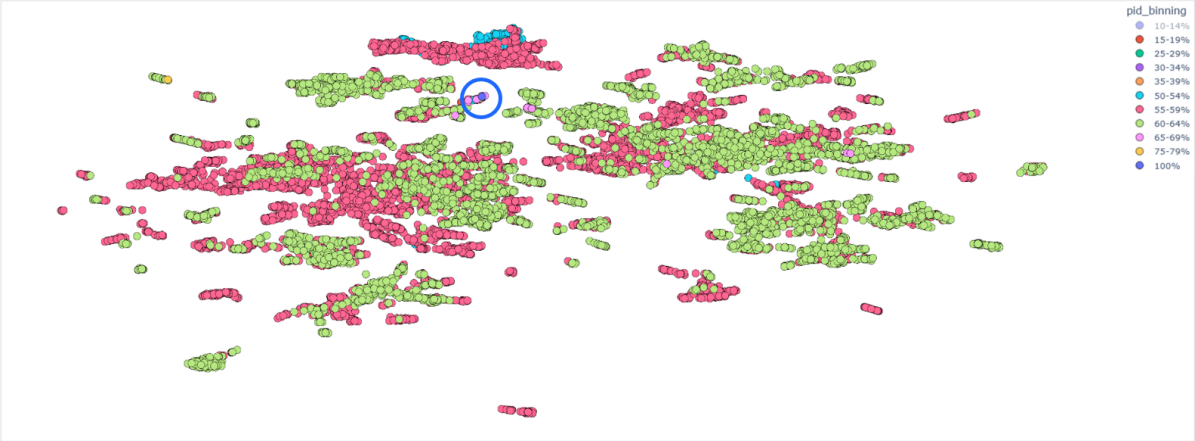


Figure 8
ProtSapce clustering, Feature: Percent Identity.
Data points colored by their Percent Identity to ftsH from E. chlorotica, binned in 5% steps.



Domain conservation

The three most important domains found were AAA+ ATPase (responsible for substrate unfolding using ATP hydrolysis), the M41 peptidase domain (involved in proteolytic degradation of damaged proteins) and Zinc binding protease domain (proteolysis of proteins). Table 3 shows the presence of these domains in our protein sequences of interest. There is a high conservation of the domains since there are no differences in the domains between the species.

Discussion and Conculsion

Our comparative genomic and structural analysis of ftsH genes across kleptoplastic and non-kleptoplastic gastropods reveals no definitive link between ftsH and the ability to retain functional chloroplasts. Despite initial hypotheses, we found no evidence of lineage-specific innovations in slug-encoded ftsH that correlate with kleptoplasty. The structural comparison of ftsH proteins showed strong similarity across all species, including algae, plants, and slugs, indicating conservation rather than specialization. Likewise, protein domain analysis consistently identified the canonical AAA+ ATPase, zinc-binding, and M41 peptidase domains in all species, regardless of their

Table 3
Domain presence across ftsH sequences in different species.

Gene ID	Length (AA)	AAA+ ATPase	Peptidase M41	Zinc-binding protease	FtsH ext
A. californica ftsH 1	816	✓	✓	✓	✓
A. californica ftsH 2	765	✓	✓	✓	✓
A. californica ftsH 3 isoform X1	746	✓	✓	✓	
A. californica ftsH 3 isoform X2	745	✓	✓	✓	
A. vulgaris ftsH 2	814	✓	✓	✓	
A. vulgaris ftsH 3	545	✓		✓	✓
C. unifasciata ftsH 1	652	✓	✓	✓	
C. unifasciata ftsH 3	738	✓	✓	✓	✓
E. chlorotica ftsH 1	778	✓	✓	✓	
E. chlorotica ftsH 2	841	✓	✓	✓	
E. chlorotica ftsH 3	339	✓	✓	✓	
E. marginata ftsH 1	650	✓	✓	✓	✓
E. marginata ftsH 2	513	✓	✓	✓	
E. marginata ftsH 3	933	✓	✓	✓	
E. marginata ftsH 4	261		✓		
E. marginata ftsH 5	374	✓		✓	
E. timida ftsH 1	776	✓	✓	✓	
E. timida ftsH 2	743	✓	✓	✓	✓
E. timida ftsH 3	761	✓	✓	✓	✓
P. ocellatus ftsH 1	280				
P. ocellatus ftsH 2	784	✓	✓	✓	
P. ocellatus ftsH 3	773	✓	✓	✓	✓
P. ocellatus ftsH 4	421	✓	✓	✓	
V. litorea ftsH	644	✓	✓	✓	✓

plastid-retention capacity. These results suggest that the core functions of ftsH are preserved and not uniquely adapted for supporting kleptoplasts. Exon structure analysis revealed high variability and low homology among gastropods, further complicating evolutionary interpretation. No exon or combination of exons was found to be uniquely associated with kleptoplastic species. While our findings do not support a role of the slugs' ftsH genes in enabling kleptoplasty, the hypothesis that chloroplast-encoded ftsH contributes to plastid autonomy remains plausible.

In summary, kleptoplasty does not appear to be driven by unique adaptations of slug-encoded ftsH. Future studies should also explore alternative mechanisms, such as chloroplast genome autonomy, post-translational regulation, or host-organelle interactions, to fully understand how sacoglossan slugs sustain photosynthetic function in stolen plastids.

Bibliography

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, *630*(8016), 493–500.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., & Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life science alliance*, *2*(5).

Bhattacharya, D., Pelletreau, K. N., Price, D. C., Sarver, K. E., & Rumpho, M. E. (2013). Genome analysis of elysia chlorotica egg dna provides no evidence for horizontal gene transfer into the germ line of this kleptoplastic mollusc. *Molecular biology and evolution*, *30*(8), 1843–1852.

Bittrich, S., Segura, J., Duarte, J. M., Burley, S. K., & Rose, Y. (2024). Rcsb protein data bank: Exploring protein 3d similarities via comprehensive structural alignments. *Bioinformatics*, *40*(6), btae370. https://doi.org/10.1093/bioinformatics/btae370

Cruz, S., LeKieffre, C., Cartaxana, P., Hubas, C., Thiney, N., Jakobsen, S., Escrig, S., Jesus, B., Kühl, M., Calado, R., & Meibom, A. (2020). Functional kleptoplasts intermediate incorporation of carbon and nitrogen in cells of the sacoglossa sea slug elysia viridis. *Scientific Reports*, *10*(1), 10548. https://doi.org/10.1038/s41598-020-66909-7

De Vries, J., Habicht, J., Woehle, C., Huang, C., Christa, G., Wägele, H., Nickelsen, J., Martin, W. F., & Gould, S. B. (2013). Is ftsh the key to plastid longevity in sacoglossan slugs? *Genome Biology and Evolution*, *5*(12), 2540–2548.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M., & Gascuel, O. (2008). Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.*, *36*(Web Server issue), W465–9.

Eddy, S. R. (2024). Hmmer: Biosequence analysis using profile hidden markov models [Accessed: 2025-04-09].

Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., BHOWMIK, D., & Rost, B. (2020). Prottrans: Towards cracking the language of life's code through self-supervised deep learning and high performance computing. *bioRxiv*. https://doi.org/10.1101/2020.07.12.199554

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). Interproscan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240.

Katoh, K., Misawa, K., Kuma, K.-i., & Miyata, T. (2002). Mafft: A novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kawaguti, S., & Yamasu, T. (1965). Electron microscopy on the symbiosis between an elysioid gastropod and chloroplasts from a green alga. *Biological Journal of Okayama University II*, 57–64.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.

Koludarov, I., Jackson, T. N., Suranse, V., Pozzi, A., Sunagar, K., & Mikheyev, A. S. (2020). Reconstructing the evolutionary history of a functionally diverse gene

family reveals complexity at the genetic origins of novelty. *bioRxiv*. https://doi.org/10.1101/583344

Lindahl, M., Spetea, C., Hundal, T., Oppenheim, A. B., Adam, Z., & Andersson, B. (2000). The thylakoid ftsh protease plays a role in the light-induced turnover of the photosystem ii d1 protein. *The Plant Cell*, *12*(3), 419–431.

McInnes, L., Healy, J., & Melville, J. (2020). Umap: Uniform manifold approximation and projection for dimension reduction. https://arxiv.org/abs/1802.03426

Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. (2011). Macse: Multiple alignment of coding sequences accounting for frameshifts and stop codons. *PloS one*, *6*(9), e22594.

Senoner, T., Olenyi, T., Heinzinger, M., Spannagl, A., Bouras, G., Rost, B., & Koludarov, I. (2025). Protspace: A tool for visualizing protein space. *Journal of Molecular Biology*, 168940. https://doi.org/https://doi.org/10.1016/j.jmb.2025.168940

Tamura, K., Stecher, G., & Kumar, S. (2021). Mega11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191.

Supplementary Material

Figure 9

3D structure alignments of V. litorea in green with the proteins from kleptoplastic slugs named above the picture.

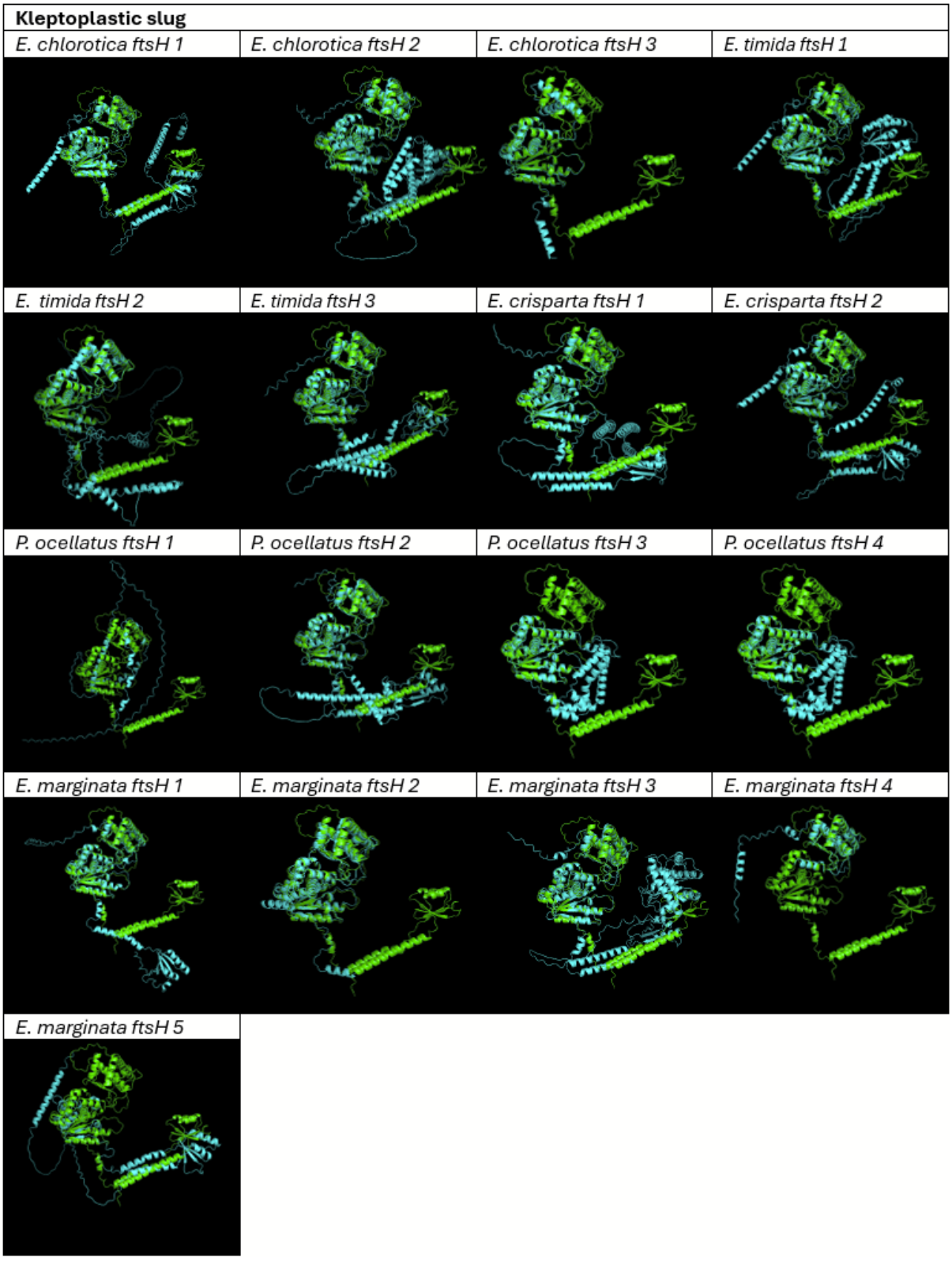


Figure 10

3D structure alignments of V. litorea in green with the proteins from non kleptoplastic slugs named above the picture.

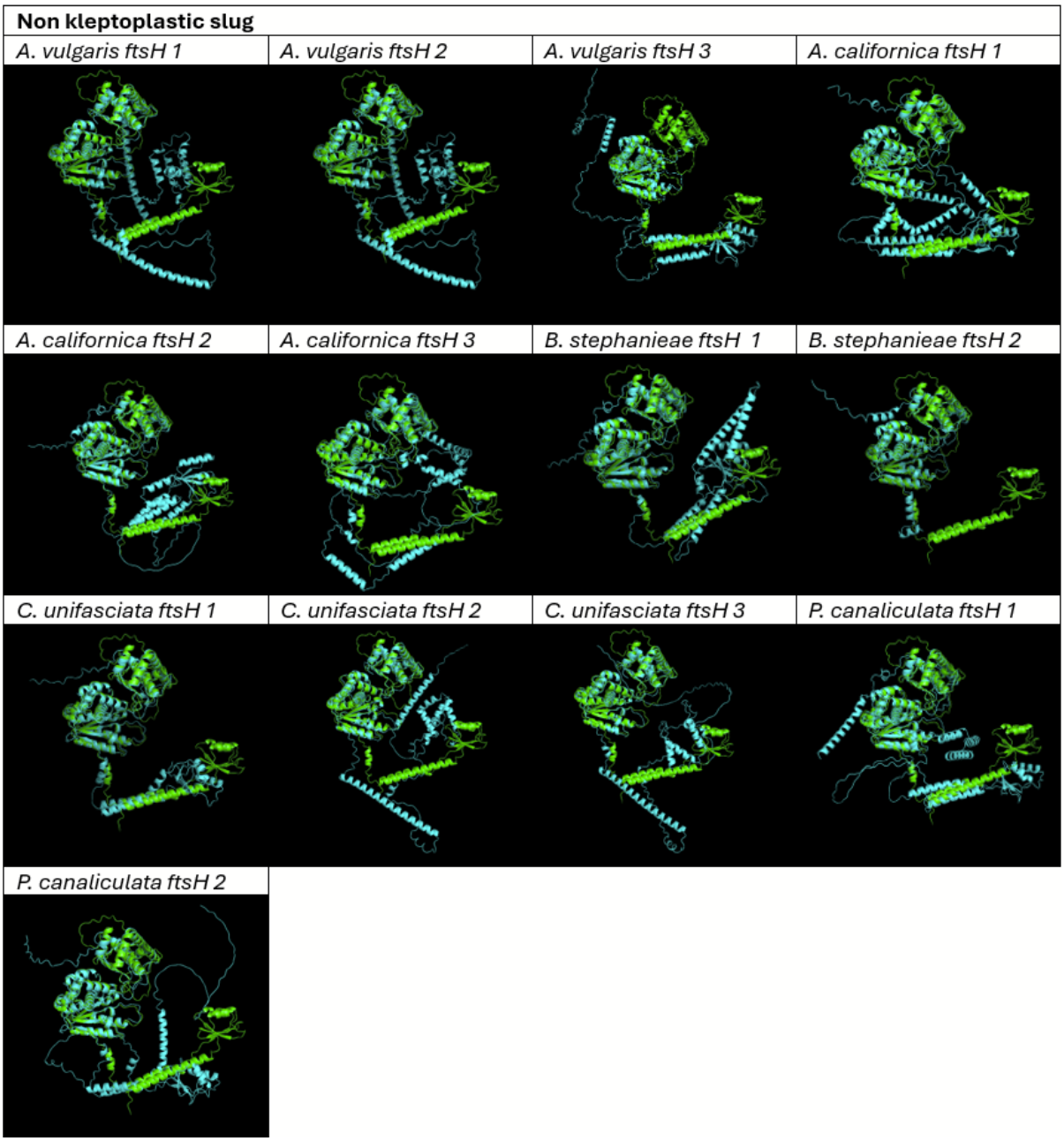


Table 4
Mapping of serialized ftsH gene names to official gene IDs used in GFF annotations.

Serialized Gene Name	Official Gene ID
A. californica ftsH 1	Paraplegin
A. californica ftsH 2	AFG3-like protein
A. californica ftsH 3	YME1L
A. vulgaris ftsH 1	ATPase 1
A. vulgaris ftsH 2	ATPase 2
A. vulgaris ftsH 3	ATPase fragment
B. stephanieae ftsH 1	jg9619
B. stephanieae ftsH 2	jg44273
C. unifasciata ftsH 1	CAG5130910.1
C. unifasciata ftsH 2	CAG5126606.1
C. unifasciata ftsH 3	CAG5131214.1
E. chlorotica ftsH 1	Echl_24602
E. chlorotica ftsH 2	218530
E. chlorotica ftsH 3	g126370
E. crisparta ftsH 1	Ecla1404g218530
E. crisparta ftsH 2	Ecla12559g126370
E. marginata ftsH 1	GFR75316.1
E. marginata ftsH 2	GFR78877.1
E. marginata ftsH 3	GFR98292.1
E. marginata ftsH 4	GFR75213.1
E. marginata ftsH 5	GFR68670.1
E. timida ftsH 1	g38
E. timida ftsH 2	g3982
E. timida ftsH 3	g1396
P. canaliculata ftsH 1	AFG3-like protein 2
P. canaliculata ftsH 2	paraplegin-like
P. ocellatus ftsH 1	GFN86709.1
P. ocellatus ftsH 2	GFN89809.1
P. ocellatus ftsH 3	GFO05900
P. ocellatus ftsH 4	GFO39444

Table 5
Source of genome and annotation data.

Species	**NCBI Assembly Accession**
A. californica	GCF_000002075.1
A. vulgaris	GCA_020796225.1
B. stephanieae	GCA_034508935.3
C. unifasciata	GCA_905116865.2
E. chlorotica	GCA_003991915.1
E. crisparta	GCA_033675545.1
E. marginata	GCA_019649035.1
E. timida	GCA_043644045.1
P. canaliculata	GCF_003073045.1
P. ocellatus	GCA_019648995.1

Table 6

Comparison of Vaucheria litorea vs various proteins with RMSD, TM-score, and Identity %; first the values gained with TM-align and second values from jFATCAT.

Vaucheria litorea vs	**RMSD**	**TM-score**	**Identity %**	**RMSD**	**TM-score**	**Identity %**
A. californica ftsH 2	3.45	0.75	41	4.00	0.64	39
A. californica ftsH 1	3.75	0.76	37	4.04	0.61	33
A. californica ftsH 3	3.40	0.71	40	3.09	0.61	39
A. californica ftsH 3	2.88	0.69	42	3.15	0.60	36
A. thaliana ftsH 8	3.57	0.81	56	3.05	0.75	53
A. vulgaris ftsH 1	2.75	0.70	42	3.20	0.61	38
A. vulgaris ftsH 2	3.57	0.77	35	3.93	0.61	33
A. vulgaris ftsH 3	4.38	0.49	44	5.01	0.49	42
B. stephanieae ftsH 1	3.85	0.76	39	3.8	0.64	39
B. stephanieae ftsH 2	2.53	0.7	39	2.43	0.7	39
C. unifasciata ftsH 1	3.49	0.70	41	3.08	0.61	38
C. unifasciata ftsH 2	3.42	0.78	37	3.20	0.76	35
C. unifasciata ftsH 3	3.42	0.74	39	3.29	0.65	38
E. crisparta ftsH 1	3.87	0.75	37	4.31	0.6	34
E. crisparta ftsH 2	3.61	0.75	41	4.78	0.63	37
E. timida ftsH 3	4.00	0.77	36	3.52	0.65	34
E. timida ftsH 1	3.31	0.72	42	5.69	0.58	39
E. timida ftsH 2	3.11	0.70	41	1.46	0.59	44
E. chlorotica ftsH 1	3.77	0.79	40	3.37	0.65	38
E. chlorotica ftsH 2	4.07	0.72	36	5.06	0.56	32
E. chlorotica ftsH 3	2.96	0.45	55	4.40	0.45	53
E. marginata ftsH 1	2.32	0.73	43	6.98	0.69	40
E. marginata ftsH 2	2.40	0.71	47	2.93	0.70	47
E. marginata ftsH 3	3.45	0.74	37	5.49	0.50	31
E. marginata ftsH 4	2.18	0.32	32	2.66	0.32	32
E. marginata ftsH 5	6.14	0.26	17	7.77	0.26	19
P. canaliculata ftsH 1	3.37	0.78	42	3.01	0.63	37
P. canaliculata ftsH 2	3.29	0.76	38	3.01	0.62	35
P. ocellatus ftsH 1	5.89	0.11	9	8.72	0.07	4
P. ocellatus ftsH 2	4.00	0.77	36	3.69	0.64	37
P. ocellatus ftsH 3	3.93	0.80	39	3.01	0.66	38
P. ocellatus ftsH 4	2.14	0.41	48	14.83	0.25	29

Table 7

Prediction of Subcellular Locations using TargetP-2.0 (Armenteros et al., 2019): Signal peptides and mitochondrial transfer peptides. We predicted the presence of N-terminal presequences for each gastropod ftsH with the non-plant option and for ftsH 8 from A. thaliana and ftsH from V. litorea with the plant option.

Name	Other	Signal peptide	Mitochondrial transfer peptide
A. californica ftsH 2	0.3688	0.0001	0.6311
A. californica ftsH 3	0.6524	0.0031	0.3445
A. californica ftsH 3	0.6524	0.0031	0.3445
A. californica ftsH 1	0.6123	0.0003	0.3874
A. vulgaris ftsH 1	0.6967	0.0069	0.2964
A. vulgaris ftsH 2	0.9777	0.0000	0.0223
A. vulgaris ftsH 3	0.0598	0.0001	0.9400
B. stephanieae ftsH 1	0.2511	0.0001	0.7489
B. stephanieae ftsH 2	0.9992	0.0001	0.0008
C. unifasciata ftsH 1	0.4583	0.0077	0.5341
C. unifasciata ftsH 3	0.0240	0.9751	0.0009
C. unifasciata ftsH 1	0.4583	0.0077	0.5341
E. chlorotica ftsH 1	0.6424	0.0001	0.3575
E. chlorotica ftsH 2	0.6502	0.1386	0.2112
E. chlorotica ftsH 3	0.1240	0.8681	0.0079
E. crisparta ftsH 1	0.9782	0.0003	0.0215
E. crisparta ftsH 2	0.265	0.0001	0.7349
E. marginata ftsH	0.5119	0.0008	0.4872
E. marginata ftsH 1	0.9997	0.0001	0.0002
E. marginata ftsH 2	0.7368	0.0002	0.2630
E. marginata ftsH 3	0.9993	0.0004	0.0003
E. marginata ftsH 4	0.9995	0.0000	0.0004
E. timida ftsH 3	0.9951	0.0020	0.0029
E. timida ftsH 1	0.4597	0.0003	0.5400
E. timida ftsH 2	0.7233	0.0013	0.2754
P. canaliculata ftsH 1	0.7977	0.0001	0.2022
P. canaliculata ftsH 2	0.7347	0.0122	0.2531
P. ocellatus ftsH 1	1.0000	0.0000	0.0000
P. ocellatus ftsH 2	0.8854	0.0013	0.1132
P. ocellatus ftsH 3	0.3658	0.0001	0.6341
P. ocellatus ftsH 4	0.9999	0.0001	0.0000
V. litorea ftsH	0.0511	0.9487	0.0002

Table 8

Gene region comparison across various chloroplasts. No ftsH found in chloroplasts from Arabidopsis thaliana, Bryopsis hypnoides and Klebsormidium nitens.

Chlorella vulgaris	Cyanidioschyzon merolae	Gonium pectorale	Pycnococcus provasolii	Vaucheria litorea
psbL	psbB	psaJ	ycf4	trnW_cca
psbF	psbT	atpI	rnpB	rpl11
cell division protein	psbN	psbJ	trnW(cca)	rpl1
psbE	psbH	pasA	trnS(uga)	rpl12
no blast match	psaE	psbD	trnS(gcu)	psaE
ftsH	ftsH	ftsH	ftsH	ftsH
trnT	trnD	psbC	trnP(ugg)	psbB
cysA	trnS	trnH	rpl23	psbT
AccD	acpP	trnF	rpl2	psbN
psaI	ycf86	psaC	rps19	psbH
trnV	accB	petL	rps3	petF