



**HAL**  
open science

# Phylogenetic and Structural Analyses of VPS13 Proteins in Archaeplastida Reveal Their Complex Evolutionary History in Viridiplantae

Sébastien Leterme, Olivier Bastien, Riccardo Aiese Cigliano, Alberto Amato, Morgane Michaud

## ► To cite this version:

Sébastien Leterme, Olivier Bastien, Riccardo Aiese Cigliano, Alberto Amato, Morgane Michaud. Phylogenetic and Structural Analyses of VPS13 Proteins in Archaeplastida Reveal Their Complex Evolutionary History in Viridiplantae. *Contact*, 2023, 6, pp.1-23. 10.1177/25152564231211976 . hal-04323023

**HAL Id: hal-04323023**

<https://hal.science/hal-04323023v1>

Submitted on 5 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Phylogenetic and Structural Analyses of VPS13 Proteins in Archaeplastida Reveal Their Complex Evolutionary History in Viridiplantae

Contact  
Volume 6: 1–23  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/25152564231211976  
journals.sagepub.com/home/ctc



Sébastien Leterme<sup>1</sup>, Olivier Bastien<sup>1</sup>, Riccardo Aiese Cigliano<sup>2</sup>,  
Alberto Amato<sup>1</sup> , and Morgane Michaud<sup>1</sup> 

## Abstract

VPS13 is a lipid transfer protein family conserved among Eukaryotes and playing roles in fundamental processes involving vesicular transport and membrane expansion including autophagy and organelle biogenesis. VPS13 folds into a long hydrophobic tunnel, allowing lipid transport, decorated by distinct domains involved in protein localization and regulation. Whereas VPS13 organization and function have been extensively studied in yeast and mammals, information in organisms originating from primary endosymbiosis is scarce. In the higher plant *Arabidopsis thaliana*, four paralogs, AtVPS13S, X, M1, and M2, were identified, AtVPS13S playing a role in the regulation of root growth, cell patterning, and reproduction. In this work, we performed phylogenetic, as well as domain and structural modeling of VPS13 proteins in Archaeplastida in order to understand their general organization and evolutionary history. We confirmed the presence of human VPS13B orthologues in some phyla and described two new VPS13 families presenting a particular domain arrangement: VPS13R in Rhodophytes and VPS13Y in Chlorophytes and Streptophytes. By focusing on Viridiplantae, we were able to draw the evolutionary history of these proteins made by multiple gene gains and duplications as well as domain rearrangements. We showed that some Chlorophytes have only three (AtVPS13M, S, Y) whereas some Charophytes have up to six VPS13 paralogs (AtVPS13M1, M2, S, Y, X, B). We also highlighted specific structural features of VPS13M and X paralogs. This study reveals the complex evolution of VPS13 family and opens important perspectives for their functional characterization in photosynthetic organisms.

## Keywords

VPS13, lipid transport, photosynthetic organisms, viridiplantae, phylogeny, evolution

## Introduction

The VPS13 protein family was first discovered in 1977 in a mutant screen for loss of Carboxypeptidase Y (CPY) activity in yeast (Jones, 1977). Since, VPS13 proteins have been identified in many organisms and suspected to be present in all Eukaryotic cells. The yeast *Saccharomyces cerevisiae* has only one paralog (ScVPS13) whereas human has four (HsVPS13A/B/C/D). In the last decades, VPS13 functional analyses have demonstrated their role in a wide array of cellular processes including vesicular transport, autophagy, mitochondria homeostasis or lipid droplet (LD) biogenesis (reviewed in Dziurdzik and Conibear, 2021; Leonzino et al., 2021) and their mutations are associated to severe neurodegenerative diseases in human (reviewed in Ugur et al., 2020; Leonzino et al., 2021). This plethora of functions is in accordance with the wide variety of subcellular localizations observed at single organelles (ER, mitochondria, endosomes, LD, Golgi...), as well as at multiple contact sites between organelles (ER-mitochondria, ER-LD, mitochondria-endosomes...) (Dziurdzik and Conibear, 2021; Neuman et al., 2022). Recently, further structural, biochemical, and *in silico*

approaches led to the characterization of VPS13 as lipid transfer proteins (LTPs) involved in bulk lipid transport and membrane expansion (Levine, 2022; Melia and Reinisch, 2022; Neuman et al., 2022). Indeed, VPS13 structural analyses by cryo-EM and 3D predictions showed that they form a long hydrophobic tunnel able to accommodate tens of glycerolipid molecules with a low specificity and that the VPS13 N-terminal (N-term) domain is able to transfer lipids between liposomes *in vitro* (Kumar et al., 2018; Li et al., 2020a; Melia and Reinisch, 2022; Neuman et al.,

<sup>1</sup>Université Grenoble Alpes, CNRS, CEA, INRAE, IRIG, LPCV, Grenoble, France

<sup>2</sup>Sequentia Biotech SL, Barcelona, Spain

Received May 3, 2023. Revised October 9, 2023. Accepted October 18, 2023.

## Corresponding Author:

Morgane Michaud, Université Grenoble Alpes, CNRS, CEA, INRAE, IRIG, LPCV, 38000 Grenoble, France.  
Email: morgane.michaud@cea.fr



2022). VPS13 are massive proteins (more than 3000 amino acids (aa)) and their structure is organized around a repetition of Repeating  $\beta$ -Groove (RBG) domains forming a hydrophobic tunnel, likely mediating lipid transport, with addition of extra-domains inserted between RBG repeats and involved in the regulation of protein localization and function (Adlakha et al., 2022; Levine, 2022; Neuman et al., 2022; Dall'Armellina et al., 2023). The RBG domain core is composed of five antiparallel  $\beta$ -strands forming a concave  $\beta$ -sheet with a hydrophobic surface followed by a variable loop that can be disorganized or folded into a helix (Levine, 2022). The number of RBG domain repeats was proposed to determine VPS13 hydrophobic tunnel length, 12 repeats giving an average length of 20 nm (Levine, 2022; Neuman et al., 2022). The canonical domains found in VPS13 proteins are (1) a N-term Chorein, involved in protein localization and able to transport lipids *in vitro*, (2) a VPS13 adaptor binding (VAB) domain, involved in protein localization to specific organelles, (3) an Autophagy-related protein 2 C-terminus homology (ATG2\_C) domain, likely involved in protein targeting *via* membrane binding, and (4) a C-term Pleckstrin Homology (PH) domain also involved in protein localization *via* protein-protein or protein-lipid interactions (Dziurdzik and Conibear, 2021; Levine, 2022; Neuman et al., 2022). The aforementioned domains are found in almost all the VPS13 proteins characterized to date. Other domains, such as WWE, Ubiquitin-associated (UBA),  $\beta$ -trefoil or  $\beta$ -sandwich domains, putatively involved in ubiquitination pathways or carbohydrate binding, are also found in HsVPS13s (Levine, 2022). Recently, clustering analyses of more than 1,000 VPS13 sequences from different organisms suggested that the last eukaryotic common ancestor (LECA) might harbor two VPS13 paralogs: ancestors of VPS13A/C/D and VPS13B (Levine, 2022). After gene duplication, VPS13D and then VPS13A and C might have emerged from the VPS13A/C/D ancestor. Interestingly, this analysis suggested that VPS13B was mainly found in Metazoa, but also in 25 protists and algae, suggesting it had been conserved in specific non-Metazoan phyla (Levine, 2022).

Whereas ScVPS13 and HsVPS13s are now well characterized, little is known about VPS13 proteins in organisms stemming from primary endosymbiotic events, that is, plants and algae. However, lipid trafficking in primary endosymbiosis derived organisms is highly complex due (1) to the presence of a chloroplast surrounded by two membranes (Petroustos et al., 2014) and (2) to the adaptation to multiple environmental stresses triggering massive lipid remodeling events (Moellering and Benning, 2011; Michaud and Jouhet, 2019). Therefore, the characterization of LTPs in such organisms is an important step toward the understanding of their cell biology and their stress response adaptation. *Arabidopsis thaliana* possesses four paralogs, namely AtVPS13S, M1, M2, and X (Levine, 2022). Interestingly, AtVPS13M1 and M2 harbor several additional domains never observed previously in other VPS13 proteins: a C2 domain, an additional PH in N-term and three  $\beta$ -tripod

domains that might be involved in protein localization mediated by protein-protein or protein-lipid interactions (Levine, 2022). AtVPS13S is the only characterized *A. thaliana* paralog at the time of writing. Initially named SHRUBBY (SHRY), AtVPS13S interacts with the transcription factor SHORT ROOT and plays a role in a pathway regulating root growth and radial patterning, even if the exact mechanism is not understood (Koizumi and Gallagher, 2013). Knock out *atvps13S* are also sterile, showing an important role in plant reproduction (Koizumi and Gallagher, 2013). In a recent pre-print, AtVPS13M2 was suggested to be involved in pollen germination and pollen tube elongation by regulating vesicular transport to the pollen tube tip (Tangpranomkorn et al., 2022, BioRxiv, DOI:10.1101/2022.11.01.514778). A VPS13-like gene has also been shown to play a role in apomixis, a plant reproductive process, in the angiosperm *Taraxacum* (Van Dijk et al., 2020). Apart from the above-mentioned studies, nothing is known about VPS13 paralogs (e.g., gene number, domain organization, functions) in organisms originating from primary endosymbiosis.

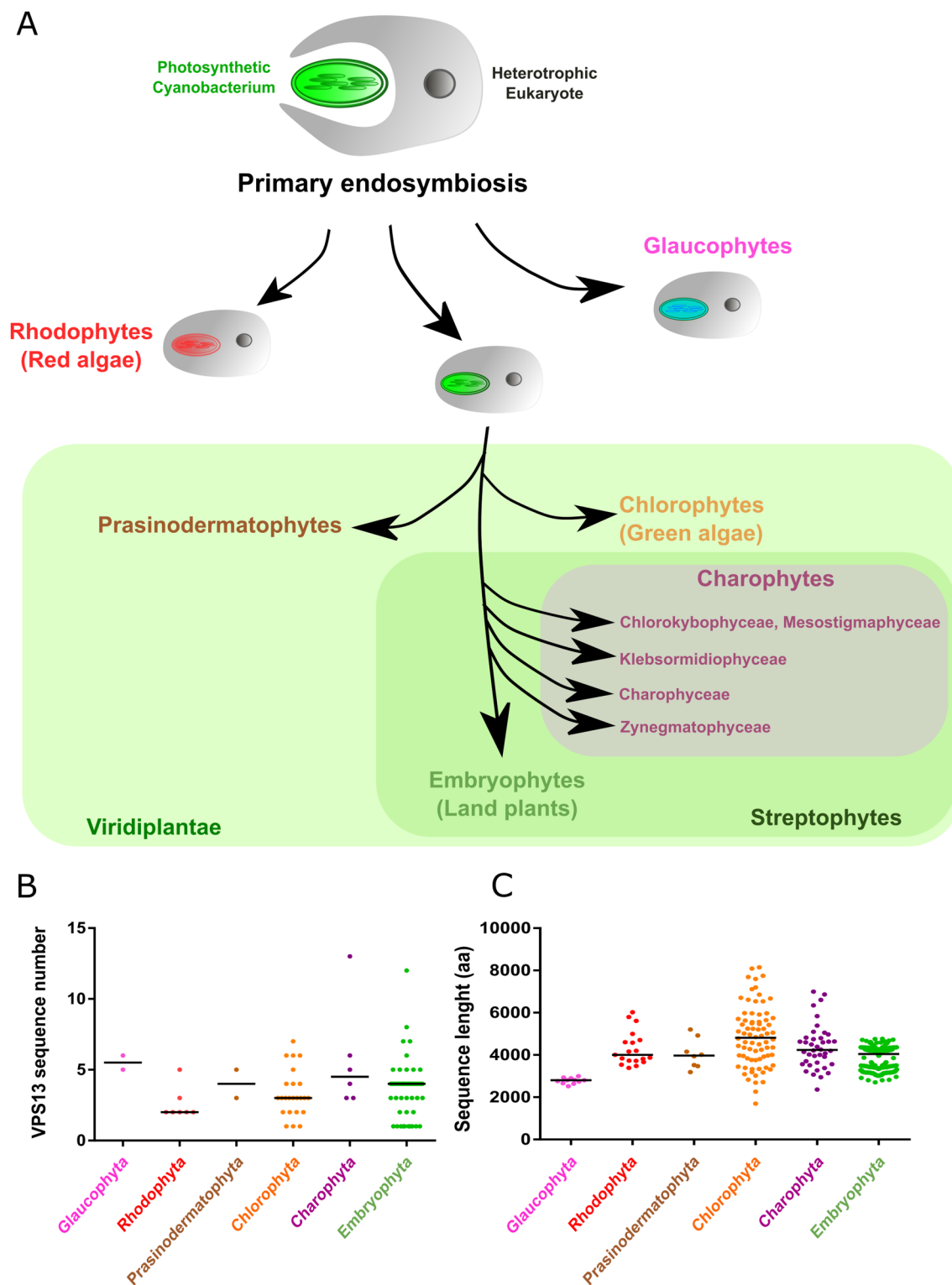
All photosynthetic organisms and some of their non-photosynthetic descendants stemmed from primary and secondary endosymbiotic events (Keeling, 2010) and literature therein). Somewhere between 1600 (Yoon et al., 2004) and 900 million years ago (Ma) (Shih and Matzke, 2013), a cyanobacterium was engulfed into a Eukaryotic heterotroph and, instead of being digested, took permanent residence into the latter. This gave rise to the so-called primary chloroplasts of the three Archaeplastida lineages: the Viridiplantae, the Glaucophytes, and the Rhodophytes (red algae) (Figure 1) (Adl et al., 2012). From a Viridiplantae ancestor emerged three clades: the Prasinodermatophytes, the Chlorophytes (green algae), and the Streptophytes (plants), which is itself subdivided into Embryophytes (land plants) and Charophytes, algae being considered as land plants ancestors (Li et al., 2020b; Domozych and Bagdan, 2022; Bowles et al., 2023).

Here, we present a series of phylogenetic, bioinformatics, and structural modeling analyses aimed at deciphering the evolutionary history of VPS13 proteins in Archaeplastida in terms of copy number, domain organization and structure. We identified and described two new VPS13 types in Chlorophytes and Charophytes (VPS13Y) as well as in red algae (VPS13R). We also confirmed the presence of HsVPS13B orthologues in diverse groups of Viridiplantae and highlighted divergences in structure prediction. We finally focused on the Viridiplantae clade for which we were able to reconstruct VPS13 paralogs evolutionary history, domain organization and structure.

## Results

### Phylogeny of VPS13 Proteins in Archaeplastida

Sequences were retrieved by BLASTP searches (Altschul, 1997) in different databases using *A. thaliana* or



**Figure 1.** Schematic representation of Archaeplastida evolutionary history and VPS13 copy number and length of the sequences analyzed in this study. (A) A primary endosymbiosis of a cyanobacterium by a heterotrophic Eukaryote gave rise to three Archaeplastida lineages: the Viridiplantae, the Glaucophytes, and the Rhodophytes (red algae). From a Viridiplantae ancestor three clades emerged: the Prasinodermatophytes, the Chlorophytes (green algae), and the Streptophytes (plants), which is itself subdivided into Charophytes, algae being considered as ancestor of land plants, and Embryophytes (land plants). Within Charophytes, the Mesostigmaphyceae and Chlorokybophyceae groups are basal and Zynegmatophyceae is considered as a sister group of land plants. Copy numbers (B) and length (C) of VPS13 sequences in the different clades analyzed in this study. Data were generated from Dataset 1. The bar represents the median.

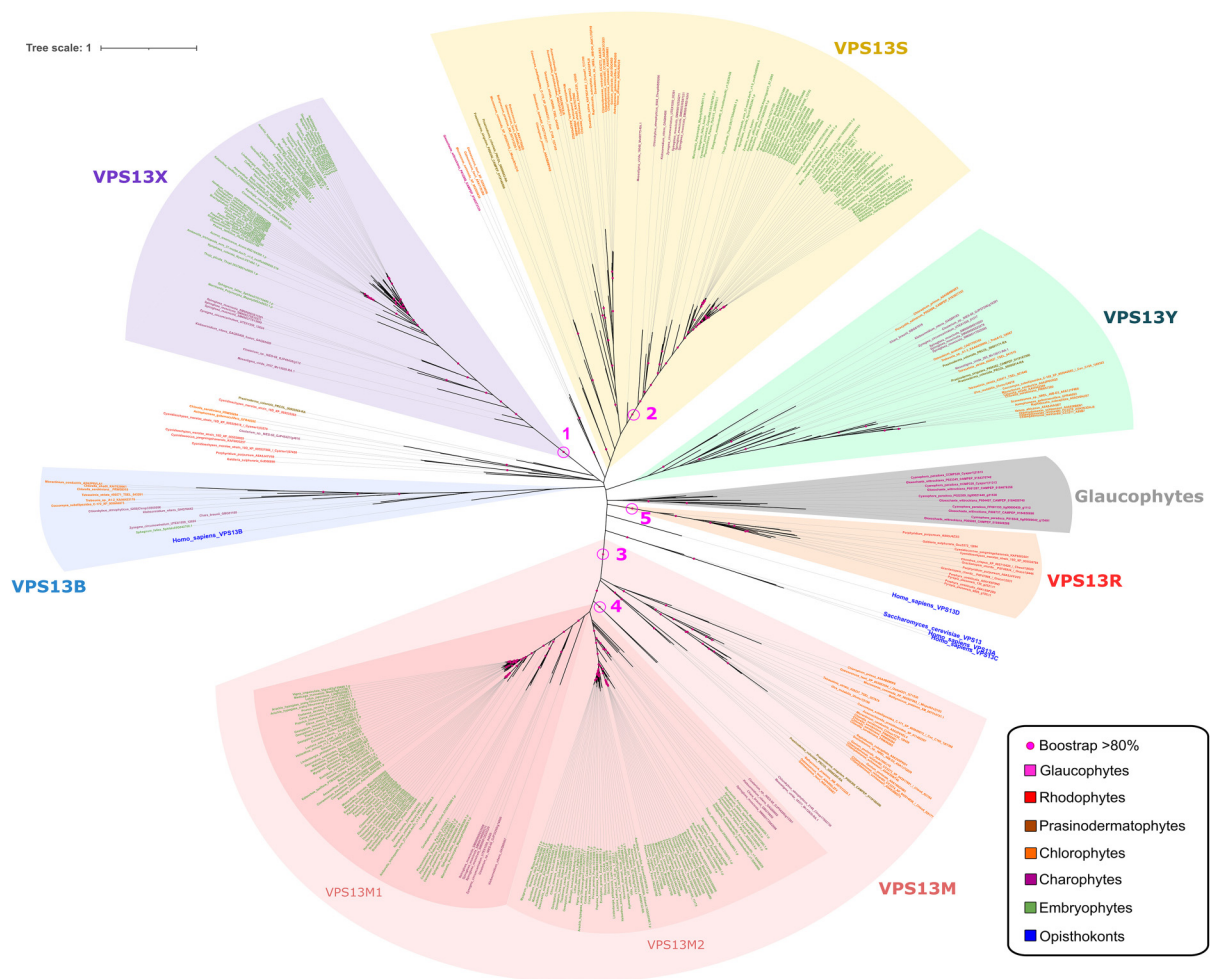
HsVPS13B sequences as queries (see Material and Methods). Only sequences with at least one VPS13 canonical domain annotation were retrieved to populate a first dataset of 537 sequences (Table S1). However, numerous sequences were fragmented (size below 2000 amino acids), likely because of poorly assembled genomes and to errors performed by automatic genome annotation that failed to properly define the correct borders of such long proteins that contain many introns/exons (63 exons for AtVPS13M1, for example). An extreme case is illustrated in Figure S1 where a *VPS13* gene in *Coffea arabica* on scaffold 440 is divided in at least seven successive gene models (evm.model.Scaffold\_440.43 to evm.model.Scaffold\_440.49) with predicted Vacuolar sorting associated protein 13 domains (IPR026847/PTHR16166) and translated to proteins ranging from 96 to 2207 aa. Some particularly long sequences were retrieved corresponding to different protein model predictions of the same gene. This is the case as example in *Chlamydomonas reinhardtii* in which two sequences, A0A2K3DAJ8 (8089 aa) and A8IIB7 (7192 aa), corresponding to the same *VPS13* gene with different exon/intron border prediction, were retrieved (Table S1). In order to perform analyses with a significant number of likely full-length sequences from Archaeplastida,  $\geq 2400$  aa long sequences were selected to create a dataset containing 383 sequences (Dataset 1). Within this dataset, major variations in gene copy number and sequence length are observed among clades (Figure 1B and C). Nevertheless, these results have to be taken with caution as the dataset was not manually curated and hence some paralogs might be missing, might correspond to different models of the same gene, or might be incomplete. The highest number of VPS13 sequences is observed in the Charophyte *Spirogloea muscicola* (13 sequences) and the Embryophyte *Tinopyrum intermedium* (12 sequences) (Figure 1B). Limited variations of sequence length are observed for Embryophytes and the Glaucophytes in contrast to other clades (Figure 1C) and sequences more than 7000 aa long are recorded in some Chlorophytes (Figure 1C) as previously reported (Levine, 2022). Noteworthy, the 13 Glaucophyte sequences are very short (below 3000 aa), although according to domain analysis (see below), most of them are full length (i.e., presence of a N-term Chorein and a C-term PH domains).

A maximum likelihood (ML) phylogenetic analysis was then performed adding HsVPS13A through D and ScVPS13 to the dataset in order to infer their relationship with those from Archaeplastida. Because of the length of the alignment, 5,000 bootstrap pseudoreplicates were performed and nodes supported by a bootstrap value above 4,000 (80%) were considered as significant. The unrooted tree presented in Figure 2 shows the complex evolutionary history of VPS13 proteins in each phylum. VPS13 proteins show high primary sequence diversity and no protein family present in all of the three Archaeplastida groups can be unambiguously defined. Embryophyte VPS13 proteins clearly split into four clades likely corresponding to the

AtVPS13M1, M2, S, and X paralogs described in *A. thaliana* (Levine, 2022). The well-supported VPS13X and VPS13S clades (Figure 2, nodes no. 1 and 2, respectively) also contain Charophyte sequences, including *Mesostigma viride* and *Chlorokybus atmophyticus*, suggesting that VPS13X and S paralogs were present in Streptophytes ancestor. A structured clade that we named “VPS13M” (Figure 2, node no. 3) contains the Embryophytes VPS13M1 and M2 as well as Charophytes, Chlorophytes, and Prasinodermatophytes sequences. As expected, Chlorophytes and Prasinodermatophytes are basal to the Streptophytes group and have only one VPS13M. Some species have more VPS13M, such as *Ostreococcus tauri* or *Chlorella sorokiniana*, but sequence alignment showed that they likely correspond to different protein model predictions for the same gene. The Streptophytes analyzed seem to have a VPS13M1 and M2 paralogs except the Charophytes *C. atmophyticus* and *M. viride* that have one VPS13M branching before the VPS13M1/M2 split (Figure 2, node no. 4). This implies that the Viridiplantae ancestor had one VPS13M which was duplicated after the emergence of Chlorokibophyceae and Mesostigmaphyceae.

Another supported clade contains only Rhodophytes’ sequences and was named “VPS13R” (Figure 2, node no. 5). Unsupported (bootstrap values below 80%) clades include (1) Glaucophytes, (2) a variably supported clade clustering several Prasinodermatophytes, Chlorophytes, and Charophytes sequences, later named “VPS13Y” (see below), (3) a clade, basal to VPS13S, composed of Prasinodermatophytes and Chlorophytes, and (4) two orphan clades (Figure 2). Interestingly, whereas ScVPS13 and HsVSP13A/C clustered together and VPS13D branched alone, HsVPS13B clustered with sequences from Charophytes, Chlorophytes and from the Bryophyte *Sphagnum fallax*, nonetheless with a weak support, corroborating the finding that VPS13B might be present in a small subset of organisms besides Metazoa (Levine, 2022). Therefore, VPS13B was likely present in the LECA.

In order to more confidently trace the evolution of the VPS13 paralogs in Archaeplastida, a Bayesian and a ML analyses were performed on a reduced dataset of 104 sequences composed of two Glaucophytes, three Rhodophytes, two Prasinodermatophytes, one organism in each of the five Charophyte groups pictured in Figure 1, four Chlorophytes and eight Embryophytes (including lower and higher plants, one gymnosperm and angiosperms, monocotyledons, and dicotyledons) as well as the human and yeast VPS13 sequences (Dataset 2). All the sequences in Dataset 2 were manually curated to eliminate redundant sequences. As VPS13 primary sequences can be highly divergent but share a conserved overall 3D structure and domain arrangement, we mathematically formalized the domain occurrence for each sequence and included such information in the Bayesian analysis (see Material and Methods). Remote homology analyses using HHPred (Zimmermann et al., 2018) were performed for all the sequences included in the



**Figure 2.** Maximum likelihood phylogenetic analysis of VPS13 proteins from Archaeplastida. Organism names are colored according to their taxonomic clade. The analysis was performed with 5000 bootstrap pseudoreplicates and nodes supported by more than 4000 bootstraps (80%) are highlighted by a pink dot. The supported nodes discussed in the text are highlighted by a pink circle and a number. The main VPS13 groups that will be discussed in this work (VPS13S, X, M (M1 and M2), R, Y, B and Glaucophyte) are highlighted. Only the groups VPS13M, X, R, and VPS13S in Streptophytes are well supported by this analysis. The other groups are more supported by the Bayesian analyses performed thereafter.

dataset in order to identify putative domains. Most of the canonical VPS13 domains such as Chorein, VAB, or ATG2\_C were found with additional domains for some sequences that will be further discussed below (Table S2). VPS13B clade was used to root the trees (Figure S2 and S3). These analyses confirmed that no group including sequences from the three Archaeplastida phyla (i.e., Viridiplantae, Glaucophyte, Rhodophyte) was found, showing that VPS13 paralogs have highly diverged. Eight clades, namely M, S, X, B, R, Glaucophyte, and opisthokont were highlighted by both the Bayesian (Figure S2) and the ML (Figure S3) analysis. However, the analyses do not allow to establish any relationship among the different VPS13 groups. The opisthokonts' sequences HsVPS13A/C/D and ScVPS13 produced a separate clade whereas VPS13B clade included a wide variety of organisms

(Chlorophytes, Charophytes, and one Embryophyte). Despite a low support (PP = 0.72 (Figure S2) and 16.5% bootstraps (Figure S3)), those sequences clustered together in all the analyses performed and the *Chara braunii* sequence GBG81180 was also identified as a VPS13B orthologue in (Levine, 2022) further suggesting the presence of this paralog in Viridiplantae. A high selection pressure may occur on VPS13B sequences during evolution in contrast to other VPS13 paralogs. However, VPS13B was possibly lost independently several times during evolution (see below) because it is currently only present in a small number of phyla other than Metazoa. These analyses confirmed the existence of a clade grouping several Rhodophyte sequences, identifying a Rhodophyte-specific VPS13 protein. The cluster was named “R” (for Rhodophyte) (Figure S2 and S3), suggesting a high

conservation of VPS13R orthologues during Rhodophyte evolution. VPS13R is not the only VPS13 found in red algae, indeed, other Rhodophyte sequences produced orphan clades (Figures S2 and S3). Such proteins may have specifically evolved in the different Rhodophyte lineages. VPS13R sequences show a common domain organization with the presence of a  $\beta$ -sandwich, and sometimes a UBA domain, downstream of the C-term PH domain (Figure S2). The  $\beta$ -sandwich and UBA domains have been acquired independently during evolution because they are also found in HsVPS13B and D, respectively, but upstream of the VAB. Glaucophyte sequences mostly clustered together, nonetheless with a low support (PP = 0.77 (Figure S2) and 45.3% (Figure S3)). VPS13 paralogs in this clade maintained a very simple organization and only the canonical Chorein, VAB, ATG\_C, and PH domains are found (Figure S2). VPS13S, M, Y, and X clades are present only in Viridiplantae and will be discussed below.

### Bayesian Analysis of Viridiplantae VPS13 Paralogs

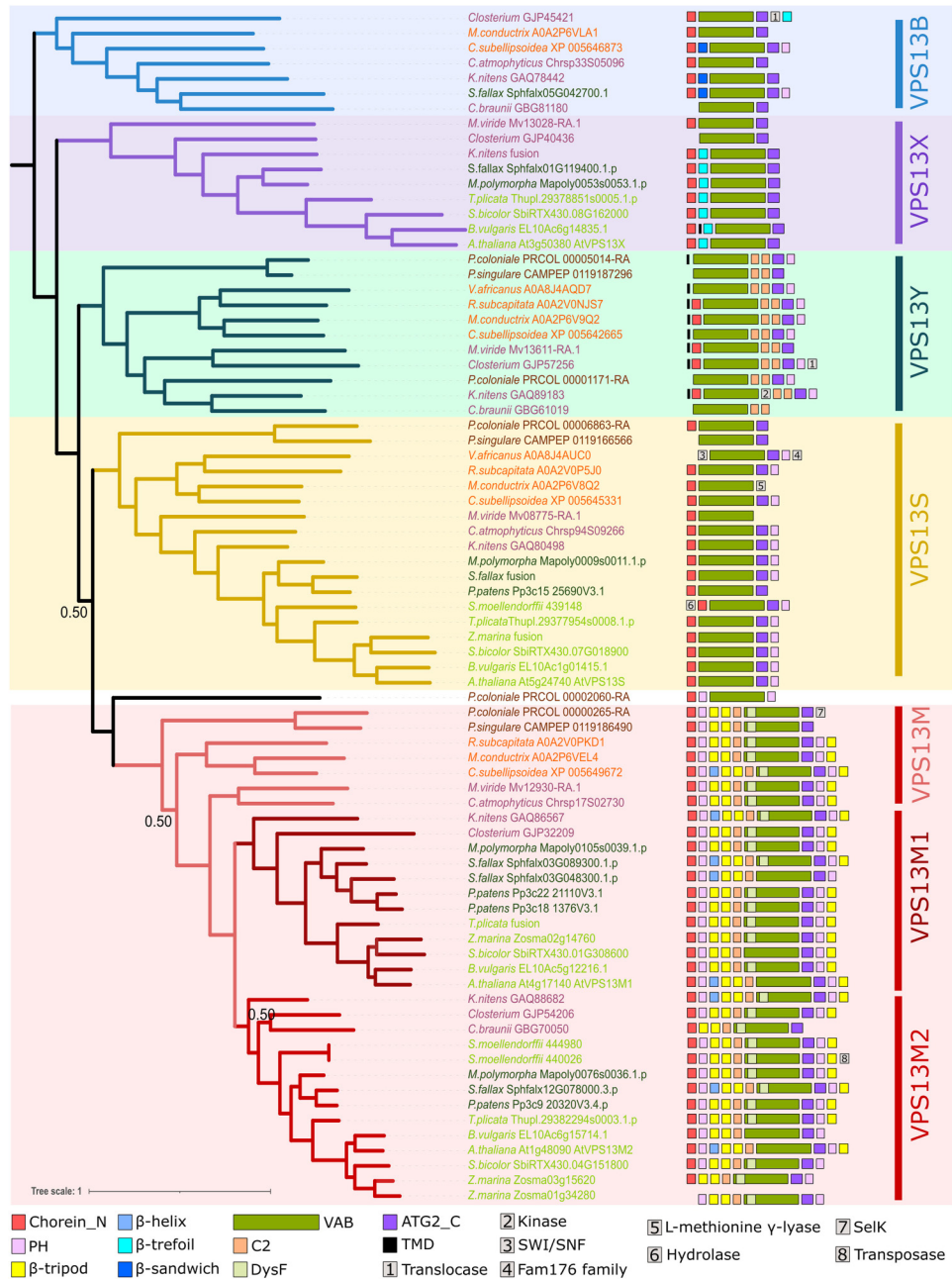
From the results presented above, Viridiplantae VPS13 family showed an interesting evolutionary history, hence a reduced dataset (Dataset 3, Table S3, 79 sequences) including Embryophytes, Streptophytes, Chlorophytes, and Prasinodermatophytes was produced and analyzed as described above. Human and yeast sequences were excluded to avoid destabilization of the analysis. The tree presented in Figure 3 shows that VPS13 protein sequences clustered into five robustly supported (PP = 1.00) clades, namely M, in turn divided into three sub-clades (M, M1, and M2), S, X, B, and Y. VPS13Y has never been reported before. We named it “Y” because no clear and unambiguous homology with other known VPS13 protein groups was revealed by this analysis. Within each VPS13 clade, the organization of the branches is in agreement with the expected taxonomy: Prasinodermatophytes and Chlorophytes are always basal to Streptophytes within which Charophytes are basal to Embryophytes, with non-vascular plants being well separated from vascular plants, showing a linear evolution of VPS13 proteins in each phylum. The only exception is VPS13B which groups incongruously with the expected taxonomy, suggesting an unresolved evolutionary pattern.

The VPS13M and S were found in all species analyzed showing that they were conserved through the evolution of Viridiplantae. VPS13S was reported as single copy in all the organisms analyzed. These analyses confirmed the presence of only one VPS13M gene in Chlorophytes and basal Charophytes (*M. viride* and *C. atmophyticus*) whereas other Streptophytes have two paralogs (M1 and M2), showing that VPS13M was likely duplicated before the emergence of Klebsormidiophyceae. Overall, only one VPS13M1 and M2 paralogs were found with some exceptions. A duplication of VPS13M1 paralog early in the evolution of Bryophytes has occurred because the two mosses

*Sphagnum fallax* and *Physcomitrium patens* possess two VPS13M1. VPS13M2 duplicated likely independently in different phyla during evolution because two paralogs were found in the monocotyledon *Zostera marina* as well as in the basal vascular plant *Selaginella moellendorffii*. VPS13X clade contains sequences belonging only to Streptophytes, including *M. viride*. Despite our advanced BLASTP searches, no VPS13X homologous sequences could be found in any Chlorophyte genome, suggesting that VPS13X arose early in the evolution of Streptophytes. Interestingly, the VPS13Y clade contains Prasinodermatophytes, Charophytes and Chlorophytes proteins only, revealing that VPS13Y was likely present in the Viridiplantae most recent common ancestor (MRCA) and lost before the emergence of Embryophytes.

The results of our analysis focused on Viridiplantae confirmed that VPS13B orthologues are likely present only in a small subset of organisms spread on different Viridiplantae clades putting forward that a VPS13B paralog was initially present and independently lost several times during evolution (Levine, 2022). Four Charophytes species out of the five analyzed here possess a VPS13B. The putative *Closterium* VPS13B orthologue sequence (GJP45421) does not always cluster with other VPS13B (Figure 2 and Figure S3) but was also clustering with HsVPS13B in the analysis performed by Levine (2022), supporting its affiliation to the VPS13B group. Among Chlorophytes, VPS13B was detected in both Trebouxiophyceae species analyzed (*Coccomyxa subellipsoidea* and *Micractinium conductrix*), but not in the Chlorophyceae (*Volvox africanus* and *Raphidocelis subcapitata*). In addition, a VPS13B orthologue was detected in the moss *Sphagnum fallax* (Embryophyte). In order to better understand when VPS13B has been lost in the different Viridiplantae phyla, Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) (Altschul, 1997) searches were performed in Viridiplantae using either HsVPS13B or the six VPS13Bs found in this study as queries on NCBI (Sayers et al., 2022). Only a few Trebouxiophyceae (*Coccomyxa*, *Micractinium*, *Trebouxia*, *Auxenochlorella*, *Chlorella*) sequences were retrieved (Supplemental File S1), confirming that VPS13B gene might have been conserved in this lineage during evolution. One sequence was also found in *Tetraselmis striata* (Figure 2) and in a basal Chlorophyte group, namely the Pedinophyceae, further suggesting that VPS13B was likely present in the MRCA of Chlorophytes and not gained by horizontal gene transfer in Trebouxiophyceae. The only other Embryophyte species possessing a VPS13B orthologue is *Sphagnum magellanicum*, indicating that VPS13B was maintained only in this plant genus.

Overall, our results suggest that the Viridiplantae MRCA had four VPS13 paralogs: VPS13M, S, B, and the newly identified Y. In Chlorophytes, VPS13B was lost in different clades but retained in several Trebouxiophyceae and likely in a small number of other groups, such as Pedinophyceae. VPS13X appeared early in the evolution of Streptophytes, its putative origin will be discussed below. After the



**Figure 3.** Bayesian phylogenetic analysis of VPS13 proteins in Viridiplantae. The domain organization, predicted using HHPred, was included as a character in the analysis and is represented on the right. Recurrent domains are represented by a colored square whereas others are represented by a number. Only the Bayesian posterior probability values below 0.95 are presented on the tree. Supported groups of VPS13 paralogs between them is summarized on a scheme below the tree. Taxonomic clades to which each sequence belongs is indicated by a color code and evolutionary relationships between them is summarized on a scheme below the tree. ATG2\_C = Autophagy-related protein 2 C-terminal homology; PH = Plekstrin Homology; TMD = transmembrane domain, VAB = VPS13 adaptor binding.



emergence of Mesostigmaphyceae and Chlorokybophyceae, VPS13M was duplicated to give rise to VPS13M1 and M2. Thus, Streptophytes' MRCA likely had five VPS13 paralogs: VPS13M, S, Y, X, and B, this hypothesis being supported by the presence of the five paralogs in the Charophytes *M. viride* and *Klebsormidium nitens*, the latter also having the VPS13M1 and M2 paralogs arising from VPS13M duplication. Then, VPS13Y was likely lost early at the emergence of Embryophytes. The presence of a VPS13B protein in two *Sphagnum* species suggests that a VPS13B paralog was present in the MRCA of Embryophytes. Then, VPS13B gene was lost in most of the Embryophyte phyla but maintained in some Bryophytes like *Sphagnum*. Thus, our results support the hypothesis that the Embryophytes' MRCA had five VPS13 paralogs, namely VPS13M1, M2, S, X, and B. Whereas our analysis confirmed the presence of HsVPS13B orthologues in non-metazoans (Levine, 2022), we were not able to establish significant relationships among the other ancestral Viridiplantae VPS13 groups (i.e., M, S, X, and Y). Of note, one *Prasinoderma coloniale* sequence (PRCOL\_00002060-RA) does not clearly cluster with one VPS13 group, being sometimes basal to VPS13M group (Figure 3 and Figure S4) or branching with other groups (Figure 2, and Figure S2 and S3). This sequence has a peculiar domain organization with a PH domain upstream of the VAB (Figure 3). Whether this paralog is an ancestral one lost during evolution in other phyla or belongs to a specific Prasinodermatophyte VPS13 group remains to be determined.

### Expression of the Different VPS13 Paralogs in Viridiplantae

Analysis of *A. thaliana* transcriptomic data available in the Bio-Analytic Resources for plant biology (BAR) browser (Winter et al., 2007) showed that the four AtVPS13 paralogs are expressed (data not shown). We next focused on species having the less widespread paralogs VPS13B, Y, and X. For the Chlorophyte *C. subellipsoidea*, putative mRNA sequences corresponding to the four VPS13 paralogs M, S, Y, and B were found in NCBI and a BLASTN search was performed against an Expressed Sequence Tags (ESTs) cluster available on the Joint Genome Institutes (JGI) (Nordberg et al., 2014). EST sequences were found with a high confidence for the four *C. subellipsoidea* VPS13 paralogs showing that they are all expressed and likely functional (Supplemental File S2). Most of the ESTs aligned with the 3' end of the transcripts, but hits were also found all along the sequences (Figure S5). This analysis also revealed errors in the predicted VPS13 gene models as several ESTs support a different intron/exon organization. In addition, an AGO1 domain was initially predicted at the C-term of CsVPS13B protein but EST analyses clearly show that the transcript ends after the PH domain (Figure S5). Thus, ESTs were used to curate CsVPS13 paralog sequences and obtain

more accurate mRNA and protein sequences for further analyses (see below) (Supplemental File S2).

We also investigated the VPS13 paralogs expression in two Charophyte species used in this study (*C. braunii* and *K. nitens*) as well as in *S. fallax*. Assembled transcriptomes were found on NCBI for *Chara braunii* (Bioproject PRJNA492241), *Klebsormidium flaccidum* (Bioproject PRJN242253) and two *Sphagnum* species: *S. recurvum* (Biosample UHLI) and *S. palustre* (Biosample RCBT). TBLASTN searches were performed using respective VPS13 protein sequences as queries. Interestingly, transcripts were identified with a high confidence for all the paralogs (e-value = 0, Table S4). In addition, transcripts covering almost the entire sequences were found for the six VPS13 paralogs in *K. flaccidum*. KfVPS13M2 and Y transcripts were likely full length whereas the others are split into different transcripts likely due to errors in the assembly. Transcripts corresponding to the five *S. fallax* VPS13 paralogs were detected in both *S. recurvum* and *S. palustre*, confirming the presence and the expression of VPS13B in the *Sphagnum* clade. In addition, the two SfVPS13M1 isoforms are also present and expressed in both *Sphagnum* species, showing that they remained functional after VPS13M1 gene duplication. In *C. braunii*, only three paralogs (VPS13B, M2, and X) were previously identified by BLASTP searches and transcripts were found for all with a high confidence (Table S4). We also found other putative hits of shorter length (around 500 aa) and lower confidence that might correspond to other VPS13 paralogs initially undetected by BLASTP searches. Thus, TBLASTN searches on *C. braunii* transcriptome were performed using all VPS13 paralogs from *K. nitens* and *Closterium* as queries. This analysis allowed the detection of putative CbVPS13X paralogs identified in both TBALSTN search and covering around 2000 aa of KnVPS13X. A putative VPS13S was detected only with KnVPS13S but with a low coverage (559 aa). However, no putative M1 sequences were retrieved, maybe due to lower expression level, poor transcript assembly, or the absence of this paralog in *C. braunii*.

Overall, this analysis showed that all the VPS13 paralogs discovered with our phylogenetic analyses were expressed in the Chlorophyte *C. subellipsoidea*, the Charophytes *K. nitens*, and *C. braunii*, and in the Bryophytes *Sphagnum* spp, showing that they might be all functional. In addition, transcriptomic analysis was used to refine VPS13 sequences in *C. subellipsoidea* and to find putative orthologues in *C. braunii* not detected by protein searches, illustrating the interest of analyzing transcriptomic data in the analysis of huge and complex proteins such as VPS13.

### Domain Organization of VPS13 Proteins in Viridiplantae

To go further into the putative evolution of VPS13 proteins, we analyzed their domain organization and structure from

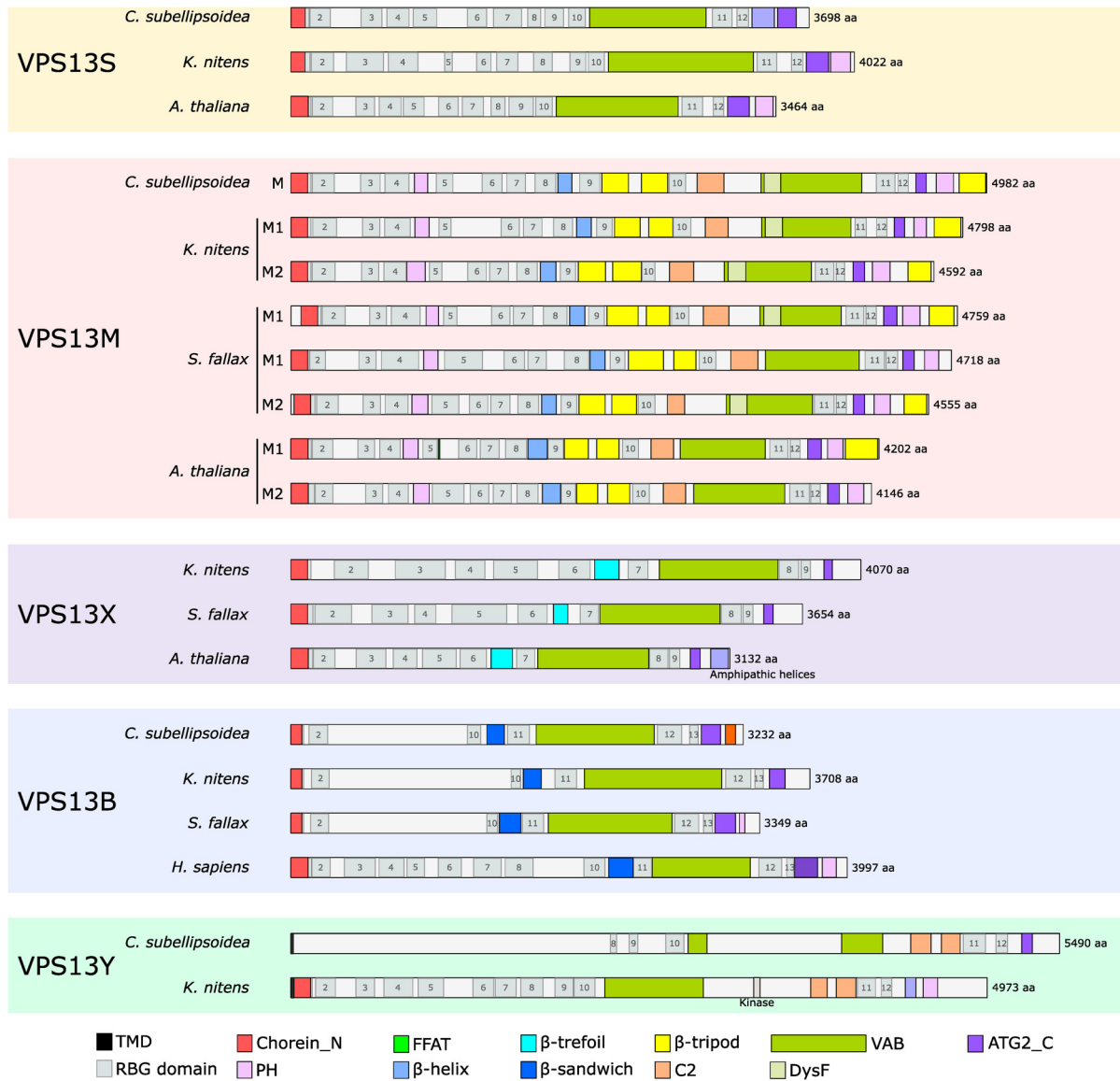
modeling. VPS13 protein structure is organized around a repetition of RBG domains forming a hydrophobic tunnel, likely mediating lipid transport, with addition of extra-domains inserted between RBG domains and involved in the regulation of protein localization and function (Adlakha et al., 2022; Levine, 2022). In order to understand if the general organization of Viridiplantae VPS13 proteins, in particular of VPS13X, VPS13B, and the newly identified VPS13Y paralogs, is conserved, a subset of key organisms (*A. thaliana*, *S. fallax*, *C. subellipsoidea*, and *K. nitens*) in which we confirmed the expression of the paralogs were chosen to carry out RBG domain predictions. RBG domains were predicted using HHPred (Zimmermann et al., 2018) and, as queries, RBG domain sequences previously identified in *A. thaliana* as well as in HsVPS13B for VPS13B orthologues (see Material and Methods) (Levine, 2022).

**VPS13S.** *A. thaliana* VPS13S is composed of 12 RBG repeats accompanied by the canonical VPS13 domains only (Levine, 2022), like most of the VPS13 proteins identified in our analysis, suggesting that the VPS13S ancestor had the following domain organization: Chorein-VAB-ATG2\_C-PH (Figure 4). Additional domains were found in N- or C-term but they could stem from gene model prediction errors as exemplified above for CsVPS13B AGO1 domain in C-term. As an example, a Hydrolase domain was found in N-term of *S. moellendorffii* VPS13S or a methionine  $\gamma$ -lyase domain in the C-term of *Micractinum conductrix* (Figure 3). An extensive analysis of transcriptomic data, if available, might shed some light on the actual occurrence of such unexpected domains. RBG domains analysis showed that VPS13S of *K. nitens* (4022 aa) and *C. subellipsoidea* (3698 aa) also contain 12 RBG domains, even if they are longer than AtVPS13S (3466 aa) therefore an increase in protein size is likely not correlated to an increase of RBG repeats or tunnel length (Figure 4). Overall, the domain organization of VPS13S proteins seems to be conserved across the Viridiplantae phylum without addition of any new domain. It is noteworthy that a similar basic organization is found in the single VPS13 paralog in yeast (Levine, 2022).

**VPS13M.** *A. thaliana* VPS13M1 and M2 proteins are composed of 12 RBG domains and are characterized by the occurrence of up to seven additional domains (Figures 3 and 4) (Levine, 2022). Indeed, a (1) PH domain is inserted between RBG4 and 5 in addition to the canonical one located downstream of the C-term ATG2\_C domain, (2) two  $\beta$ -tripod domains in tandem are inserted between RBG9 and 10 and a third one is present at the C-term end of most of the VPS13M paralogs, (3) a C2 domain is inserted right upstream of the VAB domain, (4) a  $\beta$ -helix domain, and (5) a Dysferlin/Pex24 (DysF, IPR010482) domain is predicted within the VAB of several VPS13M orthologues. The PH domain located in the N-term part

of the protein was identified in almost all the VPS13M sequences analyzed, suggesting it was present in the Viridiplantae VPS13M ancestor (Figure 3). The only organisms lacking this PH domain in the VPS13M2 N-term are the Embryophyte *Zostera marina* and the Charophyte *C. braunii*. This might be linked to a higher sequence divergence that did not allow HHpred to predict this domain or to a loss in some lineages. The tandemly repeated  $\beta$ -tripod domains are present in all the VPS13M, suggesting their presence in the Viridiplantae MRCA. Interestingly, only Angiosperms' VPS13M2 lack the C-term  $\beta$ -tripod domain (Figure 3). Noteworthy, this domain is present in all VPS13M1, in Chlorophytes' VPS13M as well as in Charophytes' VPS13Ms excluding *C. braunii* VPS13M2, lower plants and the Gymnosperm *Thuja plicata*. A loss of the C-term  $\beta$ -tripod domain in Angiosperms is conceivable, possibly to allow the emergence of a new function (see Discussion). Interestingly, the two Prasinodermatophytes VPS13Ms do not possess a predicted  $\beta$ -tripod domain in C-term, neither a PH domain, suggesting a particular evolution of the C-term in this phylum or an error in gene model annotation. The additional C2 domain upstream of the VAB is predicted in all phyla, including Prasinodermatophytes, suggesting it was present in the VPS13M ancestral protein. In the analysis performed by Levine (2022), a  $\beta$ -helix domain was identified between RBG8 and 9. This domain was also detected when HHpred predictions were performed on a reduced portion of the proteins (between 150 and 180 aa) and was confirmed by AlphaFold2 (Jumper et al., 2021; Mirdita et al., 2022) structural prediction for a small subset of VPS13M, M1, and M2 (see below and Figure 5). Thus, the  $\beta$ -helix domain was likely present in all VPS13M and in the common ancestor of VPS13M. Noteworthy, the DysF domain is predicted upstream of or overlapping the VAB domain for almost all the VPS13M orthologues, excluding *A. thaliana* (Figures 3 and 4). Structural analyses (see below) showed that this domain is indeed inserted in the second module of the VAB domain, highlighting an additional feature of VPS13M sequences. Because of its presence in all the Viridiplantae phylum, it was likely present in the VPS13M ancestor and lost in some lineages such as in *A. thaliana*. A FFAT (two phenylalanines in an acidic tract) motif is also present in AtVPS13M1, but not AtVPS13M2 (Slee and Levine, 2019), downstream of RBG5 but we were not able to perform automatic analysis to easily detect FFAT motifs in other VPS13M orthologues in order to assess its conservation. All VPS13M proteins analyzed are composed of 12 RBG domain repeats, showing the high conservation of the VPS13M clade (Figure 4).

Overall, our results indicate that VPS13M ancestor had 12 RBG domains and the following organization: Chorein-PH- $\beta$ -helix- $\beta$ -tripod- $\beta$ -tripod-C2-DysF-VAB-ATG2\_C-PH- $\beta$ -tripod. VPS13M was likely duplicated shortly after the emergence of the Charophyceae of the Mesostigmaphyceae and Chlorokybophyceae groups, giving rise to VPS13M1 and M2. Finally, the C-term  $\beta$ -tripod domain was then lost in Angiosperms' VPS13M2.



**Figure 4.** RBG domain analyses in Viridiplantae VPS13 paralogs. On each sequences, domains and RBG repeats detected using HHPred were indicated, apart for AtVPS13 and HsVPS13B sequences for which the information was retrieved from Levine (2022). Protein size is indicated on the right and expressed as amino acids (aa). ATG2\_C = Autophagy-related protein 2 C-terminus homology; FFAT = two phenylalanine in an acidic track; PH = Plekstrin Homology; TMD = transmembrane domain; VAB = VPS13 adaptor binding.

**VPS13X.** VPS13X has a particular structure in *A. thaliana* as it is composed of only nine RBG repeats and does not possess a canonical PH domain in C-term but amphipathic helices instead (Levine, 2022). Also, a  $\beta$ -trefoil domain was found between RBG6 and 7. Most of the VPS13X proteins analyzed have a similar organization, apart in *Closterium* where no Chorein and  $\beta$ -trefoil domains were detected, maybe because we did not retrieve a full-length protein (Figure 3). *K. nitens* and *S. fallax* VPS13X are also composed of nine RBG domains and are significantly longer than AtVPS13X (3132 aa), being 4792 and 3654 aa long, respectively (Figure 4). The origin of VPS13X is mysterious (see

Discussion) but we can assess that it appeared early in the Streptophyte lineage.

**VPS13B.** In human, VPS13B is composed of 13 RBG repeats and contains the four VPS13 canonical domains and a  $\beta$ -sandwich domain inserted between the RBG10 and 11 (Levine, 2022; Dall'Armellina et al., 2023). In Viridiplantae, the Chorein, VAB, and ATG2\_C are also found. The PH domain was found only in two sequences (*C. subellipsoidea* and *S. fallax*) and the  $\beta$ -sandwich was found by a closer HHPred search and structural prediction performed in *C. subellipsoidea*, *S. fallax*, and *K. nitens*.

The  $\beta$ -sandwich may be considered as an ancestral domain maintained in Viridiplantae VPS13B orthologues (Figure 4). RBG domain prediction shows that RBG1, 2, 10, 11, and 13 were easily detected using HsVPS13B RBG sequences as queries. Surprisingly, we failed to detect any RBG domains in the N-term portion of the proteins between RBG2 and 10 in contrast to HsVPS13B (Figure 4). We hypothesized that this might be linked to a higher divergence of RBG repeats in this region in Viridiplantae and performed a HHPred search for RBG domains using RBGs from AtVPS13 paralogs without any success. Neither specific domains nor motifs were found. Structural predictions also failed to reveal clear RBG domain folding (see below). As AlphaFold2 predictions can be challenging for these proteins (Neuman et al., 2022), we performed secondary structure predictions for *C. subellipsoidea*, *K. nitens*, and *S. fallax* (from N-term to the VAB domain) using Quick2D, which gathers predictions from several softwares (Zimmermann et al., 2018). This analysis showed the presence of several predicted disordered regions as well as some potential  $\beta$ -strands upstream of the RBG10 for CsVPS13B and in KnVPS13B (Supplemental File 3), corroborating the hypothesis that the region between RBG2 and 10 might not fold into the well-described RBG domains. For *C. subellipsoidea*, we were able to detect ESTs and to curate the VPS13B sequence to obtain the probable expressed protein (see above) confirming that the unstructured region exists and is not linked to automatic annotation errors. In contrast to CsVPS13B and KnVPS13B, the secondary structure analysis predicted several successive  $\beta$ -strands followed by a helix in SfVPS13B sequence, suggesting the occurrence of four additional domains folded into  $\beta$ -sheets followed by helices not predicted by HHPred or AlphaFold2 (Supplemental File 3).

Overall, the VPS13B C-term part downstream of the RBG10 as well as the Chorein and RBG2 domains seem to be conserved through evolution whereas the region between RBG2 and 10 are highly divergent in terms of RBG repeats and structure in Viridiplantae compared to HsVPS13B. From this analysis, it is difficult to infer the putative ancestral state (with 13 RBGs as in human or six as in photosynthetic organisms, see Discussion) of VPS13B orthologues.

**VPS13Y.** The newly identified VPS13Y was found in Prasinodermatophytes, Chlorophytes, and Charophytes only. VPS13Y characteristic feature is the presence of two C2 domains in tandem downstream of the VAB domain (Figures 3 and 4). VPS13 canonical ATG2\_C and PH domains in C-term are also found in most of the VPS13Y proteins. No Chorein domain was detected by HHPred in N-term for some species, likely linked to errors in gene model prediction or sequence divergence. Interestingly, using TMHMM2.0 software (Krogh et al., 2001), a transmembrane (TM) helix was predicted at the N-term for most of the VPS13Y sequences. No TM domains were predicted

in other VPS13 paralogs apart in *Beta vulgaris* VPS13X, implying that membrane anchoring might be a VPS13Y group specific feature. Thus, the ancestral VPS13Y protein might present the following domain organization: TM-Chorein-VAB-C2-C2-ATG2\_C-PH. An additional Kinase domain was found in an extra-region located between the VAB and the first C2 KnVPS13Y. In *K. nitens*, we were able to predict the presence of 12 RBG domains (Figure 4), some of them being more divergent and predicted with the help of structural modeling. Unexpectedly, we failed to predict RBG domains in the N-term part of CsVPS13Y, even using KnVPS13Y RBG domains as a query (Figure 4). Thus, we performed secondary structure as well as structural prediction as described for VPS13B paralogs. Both methods predicted the presence of helices as well as disordered regions without  $\beta$ -strands, even in the N-term, in accordance with the absence of predicted Chorein domain by HHPred (see below and Supplemental File 3). The presence of the predicted TM domain, like in other VPS13Y orthologues, and transcriptomic data validate that the sequence is full length, raising question about the function of CsVPS13Y as an LTP. Similarly, the *Prasimoderma coloniale* VPS13Y (PRCOL\_00005014-RA) sequence has two predicted TM domains in N-term, confirming that the sequence is full length, but no Chorein domain is predicted. Thus, VPS13Y C-term seems to be well conserved whereas the N-term is more divergent, as observed for VPS13B.

### VPS13 Protein Structures in Viridiplantae

In order to further investigate the structural organization of the different Viridiplantae VPS13 paralogs, structural predictions were performed with ColabFold, an online AlphaFold2 tool (Jumper et al., 2021; Mirdita et al., 2022) as well as with RoseTTAFold (Baek et al., 2021). Structural prediction of overlapping VPS13 fragments was performed and aligned along the RBG domains, which were robustly predicted, and assembled using ChimeraX (Pettersen et al., 2021) (see Material and Methods). We succeeded in obtaining full-length structures for AtVPS13S, X, and M1. AtVPS13M2 structural prediction was not performed because of its close relationship to AtVPS13M1. AtVPS13M1 structural prediction was a challenging task because of the multiple additional domains that often led to steric clashes when different fragments were assembled. Thus, a combination of six fragments predicted with AlphaFold2 and also RoseTTAFold (Baek et al., 2021) were used in order to build a putative structure without steric clashes.

Structural prediction performed using the sequence from *A. thaliana* shows that VPS13S, M, and X harbor the well-described tunnel structure composed of RBG repeats with the canonical VAB, ATG\_C and PH domains located at the C-term extremity of the tunnel (Figure 5A, Supplemental Files S4-6). In AtVPS13M1, whereas the

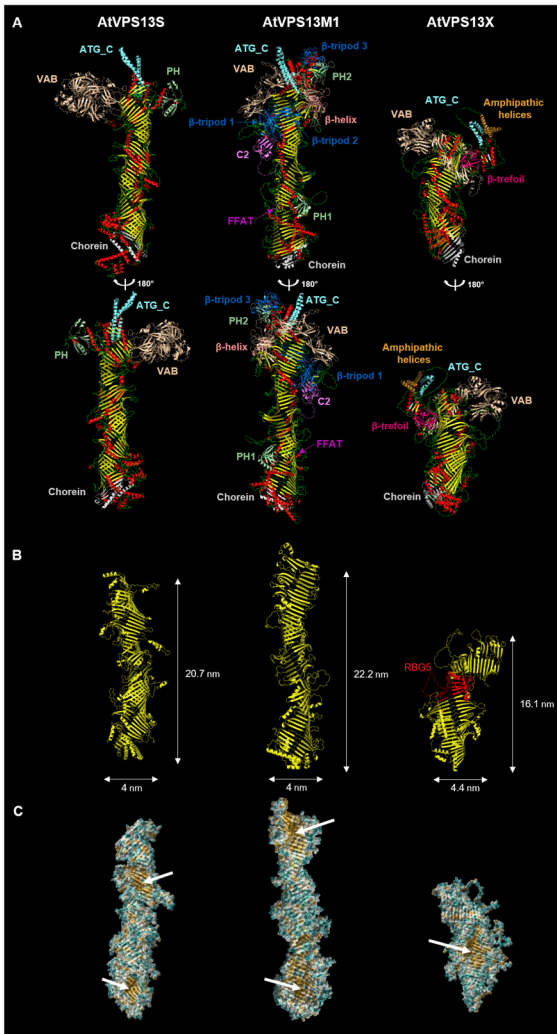
PH1 domain and the FFAT motif are located within the first third of the tunnel, the other domains are all close to the C-term part. The different domains protrude from the tunnel without disturbing its structure and are linked by flexible loops. Thus, the orientation of the domains with respect to each other should be taken with caution, as it may not reflect the true organization of AtVPS13M1. In addition, their position might be modified by binding to partners. Accessory domains are suspected to play a role in VPS13 protein binding to membranes and positioning between two membranes. Thus, the particular location of the PH1 and C2 domains, known to bind lipids, in the central part of the tunnel raised questions about their mode of action. In VPS13X, the additional amphipathic helices and the  $\beta$ -trefoil domain are located close to the C-term part of the tunnel, likely for partners binding at the membrane surface.

Tunnel length of AtVPS13S and M1 is around 20.7 and 22.2 nm respectively, in accordance with what has been described for other VPS13 proteins harboring 12 RBG units (Figure 5B) (Neuman et al., 2022). As expected because of the presence of nine RBG repeats, VPS13X is shorter, measuring around 16.1 nm. Surprisingly, whereas AtVPS13S and M1 form a rather straight tunnel, AtVPS13X tunnel is bended, forming an almost 90° angle between the N- and the C-term tunnel extremities. Such bending was obtained whatever the overlapping fragments used for the predictions and assembly. The tunnel curvature seems to occur at the level of the RBG5 unit, likely because it is composed of seven instead of five  $\beta$ -strands (Levine, 2022). In addition, AtVPS13X tunnel is wider with an average width of 4.4 nm in contrast to AtVPS13S and M1 that are around 4 nm. Thus, the VPS13X protein harbors an unusual structure with a shorter, wider, and bended tunnel. Hydrophobicity analysis reveals the presence of an hydrophobic surface, likely involved in lipid transport, all along the tunnel concave face for the three VPS13s (Figure 5C), highlighting the conservation of the biochemical properties of Viridiplantae VPS13M, X, and S paralogs compared to opisthokonts' VPS13s (Neuman et al., 2022; Dall'Armellina et al., 2023).

We attempted to predict the structure of VPS13Y from *K. nitens* and *C. subellipsoidea*. For KnVPS13Y, the N-term segment 1–1200 folded into a canonical hydrophobic tunnel, as described above (Figure S6), in accordance with the prediction of RBG repeats in KnVPS13Y (Figure 4). However, we were not able to assemble the full-length structure due to the presence of long unstructured regions in the tunnel and between the VAB and the C2 domains. In contrast, CsVPS13Y did not fold into a  $\beta$ -stranded structure but rather into a low confidence structure mainly composed of helices (Figure S6). Thus, it seems that VPS13Y has differentially evolved in the Viridiplantae phylum, keeping a canonical structure likely involved in lipid transport in *K. nitens* whereas diverging in *C. subellipsoidea*.

The analyzed VPS13B sequences did not contain well-predicted RBGs and as expected, did not fold into the well-known VPS13 hydrophobic tunnel in *C. subellipsoidea*, *K. nitens*, and *S. fallax* (Figure S7). The Chorein-RBG2 domains fold into the canonical  $\beta$ -sheet structure with an N-term helix, but the rest of the sequence is predicted to fold mostly into  $\alpha$ -helices instead of  $\beta$ -sheets (Figure S7).

One characteristic feature of VPS13 proteins is the presence of the VAB domain involved in VPS13 targeting to membrane organelles (Bean et al., 2018; Dziurdzik et al., 2020). In yeast and humans, the VAB domain is composed of six repeated modules harboring a similar  $\beta$ -sandwich structure (Adlakha et al., 2022; Dall'Armellina et al., 2023). In yeast, the modules are composed of nine  $\beta$ -strands (Adlakha et al., 2022). We performed structural prediction of the VAB domains of each representative VPS13 paralogs (Figure 6). For VPS13Ms, we also included the C2 domain upstream of the VAB. Most of the sequences folded into a six module-structure composed of 5 to 10  $\beta$ -strands, most of them being composed of 7 or 8 strands. The presence of  $\alpha$ -helices is predicted in the module 4 of some paralogs such as AtVPS13S and X (Figure 6). The modules size ranges from 98 to 311 aa with an average length of 147 aa. Surprisingly, whereas six modules are clearly predicted for AtVPS13S, X, KnVPS13Y and SfVPS13B, only five are found in AtVPS13M1 and M2. Some  $\beta$ -strands, that might be reminiscent to the module 1 (Figure 6), are predicted upstream of the module 2. In order to better understand the structure of VPS13M VAB domains, structural predictions were also performed for the monocotyledon *Zostera marina* VPS13M1, the Charophyte *K. nitens* VPS13M1 and M2, the Chlorophyte *C. subellipsoidea* and the Prasinodermatophyte *P. singulare* VPS13M. It should be noticed that all these sequences have a predicted DysF domain, unlike the *A. thaliana* sequences (Figure 3). All the VPS13M sequences folded into a five-modules VAB domain with three to five  $\beta$ -strands upstream (Figure 6 and Figure S8). The C2 forms a domain clearly independent from the VAB. Interestingly, the predicted DysF forms a distinct domain inserted in the VAB module 2 (Figure 6). Its presence in all the Viridiplantae phylum is indicative of an ancestral domain lost in some species, such as *A. thaliana* in both VPS13M1 and M2 proteins. In yeast and mammals, a conserved Asn residue, important for VPS13 binding to partners, is present close to the N-term of each module (Bean et al., 2018; Dziurdzik et al., 2020; Dall'Armellina et al., 2023). Therefore, we aligned the sequences of all the VAB modules of the 12 VPS13 analyzed in Figures 6 and Figure S8, including the  $\beta$ -strands upstream of VPS13M module 2 (72 sequences in total). The results show that an Asn residue is indeed found within the seven first aa in almost all the modules (Supplemental Figure S9) with the exception of, as expected, most of the  $\beta$ -strands located upstream of VPS13M module 2 and, more surprisingly, in the module four of the three VPS13M1 analyzed.



**Figure 5.** Structural modeling and characteristics of VPS13S, M1, and X in *Arabidopsis thaliana*. The 3D structures were obtained mainly using AlphaFold2 (Jumper et al., 2021) and RoseTTAFold (Baek et al., 2021) for AtVPS13M1. Overlapping fragments were aligned on RBG repeats and assembled using ChimeraX. (A) Predicted structures of the full length AtVPS13S, M1, and X proteins.  $\beta$ -strands,  $\alpha$ -helices, and loops are colored in yellow, red, and green, respectively. The different VPS13 domains are highlighted by different colors. VPS13 proteins folded into a long tunnel around which accessory domains are organized. (B) Representation of the AtVPS13S, M1, and X RBG repeats only showing the formation of the canonical VPS13 tunnel shape. For VPS13X, the tunnel is bended at the level of RBG5 colored in red. (C) Hydrophobic (in orange) and hydrophilic (in blue) surface representation of the RBG repeats structure of (B), highlighting the hydrophobic surface of the tunnel (arrows).

As the conserved Asn is found in VPS13M2 as well as in CsVPS13M and PsVPS13M module 4, its absence could be a feature of VPS13M1 acquired after VPS13M duplication. VPS13Y VAB domains from *C. subellipsoidea* and *K. nitens* harbor canonical six-module structures (Figure 6 and Figure S8). Noteworthy is the presence of long, mostly

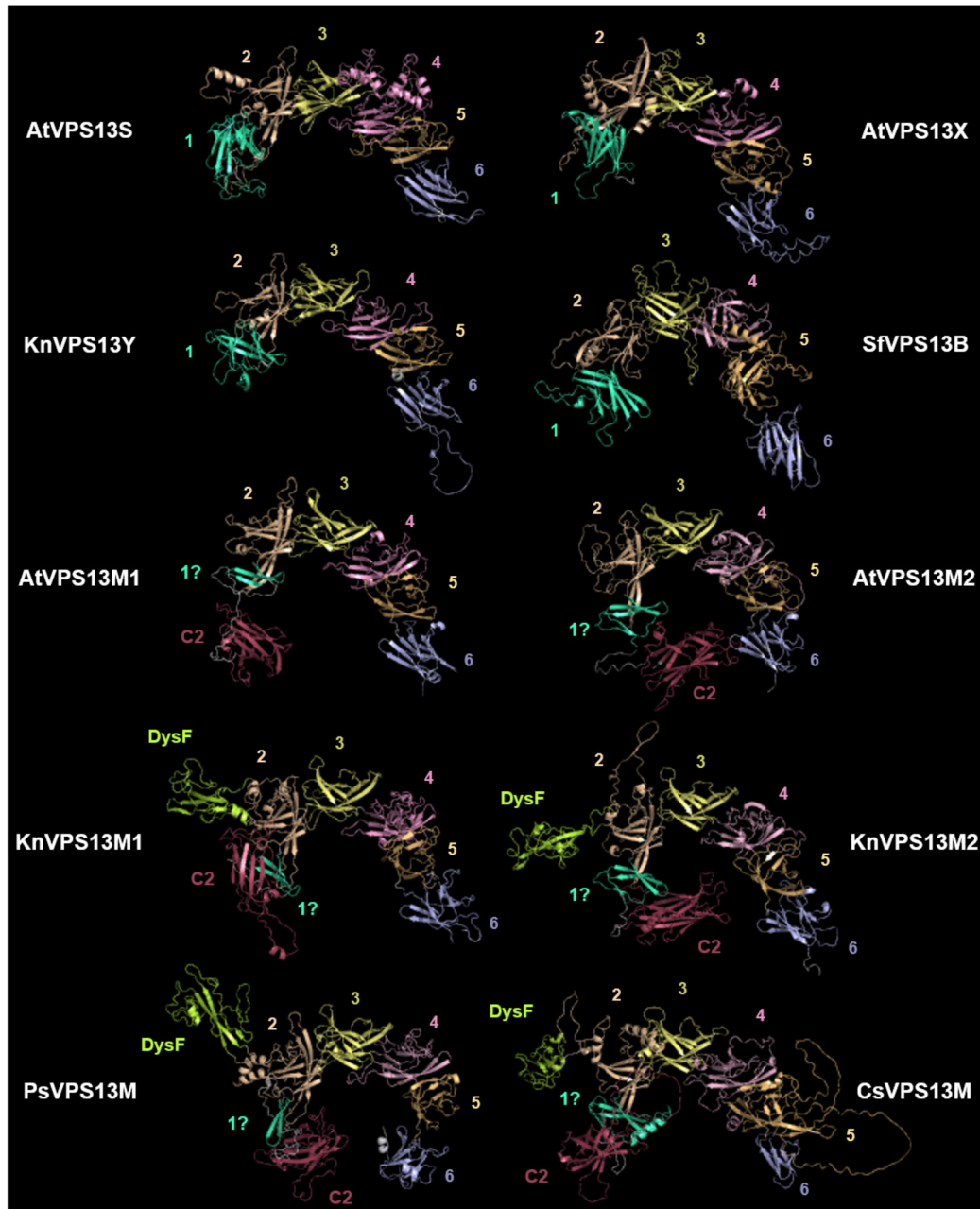
unfolded, sequences in modules 3 and 4 of CsVPS13 (Figure S8) which explain the prediction of a fragmented VAB by HHPred (Figure 4). *S. fallax* and *C. subellipsoidea* VPS13B VAB also fold into a six-module unit in contrast to *K. nitens* VPS13B for which no clear structure was predicted (Figure 6 and Figure S8). Thus, while AlphaFold2 and RoseTTAFold failed to predict a canonical tunnel structure for the N-term part (Figures S6 and S7), the presence of a canonical VAB domain in CsVPS13Y, CsVPS13B, and SfVPS13B suggests that they are indeed VPS13 proteins.

## Discussion

### Evolutionary History of VPS13 Proteins Viridiplantae

In this work, we performed phylogenetic analyses of VPS13 proteins in Archaeplastida. Our results revealed the complex evolutionary history of this protein family with several losses, acquisitions, and duplications. As VPS13 proteins are mainly characterized by a specific domain organization and structural similarities rather than conservation of the amino acid sequences, in addition to ML analyses, we performed Bayesian analyses in which domains were inferred as characters in a reduced dataset of Rhodophytes, Glaucophytes, and Viridiplantae and then, only in Viridiplantae where VPS13 paralogs seemed more conserved. Our results showed that one VPS13 group, VPS13B, mainly represented in Metazoa (Velayos-Baeza et al., 2004; Levine, 2022), is found also in different phyla of Viridiplantae and was likely present in the LECA corroborating Levine's hypothesis (Levine, 2022). In Viridiplantae, four groups, VPS13M, S, X, and Y were clearly identified and interestingly, none of the Rhodophytes' and Glaucophytes' sequences clustered with these groups suggesting a high divergence of VPS13 sequences after the primary endosymbiosis. Rhodophytes' sequences produced one robust clade that we named VPS13R grouping proteins with a similar domain organization. The other red algal sequences, though, are spread in different clades of the tree. Glaucophytes' sequences clustered together in the different analyses performed but with a weak support and no clear groups or domain organization was detected, likely because of the reduced dataset available for this clade.

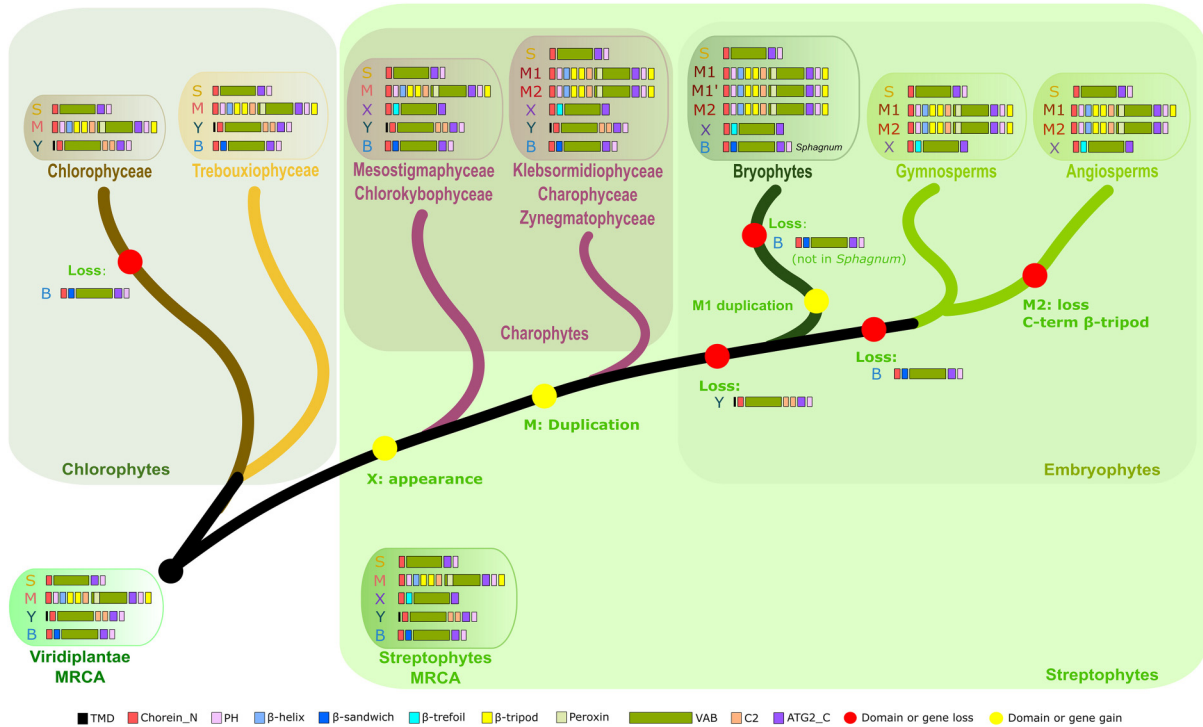
The evolutionary history of VPS13 proteins in Viridiplantae is more linear and summarized in Figure 7. The Viridiplantae MRCA had likely four VPS13 paralogs: VPS13S, M, Y, and B with the domain organization presented in Figure 7. VPS13B paralogs disappeared in Prasinodermatophytes and most of the Chlorophyte lineages apart from the Trebouxiophyceae and some others. Of note, in *P. coloniale*, two VPS13Y paralogs seem to be present as well as another sequence, without any clear link with the described Viridiplantae VPS13s, suggesting that Prasinodermatophytes might have acquired other groups of



**Figure 6.** Structural modeling and characteristics of the VAB domain of Viridiplantae VPS13 paralogs. The sequences used are: *Arabidopsis thaliana* VPS13S (AtVPS13S), VPS13X (AtVPS13X), VPS13M1 (AtVPS13M1), and VPS13M2 (AtVPS13M2), *Coccomyxa subellipsoidea* VPS13M (CsVPS13M), *Klebsormidium nitens* VPS13Y (KnVPS13Y), VPS13M1 (KnVPS13M1), and VPS13M2 (KnVPS13M2), *Prasinoderma singulare* VPS13M (PsVPS13M), and *Sphagnum fallax* VPS13B (SfVPS13B). The 3D structures were obtained mainly using AlphaFold2 (Jumper et al., 2021) but also RoseTTAFold (Baek et al., 2021) for PsVPS13M. The C2 domain of VPS13M paralogs were included in the prediction. The different modules of the VAB domains are represented by colors and numbered from 1 to 6, the C2 and DysF domains found in VPS13Ms are in raspberry and limon, respectively.

VPS13s. Early in the emergence of Streptophytes, VPS13X appeared. VPS13M gene was then duplicated to give rise to VPS13M1 and M2 groups shortly before the emergence of the Klebsormidiophyceae. This hypothesis is supported by the presence of five paralogs (VPS13B, S, M, X, and Y) in *M. viride* and six (VPS13B, S, M1, M2, X, and Y) in *K.*

*nitens*. VPS13Y paralog was lost before the emergence of land plants around 500–600 Ma (Becker, 2013), and VPS13B was then lost in almost all lineages except the Bryophyte *Sphagnum* which maintained a VPS13B gene. Thereafter, VPS13M1 was duplicated in Bryophytes and the Angiosperms' VPS13M2 proteins lost their C-term



**Figure 7.** Schematic representation of VPS13 evolutionary history in Viridiplantae. VPS13 paralogs with their domain organization are represented in the different phyla analyzed in this study as well as the probable paralogs in most recent common ancestors (MRCA) of Viridiplantae and Streptophytes. Red and yellow dots indicate events of gene or domain losses and acquisitions, respectively. The MRCA of Viridiplantae had likely four VPS13 paralogs: VPS13S, M, Y, and B with the presented putative domain organization. After the separation between Chlorophytes and Streptophytes, VPS13B paralogs disappeared in most of the Chlorophyte lineages but the Trebouxiophyceae. Early in the emergence of Streptophytes, VPS13X appeared. VPS13M gene was duplicated to give rise to VPS13M1 and M2 groups after the emergence of Mesostigmaphyceae and Chlorokybophyceae. VPS13Y paralog was lost before the emergence of land plants and VPS13B was then lost in almost all lineages but the Bryophyte *Sphagnum* which maintained a *VPS13B* gene. Thereafter, the Angiosperms' VPS13M2 proteins lost their C-term β-tripod domain. VPS13M2 duplication, not represented in this scheme, also occurred in some Embryophyte species (*Zostera marina* and *Selaginella moellendorffii*). ATG2\_C = Autophagy-related protein 2 C-terminus homology; PH = Plekstrin Homology; TMD = transmembrane domain; VAB = VPS13 adaptor binding.

β-tripod domain (Figure 7). In addition, VPS13M2 was also duplicated in some angiosperm species (*S. moellendorffii* and *Z. marina*). These results illustrate the complex evolution of VPS13 proteins in terms of paralog number (gain and loss) and domain organization.

### Characteristics of VPS13 Proteins in Viridiplantae

The retrieval of putative full-length VPS13 sequences was a challenging task because some genomes were poorly annotated and because automatic annotation software often failed to predict correct intron/exon borders for such long genes and some of them appeared split up. Extra unusual domains in N- or C-term are also frequently observed, likely linked to error in gene border annotations. Thus, for such genes, analysis of transcriptomic data, as we did for *C. subellipsoidea*, is the best method to obtain accurate VPS13 gene models. By selecting sequences longer than 2400 aa, we retrieved 383 sequences covering the three Archaeplastida lineages and showed that the gene number is highly variable

ranging from 1 to 13. As mentioned above, the lower copy number is probably underestimated because we failed to detect split up genes. In Viridiplantae, our most curated dataset, the VPS13 gene copy number varies from three in Chlorophyceae to six in the Charophyte *K. nitens* and the Bryophyte *S. fallax* (Figure 7). Huge variations in sequence length were also observed.

By analyzing RBG repeats, we showed that sequence length is not always correlated to an increase in RBG repeats. Tunnel length and the addition of new domains are not directly correlated to the protein sequence length neither, as exemplified for VPS13X proteins ranging from 3122 to 4070 aa (Figure 4) though harboring the same RBG number and domain organization. Our analysis showed that general VPS13 protein domain and structure organization is widely conserved among Viridiplantae with the occurrence of a N-term Chorein, a VAB and an ATG2\_C domain, most of them also having a C-term PH domain (Figure 3). Structural prediction revealed that the VAB domain structure composed of six repeated modules



with an invariant Asn residue in N-term (Adlakha et al., 2022; Dall'Armellina et al., 2023) is also conserved in Viridiplantae VPS13s with some variations in VPS13M paralogs that will be discussed below. Additional domains, such as  $\beta$ -tripod,  $\beta$ -helix,  $\beta$ -sandwich, C2, or UBA domains, are found in different VPS13 groups and are likely real and functional. Others are only present in sparse sequences and might be linked to annotation errors. RBG repeats are also predicted for most of the paralogs we have analyzed, maximum number being 12 for Viridiplantae. In accordance with the presence of RBG domains, structural predictions showed that Viridiplantae VPS13s form long tunnels with a hydrophobic surface as described in opisthokonts (Neuman et al., 2022; Dall'Armellina et al., 2023). Additional localizations and/or functions might have arisen from the addition of new domains, not found in opisthokonts, such as  $\beta$ -tripod or C2 domains.

### VPS13S

VPS13S has the simplest organization and it is tempting to speculate that VPS13S proteins harbor an ancestral structure and domain organization. It is particularly worth mentioning that VPS13S domain organization is highly similar to the one found in yeast ScVPS13. Both have 12 RBG repeats and harbor the same canonical functional domains. ScVPS13 is the only VPS13 protein found in *S. cerevisiae* and its presence is important for several basic cellular processes like prospore membrane expansion (Park and Neiman, 2012; Park et al., 2013) or autophagy (Dabrowski et al., 2023). It is therefore tempting to hypothesize that at least one simple form of VPS13 protein without additional non-canonical domains or alternative RBG domain organization is required in an organism to fulfill basic cellular functions. This hypothesis is corroborated by the fact that AtVPS13S plays important roles in *Arabidopsis* reproduction as its absence causes male as well as female sterility (Koizumi and Gallagher, 2013). AtVPS13S also interacts with the transcription factor SHORT ROOT and plays a role in a pathway regulating root growth and radial patterning (Koizumi and Gallagher, 2013). However, the exact mechanism by which AtVPS13S acts on this pathway is still elusive. AtVPS13S seems to act in gibberellic acid (GA) signaling, downstream of its synthesis (Koizumi and Gallagher, 2013). As GA is a mostly hydrophobic phytohormone, it is conceivable to suppose that AtVPS13S could be involved in its transport. AtVPS13S probably has other functions in cells, as this paralog is well conserved in Viridiplantae and GA signaling has emerged in vascular plants (Nishiyama et al., 2018).

### VPS13M

VPS13M has the most complex organization with seven domains added during evolution, likely to regulate its localizations and/or functions. In VPS13M ancestor, three  $\beta$ -tripod

domains, an additional PH, a C2 domain, a  $\beta$ -helix, as well as a DysF domain were present (Figure 4). A similar  $\beta$ -tripod domain was described in bacterial insecticidal proteins and shown to be involved in specific cell surface receptor recognition prior to the formation of membrane pores mediated by other protein domains (Zaitseva et al., 2019; Levine, 2022). Interestingly, the  $\beta$ -tripod domain seems to be widespread in Viridiplantae as 1085 out of the 1743 identified sequences containing a  $\beta$ -tripod domain, were from Viridiplantae (Zaitseva et al., 2019). Thus, the addition of  $\beta$ -tripod domains in VPS13M proteins might be involved in the regulation of protein localization and/or interaction with protein partners. A similar role could be attributed to the additional PH domain in N-term as well as to the C2 domain downstream the VAB domain, which could be involved in protein-protein or protein-lipid interactions. A  $\beta$ -helix was also found in VPS13M and it was hypothesized that it could be linked to the ubiquitination pathway because of its occurrence in F-box proteins (Levine, 2022). A certain degree of divergence was also revealed in VPS13M VAB domains by structural predictions (Figure 6). Indeed, only five full modules instead of six were identified and a DysF domain is inserted in the second module. The function of DysF domains is not clearly established but it is found in several proteins in yeast and humans involved in membrane remodeling (Bulankina and Thoms, 2020; Ferreira and Carvalho, 2021). In yeast, the DysF domain of Pex30 is essential for its function in LD biogenesis and at different contact sites (Ferreira and Carvalho, 2021). Therefore, the DysF domain could give to VPS13M new location determinants or new functions. The DysF domain was curiously lost in some sequences including in *A. thaliana* VPS13M1 and M2. In addition, in the VPS13M1 paralogs, no Asn residue was detected at the N-term of the VAB module 4. Asn residues from RBG5 and 6 are important for ScVPS13 and HsVPS13D partner binding (Dziurdzik et al., 2020), whereas HsVPS13B lacks this invariant residue in modules 2 and 3 (Dall'Armellina et al., 2023). Thus, the impact of the lack of Asn residue in the VPS13M VAB module 4 remains to be investigated. Structural prediction of AtVPS13M1 shows that it harbors the canonical hydrophobic tunnel structure with all the extra-domains protruding outside without affecting its folding (Figure 6). Questions remain about the organization of the extra-domains in relation to each other. In addition, these domains are likely involved in partners and membrane binding to regulate VPS13M localization and function. Although VPS13s in opisthokonts and yeast have simpler domain organizations, the mechanisms involved in the regulation of VPS13 localization at different sites is already complex (Bean et al., 2018; Dziurdzik and Conibear, 2021). Thus, it will be of high interest to investigate how VPS13M paralogs can integrate all the cellular signals to fulfill their function.

The evolutionary origin of VPS13M was not inferred from our analyses, but the complex domain organization depicted

above is already found in Prasinodermatophytes. Up to now,  $\beta$ -tripod-containing VPS13 proteins were identified only in Viridiplantae, neither Rhodophytes or Glaucophytes possess  $\beta$ -tripod domains in VPS13 proteins. This suggests that the  $\beta$ -tripod domain emerged after the separation of Rhodophytes and Viridiplantae 1.5–2 Ma (Bowles et al., 2023). In addition, our analysis did not reveal any relationships among VPS13M and another VPS13 groups. As it possesses 12 RBG repeats like VPS13S, it would be parsimonious to hypothesize that it could come from a duplication of VPS13S ancestor followed by the rapid acquisition of the extra-domains.

From an interrogation of the InterPro database (Paysan-Lafosse et al., 2023), we found that IPR009291 domain (Vacuolar protein sorting-associated protein 62, aka  $\beta$ -tripod domain) is present in  $\approx 6000$  sequences belonging mainly to Viridiplantae ( $\approx 3000$  sequences) and fungi ( $\approx 2000$  sequences). In fungi, none of the sequences containing a  $\beta$ -tripod domain was associated to a known VPS13 domain. A few other lineages harboring  $\beta$ -tripod domain containing proteins, such as the Chlorarachniophytes *Lotharella* or the Rhodophyte *Porphyridium purpureum*, did not show association with VPS13 domains either. This makes the association of the  $\beta$ -tripod with the VPS13 canonical domains a Viridiplantae-specific feature. In order to check whether the  $\beta$ -tripod domain was retained through the secondary green endosymbiotic event, we retrieved sequences from *Euglena gracilis* and *Bigeloviella natans* (Supplemental File S7). However, none of them contained a predicted  $\beta$ -tripod domain, suggesting that VPS13M paralogs were not retained after the secondary endosymbiosis. Thus, the origin of the Viridiplantae VPS13M paralog with its high complexity remains unknown.

Our analyses revealed the highly dynamic and still ongoing evolution of the VPS13M group. VPS13M was duplicated early in the Streptophyte lineage into VPS13M1 and M2. Other duplication events were observed like VPS13M1 in Bryophytes and VPS13M2 in *Z. marina* and *S. moellendorffii* (Figure 3). In most of the phyla, both proteins maintained a conserved domain organization with 12 RBG and the domains described above. Interestingly, some VPS13M have also lost their DysF domain inserted in the VAB, like *A. thaliana* VPS13M1 and M2. In Angiosperms, the C-term  $\beta$ -tripod domain was lost, likely to allow the emergence of a new function. In a recent pre-print, functional analysis of AtVPS13M1 and M2 suggested that AtVPS13M2, but not M1, could play a role in pollen germination and pollen tube growth (Tangpranomkoorn, 2022, BioRxiv, DOI:10.1101/2022.11.01.514778). A rapid polarized extension of the pollen tube is required in order to deliver sperm cells to the ovule. Its extension requires a massive vesicular lipid and protein transport to the pollen tube tips to allow the delivery of components required to sustain its growth, including lipids (Adhikari et al., 2020). AtVPS13M2 accumulates at the pollen tube tip secretory

vesicles and might play a role in the establishment of polarity and in cell wall deposition during pollen tube growth (Tangpranomkoorn, 2022, BioRxiv, DOI:10.1101/2022.11.01.514778). Pollen tube growth is common in Angiosperms but also found in some Gymnosperm species that have usually shorter pollen tubes that grow more slowly (Fernando et al., 2005; Adhikari et al., 2020). The only Gymnosperm species included in our dataset is *Thuja plicata* in which pollen germination and pollen tube growth was reported (Land, 1902). TpVPS13M2 still has a C-term  $\beta$ -tripod domain, thus it is difficult to speculate about the link between the presence of this domain and an essential role in the pollen germination process. An extended investigation of VPS13M2 sequences and a functional analysis in Gymnosperms might help to understand when the C-term  $\beta$ -tripod domain was lost and the reason for the VPS13M2 specialization in Angiosperms.

### VPS13B

From previous analyses of VPS13 RBG domain organization, it was suggested that the LECA likely had two paralogs: VPS13B and an ancestral VPS13 which gave rise to the other groups of VPS13 (Levine, 2022). Our analyses further confirmed the ancestry of VPS13B as it was found in some Viridiplantae. However, VPS13B is maintained in a small subset of phyla, raising questions about the function of this protein. In human, VPS13B mutations are associated to the Cohen Syndrome, a developmental disorder associated with variable clinical phenotypes such as microcephaly, pigmentary retinopathy or hypotonia (Douzgou and Petersen, 2011). At the cellular level, VPS13B localizes to the Golgi and is involved in endosome tethering (Seifert et al., 2011; Koike and Jahn, 2019). Roles in the maintenance of Golgi integrity (Seifert et al., 2011; Da Costa et al., 2020), autophagy and neurite outgrowth (Seifert et al., 2015) as well as in acrosome formation in mice (Da Costa et al., 2020), were demonstrated. Most of these functions are associated to extensive membrane trafficking. Whereas maintenance of Golgi and endosome tethering functions are basal processes likely required for the maintenance of all Eukaryotic cells, VPS13B function might be essential in some developmental processes specific to animals such as brain development or acrosome formation during spermatogenesis, explaining its maintenance in this clade. However, we could not find evident functions that might be specific to the small subset of Viridiplantae species in which VPS13B orthologues are present, suggesting that new functions might have evolved independently in each phylum. We were able to find transcripts supporting the expression of VPS13B in *C. subellipsoidea* and in two *Sphagnum* species, proving their expression and functionality, possibly as LTPs because of the conservation of the Chorein + RBG2 domain, able to bind lipids *in vitro* (Kumar et al., 2018).

Another intriguing point is the domain organization of VPS13B orthologues. Whereas the Chorein and RBG2 domains as well as the C-term part of VPS13B orthologues, including the VAB domain, appeared well conserved in terms of domain and structure, we failed to detect seven of the RBG domains predicted in HsVPS13 between RBG2 and RBG10 (Levine, 2022; Dall'Armellina et al., 2023). In contrast, this region is mostly composed of helices and unstructured loops, raising questions about their general structure and function conservation (Figure S7). A reduced number of RBG domains might imply a reduction of the tunnel length, maybe to accommodate the protein to contact sites of shorter length. However, if this is the case, why was the sequence maintained between the Chorein and the VAB domains during evolution? This might be the sign that VPS13B is slowly losing its bridge-like LTP function in Viridiplantae. Alternatively, the structure might have been modified to accommodate for the interaction with other protein partners or to gain new functions. A closer structural analysis coupled with functional studies by reverse genetics approaches might help to decipher the role of VPS13B paralogs in non-Metazoa and their evolutionary history.

From our analyses, two main hypotheses can be posed concerning the evolution of VPS13B: (1) the VPS13B ancestor had six RBG repeats which were then multiplied in Humans to obtain a VPS13B with 13 RBG repeats or (2) VPS13B ancestor encompassed 13 RBG repeats that have progressively diverged in Viridiplantae. The second hypothesis seems more probable because an ancestor with six RBG would imply the addition of a 1000-aa sequence, not clearly folded, in Viridiplantae VPS13B. The analyses of the N-term domain organization of a higher number of VPS13B orthologues, including basal Metazoan, might help to depict the organization of ancestral VPS13B and their evolution.

### VPS13X

The origin of VPS13X is puzzling as its sequence and domain organization is highly divergent compared to other paralogs. In agreement with this, VPS13X was only recently identified by remote homology analyses (Levine, 2022) and not in previous studies using BLASTP searches (Velayos-Baeza et al., 2004; Michaud et al., 2017; Leterme and Michaud, 2022). VPS13X is composed of nine RBG domains, having the smallest number of repeats in currently characterized VPS13s, and is the only group of VPS13 in which a C-term PH domain is absent, amphipathic  $\alpha$ -helices are found instead (Levine, 2022). Thus, VPS13X looks like a basal ancestral protein, with an extra  $\beta$ -sandwich domain between RBG6 and 7. However, we were not able to detect any VPS13X orthologues in species other than Streptophytes, even by PSI-BLASTP using either Charophytes or Embryophytes VPS13X as queries (data not shown), hinting that VPS13X appeared soon after the divergence of Streptophytes and

Chlorophytes or that it was lost in the genome of all the other species sequenced so far. As it was not found in other organisms, the hypothesis of a horizontal gene transfer is also unlikely. A previous extended RBG analysis proposed that RBG3 and 6 domains are unrelated to specific other known VPS13 RBG domains, but might have evolved from penultimate domains, and that RBG5 is unique to VPS13X as it is composed of seven instead of five  $\beta$ -strands (Levine, 2022). Structural predictions show that VPS13X forms a shorter, larger, and bended hydrophobic tunnel in contrast to other VPS13 (Figure 4). In addition, our analysis did not reveal any robust relationship with other VPS13 from Archaeplastida. Thus, if VPS13X arose from a VPS13 paralog gene duplication, a rapid evolution of the gene occurred with a loss of PH and RBG domains. Proteins can mutate at different rates depending on several factors like selection pressure (such as sperm proteins (Torgerson et al., 2002)), folding stability and chaperoning efficiency (Agozzino and Dill, 2018). A new protein stemming from a gene duplication may have folding instabilities and be more prone to mutation. An extreme consequence would be the loss of the protein. In comparison, VPS13M1 and M2 genes appeared by duplication of VPS13M ancestor almost at the same time as VPS13X appeared after the divergence of Streptophytes and Chlorophytes. Nonetheless, a high similarity is still observed between VPS13M1 and M2, which is not the case of VPS13X. An alternative hypothesis might be that VPS13X arose from VPS13B. In fact, VPS13B proteins harbor a  $\beta$ -sandwich fold between RBG10 and 11 and VPS13X proteins have a  $\beta$ -trefoil fold inserted between RBG6 and 7 (structurally equivalent positions). Interestingly, the  $\beta$ -sandwich and the  $\beta$ -trefoil folds are both structurally related as it has been proposed that  $\beta$ -trefoils might have evolved from  $\beta$ -sandwiches (Longo et al., 2022). Furthermore, HsVPS13B and AtVPS13X both have central RBG domains that share low similarity to the penultimate RBG domains, in contrast with other previously analyzed VPS13 paralogs (Levine, 2022). It is also noteworthy that the AtVPS13X RBG5 is composed of seven  $\beta$ -strands instead of the usual five, which might be a relic of the RBG domain loss necessary for the transition from 13 to 9 RBG repeats in VPS13X. Additionally, the C-term PH domain is only weakly predicted in VPS13B proteins while it is absent in VPS13X proteins. It is therefore possible that VPS13X proteins derived from VPS13B proteins by losing several central RBG domains, by converting the  $\beta$ -sandwich fold into a  $\beta$ -trefoil fold and by losing the C-term PH domain. However if so, this event implies the duplication of VPS13B gene early in Streptophytes as both VPS13B and X are found in Charophytes and in the Bryophyte *Sphagnum*. Then, VPS13X was maintained whereas VPS13B was lost. At this stage, the number of VPS13B sequences obtained from Streptophytes is too low to perform an informative comparative analysis of VPS13X and B structures and sequences. Finally, VPS13X might also have emerged from the fusion of a C-term VPS13 paralog and the

N-term of another RBG protein family, such as SHIP164 or Tweek, present in *A. thaliana* (Levine, 2022), but an extensive analysis of these families in photosynthetic organisms may be required to corroborate or deny this hypothesis.

The function of VPS13X is currently unknown but structural predictions suggest that it might also play a role in lipid and membrane trafficking processes. In accordance with the reduced number of RBG domains, AtVPS13X forms a shorter hydrophobic tunnel, likely around 16 nm, which is predicted to be larger and bended (Figure 4), inferring that it may play a role at contact sites other than those of the longer VPS13 paralogs (around 20 nm) and between membranes that are perpendicularly apposed. Its occurrence only in Streptophytes suggests a role in a process specific to this clade. Expansion of LTPs in plants is observed for several families, the strongest example being the apparition of a plant-specific LTP family: the non-specific (ns) LTPs. This family emerged in Embryophytes after the split with Charophytes and more than 100 genes are currently found in genomes of higher plants, highlighting the extremely rapid expansion of this family (Edstam et al., 2011). Noteworthy, the number of other LTP homologous families, such as the StARTs (steroidogenic acute regulatory protein-related lipid transfer) or the ORPs (Oxysterol-binding protein related proteins), is often higher in plants compared to yeast or animals (Leterme and Michaud, 2022). This is likely linked to the emergence of new cell functions and structures such as chloroplasts, plasmodesmata, or complex cell walls. Indeed, the synthesis of plant cell wall structures, such as cutin, wax, or sporopollenin, requires a massive secretion of hydrophobic components and many nsLTPs have been involved in such processes (Edstam et al., 2011; Edqvist et al., 2018). In addition, the terrestrialization of plants was accompanied by their ability to adapt to a wide variety of environmental variations and biotic stresses (Bowles et al., 2023). Plants have evolved several mechanisms involving massive membrane remodeling in order to adapt to nutrient stress or to establish symbiosis with microorganisms (Moellering and Benning, 2011; Michaud and Jouhet, 2019). Genome analyses of different Charophytes revealed that they already encode several key genes involved in land plant specific processes, including symbiosis with microorganisms (Nishiyama et al., 2018; Domozych and Bagdan, 2022; Bowles et al., 2023). Thus, VPS13X might have emerged in Charophytes to ensure a key function in Streptophyte evolution likely linked to lipid and membrane trafficking. No functional studies have been carried out yet in *A. thaliana* or other species and this will be a key step to understand the role and evolution of the VPS13X family.

## VPS13Y

VPS13Y is a newly described group of VPS13 characterized by the presence of a putative TM domain in N-term end as well as two C2 domains downstream of the VAB. Whereas

the N-term region of KnVPS13Y is predicted to fold into a  $\beta$ -sheet tunnel, as expected for a VPS13 protein, such folding was not predicted by structural nor by secondary structure analyses for CsVPS13Y (Figure S6). Instead, it folds into helices, as observed for some VPS13B paralogs, raising questions about VPS13Y function and conservation. Inferring a function to this group is a challenging task as no other described VPS13 has a similar domain organization. C2 domains are involved in protein–lipid binding and as VPS13Y proteins putatively have N-term TM domain, they might act as membrane tethers, like Extended-synaptotagmin LTPs (Saheki and De Camilli, 2017). Because of the absence of the Chorein domain, it is also possible that CsVPS13Y has lost its ability to act as a LTP.

VPS13Y has been detected in Prasinodermatophytes, Chlorophytes, and Charophytes, suggesting it was present in the Viridiplantae MRCA and lost in Embryophytes, the reason being unknown.

Overall, our results provided the first extensive description of VPS13 paralogs in Archaeplastida and revealed a conservation of their general organization. It also highlighted the complex evolutionary history of Viridiplantae VPS13 proteins, with several paralogs and losses and appearances of functional domains. Future efforts are now required in order to decipher the localization, structure, and function of the different VPS13 paralogs in Archaeplastida.

## Material and Methods

### Sequence Retrieval and Selection

For Viridiplantae, all the sequences from Embryophytes (land plants) were retrieved from the Phytozome database (Goodstein et al., 2012), which contains assembled genomes from more than 300 Viridiplantae species. For Glaucophytes, Rhodophytes, Chlorophytes, Prasinodermatophytes, and Charophytes, sequences were retrieved from five databases: Uniprot (The UniProt Consortium et al., 2023), Joint Genome Institute (JGI) (Nordberg et al., 2014), Phycocosm (Grigoriev et al., 2021), Eukprot V3 (Richter et al., 2022), and the National Center for Biotechnology Information (NCBI) (Sayers et al., 2022). To find as many VPS13 homologs as possible, BLAST searches (Altschul et al., 1990) were primarily performed using the well-described *Arabidopsis thaliana* sequences (M1: At4g17140, M2: At1g48090, S: At5g24740, and X: At3g50380) as queries. For Embryophytes, BLASTPs were performed with AtVPS13M1 and AtVPS13X aa sequences, whereas for other organisms, BLAST searches were performed from the four *A. thaliana* sequences because of the putative higher divergence of sequences compared to the plant ones. As orthologues of HsVPS13B were recently supposed to be present in some protists (Levine, 2022), HsVPS13B was BLASTP searched as well. Sequences with at least one VPS13 domain annotation, such as VPS13\_N, VAB, Chorein or VPS13\_C, were retrieved to

generate data in Table S1. Then, a curation was performed by selecting sequences that were at least 2400 aa-long to constitute a dataset of 383 sequences (with addition of the four HsVPS13 and ScVPS13), used for ML phylogenetic analyses (Dataset 1). Some sequences were split up in two or three pieces and were merged to obtain likely full-length sequences: *Closterium* fusion = GJP45422 + GJP45421 + GJP45420; *Klebsormidium nitens* fusion = GAQ85408 + GAQ85409; *Sphagnum fallax* fusion = Sphfalx12G108100.1 + Sphfalx12G108200.1, *Thuja plicata* fusion = Thupl.29379161s0002.1.p + Thupl.29379161s0001.1.p. Then, a subset of organisms was chosen in each taxonomic clade to obtain a reduced dataset of 104 sequences to perform ML and Bayesian phylogenetic analysis (Dataset 2). We selected species for which we had the highest number of sequence and eliminate protein models potentially emerging from the same gene but corresponding to splicing variant by multiple alignments of sequences within each species. When two proteins likely originating from the same gene were discovered, the longest isoform was kept for the analysis. Presence of at least two canonical VPS13 domains (Chorein, VAB, ATG2\_C, PH) was also used as a criterion to select sequences. HsVPS13A/B/C/D and ScVPS13 were included in this dataset. A final dataset, containing only Viridiplantae sequences of Dataset 2, was created in order to perform ML and Bayesian phylogenetic analysis in this phylum, excluding HsVPS13A/B/C/D and ScVPS13 sequences (Dataset 3, 79 sequences).

### Domain and RBG Predictions

Domain prediction was performed using HHPred (Zimmermann et al., 2018) on all the sequences of Dataset 2 in order to infer domain information in the Bayesian analysis. The full-length sequences were used for analysis and domains predicted with at least 50% confidence were kept. Because of the huge protein size, HHPred usually failed to predict all the putative domains and 2nd and 3rd analysis were performed in regions in which no domains were predicted at a first place. Once again, domains predicted with at least 50% confidence were kept. Domains found in the sequences are summarized in Tables S2 and S3 in the form it was coded for the Bayesian analysis.

RBG domain prediction was performed by HHPred in the targeted sequences by using as queries RBG domain sequences of AtVPS13 paralogs as well as HsVPS13B described by Levine (2022). For CsVPS13Y, KnVPS13Y predicted RBG domains were also used. When a low similarity was found, some RBG were confirmed by performing structural predictions with ColabFold (Mirdita et al., 2022).

### Maximum Likelihood Analyses

ML analyses were performed with datasets 1, 2, and 3. The full-length protein sequences were aligned with the software

MAFFT (Katoh and Standley, 2013) performing 16 iterations of refinement. The obtained alignment files were processed with the trimAL (Capella-Gutiérrez et al., 2009) program to remove all sites with a percentage of gaps higher than 10% (gap score = 10). After trimming, the size of sequences in the final alignments is around 900–1000 aa and include mainly amino acids from conserved domains (Chorein-RBG2-RBG10-VAB-RBG11-RBG12-ATG\_C-PH). The phylogenetic reconstruction was performed with the software FastME (Lefort et al., 2015) using the LG model and 5000 bootstrap iterations. Nodes with at least 4000 bootstraps (80%) were considered as well supported.

### Bayesian Inference

Two independent Metropolis-coupled Markov Chain Monte Carlo (MC<sup>3</sup>) analyses were performed using the final curated datasets in the MrBayes3.2 software (Ronquist et al., 2012). Two partitioned datasets were used for Bayesian inference, each including a binary column indicating the result of domain annotation reported above and a second partition for the amino acid sequence, for the Archaeplastida (Table S2) and for Viridiplantae (Table S3) containing 104 (Standard:1-42,PROTEIN:43-1166) and 79 (Standard:1-25,PROTEIN:26-1332) entries, respectively. The domain annotation was mathematically formalized into a 1/0 code (presence = 1, absence = 0) and each domain accounted as a position in the alignment. One cold chain and either seven (temperature parameter 0.01, Dataset 2) or three “heated” (temperature parameter 0.1, Dataset 3) chains drove the analyses. The analyses were allowed to switch among all the substitution models implemented in the software to identify the best model without any *a priori* (aamodelpr = mixed). 90,000 and 35,000 generations with sampling every 100 generations were set for Dataset 2 and Dataset 3, respectively. The standard deviation of the split frequency between the two parallel analyses at the end of each analysis were 0.068996 and 0.006487 for Dataset 2 and Dataset 3, respectively. The first 25% of the trees produced were discarded in order to let the analyses stabilize (burnin = 0.25). The convergence of the runs was estimated and the potential scale reduction factor (PSRF +) were all around 1 for all the parameters indicating that the associated chains converged to one target distribution. Large PSRF (PSRF + > 1) indicate convergence failure. Nodes with a posterior probability  $\geq 0.95$  were considered as well supported. Alignments in fasta and trees in newick format are available upon request.

### Expression Analysis

For *C. subellipsoidea*, BLASTN searches on JGI were performed against the EST cluster available using the different VPS13 paralog hypothetical mRNA sequences as queries (CsVPS13M: XM\_005649615.1, CsVPS13S: XM\_005645274.1, CsVPS13Y: XM\_005642608.1, CsVPS13B:

XM\_005646816.1). ESTs with at least 90% identity were retrieved and aligned with the corresponding hypothetical mRNA sequences using SnapGene® software version 5.0.8. Then, in order to obtain curated version of the mRNA and protein sequences, modification on hypothetical mRNAs were performed by incorporating variations (insertions, deletions, substitutions) observed within at least two ESTs.

*C. braunii* (Bioproject PRJNA492241), *K. nitens* (Bioproject PRJN242253), *Sphagnum recurvum* (Biosample UHLI), and *Sphagnum palustre* (Biosample RCBT) transcriptome assemblies were retrieved from NCBI. TBLASTN searches were performed using respective VPS13 protein sequences as queries and sequences that covered at least 10% of the query and with an e-value below 0.001 were retrieved (maximum 20 hits). Then, according to the percentage of identity to the VPS13 paralogs used as query, we annotated each hit in order to determine which paralogs are expressed in each species.

### Structural Predictions

Structural predictions were performed using ColabFold, an easily accessible web interface to perform AlphaFold2 3D predictions (Jumper et al., 2021; Mirdita et al., 2022). RoseTTAFold (Baek et al., 2021) was also used for some complex structure, such as AtVPS13M1, or some VAB sequences that were not properly folded with AlphaFold2 predictions. For full-length structures determination, overlapping fragments of maximum 1400 aa were designed with overlapping RBG repeats at each extremity for structural prediction. Fragments structure were aligned on overlapping RBG and joined between the two most-well aligned residues using ChimeraX (Pettersen et al., 2021). Four fragments were assembled for AtVPS13S (1-1107; 1108-1801; 1802-2840; 2841-3464) and AtVPS13X (1-989; 990-1676; 1677-2622; 2623-3132). For AtVPS13M1, six fragments were predicted (1-955; 956-1837; 1838-2414; 2415-3469; 3470-3737; 3738-4202) and assembled in order to avoid steric clashes between all the additional domains. Assembled structure of AtVPS13S, M1, and X are provided in Supplemental Files S4, S5 and S6, respectively. VAB modules sequences were determined from the structures. Structures were visualized and annotated in PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

### Declaration of Conflicting Interests


The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the French National Research Agency in the framework of the “investissement d’avenir” program (ANR-15-IDEX-02) and of the ANRJJC MiCoSLiT (ANR-19-C13-0013) to MM. It

also received funding from GRAL, a program from the Chemistry Biology Health (CBH) Graduate School of University Grenoble Alps (ANR-17-EURE-0003).

### ORCID iDs

Alberto Amato  <https://orcid.org/0000-0002-9126-5535>

Morgane Michaud  <https://orcid.org/0000-0002-4853-8411>

### Supplemental Material

Supplemental material for this article is available online.

### References

- Adhikari PB, Liu X, Kasahara RD (2020). Mechanics of pollen tube elongation: A perspective. *Front Plant Sci* 11, 589712. doi: 10.3389/fpls.2020.589712
- Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, et al. (2012). The revised classification of eukaryotes. *J Eukaryot Microbiol* 59, 429–493. doi: 10.1111/j.1550-7408.2012.00644.x
- Adlakha J, Hong Z, Li P, Reinisch KM (2022). Structural and biochemical insights into lipid transport by VPS13 proteins. *J Cell Biol* 221, e202202030. doi: 10.1083/jcb.202202030
- Agozzino L, Dill KA (2018). Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *Proc Natl Acad Sci* 115, 9092–9097. doi: 10.1073/pnas.1810194115
- Altschul S (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. doi: 10.1126/science.abj8754
- Bean BDM, Dziurdzik SK, Kolehmainen KL, Fowler CMS, Kwong WK, Grad LI, Davey M, Schluter C, & Conibear E (2018). Competitive organelle-specific adaptors recruit Vps13 to membrane contact sites. *J Cell Biol*, 217(10), 3593–3607. doi: 10.1083/jcb.201804111
- Becker B (2013). Snow ball earth and the split of streptophyta and chlorophyta. *Trends Plant Sci* 18, 180–183. doi: 10.1016/j.tplants.2012.09.010
- Bowles AMC, Williamson CJ, Williams TA, Lenton TM, Donoghue PCJ (2023). The origin and early evolution of plants. *Trends Plant Sci* 28, 312–329. doi: 10.1016/j.tplants.2022.09.009
- Bulankina AV, Thoms S (2020). Functions of vertebrate ferlins. *Cells* 9, 534. doi: 10.3390/cells9030534
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009). Trimal: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Dabrowski R, Tulli S, Graef M (2023). Parallel phospholipid transfer by Vps13 and Atg2 determines autophagosome biogenesis dynamics. *J Cell Biol* 222, e202211039. doi: 10.1083/jcb.202211039

- Da Costa R, Bordessoules M, Guilleman M, Carmignac V, Lhussiez V, Courrot H, Bataille A, Chlémaire A, Bruno C, Fauque P, et al. (2020). Vps13b is required for acrosome biogenesis through functions in Golgi dynamic and membrane trafficking. *Cell Mol Life Sci* 77, 511–529. doi: 10.1007/s00018-019-03192-4
- Dall'Armellina F, Stagi M, Swan LE (2023). In silico modeling human VPS13 proteins associated with donor and target membranes suggests lipid transfer mechanisms. *Proteins: Struct Funct Bioinform* 91, 439–455. doi: 10.1002/prot.26446
- Domozych DS, Bagdan K (2022). The cell biology of charophytes: Exploring the past and models for the future. *Plant Physiol* 190, 1588–1608. doi: 10.1093/plphys/kiac390
- Douzgou S, Petersen MB (2011). Clinical variability of genetic isolates of Cohen syndrome. *Clin Genet* 79, 501–506. doi: 10.1111/j.1399-0004.2011.01669.x
- Dziurdzik SK, Bean BDM, Davey M, Conibear E (2020). A VPS13D spastic ataxia mutation disrupts the conserved adaptor-binding site in yeast Vps13. *Hum Mol Genet* 29, 635–648. doi: 10.1093/hmg/ddz318
- Dziurdzik SK, Conibear E (2021). The Vps13 family of lipid transporters and its role at membrane contact sites. *Int J Mol Sci* 22(6), 2905. doi: 10.3390/ijms22062905
- Edqvist J, Blomqvist K, Nieuwland J, Salminen TA (2018). Plant lipid transfer proteins: Are we finally closing in on the roles of these enigmatic proteins? *J Lipid Res* 59, 1374–1382. doi: 10.1194/jlr.R083139
- Edstam MM, Viitanen L, Salminen TA, Edqvist J (2011). Evolutionary history of the non-specific lipid transfer proteins. *Mol Plant* 4, 947–964. doi: 10.1093/mp/ssr019
- Fernando DD, Lazzaro MD, Owens JN (2005). Growth and development of conifer pollen tubes. *Sex Plant Reprod* 18, 149–162. doi: 10.1007/s00497-005-0008-y
- Ferreira JV, Carvalho P (2021). Pex30-like proteins function as adaptors at distinct ER membrane contact sites. *J Cell Biol* 220, e202103176. doi: 10.1083/jcb.202103176
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Grigoriev IV, Hayes RD, Calhoun S, Kamel B, Wang A, Ahrendt S, Dusheyko S, Nikitin R, Mondo SJ, Salamov A, et al. (2021). Phycocosm, a comparative algal genomics resource. *Nucleic Acids Res* 49, D1004–D1011. doi: 10.1093/nar/gkaa898
- Jones EW (1977). Proteinase mutants of *saccharomyces cerevisiae*. *Genetics* 85, 23–33. doi: 10.1093/genetics/85.1.23
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Katoh K, Standley DM (2013). MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30, 772–780. doi: 10.1093/molbev/mst010
- Keeling PJ (2010). The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond. B, Biol Sci* 365, 729–748. doi: 10.1098/rstb.2009.0103
- Koike S, Jahn R (2019). SNAREs define targeting specificity of trafficking vesicles by combinatorial interaction with tethering factors. *Nat Commun* 10, 1608. doi: 10.1038/s41467-019-09617-9
- Koizumi K, Gallagher KL (2013). Identification of SHRUBBY, a SHORT-ROOT and SCARECROW interacting protein that controls root growth and radial patterning. *Development* 140, 1292–1300. doi: 10.1242/dev.090761
- Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001). Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. Edited by F. Cohen. *J Mol Biol* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Kumar N, Leonzino M, Hancock-Cerutti W, Horenkamp FA, Li P, Lees JA, Wheeler H, Reinisch KM, De Camilli P (2018). VPS13A And VPS13C are lipid transport proteins differentially localized at ER contact sites. *J Cell Biol* 217, 3625–3639. doi: 10.1083/jcb.201807019
- Land WJG (1902). A morphological study of Thuja. *Bot Gaz* 34, 249–259. doi: 10.1086/328288
- Lefort V, Desper R, Gascuel O (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program: Table 1. *Mol Biol Evol* 32, 2798–2800. doi: 10.1093/molbev/msv150
- Leonzino M, Reinisch KM, De Camilli P (2021). Insights into VPS13 properties and function reveal a new mechanism of eukaryotic lipid transport. *Biochim Biophys Acta Mol Cell Biol Lipids* 1866, 159003. doi: 10.1016/j.bbalip.2021.159003
- Leterme S, Michaud M (2022). Non-vesicular glycerolipids transport in plant cells. In: *Lipids in Plants and Algae: From Fundamental Science to Industrial Applications* 101, pp. 121–189. *Advance in botanical research*. doi: 10.1016/bs.abr.2021.07.001
- Levine TP (2022). Sequence analysis and structural predictions of lipid transfer bridges in the repeating beta groove (RBG) superfamily reveal past and present domain variations affecting form, function and interactions of VPS13, ATG2, SHIP164, hobbit and tweek. *Contact* 5, 251525642211343. doi: 10.1177/25152564221134328
- Li P, Lees JA, Lusk CP, Reinisch KM (2020a). Cryo-EM reconstruction of a VPS13 fragment reveals a long groove to channel lipids between membranes. *J Cell Biol* 219, e202001161. doi: 10.1083/jcb.202001161
- Li L, Wang S, Wang H, Sahu SK, Marin B, Li H, Xu Y, Liang H, Li Z, Cheng S, et al. (2020b). The genome of prasinoderma coloniale unveils the existence of a third phylum within green plants. *Nat Ecol Evol* 4, 1220–1231. doi: 10.1038/s41559-020-1221-7
- Longo LM, Kolodny R, McGlynn SE (2022). Evidence for the emergence of  $\beta$ -trefoils by 'peptide budding' from an IgG-like  $\beta$ -sandwich. *PLOS Comput Biol* 18, e1009833. doi: 10.1371/journal.pcbi.1009833
- Melia TJ, Reinisch KM (2022). A possible role for VPS13-family proteins in bulk lipid transfer, membrane expansion and organelle biogenesis. *J Cell Sci* 135, jcs259357. doi: 10.1242/jcs.259357
- Michaud M, Jouhet J (2019). Lipid trafficking at membrane contact sites during plant development and stress response. *Front Plant Sci* 10, 2. doi: 10.3389/fpls.2019.00002
- Michaud M, Prinz WA, Jouhet J (2017). Glycerolipid synthesis and lipid trafficking in plant mitochondria. *FEBS J* 284, 376–390. doi: 10.1111/febs.13812
- Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M (2022). Colabfold: Making protein folding

- accessible to all. *Nat Methods* 19, 679–682. doi: 10.1038/s41592-022-01488-1
- Moellering ER, Benning C (2011). Galactoglycerolipid metabolism under stress: A time for remodeling. *Trends Plant Sci* 16, 98–107. doi: 10.1016/j.tplants.2010.11.004
- Neuman SD, Levine TP, Bashirullah A (2022). A novel superfamily of bridge-like lipid transfer proteins. *Trends Cell Biol*, 32(11), 962–974. doi: 10.1016/j.tcb.2022.03.011
- Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB, Vanderstraeten L, Becker D, Lang D, et al. (2018). The chara genome: Secondary complexity and implications for plant terrestrialization. *Cell* 174, 448–464.e24. doi: 10.1016/j.cell.2018.06.033
- Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, Smirnova T, Grigoriev IV, Dubchak I (2014). The genome portal of the department of energy joint genome institute: 2014 updates. *Nucleic Acids Res* 42, D26–D31. doi: 10.1093/nar/gkt1069
- Park J-S, Neiman AM (2012). *VPS13* Regulates membrane morphogenesis during sporulation in *Saccharomyces cerevisiae*. *J Cell Sci* 125, 3004–3011. doi: 10.1242/jcs.105114
- Park J-S, Okumura Y, Tachikawa H, Neiman AM (2013). *SPO71* Encodes a developmental stage-specific partner for Vps13 in *saccharomyces cerevisiae*. *Eukaryot Cell* 12, 1530–1537. doi: 10.1128/EC.00239-13
- Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, et al. (2023). Interpro in 2022. *Nucleic Acids Res* 51, D418–D427. doi: 10.1093/nar/gkac993
- Petroutsos D, Amiar S, Abida H, Petroutsos D, Amiar S, Abida H, Dolch LJ, Bastien O, Rébeillé F, Jouhet J, Falconet D, Block MA, McFadden GI, et al. (2014). Evolution of galactoglycerolipid biosynthetic pathways – from cyanobacteria to primary plastids and from primary to secondary plastids. *Prog Lipid Res* 54, 68–85. doi: 10.1016/j.plipres.2014.02.001
- Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* 30, 70–82. doi: 10.1002/pro.3943
- Richter DJ, Berney C, Strassert JFH, Poh YP, Herman EK, Muñoz-Gómez SA, Wideman JG, Burki F, de Vargas C (2022). Eukprot: A database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Commun J* 2, e56. doi: 10.24072/pcjournal.173
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61, 539–542. doi: 10.1093/sysbio/sys029
- Saheki Y, De Camilli P (2017). The extended-synaptotagmins. *Biochim Biophys Acta* 1864, 1490–1493. doi: 10.1016/j.bbamcr.2017.03.013
- Sayers EW, Bolton EE, Brister JR, Canese K., Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, et al. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Res* 50, D20–D26. doi: 10.1093/nar/gkab1112
- Seifert W, Kühnisch J, Maritzen T, Horn D, Haucke V, Hennies HC (2011). Cohen syndrome-associated protein, COH1, is a novel, giant Golgi matrix protein required for Golgi integrity. *J Biol Chem* 286, 37665–37675. doi: 10.1074/jbc.M111.267971
- Seifert W, Kühnisch J, Maritzen T, Lommatzsch S, Hennies HC, Bachmann S, Horn D, Haucke V (2015). Cohen syndrome-associated protein COH1 physically and functionally interacts with the small GTPase RAB6 at the Golgi complex and directs neurite outgrowth. *J Biol Chem* 290, 3349–3358. doi: 10.1074/jbc.M114.608174
- Shih PM, Matzke NJ (2013). Primary endosymbiosis events date to the later proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc Natl Acad Sci U S A* 110, 12355–12360. doi: 10.1073/pnas.1305813110
- Slee JA, Levine TP (2019). Systematic prediction of FFAT motifs across eukaryote proteomes identifies nucleolar and eisosome proteins with the predicted capacity to form bridges to the endoplasmic reticulum. *Contact* 2, 251525641988313. doi: 10.1177/2515256419883136
- The UniProt Consortium, Bateman A, Martin M-J, Orchard S, Magrane M, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bye-A-Jee H, Cukura A, et al. (2023). Uniprot: The universal protein knowledgebase in 2023. *Nucleic Acids Res* 51, D523–D531. doi: 10.1093/nar/gkac1052
- Torgerson DG, Kulathinal RJ, Singh RS (2002). Mammalian sperm proteins are rapidly evolving: Evidence of positive selection in functionally diverse genes. *Mol Biol Evol* 19, 1973–1980. doi: 10.1093/oxfordjournals.molbev.a004021
- Ugur B, Hancock-Cerutti W, Leonzino M, De Camilli P (2020). Role of VPS13, a protein with similarity to ATG2, in physiology and disease. *Curr Opin Genet Dev* 65, 61–68. doi: 10.1016/j.gde.2020.05.027
- Van Dijk PJ, Op den Camp R, Schauer SE (2020). Genetic dissection of apomixis in dandelions identifies a dominant parthenogenesis locus and highlights the complexity of autonomous endosperm formation. *Genes* 11, 961. doi: 10.3390/genes11090961
- Velayos-Baeza A, Vettori A, Copley RR, Dobson-Stone C, Monaco AP (2004). Analysis of the human VPS13 gene family. *Genomics* 84, 536–549. doi: 10.1016/j.ygeno.2004.04.012
- Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ (2007). An “electronic fluorescent pictograph” browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* 2, e718. doi: 10.1371/journal.pone.0000718
- Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol* 21, 809–818. doi: 10.1093/molbev/msh075
- Zaitseva J, Vaknin D, Krebs C, Doroghazi J, Milam SL, Balasubramanian D, Duck NB, Freigang J (2019). Structure-function characterization of an insecticidal protein GNIPIAa, a member of an MACPF and  $\beta$ -tripod families. *Proc Natl Acad Sci* 116, 2897–2906. doi: 10.1073/pnas.1815547116
- Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 430, 2237–2243. doi: 10.1016/j.jmb.2017.12.007