EXPLORING THE ORIGINS OF SEPTIN GTPASES

by

BRENT SHUMAN

(Under the Direction of Michelle Momany)

ABSTRACT

Septins are a class of cytoskeletal GTPase used in cytokinesis and conserved in most eukaryotes. Septins form hetero-hexamers or hetero-octamers in a predetermined order strictly based on their ancestry. When comparing fungal and animal septin monomer order within polymers, some septins are seated in the same relative position as their cross-kingdom septin with the most recent common ancestor while others do not. Many attempts and revisions to classify the evolutionary history of septins and group them according to their functions have been made. The first study documents the history of septin classification and compares the septins within the most commonly researched model organisms used for septin research. The second study expands the collection of non-opisthokont (non-animal or fungal species) septins and revises septin group memberships. Using ancestral sequence reconstruction, secondary structure prediction, and sequence alignment, the mechanism for the origins of heteropolymerization from an ancestral homodimer is hypothesized. The final study began while studying the techniques used in the above studies. This study shows intra-species contamination can be detected using B-allele frequency visualization, and documents that low levels of contamination are sufficient to disrupt phylogenetic and population genomic analyses.

INDEX WORDS: Septin, GTPase, Cytoskeleton, Orthology, Polymerization, Evolution, Opisthokont, Protist, Phylogenetics, Arginine-Finger, Contamination, Ploidy, Admixture, SNPs, B-Allele Frequency

EXPLORING THE ORIGINS OF SEPTIN GTPASES

by

BRENT SHUMAN

B.S., Winthrop University, 2018

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2024

EXPLORING THE ORIGINS OF SEPTIN GTPASES

by

BRENT SHUMAN

Major Professor:	Michelle Momany
Committee:	Chang-Hyun Khang
	Douda Bensasson
	Zachary Lewis

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2024

DEDICATION

To my wife, Cassidy: without you I would have quit years ago.

ACKNOWLEDGEMENTS

Thank you, Michelle. I could not have asked for a better mentor. You have encouraged and mentored me to be a better scientist, teacher, and human, in reverse order.

Thank you, Chang-Hyun and Zach, for providing a base of fungal knowledge I will keep with me always.

Thank you, Douda, for introducing me to bioinformatics and for modeling how to ask deep questions about phylogenetics and population genetics.

Thank you, Paola, for your mentoring, and training in teaching and professional development.

Thank you, Alex and Earl, for mentoring me in the ways of the fungus, and setting strong examples to follow behind.

Thank you, Justina, for asking questions people think they know the answer to, but, actually, do not. I have learned much from our conversations.

Thank you, Brandi and Manny, for being exquisite friends.

Thank you, Mom and Dad, for loving me, supporting me, and fostering my curiosity about the world. I love you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

This dissertation is divided into two major sections. The first section, encompassing Chapters 2 and 3, documents the history of and expands the field's understanding of septin origins and expansion. The second section, Chapter 4, began during a laboratory rotation to prepare for the analyses in the first section, and documents the observation of intra-species contamination of whole genome sequences and its impact on downstream analyses.

Septin Classification

Septin GTPases were first discovered in yeast where they were found to be essential for cytokinesis (Hartwell et al., 1970). Of the organisms that have septins, most have at least three, and some have up to seventeen orthologs (Pan et al., 2006, Willis et al., 2016). The first attempt to classify septins by Kinoshita (2003a) found there were four phylogenetic groups in humans each named after one of the group members (Sept2, Sept3, Sept6, and Sept7). Except for Sept7, each group contained more than one protein, and further analysis with septin sequences from more animals confirmed Sept7 was a separate group (Kinoshita, 2003b). Later, crystal structures of human septins confirmed one member from each group joins together to form a heteropolymer with the group-order being Sept2-6-7-3-3-7-6-2 (Mendonça et al., 2019). After expanding the search to include more fungi and animals, Pan et al. (2006) re-classified septins into 5 groups. Core septins within Groups 1-4 are incorporated into filaments while there is no evidence to suggest that non-core septin members of Group 5 act canonically (Hernandez-Rodriguez et al., 2014). All animal septins were placed in Momany Groups 1a, 1b and 2b while fungal septins

representatives were present Groups 1a, 2a, 3, 4, and 5 (Pan et al., 2006). Since then, a non-insignificant number of non-opisthokont septins were discovered, yet their fit within the septin-family phylogeny is not as clear as the opisthokonts (Cao et al. 2007, and Wloga et al., 2008, Nishihama et al., 2011, and Onishi and Pringle, 2016).

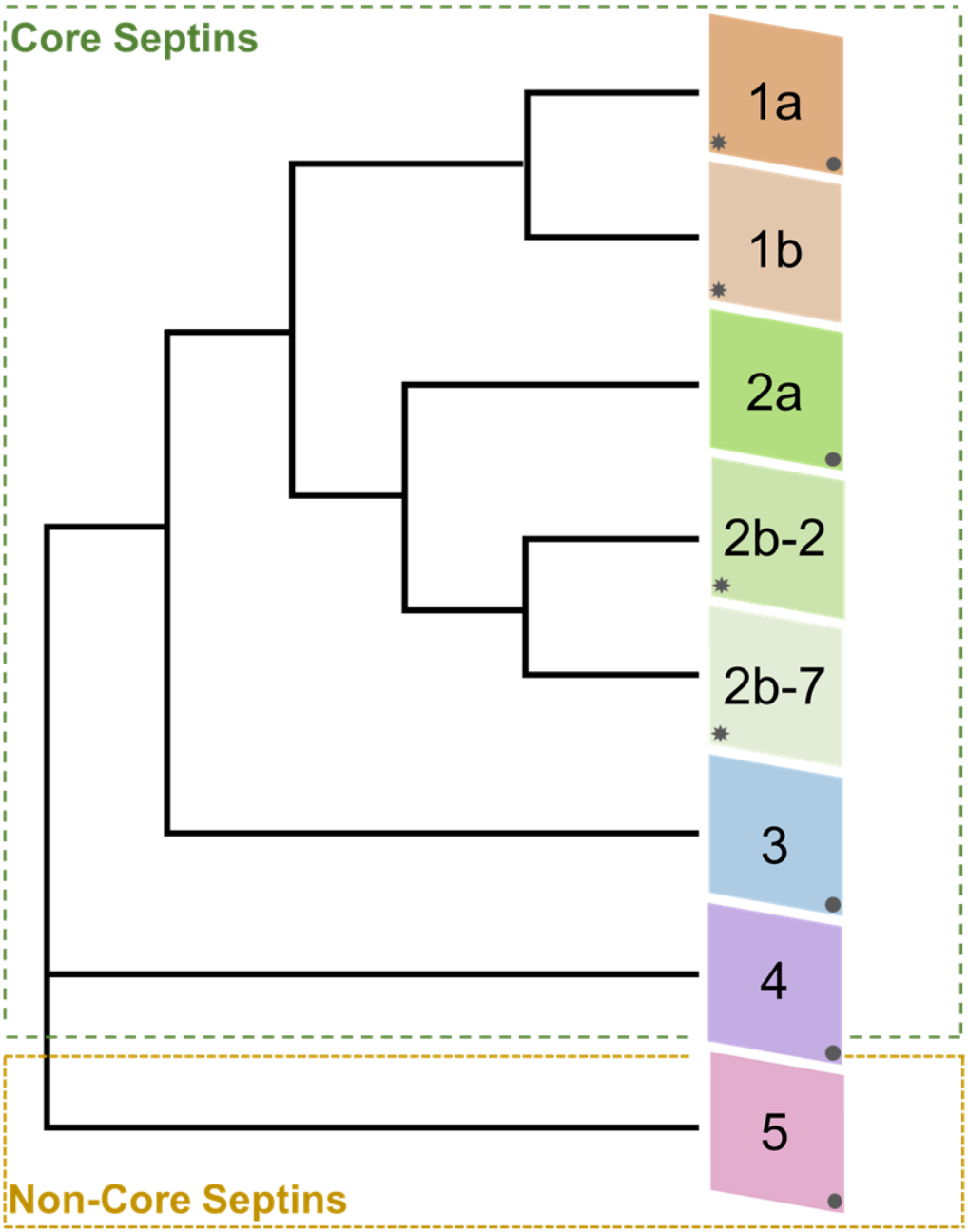

Figure 1.1. Cladogram showing the Momany septin groups. Septins can be divided into 5 groups with additional subgroups. A star in the bottom left corner of a group indicates presence

of animal septins within-group. A circle in the bottom right corner of a group indicates presence of fungal septins within-group. This cladogram is based on the phylogenetic tree from Pan et al. (2006), but, unlike that tree, Groups 2b-2 and 2b-7 were distinguished to point out Sept2 and Sept7 as monophyletic subgroups within Group 2b.

Idiosyncratic Heteropolymerization of Septins

Microfilaments, microtubules, and septin filaments are all cytoskeletal elements that form polymers, but they use various strategies to do so. Microfilament polymers are formed by homomeric G-actin subunits; microtubule polymers are formed by heterodimeric subunits composed of an alpha-tubulin bound to a beta-tubulin. Septin filaments are composed of hetero-hexameric or hetero-octameric subunits composed of three or four separate sub-group members mirrored across a central homodimer.

Actin forms homopolymers such that the binding faces on either side of a G-actin monomer bind with the same gene product's opposite interface on the neighboring G-actin monomer. This creates polarity within the polymer whereby the barbed end from one monomer binds to the pointed end of the neighboring monomer. All interfaces necessary for polymerization are encoded within one gene.

Unlike actin, alpha-tubulin and beta-tubulin have interfaces specific for binding a separate gene product. Alpha- and beta-tubulin thus have two classes of binding sites: the intra-heterodimer interfaces which facilitate alpha-beta-tubulin dimerization and the inter-heterodimer interfaces which facilitate microtubule formation. Some organisms have more than one isotype of either tubulin, but interfaces necessary for polymerization are encoded within two gene groups.[1]

[1] Of note, microtubules require a gamma-tubulin seed to begin polymerization, but since gamma-tubulin is not incorporated into the mature microtubule, those interfaces are left out in this comparison CITATION

Septins are heteropolymers unlike actin, and, unlike microtubules, contain three or four different proteins within one heteropolymer (Sellin et al., 2011). Consider again the human septin octamer Sept2-Sept6-Sept7-Sept3-Sept3-Sept7-Sept6-Sept2[2] with the next hetero-octamer identical to the one before it, joining at Sept2. Self-self and self-nonself binding occurs, and no two septins share the same binding partners. All interfaces necessary for polymerization are encoded within four gene groups.[3] This points to the specificity of interaction and strength of selection needed to maintain strict monomer order within opisthokont septin heteropolymers.

On the Origins of Septins

The majority of discovered septins are within clade Opisthokonta which bifurcates into the clade Holozoa including all animals, and the clade Holomycota including all fungi (Auxier et al. 2019). High septin copy number across opisthokonts allows for considerable modularity of hexamer vs octamer expression (Sellin et al., 2011, Hernandez-Rodriguez et al., 2014) and within-group substitution in polymers as exemplified in the mitotic/meiotic septin swap in *Saccharomyces cerevisiae* (De Virgilio et al., 1996). The complexity of this septin polymerization has only been shown in opisthokonts possibly due to the small number of non-opisthokont septins that have been identified (Onishi and Pringle, 2016). While some non-opisthokonts have multiple septins whose localizations have been visualized, it is unknown whether they are able to form heteropolymers (Wloga et al., 2008).

It is hypothesized that septins arose from a horizontal gene transfer event from a bacterial GTPase into an early eukaryote (Leipe, 2002). Because of the underrepresentation of septin phylogenetics and biochemistry from protist lineages, it is difficult to know when septins first

[2] If this were a hetero-hexamer instead, the protein sequence would be Sept2-Sept6-Sept7-Sept7-Sept6-Sept2
[3] Three genes if referencing a hetero-hexamer.

arose in eukaryotes, which secondary structures were present, and whether GTPase function was necessary for polymerization.

Contamination Whole Genome Sequences

Robust genomic analysis requires quality genome assemblies. Contamination of whole genome sequence reads is a known problem that can come from both *in vitro* and *in silico* sources (Cornet and Baurain, 2022). Both inter- and intra-species contamination can mislead researchers to observe unique genomic events such as horizontal gene transfer across domains and sexual reproduction in asexual organisms (Goig et al., 2020, Wilson et al., 2018). While tools exist to detect contamination, they largely apply to prokaryotic organisms or inter-species contamination (Cornet and Baurain, 2022).

Summary

Chapter 2 reviews the history of septin subgroup classification with special emphasis on the conservation of septin-specific domains and motifs as well as the interface residues that facilitate septin-septin heteropolymerization. Chapter 3 catalogs novel non-opisthokont septins, revises the number and distribution of septin subgroups, comments on the origin of heteropolymerization and predicts the structure and function of ancestral septins. Chapter 4 demonstrates the use of B-allele frequency visualization to identify intra-species contamination, and examines the effects of intra-species contamination on ploidy estimation, SNP availability, phylogenetics and population genomic analysis.

REFERENCES

Auxier, B., Dee, J., Berbee, M. L., and Momany, M. (2019). Diversity of Opisthokont Septin Proteins Reveals Structural Constraints and Conserved Motifs. *BMC Evol. Biol.* 19 (1), 4. doi:10.1186/s12862-018-1297-8

Cao, L., Ding, X., Yu, W., Yang, X., Shen, S., and Yu, L. (2007). Phylogenetic and Evolutionary Analysis of the Septin Protein Family in Metazoan. *FEBS Lett.* 581 (28), 5526–5532. doi:10.1016/j.febslet.2007.10.032

Cornet, L., Baurain, D. (2022). Contamination detection in genomic data: more is not enough. Genome Biol 23, 60. https://doi.org/10.1186/s13059-022-02619-9

De Virgilio C., DeMarini D. J. , Pringle J. R. (1996). SPR28, a sixth member of the septin gene family in *Saccharomyces cerevisiae* that is expressed specifically in sporulating cells. *Micro*. 142(10):2897-2905. doi: 10.1099/13500872-142-10-2897.

Goig, G.A., Blanco, S., Garcia-Basteiro, A.L., Comas, I. (2020). Contaminant DNA in bacterial sequencing experiments is a major source of false genetic variability. *BMC Biol* 18, 24. https://doi.org/10.1186/s12915-020-0748-z

Hartwell, L. H., Culotti, J., and Reid, B. (1970). Genetic Control of the Cell-Division Cycle in Yeast, I. Detection of Mutants. *Proc. Natl. Acad. Sci.* 66 (2), 352–359. doi:10.1073/pnas.66.2.352

Hernández-Rodríguez, Y., Masuo, S., Johnson, D., Orlando, R., Smith, A., Couto-Rodriguez, M., et al. (2014). Distinct Septin Heteropolymers Co-exist during Multicellular Development in the Filamentous Fungus *Aspergillus nidulans. PLoS One* 9 (3), e92819. doi:10.1371/journal.pone.0092819

Kinoshita, M. (2003a). Assembly of Mammalian Septins. *J. Biochem*. 134 (4), 491–496.

doi:10.1093/jb/mvg182

Kinoshita, M. (2003b). The Septins. *Genome Biol.* 4 (11), 236. doi:10.1186/gb-2003-4-11-236

Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002). Classification and Evolution of P-Loop GTPases and Related ATPases. *J. Mol. Biol.* 317 (1), 41–72. doi:10.1006/jmbi.2001.5378

Mendonça, D. C., Macedo, J. N., Guimarães, S. L., Barroso da Silva, F. L., Cassago, A., Garratt, R. C., et al. (2019). A Revised Order of Subunits in Mammalian Septin Complexes. *Cytoskeleton (Hoboken)* 76 (9-10), 457–466. doi:10.1002/cm.21569=

Nishihama, R., Onishi, M., and Pringle, J. R. (2011). New insights into the phylogenetic distribution and evolutionary origins of the septins. Biol Chem 392, 681–687. doi: 10.1515/BC.2011.086

Onishi, M., and Pringle, J. R. (2016). The nonopisthokont septins: How many there are, how little we know about them, and how we might learn more. Methods Cell Biol 136, 1–19. doi: 10.1016/bs.mcb.2016.04.003

Pan, F., Malmberg, R. L., and Momany, M. (2007). Analysis of Septins across Kingdoms Reveals Orthology and New Motifs. *BMC Evol. Biol.* 7, 103. doi:10.1186/1471-2148-7-103

Sellin M.E., Sandblad L., Stenmark S., Gullberg M. (2011). Deciphering the rules governing assembly order of mammalian septin complexes. *Mol Biol Cell.* 22(17):3152-64. doi: 10.1091/mbc.E11-03-0253.

Willis, A., Mazon-Moya, M., , Mostowy, S. (2016) Chapter 13 - Investigation of septin biology in vivo using zebrafish. *Meth in Cell Bio.* 136, 221-241. https://doi.org/10.1016/bs.mcb.2016.03.019.

Wilson, G.C., Nowell, R.W., and Barraclough, T.G. (2018). Cross-Contamination Explains "Inter and Intraspecific Horizontal Genetic Transfers" between Asexual Bdelloid Rotifers. *Curr Biol.* 28(15), 2436-2444. https://doi.org/10.1016/j.cub.2018.05.070.

Wloga, D., Strzyzewska-Jówko, I., Gaertig, J., and Jerka-Dziadosz, M. (2008). Septins Stabilize Mitochondria in *Tetrahymena thermophila*. *Eukaryot. Cel* 7 (8), 1373–1386. doi:10.1128/EC.00085-08

CHAPTER 2

SEPTINS FROM PROTISTS TO PEOPLE[4]

[4] Shuman, B. and M. Momany. 2022. *Front. Cell Dev. Biol.* 9:824850. Reprinted here under the Creative Commons CC-BY license

Abstract

Septin GTPases form nonpolar heteropolymers that play important roles in cytokinesis and other cellular processes. The ability to form heteropolymers appears to be critical to many septin functions and to have been a major driver of the high conservation of many septin domains. Septins fall into five orthologous groups. Members of Groups 1–4 interact with each other to form heterooligomers and are known as the "core septins." Representative core septins are present in all fungi and animals so far examined and show positional orthology with monomer location in the heteropolymer conserved within groups. In contrast, members of Group 5 are not part of canonical heteropolymers and appear to interact only transiently, if at all, with core septins. Group 5 septins have a spotty distribution, having been identified in specific fungi, ciliates, chlorophyte algae, and brown algae. In this review we compare the septins from nine well-studied model organisms that span the tree of life (*Homo sapiens*, *Drosophila melanogaster*, *Schistosoma mansoni*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Aspergillus nidulans*, *Magnaporthe oryzae*, *Tetrahymena thermophila*, and *Chlamydomonas reinhardtii*). We focus on classification, evolutionary relationships, conserved motifs, interfaces between monomers, and positional orthology within heteropolymers. Understanding the relationships of septins across kingdoms can give new insight into their functions.

Introduction

Septin GTPases form nonpolar heteropolymers and are a component of the cytoskeleton (Mostowy and Cossart 2012). The first septins were identified in *Saccharomyces cerevisiae* in the classic cell division cycle (cdc) mutant screen. The four prototypical septin mutants, cdc3, cdc10, cdc11, and cdc12, made elongated buds and did not complete cytokinesis at restrictive temperature (Hartwell et al., 1970; Hartwell et al., 1973). In the 50 years since their discovery, septins have been identified in organisms across kingdoms and much work has been done to understand septin evolution, structure, assembly, and function (Angelis and Spiliotis 2016; Momany and Talbot 2017; Spiliotis and McMurray 2020; Spiliotis and Nakos 2021; Woods and Gladfelter 2021). The purpose of this review is to allow researchers to understand the relationships of septins across kingdoms so that they can use insights from other systems to inform their own work. We will do this by comparing septins from nine model systems (Figure 2.1), focusing on classifications based on evolutionary relationships and conserved motifs, including interfaces. We chose these nine organisms because they span the tree of life and are well-studied, but we urge researchers to remember that there is great diversity in septin structure and function and we are covering only a small, though hopefully representative, slice. Much of the knowledge presented in this review builds on many structural and functional studies conducted in many labs over many years. We apologize to our colleagues whose work might be missed.

Momany Group[1]	Kinoshita Group[2]	Hs[3]	Dm	Sm	Ce	Sc	An	Mo	Tt	Cr
1a	SEPT3	SEPT3 SEPT9 SEPT12				Cdc10	AspD	Sep4		
1b	SEPT6	SEPT6 SEPT8 SEPT10 SEPT11 SEPT14	Sep2 Sep5	SEPT10	UNC61					
2a						Cdc3	AspB	Sep3		
2b	SEPT2	SEPT1 SEPT2 SEPT4 SEPT5	Sep1 Sep4	SEPT5						
	SEPT7	SEPT7	Pnut	SEPT7.1 SEPT7.2	UNC59					
3						Cdc11 Shs1 Spr28	AspA	Sep5		
4						Cdc12 Spr3	AspC	Sep6		
5							AspE	Sep7	Sep2 Sep3	Sep1

Figure 2.1. Groupings of septins across Kingdoms. Septin designations for representative model organisms across kingdoms are shown. Protein sequences were retrieved from UniProt, WormBase, NCBI, or FungiDB.

[1] All 161 then-available septin sequences were classified into 5 groups by the Momany lab (Pan, Malmberg et al. 2007).

[2] All 12 then-available human septins were classified into 4 homologous groups, members of which were proposed to substitute for each other in a polymer ("Kinoshita Rule") by the Kinoshita lab (Kinoshita 2003).

[3]Hs, *Homo sapiens*: SEPT1(*Q8WYJ6*), SEPT2(*Q15019*), SEPT3(*Q9UH03*), SEPT4(*Q6ZU15*), SEPT5(*Q99719*), SEPT6(*Q14141*), SEPT7(*Q16181*), SEPT8(*Q92599*), SEPT9(*Q9UHD8*), SEPT10(*Q9P0V9*), SEPT11(*Q9NVA2*), SEPT12(*Q8IYM1*), SEPT14(*Q6ZU15*); Dm, *Drosophila melanogaster*: Sep1(*P42207*), Sep2(*P54359*), Sep4(*Q0KHR7*), Sep5(*Q7KLG8*), Pnut(*P40797*); Sm, *Schistosoma mansoni*: SEPT5(*KC916723*), SEPT7.1(*KC916724*), SEPT7.2(*KC916725*), SEPT10(*KC916726*); Ce, *Caenorhabditis elegans*: UNC-59 (*CE20165*), UNC-61(*CE47829*); Sc, *Saccharomyces cerevisiae*: Cdc3(*YLR314C*), Cdc10(*YCR002C*), Cdc11(*YJR076C*), Cdc12(*YHR107C*), Shs1(*YDL225W*), Spr3(*YGR059W*), Spr28(*YDR218C*); An, *Aspergillus nidulans*: AspA(*AN4667*), AspB(*AN6688*), AspC(*AN8182*), AspD(*AN1394*), AspE(*AN10595*); Mo, *Magnaporthe oryzae*: Sep3(*MGG_01521*), Sep4(*MGG_06726*), Sep5(*MGG_03087*), Sep6(*MGG_07466*), Sep7(*MGG_02626*); Tt, *Tetrahymena thermophila*: Sep2(I7M2Q5), Sep3(Q240L4); Cr, *Chlamydomonas reinhardtii*: Sep1(*EDP03113*).

Classification and Evolution

The first attempt to classify the fungal septins was in 2001 when the Momany lab analyzed 27 septins from eight fungi (Momany et al., 2001). They concluded that most fungal septins could be grouped with one of the four *S. cerevisiae* prototypical, or core, septins, and referred to each class based on the *S. cerevisiae* septin member (Cdc3, Cdc10, Cdc11, or Cdc12). They also noted a few fungal septins did not appear to group with others and suggested these might be part of a fifth class.

The first attempt to classify the mammalian septins was in 2003 when Makoto Kinoshita placed the 12 human septins known at the time into four groups based on sequence similarity (Kinoshita 2003a). He also proposed that group members were interchangeable within heteropolymers, an idea that is often now called the "Kinoshita rule." This rule was very important in shaping septin research because while it was clear that septin monomers interacted with each other to form nonpolar heterohexamers and heterooctamers, the principles that governed their assembly were not yet known. These four Kinoshita septin groups were named after their best-studied members: SEPT2 (group also contains SEPT1, SEPT4, and SEPT5), SEPT3 (also contains SEPT9 and SEPT12), SEPT6 (also contains SEPT8, SEPT10, and SEPT11), and SEPT7 (the only group with a single representative) (Figure 2.1). In 2003 Kinoshita widened his scope with a phylogenetic analysis of 33 septins from two yeasts and three animals (Kinoshita 2003b). He concluded that septins fell into 2-4 groups within these species. He also concluded that there were no orthologs between animal and fungal septins and suggested that this meant lessons learned from septins in one kingdom might not be informative for those in the other kingdom. Contrary to this view and as described below, later analysis of a wider range of genome sequences showed that there was orthology between animal and fungal

septins and that some lessons learned could be applied across kingdoms. As more genome sequences became available, other researchers analyzed more septins and found that the Kinoshita classification system worked well for other animals in addition to humans. In work largely focused on mammalian septins, Martinez and Ware (Martinez et al., 2004) renamed the Kinoshita groups "Group I-IV" based on sequence similarity, though that designation doesn't seem to have been used much since. Later the Yu lab analyzed 78 septins from nine metazoan species, found they all fell into the four Kinoshita groups, and continued to use the Kinoshita Group names to describe them (Cao et al., 2007).

In 2007 the Momany lab performed an extensive phylogenetic analysis including all 161 septin sequences that were publicly available at the time (Pan et al., 2007). They found septins in animals, fungi, and microsporidia (which were later reclassified as fungi), but none in plants or protists. These septins were placed into five classes and were named Groups 1–5 (Figure 2.1). Animal septins were found only in Momany Groups 1 and 2, but those groups contained subgroups that were generally consistent with the Kinoshita classification (Momany Group 1a = Kinoshita Group SEPT3, 1b = SEPT6, 2b = SEPT2 and SEPT7). The Momany classification Group 2b contained both Kinoshita Group SEPT2 and Kinoshita Group SEPT7 septins.

In the Momany classification, fungal septins were found in all five groups, with Groups 3–5 being unique to fungi (Figure 2.1). The Group 5 septins were the most different from those in other groups and were only found in filamentous fungi leading the authors to speculate that Group 5 septins either diverged very recently in filamentous fungi or very early with subsequent loss in most species. Interestingly later work showed that the Group 5 septins don't participate in the heteropolymer like the prototypical core septins (Hernandez-Rodriguez et al., 2014).

As more high-quality genomes became publicly available, septin sequences were identified in organisms outside of animals and fungi including protists, ciliates, chlorophyte algae and brown algae (Merchant et al., 2007; Wloga et al., 2008; Nishihama et al., 2011). In 2011 the Pringle lab proposed that Group 5 septins were ancestral to animals and fungi based on their identification in Alveolata, Heterokonta, and Planta and additional phylogenetic research in fungal septins supports that proposition (Nishihama et al., 2011; Auxier et al., 2019). It is currently unclear if Group 5 septins are a single group or if they might contain multiple groups, a sort of "none of the above" category relative to the core septins in Groups 1–4. An analysis of septins including those from green algae, brown algae, ciliates, and diatoms published in 2013 by the Kawano lab (Yamazaki et al., 2013) suggests that Group 5 fungal septins are a distinct subgroup, separate from septins in green algae and ciliates.

Heteropolymers

With the publication of the first septin crystal structure, it became clear that core septin monomers form heteropolymers by interacting with each other via two distinct interface regions, called the G-interface and the NC-interface (Figures 2.2 C,D) (Sirajuddin et al., 2007). Several highly conserved regions first noted in Pan et al. (Pan et al., 2007) were later shown to participate in monomer contacts across interfaces that defined four interacting groups at the NC-interface and five at the G-interface (Auxier et al., 2019; Castro et al., 2020) (Figure 2.2).
It should be noted that there are far more protein structures available for human septins than for others. Structures of septins from other species would certainly help to clarify many of the outstanding questions in the field. Still, based on known structures and other biochemical and genetic analyses it is clear that canonical octamer and hexamer heteropolymers are assembled following the same general plan; that is to say evolutionarily-related septins occupy the same

positions within heteropolymers and bind using the same interfaces (McMurray and Thorner 2019; Mendonca et al., 2019; Soroor et al., 2021). This allows for the comparison of septins across evolutionary groups because they occupy equivalent positions within heteropolymers acting as "positional orthologs," though it should be noted that positional orthology does not necessarily mean that all activities are equivalent. Since SEPT2 and SEPT7 are within the same subgroup in the Momany classification, but are at different positions within polymers, we show them here as Groups 2b-2 (SEPT2) and 2b-7 (SEPT7). Octamer heteropolymers are formed with a central Group 1a septin dimer, flanked by Group 2 septins (2b-7 in animals or 2a in fungi), flanked by a Group 1b septin in animals or Group 4 septin in fungi, flanked by Group 2b-2 in animals or three in fungi (Figure 2.2D). The canonical hexamers are identical, except they lack the central Group 1a septin dimer.

Interestingly, *Caenorhaabditis elegans* has only two septins, UNC61 and UNC59. In all classifications, UNC61 was placed with the SEPT6/Group 1b septins. The placement of UNC59 has been less consistent. In the Momany lab analysis UNC59 was in a small Group 2b subclade with two septins from Caenorhabditis briggsae. This clade was an outlier, though it was closer to the subclade containing human SEPT2 than to that containing human SEPT7 (Pan et al., 2007). In the Kinoshita and Yu analyses (Kinoshita 2003a; Cao et al., 2007) UNC59 was again something of an outlier but fell into the same clade as human SEPT7.

In other model organisms examined here the Group 2b-7 septin is positioned in the interior of the heterohexamer or heterooctamer (Figure 2.2D). In contrast, UNC-59 appears to be the exposed subunit of the heterotetramer (John et al., 2007). Despite its lack of positional orthology, we show UNC59 as a member of Group 2b-7 in Figures 2.1–2.4 because UNC59 was first identified by its homology to the *Drosophila melanogaster* SEPT7 ortholog PNUT (Nguyen

et al., 2000). It will be very interesting to see which features of this unusual septin allow it to function as a heterotetramer rather than the more common heterohexamer or heterooctamer.

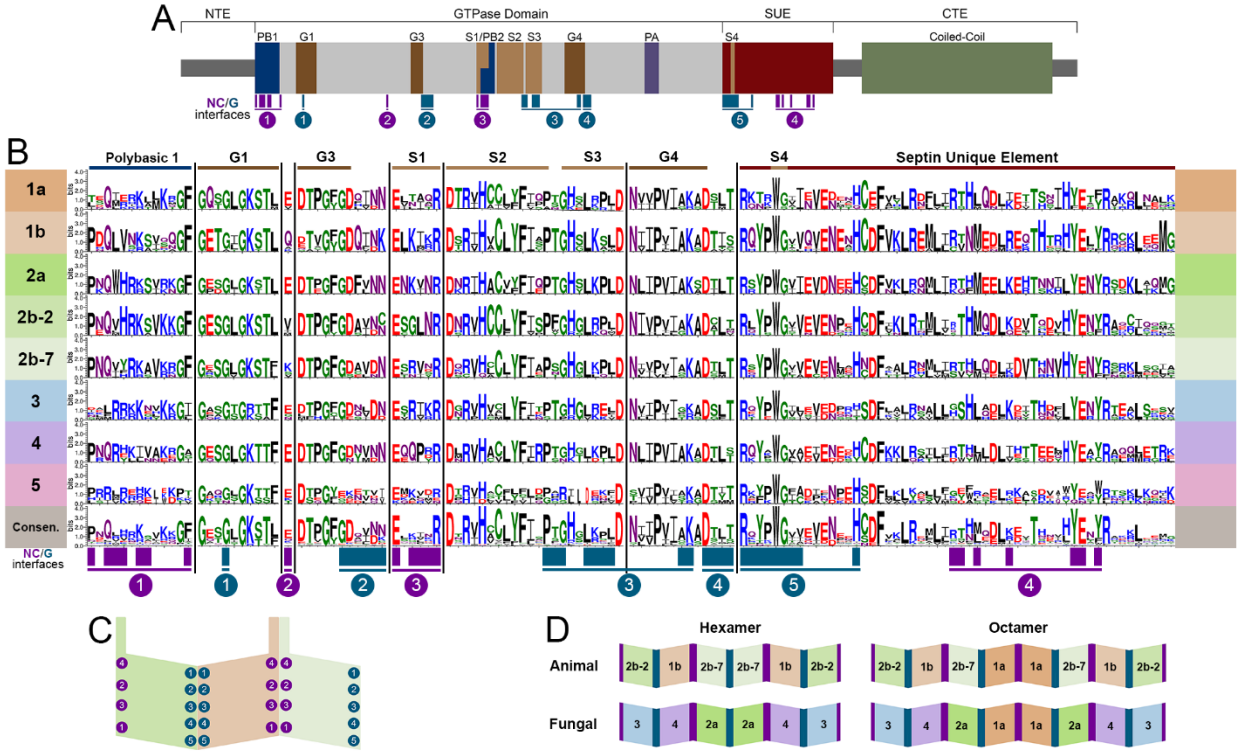

Figure 2.2 Conserved septin domains and interfaces. (A) Septin conserved domains shown to scale modeled on human SEPT7. N-Terminal Extension (NTE), polybasic region (PB), GTPase domain motifs (G1, G3, G4), septin motifs (S1, S2, S3, S4), polyacidic region (PA), septin unique element (SUE), coiled-coil, and C-Terminal Extension (CTE). NC- and G-interface residues shown in magenta and teal. **(B)** Septins were aligned using Clustal Omega (Madeira, Park et al. 2019) and WebLogo 3 (Crooks, Hon et al. 2004) was used to highlight conserved regions of septins listed in the table. **(C)** Representation of the SEPT2/6/7 trimer with interacting group designations adapted from Auxier et al. (Auxier, Dee et al. 2019). **(D)** Canonical hexamers and octamers from both animals and fungi highlighting positional orthologs.

Motifs

Not surprisingly based on their evolutionary origin from a common ancestor, septins across species share several well-conserved motifs (Figures 2.2–2.4) including those shown in Figure 2.2A: the polybasic regions (PB1 and PB2), GTPase motifs (G1, G3, G4), septin

conserved motifs (S1, S2, S3, S4), polyacidic region (PA), and the septin unique element (SUE) (Zhang et al., 1999; Leipe et al., 2002; Versele et al., 2004; Pan et al., 2007; Valadares et al., 2017; Omrane et al., 2019). Comparing the nine representative model organisms highlights interesting features of the conserved motifs across the five septin groups (Figures 2.1–2.4):

N-Terminal Extension

The length of the NTE varies widely (Figure 2.4). Interestingly the extended NTE of SEPT9 has been shown to interact with the acidic tails of microtubules and crosslink F-actin, bridging three mammalian cytoskeletal elements (Bai et al., 2013; Smith et al., 2015). This does not appear to be a feature common to other Group 1a septins (even within humans), and may have evolved independently. Among the nine model organisms we examine here, fungal-specific Groups 3 and 4 have very short NTEs with the exception of the *S. cerevisiae* meiotic septin Spr3. Representative Group 5 septins also have longer NTEs than most, but their function is currently unknown.

Polybasic Regions (PB1 and PB2)

The PBs have been shown to bind phosphoinositides in fungi and humans (Zhang et al., 1999; Casamayor and Snyder 2003; Omrane et al., 2019). PB1 of these model organisms vary in the number of basic residues (between 1 and 7) and their relative positions within the conserved region (Figure 2.2B). However, none of these septins completely lack basic residues within this N-terminal region. Though less common, acidic residues are present in PB1 of some septins. It has been shown that adding an acidic residue can reduce the ability of PB1 to bind phosphatidylinositides (Casamayor and Snyder 2003).

First identified in 2019 in human septins, PB2 has also been shown to bind phosphoinositides, and in SEPT9 both polybasic regions are thought to increase the selectivity of

the septin to particular membrane curvatures (Omrane et al., 2019). PB2 overlaps with the previously identified highly conserved Sep1 (S1) by four residues where it takes part in an NC-interface interacting group (Figure 2.2). Among the model organisms we analyzed, the presence of a PB2 region is highly conserved across groups; however, except for the highly conserved central arginine residue, the sequence of PB2 is conserved within groups, not across them (Figure 2.3B).

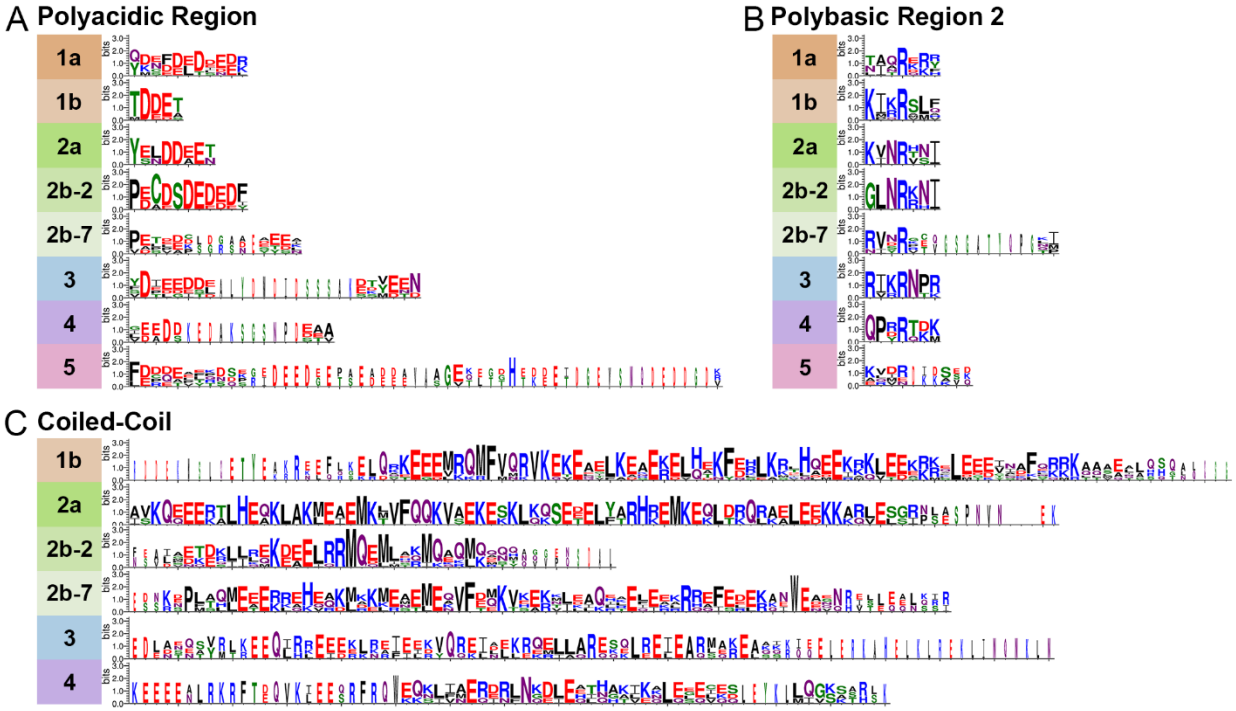

Figure 2.3. Variable Septin domains. Septins were aligned using Clustal Omega (Madeira, Park et al. 2019) and WebLogo 3 (Crooks, Hon et al. 2004) was used to view variable septin domains. **(A)** The polyacidic, **(B)** polybasic 2, and **(C)** coiled-coil regions vary between groups in size and residues.

GTPase Motifs (G1, G3, G4)

Septins contain three of the five motifs that define P-loop GTPases (Leipe et al., 2002). The G1, G3, and G4 motifs interact with the GTP nucleotide and/or its cofactor. As expected, based on this critical function, the GTPase domains are highly conserved across septins. The

Group 5 septins show the most variation in GTPase domains, though the conservation is still high.

Sep Motifs (S1-S4)

The Sep motifs were identified based on high levels of conservation across multiple species (Pan et al., 2007). Though their function was not clear at the time, it was later found that Sep1, Sep3, and Sep4 were likely involved in contacts between septin monomers within heteropolymers (Auxier et al., 2019). Once more the most variation is seen within Group 5 Sep motifs consistent with data suggesting they do not interact with other septins in the canonical manner (Hernandez-Rodriguez et al., 2014).

Polyacidic Region

First identified by the Araujo and Garratt labs in 2017 (Valadares et al., 2017), the polyacidic region is thought to interact with PB1 or PB2 at the center of the octamer depending on the conformation of the NC-interface (Castro et al., 2020). As shown for our nine model organisms the septin PA is highly variable in length and acidity across groups, but is conserved within groups (Figures 2.3A, 2.4).

Septin Unique Element

The SUE is a large (53 amino acids) region first identified based on conservation in many septins (Versele et al., 2004). It was later shown to contain contact regions for both NC- and G-interfaces (Auxier et al., 2019). No other functions have been described for this motif.

C-Terminal Extension

The CTE of septins varies more than the central regions of the protein (Figure 2.4). One of the CTE features that has drawn notice is the presence or absence of a coiled-coiled. Coiled-coils are important for septin higher-order structure formation, and recent crystal structures of the

coiled-coils reveal they interact both within and across polymers (Bertin et al., 2010; Leonardo et al., 2021). Septins from Groups 1a and five do not have predicted coiled-coils (Pan et al., 2007) (Figure 2.3C). The truncated CTE and lack of a coiled-coil is especially interesting for Group1a septins since they form the central dimer of the heterooctamer (Figure 2.2D). This central dimer is formed via the NC-interface which in all other heteropolymeric septins is the interface from which the coiled-coils emerge (Figure 2.2C).

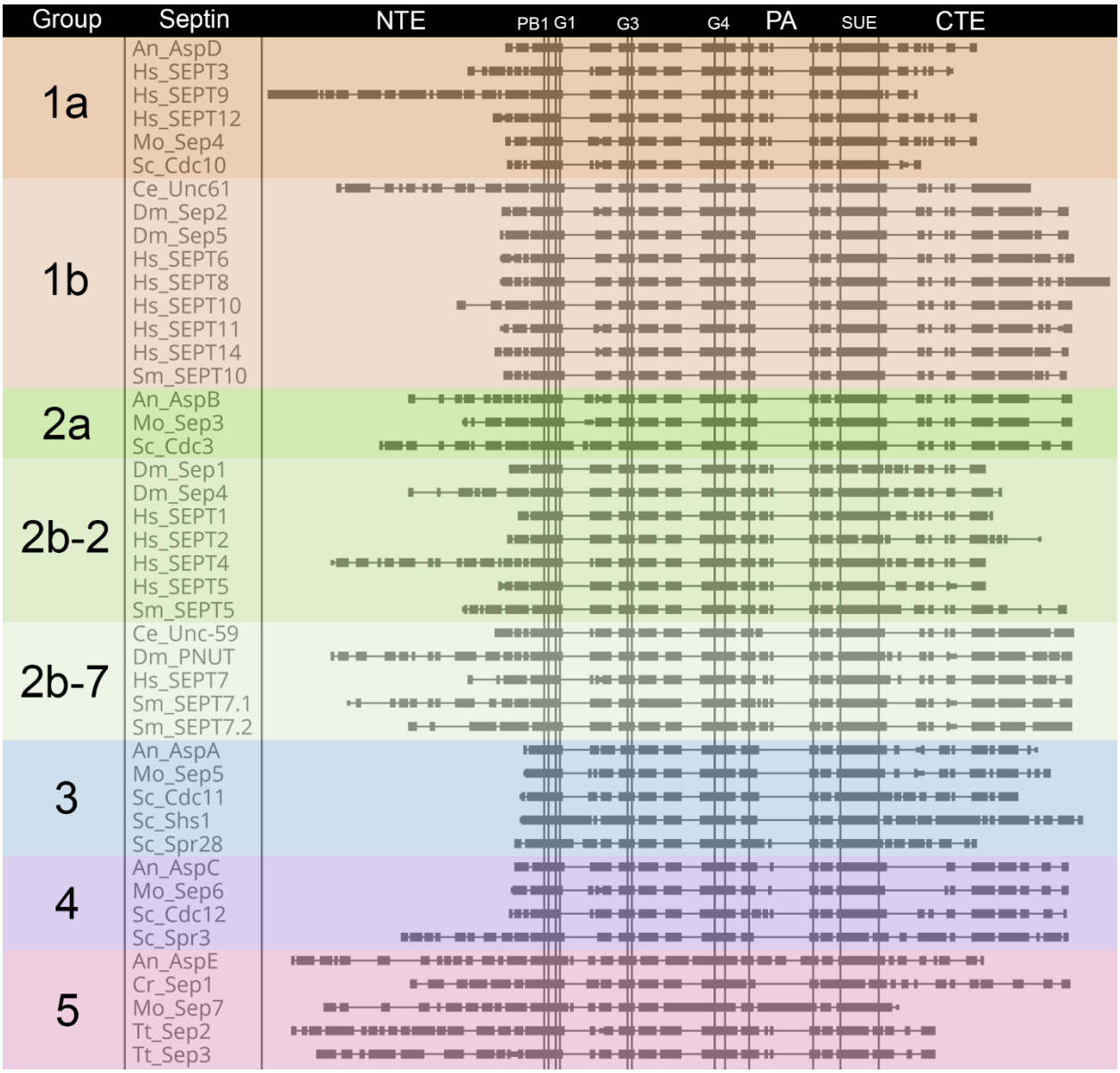

Figure 2.4. Comparison of septin domain conservation. Large scale view of septins highlighting areas with variable sequence lengths. Septin proteins were aligned using Clustal Omega (Madeira, Park et al. 2019), and placed in alphabetical order within groups. Domains are as are shown for Fig. 2.2 and are indicated above the alignment. Vertical lines denote boundaries of domains. Rectangles indicate blocks of conservation. Thin horizontal lines indicate gaps introduced by Clustal to facilitate alignment.

Another feature of note is the presence of an amphipathic helix in many septin CTEs. The distinct hydrophobic and hydrophilic faces of amphipathic helices are thought to allow septins to sense membrane curvature. So far amphipathic helices have been identified in septins from groups 1b, 2b-2, 2b-7, 3, and 4 (Cannon et al., 2019; Woods et al., 2021).

Conclusion

Septins have duplicated and diversified from their common evolutionary ancestor to perform an amazing array of important functions across species from protists to people. The ability to form heteropolymers appears to be critical to many of those functions and to have been a major driver of the high conservation of many critical domains. Because of this, comparisons can be made between animal and fungal septins that are positional orthologs, though more structural studies on septins from other species are certainly needed to complement the studies on human septins.

The significance of the divergence in polymer composition is still unknown. Perhaps the most striking example of this divergence is in the Group 5 septins which are present in some fungi and protists, absent in animals, and which appear to interact with core septins only transiently. More studies on Group 5 septins from fungi and their non-opisthokont relatives may prove critical to understanding the origins of septins and their ability to interact with each other, with other proteins, and with the membrane. Studying positional orthologs across kingdoms

should reveal the history of how septins have been added or removed from polymers and possible conserved functions of septins with shared evolutionary history.

References

Angelis, D., and Spiliotis, E. T. (2016). Septin Mutations in Human Cancers. *Front. Cel Dev. Biol.* 4, 122. doi:10.3389/fcell.2016.00122

Auxier, B., Dee, J., Berbee, M. L., and Momany, M. (2019). Diversity of Opisthokont Septin Proteins Reveals Structural Constraints and Conserved Motifs. *BMC Evol. Biol.* 19 (1), 4. doi:10.1186/s12862-018-1297-8

Bai, X., Bowen, J. R., Knox, T. K., Zhou, K., Pendziwiat, M., Kuhlenbäumer, G., et al. (2013). Novel Septin 9 Repeat Motifs Altered in Neuralgic Amyotrophy Bind and Bundle Microtubules. *J. Cel Biol* 203 (6), 895–905. doi:10.1083/jcb.201308068

Bertin, A., McMurray, M. A., Thai, L., Garcia, G., Votin, V., Grob, P., et al. (2010). Phosphatidylinositol-4,5-bisphosphate Promotes Budding Yeast Septin Filament Assembly and Organization. *J. Mol Biol* 404 (4), 711-731. https://doi.org/10.1016/j.jmb.2010.10.002

Cannon, K. S., Woods, B. L., Crutchley, J. M., and Gladfelter, A. S. (2019). An Amphipathic helix Enables Septins to Sense Micrometer-Scale Membrane Curvature. *J. Cel Biol* 218 (4), 1128–1137. doi:10.1083/jcb.201807211

Cao, L., Ding, X., Yu, W., Yang, X., Shen, S., and Yu, L. (2007). Phylogenetic and Evolutionary Analysis of the Septin Protein Family in Metazoan. *FEBS Lett.* 581 (28), 5526–5532. doi:10.1016/j.febslet.2007.10.032

Casamayor, A., and Snyder, M. (2003). Molecular Dissection of a Yeast Septin: Distinct Domains Are Required for Septin Interaction, Localization, and Function. *Mol. Cel Biol* 23 (8), 2762–2777. doi:10.1128/mcb.23.8.2762-2777.2003

Castro, D. K. S. D. V., da Silva, S. M. O., Pereira, H. D., Macedo, J. N. A., Leonardo, D. A., Valadares, N. F., et al. (2020). A Complete Compendium of crystal Structures for the Human SEPT3 Subgroup Reveals Functional Plasticity at a Specific Septin Interface. *IUCrJ* 7 (Pt 3), 462–479. doi:10.1107/S2052252520002973

Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator: Figure 1. *Genome Res.* 14 (6), 1188–1190. doi:10.1101/gr.849004

Hartwell, L. H., Culotti, J., and Reid, B. (1970). Genetic Control of the Cell-Division Cycle in Yeast, I. Detection of Mutants. *Proc. Natl. Acad. Sci.* 66 (2), 352–359. doi:10.1073/pnas.66.2.352

Hartwell, L. H., Mortimer, R. K., Culotti, J., and Culotti, M. (1973). Genetic Control of the Cell Division Cycle in Yeast: V. Genetic Analysis of Cdc Mutants. *Genetics* 74 (2), 267–286. V. Genetic Analysis of cdc Mutants. doi:10.1093/genetics/74.2.267

Hernández-Rodríguez, Y., Masuo, S., Johnson, D., Orlando, R., Smith, A., Couto-Rodriguez, M., et al. (2014). Distinct Septin Heteropolymers Co-exist during Multicellular Development in the Filamentous Fungus *Aspergillus nidulans. PLoS One* 9 (3), e92819. doi:10.1371/journal.pone.0092819

John, C. M., Hite, R. K., Weirich, C. S., Fitzgerald, D. J., Jawhari, H., Faty, M., et al. (2007). The *Caenorhabditis elegans* Septin Complex Is Nonpolar. *EMBO J.* 26 (14), 3296–3307. doi:10.1038/sj.emboj.7601775

Kinoshita, M. (2003a). Assembly of Mammalian Septins. *J. Biochem*. 134 (4), 491–496. doi:10.1093/jb/mvg182

Kinoshita, M. (2003b). The Septins. *Genome Biol.* 4 (11), 236. doi:10.1186/gb-2003-4-11-236

Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002). Classification and Evolution of P-Loop GTPases and Related ATPases. *J. Mol. Biol.* 317 (1), 41–72. doi:10.1006/jmbi.2001.5378

Leonardo, D. A., Cavini, I. A., Sala, F. A., Mendonça, D. C., Rosa, H. V. D., Kumagai, P. S., et al. (2021). Orientational Ambiguity in Septin Coiled Coils and its Structural Basis. *J. Mol. Biol.* 433 (9), 166889. doi:10.1016/j.jmb.2021.166889

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Res.* 47 (W1), W636–W641. doi:10.1093/nar/gkz268

Martínez, C., Sanjuan, M. A., Dent, J. A., Karlsson, L., and Ware, J. (2004). Human Septin-Septin Interactions as a Prerequisite for Targeting Septin Complexes in the Cytosol. *Biochem. J.* 382 (Pt 3), 783–791. doi:10.1042/BJ20040372

McMurray, M. A., and Thorner, J. (2019). Turning it inside Out: The Organization of Human Septin Heterooligomers. *Cytoskeleton (Hoboken)* 76 (9-10), 449–456. doi:10.1002/cm.21571

Mendonça, D. C., Macedo, J. N., Guimarães, S. L., Barroso da Silva, F. L., Cassago, A., Garratt, R. C., et al. (2019). A Revised Order of Subunits in Mammalian Septin Complexes. *Cytoskeleton (Hoboken)* 76 (9-10), 457–466. doi:10.1002/cm.21569

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., et al. (2007). The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. *Science* 318 (5848), 245–250. doi:10.1126/science.1143609

Momany, M., and Talbot, N. J. (2017). Septins Focus Cellular Growth for Host Infection by Pathogenic Fungi. *Front. Cel Dev. Biol.* 5, 33. doi:10.3389/fcell.2017.00033

Momany, M., Zhao, J., Lindsey, R., and Westfall, P. J. (2001). Characterization of the *Aspergillus nidulans* Septin (Asp) Gene Family. *Genetics* 157 (3), 969–977. doi:10.1093/genetics/157.3.969

Mostowy, S., and Cossart, P. (2012). Septins: the Fourth Component of the Cytoskeleton. *Nat. Rev. Mol. Cel Biol* 13 (3), 183–194. doi:10.1038/nrm3284

Nguyen, T. Q., Sawa, H., Okano, H., and White, J. G. (2000). The *C. elegans* Septin Genes, Unc-59 and Unc-61, Are Required for normal Postembryonic Cytokineses and Morphogenesis but Have No Essential Function in Embryogenesis. *J. Cel Sci* 113, 3825–3837. doi:10.1242/jcs.113.21.3825

Nishihama, R., Onishi, M., and Pringle, J. R. (2011). New Insights into the Phylogenetic Distribution and Evolutionary Origins of the Septins. *Biol. Chem.* 392 (8-9), 681–687. doi:10.1515/BC.2011.086

Omrane, M., Camara, A. S., Taveneau, C., Benzoubir, N., Tubiana, T., Yu, J., et al. (2019). Septin 9 Has Two Polybasic Domains Critical to Septin Filament Assembly and Golgi Integrity. *iScience* 13, 138–153. doi:10.1016/j.isci.2019.02.015

Pan, F., Malmberg, R. L., and Momany, M. (2007). Analysis of Septins across Kingdoms Reveals Orthology and New Motifs. *BMC Evol. Biol.* 7, 103. doi:10.1186/1471-2148-7-103

Sirajuddin, M., Farkasovsky, M., Hauer, F., Kühlmann, D., Macara, I. G., Weyand, M., et al. (2007). Structural Insight into Filament Formation by Mammalian Septins. *Nature* 449 (7160), 311–315. doi:10.1038/nature06052

Smith, C., Dolat, L., Angelis, D., Forgacs, E., Spiliotis, E. T., and Galkin, V. E. (2015). Septin 9 Exhibits Polymorphic Binding to F-Actin and Inhibits Myosin and Cofilin Activity. *J. Mol. Biol.* 427 (20), 3273–3284. doi:10.1016/j.jmb.2015.07.026

Soroor, F., Kim, M. S., Palander, O., Balachandran, Y., Collins, R. F., Benlekbir, S., et al. (2021). Revised Subunit Order of Mammalian Septin Complexes Explains Their *In Vitro* Polymerization Properties. *MBoC* 32 (3), 289–300. doi:10.1091/mbc.e20-06-0398

Spiliotis, E. T., and McMurray, M. A. (2020). Masters of Asymmetry - Lessons and Perspectives from 50 Years of Septins. *MBoC* 31 (21), 2289–2297. doi:10.1091/mbc.e19-11-0648

Spiliotis, E. T., and Nakos, K. (2021). Cellular Functions of Actin- and Microtubule-Associated Septins. *Curr. Biol.* 31 (10), R651–R666. doi:10.1016/j.cub.2021.03.064

Valadares, N. F., d' Muniz Pereira, H., Ulian Araujo, A. P., and Garratt, R. C. (2017). Septin Structure and Filament Assembly. *Biophys. Rev.* 9 (5), 481–500. doi:10.1007/s12551-017-0320-4

Versele, M., Gullbrand, B., Shulewitz, M. J., Cid, V. J., Bahmanyar, S., Chen, R. E., et al. (2004). Protein-Protein Interactions Governing Septin Heteropentamer Assembly and Septin Filament Organization in *Saccharomyces Cerevisiae*. *MBoC* 15 (10), 4568–4583. doi:10.1091/mbc.e04-04-0330

Wloga, D., Strzyzewska-Jówko, I., Gaertig, J., and Jerka-Dziadosz, M. (2008). Septins Stabilize Mitochondria in *Tetrahymena thermophila*. *Eukaryot. Cel* 7 (8), 1373–1386. doi:10.1128/EC.00085-08

Woods, B. L., Cannon, K. S., Vogt, E. J. D., Crutchley, J. M., and Gladfelter, A. S. (2021). Interplay of Septin Amphipathic Helices in Sensing Membrane-Curvature and Filament Bundling. *MBoC* 32 (20), br5. doi:10.1091/mbc.e20-05-0303

Woods, B. L., and Gladfelter, A. S. (2021). The State of the Septin Cytoskeleton from Assembly to Function. *Curr. Opin. Cel Biol.* 68, 105–112. doi:10.1016/j.ceb.2020.10.007

Yamazaki, T., Owari, S., Ota, S., Sumiya, N., Yamamoto, M., Watanabe, K., et al. (2013). Localization and Evolution of Septins in Algae. *Plant J.* 74 (4), 605–614. doi:10.1111/tpj.12147

Zhang, J., Kong, C., Xie, H., McPherson, P. S., Grinstein, S., and Trimble, W. S. (1999). Phosphatidylinositol Polyphosphate Binding to the Mammalian Septin H5 Is Modulated by GTP. *Curr. Biol.* 9 (24), 1458–1467. doi:10.1016/s0960-9822(00)80115-3

CHAPTER 3

THE EVOLUTIONARY ORIGINS AND ANCESTRAL FEATURES OF SEPTINS[5]

[5] Delic S., Shuman, B., Lee, S., Bahmanyar, S., Momany, M., and Onishi, M. Submitted to *Front. Cell Dev. Biol. - Cell Growth and Division*, 05/01/24

Abstract

Septins are a family of membrane-associated cytoskeletal guanine-nucleotide binding proteins that play crucial roles in various cellular processes, such as cell division, phagocytosis, and organelle fission. Despite their importance, the evolutionary origins and ancestral function of septins remain unclear. In opisthokonts, septins form five distinct groups of orthologs, with subunits from multiple groups assembling into heteropolymers, thus supporting their diverse molecular functions. Recent studies have revealed that septins are also conserved in algae and protists, indicating an ancient origin from the last eukaryotic common ancestor. However, the phylogenetic relationships among septins across eukaryotes remained unclear. Here, we expanded the list of non-opisthokont septins, including previously unrecognized septins from rhodophyte red algae and glaucophyte algae. Constructing a rooted phylogenetic tree of 254 total septins, we observed a bifurcation between the major non-opisthokont and opisthokont septin clades. Within the non-opisthokont septins, we identified three major subclades: Group 6 representing chlorophyte green algae (6A mostly for species with single septins, 6B for species with multiple septins), Group 7 representing algae in chlorophytes, heterokonts, haptophytes, chrysophytes, and rhodophytes, and Group 8 representing ciliates. Glaucophyte and some ciliate septins formed orphan lineages in-between all other septins and the outgroup. Combining ancestral-sequence reconstruction and AlphaFold predictions, we tracked the structural evolution of septins across eukaryotes. In the GTPase domain, we identified a conserved GAP-like arginine finger within the G-interface of at least one septin in most algal and ciliate species. This residue is required for homodimerization of the single Chlamydomonas septin, and its loss coincided with septin duplication events in various lineages. The loss of the arginine finger is often accompanied by the emergence of the α0 helix, a known NC-interface interaction motif,

potentially signifying the diversification of septin-septin interaction mechanisms from homo-dimerization to hetero-oligomerization. Lastly, we found amphipathic helices in all septin groups, suggesting that membrane binding is an ancestral trait. Coiled-coil domains were also broadly distributed, while transmembrane domains were found in some septins in Group 6A and 7. In summary, this study advances our understanding of septin distribution and phylogenetic groupings, shedding light on their ancestral features, potential function, and early evolution.

Introduction

Septins are a family of paralogous cytoskeletal guanine-nucleotide binding proteins (with some possible exceptions: Hussain et al., 2023) that associate with one another in defined stoichiometries in a defined order to create nonpolar filaments. The first four septin genes (*CDC3*, *CDC10*, *CDC11*, and *CDC12*) were identified in a cell-cycle defective screen in *Saccharomyces cerevisiae* (Hartwell, 1971; Hartwell et al., 1974). Detailed molecular characterization of these septins showed that each gene encodes a distinct septin subunit that associates with other septin subunits in a defined order to create filaments and other higher-order structures such as rings on the plasma membrane (Byers and Goetsch, 1976; Field et al., 1996; Longtine et al., 1996; McMurray and Thorner, 2008). It was later shown that septin assembly and filamentation are influenced by lipid composition of membranes (Bertin et al., 2010).

A septin subunit is comprised of a core GTPase domain and variable N- and C-terminal extensions (NTE and CTE). The GTPase domain is responsible for binding and/or hydrolyzing GTP depending on the subunit, as well as mediating septin-septin interactions and polymerization (Sirajuddin et al., 2007; Hussain et al., 2023). The N-terminal domain of septins often contains a polybasic domain (PB1) directly upstream of the start of the GTPase domain, which plays critical roles in lipid recognition and septin polymerization (Zhang et al., 1999; Omrane et al., 2019; Cavini et al., 2021). Depending on the septin subunit, the C-terminal domain can contain a coiled-coil domain which has been proposed to mediate lateral pairing of septin filaments (Leonardo et al., 2021). Additionally, some subunits also possess an amphipathic helix (AH) which has been shown to allow septins to bind to membranes and recognize micron-scale curvature (Bridges et al., 2016; Cannon et al., 2019). The structure of septin protomers has been described using the human SEPT2/6/7 heterohexameric complex,

which unequivocally identified two binding interfaces for septin subunits (Sirajuddin et al., 2007): The G-interface is defined as the face of the subunit with the GTP-binding pocket, where *trans* interactions with an opposing subunit stimulates GTP hydrolysis, whereas the NC-interface is the opposite face of the subunit. Both interfaces can be involved in homomeric and heteromeric dimerization events.

Previous phylogenetic analyses of opisthokont septins identified conserved residues within the G- and NC-interfaces that drive subunit assembly into heteropolymers (Pan et al., 2007; Auxier et al., 2019; Shuman and Momany, 2021). Additionally, these analyses provided an evolutionary basis for the modularity of septin paralogs in support of Kinoshita's rule, which states that septins belonging to the same phylogenetic group can replace one another within the canonical protomer, maintaining the same defined order of subunits (Kinoshita, 2003b; Pan et al., 2007) For example, human SEPT3, 9, and 12 all belong to Group 1A and can replace one another as the central dimer within a hetero-octamer. Thus, these phylogenetic analyses can provide structural and biochemical insights into the assembly of septins.

Most of the cellular, biochemical, and phylogenetic characterizations of septin proteins have been from the opisthokont (animal & fungal) lineage. The presence of septins outside of opisthokonts was initially noted by Versele & Thorner, who mentioned the presence of bona fide septins in *Chlamydomonas reinhardtii* & *Nannochloris spp.* (Versele and Thorner, 2005). Subsequent studies in the green algae *Nannochloris bacillaris* and *Marvania geminata* and the ciliate *Tetrahymena thermophilus* characterized the localization of septins outside of the opisthokont paradigm. In the former, immunofluorescence studies using an antibody against the single septin in *N. bacillaris* showed its localization at the division site of both algae (Yamazaki et al., 2013). In the latter, septins were reported to localize to the mitochondria scission sites and

proposed to regulate mitochondrial stability via autophagy pathways (Wloga et al., 2008). Additional septins have since been identified in some other algae and protists (Nishihama et al., 2011; Yamazaki et al., 2013; Onishi and Pringle, 2016; Brawley et al., 2017; Goodson et al., 2021) however, the phylogenetic relationship and implications for subunit assembly of these non-opisthokont septins remained unclear.

In this work, we provide an update to the distribution of septins across the eukaryotic tree of life and a rigorous phylogenetic analysis to compare their relationship to previously identified septin groups. We trace the evolution of structural motifs within the septin GTPase domains by combining ancestral sequence reconstruction and machine-learning 3D structural prediction. Lastly, we trace the gains and losses of septin-associated features in the NTE and CTE, such as the polybasic domain, coiled-coil, AH, and putative transmembrane domains to assess their evolutionary origins.

Materials and Methods

Identification of New Septin Sequences

To identify new non-opisthokont septin sequences, we utilized both the Joint Genome Institute Phycocosm webpage (https://phycocosm.jgi.doe.gov/) and the NCBI Genome database (https://blast.ncbi.nlm.nih.gov/). We used the initial set of queries consisting of *Chlamydomonas*, *Symbiodinium*, and *Paramecium* septins. These searches identified several septins in the phyla in which they have not been reported. To enhance the chance of finding new sequences in these and other divergent branches, we added *Porphyra*, *Ectocarpus*, and *Cyanophora* to the list of queries and performed additional searches (Table 3.1; Supplementary File 1). BLASTP searches were performed on November 14, 2021 using a BLOSUM62 matrix, E-value cutoff of $1x10^{-5}$, word size of 3, and filtered low complexity regions. The JGI database searches used proteomes from

Excavata, Archeaplastida, Rhizaria, Heterokonta, and Alveolata (Supplementary File 2). Due to the limited availability of information for ciliate species on JGI, additional searches were performed using the NCBI database, specifically focusing on Alveolata (taxid:33630) (Supplementary File 2). Identified sequences were further examined manually for the presence of G-motifs (G1, G3, and G4) and S-motifs (S1-S4) to confirm that they are bona fide septins. Opisthokont septins were selected from (Auxier et al., 2019).

Phylogenetic Analysis and Ancestral Sequence Reconstruction

Phylogenetic trees were constructed following the methodology described by (Auxier et al., 2019). A total of 131 opisthokont and 123 non-opisthokont septins were used; as an outgroup, several prokaryotic YihA proteins were also included (Supplementary File 3). Sequences were first aligned using the constraint-based alignment tool (COBALT) (Papadopoulos and Agarwala, 2007), which incorporates information about protein domains in a progressive multiple alignment. This tool biases the alignment within the septin GTPase domain. To remove regions of randomly similar sequences from the alignment, we employed ALISCORE and ALICUT (Misof and Misof, 2009; Kück et al., 2010; Kueck, 2017). ALISCORE identifies regions of ambiguous alignment, which were subsequently removed using ALICUT. This process resulted in a reduced MSA file containing highly conserved regions within the GTPase domain (Supplementary File 4), which was then used to generate the phylogenetic tree.

Tree generation was performed using the CIPRES gateway (Miller et al., 2010), employing RAxML-HPC v.8 on XSEDE with the PROTCAT substitution model and the LG protein matrix and a rapid 1000 bootstrap analysis. The generated trees were visualized using the Rstudio package "ggtree." Bootstrap values displayed on the trees have been limited to values greater than 25.

For ancestral sequence reconstruction (ASR), we utilized the FASTML server for maximum-likelihood computing of the ancestral states (Ashkenazy et al., 2012). Due to limitations with the FASTML server, we reduced our list of septin sequences from 254 to 200 by removing some sequences from some fungal species and all sequences from the genus *Paramecium* except for the species *tetraurelia*. The resulting 200 sequences (Supplementary File 5) were aligned using COBALT alignment. As ASR provides meaningful interpretation when the entire protein sequence is provided, we did not utilize ALISCORE and ALICUT processing. To generate a new phylogenetic tree, we used the IQTree webserver (http://iqtree.cibiv.univie.ac.at/) with an automatic amino acid replacement matrix, 1000 ultrafast bootstraps, and all other default parameters (Trifinopoulos et al., 2016; Minh et al., 2020). This tree reproduced the same phylogenetic groupings and general branching patterns as our more rigorous ALISCORE and ALICUT processed tree. Nodes of interest, including parental nodes for the septin phylogenetic groups, opisthokont and protist divide, and the last eukaryotic common ancestor (LECA) node, were defined based on the joint reconstruction output file and labeled in Supplementary File 5. The protein sequences at these nodes were extracted and referred to as the ancestral septins.

AlphaFold Predictions and Search for Polybasic Domains in N-terminal Extension

AlphaFold predictions were executed using the Colabfold Google notebook v1.3.0. The specific parameters can be found within the "config.json" file in each respective folder. Due to computational limitations of AlphaFold with extremely long sequences, some sequences required trimming. The objective of trimming was to preserve the entire GTPase domain and the CTE while reducing the sequence length to a manageable size (approximately 800 amino acids). Generally, the protein sequence was truncated from the N-terminal end. Predictions primarily used an MMseqs2 MSA. Five models with three recycles each were generated and the highest-

ranking model was selected (Supplementary File 6). The resulting 3D structures were visualized using ChimeraX. Topology diagrams were drawn in Adobe Illustrator, following the convention used in (Cavini et al., 2021). For AlphaFold predictions of *K. flaccidum* and *I. multifiliis* septins, we used version 1.5.2 of the ColabFold notebook. The structures were visualized using ChimeraX and colored according to AlphaFold confidence.

To search for potential polybasic domains in the NTE of our reconstructed ancestral sequences, we developed a Python script that uses a sliding 10-amino-acid window to calculate the local average isoelectric point and plots this value against the first amino acid position across the entire protein length. To focus solely on the NTE, which is where PB1 in extant septins is primarily located, we aligned the ancestral septins to the GTPase domain of *S. cerevisiae* Cdc3 using CLUSTALω. Only residues before the start of the GTPase domain were plotted. To visualize the multiple sequence alignment (MSA) of the ancestral septins, a CLUSTALω alignment was performed without the Cdc3 GTPase domain to compare the amino acid composition between GTPase domain-adjacent polybasic domains. The MSA was visualized using the R package "ggmsa," and the amino acids were colored according to their properties.

Identification of amphipathic helices in extant septin sequences

For high-throughput prediction of amphipathic helices, we developed a Python script that consists of two steps of analysis: (1) secondary structure prediction by s4pred (Moffat and Jones, 2021) followed by (2) amphipathicity assessment of α-helices. In (1), secondary structure prediction was performed for the amino acid sequence of a given septin protein using the run_model.py script provided in https://github.com/psipred/s4pred. In (2), either a "fully-helical" or "partially-helical" segment of an amino-acid sequence was extracted by a sliding 18 amino-acid window. In a "partially-helical" segment, at least 6 amino acids at both ends of the 18 amino

acid window must be fully helical. For example, while a segment with a prediction "HHHHHHCCCCCCHHHHHH" (6x H – 6x C – 6x H) was permitted, those with "HHHHHCCCCCCHHHHHHH" (5x H – 6x C – 7x H) were not. We included "partially-helical" segments for further assessment because some membrane-bound Ahs could be predicted as "partially helical," where two helices are broken apart by non-helical sequence (e.g., Sun2 AH: Lee et al., 2023). For each helical segment, the amphipathicity was calculated and assessed similarly to HeliQuest software (Gautier et al., 2008), but with modifications. First, the mean hydrophobic moment value $<\mu H>$ was calculated as previously described (Eisenberg et al., 1982) using the hydrophobicity scale values (Fauchere and Pliska, 1983) based on an assumption that all helices rotate with a 100 degree step. Then, the discriminant factor $D = 0.944 \times <\mu H> + 0.33 \times z$ (where z is the net charge) was calculated accordingly to HeliQuest. Finally, the helical segment was considered amphipathic if all of the criteria below were satisfied: i) $D > 0.68$ OR ($<\mu H> > 0.4$ AND $z = 0$); ii) The hydrophobic face contains at least 3 consecutive bulky hydrophobic residues (L, V, F, I, W, M, Y) (e.g. a hydrophobic face "SYALLVT" is satisfactory); iii) "Core" of the hydrophobic face does NOT contain any charged residue ("core": the area of 90° centered around the pole). This search resulted in the identification of 4809 possible AH domains, with the vast majority showing overlap with one another (Supplementary File 7).

We then filtered the data to exclude AHs that are positioned inside of an septin GTPase domain. The GTPase domain of Cdc3 from *S. cerevisiae* was used as a reference to define the start and end residues for the GTPase domain of the other 254 extant sequences. The list of possible AHs of 18 amino acids in length was then screened by excluding those that overlapped with the GTPase domain. Sequences satisfying these criteria were considered to possess an AH

(Supplementary File 8) and were highlighted in a cladogram generated using the R package "ggtree." To generate helical wheel diagrams, individual AH sequences from the dataset were used as input to run the HeliQuest program (Gautier et al., 2008).

Search for coiled-coil and putative transmembrane domains in extant septin sequences

To identify septins with coiled-coil domain and/or putative transmembrane domains in the set of 254 extant septins, we used the existing annotations on the UniProt database (The UniProt Consortium, 2023) release 2023_04. A BLASTP search using our list of 254 septins as query against the UniprotKB database retrieved 206 hits, for which “Coiled coil” and “Transmembrane” annotations were downloaded from the database. According to the UniProt documentation, these annotations are based on the COILS program (Lupas et al., 1991) with a minimum size of 28 amino acids for coiled-coil domains, and TMHMM and Phobius predictions (Krogh et al., 2001; Käll et al., 2004) for transmembrane domains. For the remaining 48 sequences, manual searches for coiled-coil and transmembrane domains were performed using Cocopred (Feng et al., 2022) and Phobius. These predictions are conservative and unlikely to identify all possible coiled-coil and transmembrane domains; for example, the present analysis identified fewer coiled-coil-containing septins than Auxier et al. (2019), which used the hidden-Markov-model-based Marcoil program. Results of these searches are summarized in Supplementary File 8.

Results

Identification of new septin sequences

To search for septin sequences outside of opisthokonts, we compiled a small query list of previously identified septin sequences from algal and protist species (Table 3.1). These sequences were selected based on their evolutionary diversity, aiming to enhance the chance of

identifying septins from various taxa. We conducted BLASTP searches using the BLOSUM62 matrix and an E-value cutoff of $1x10^{-5}$, utilizing the protein databases available on the Joint Genome Institute's (JGI) Phycocosm webpage and the Alveolata database on the NCBI BLAST website (see Materials and Methods). These searches revealed previously unreported sequences in multiple taxa under the supergroups Archaeplastida and Chromista (Figure 3.1). Our searches also reproduced a previous failure to identify any septin sequences in the entire supergroups of Amoebozoa and Excavata (Fig. 1; Onishi & Pringle, 2016). At lower phylogenetic levels, septins were also not detected in Viridiplantae (land plants) (Fig. 3.1).

Table 3.1. Query sequences used in BLASTP searches

Phylum	Species	Identifier
Chlorophyta (green algae)	*Chlamydomonas reinhardtii*	Cre12.g556250
Glaucophyta	*Cyanophora paradoxa*	13652g13185t1
Rhodophyta (red algae)	*Porphyra umbilicalis*	6951
Phaeophyceae (brown algae)	*Ectocarpus siliculosus*	CBN74010
Ciliophora (ciliates)	*Paramecium tetraurelia*	CAI38984
Dinoflagellates	*Symbiodinium minutum*	symbB1.v1.2.007989.t1 [a]

[a] This transcript encodes a very long 4484-aa predicted protein. See Onishi & Pringle (2016) for details. The 560-aa amino-terminal sequence containing the septin GTPase domain was used as query.

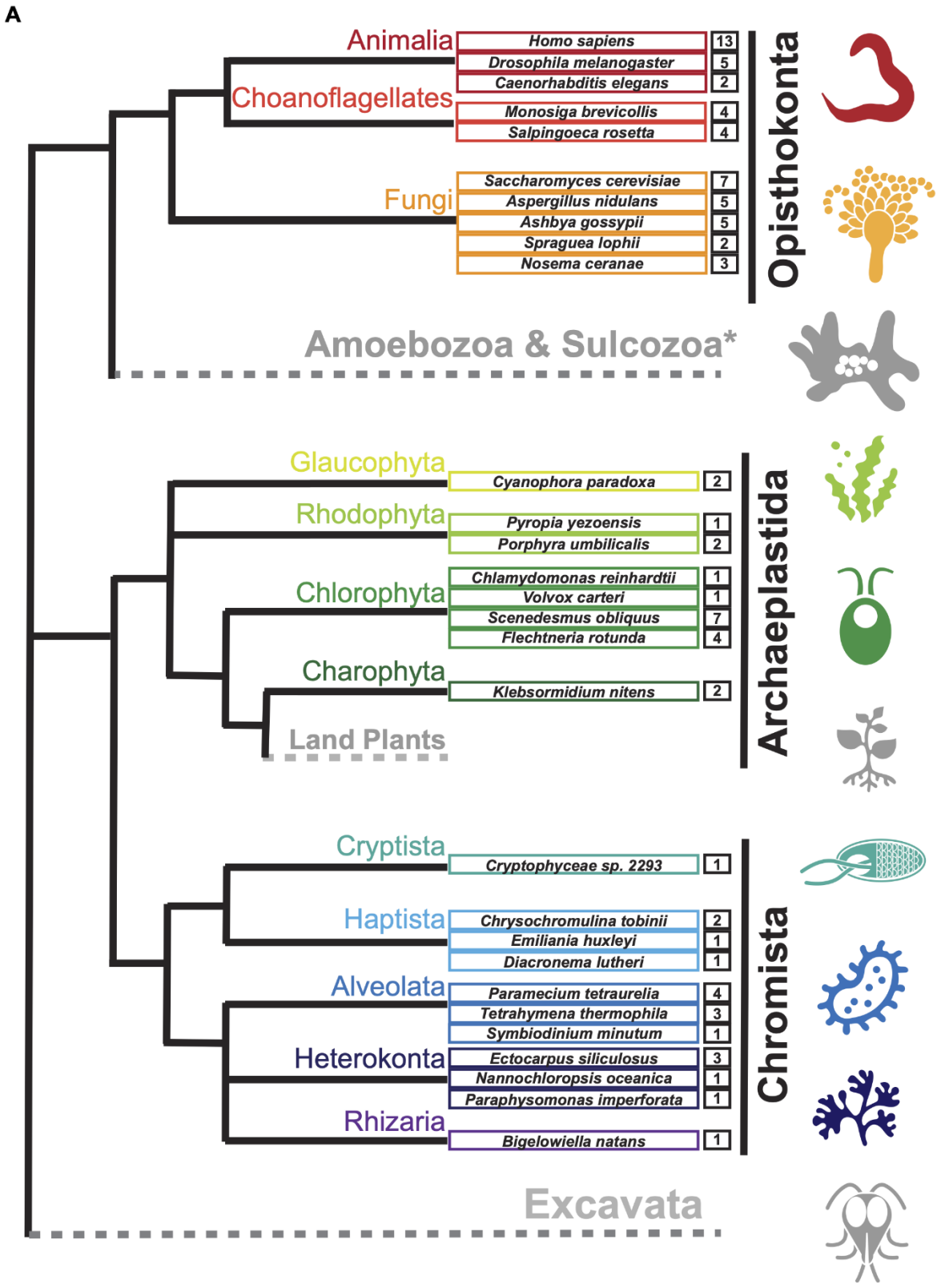

Figure 3.1. Distribution of septins in non-opisthokont phyla. (**A**) Unrooted taxonomic tree of eukaryotes (based on (Cavalier-Smith, 2018). Gray and dotted branches indicate lineages in

which no septin sequence was identified, while black and colored branches represent lineages with identified septins. Representative species are shown and color-matched to their respective lineages, and the total numbers of septin paralogs identified in their genomes are indicated. *Possible septins were identified in *Planoprotostelium fungivorum*; because this is the only example of species with septins within Amoboezoa and Sulcozoa, we could not determine whether they are a result of unique gene retention, horizontal gene transfer, or contamination.

New septin phylogenetic groups

The discovery of new septin sequences in distant branches of eukaryotes raised questions about their phylogenetic relationship with other septins. Previous studies have classified septins into five groups, but these groupings were defined predominantly based on septin sequences within the opisthokont lineage. We thus combined these new non-opisthokont septin sequences with a preexisting list of opisthokont septins (Auxier et al., 2019) and used the resulting 254 sequences to generate a consensus RAxML tree (Fig. 3.2) and a simplified cladogram (Fig. 3.3A). Briefly, the 254 sequences and four prokaryotic YihA NTPases (used here as an outgroup; (Weirich et al., 2008)) were aligned using NCBI's COBALT alignment tool and processed using ALISCORE and ALICUT to remove ambiguous regions of alignment.

Consistent with results from previous reports (Momany et al., 2001; Kinoshita, 2003a; Pan et al., 2007; Shuman and Momany, 2021), our phylogenetic analysis grouped the opisthokont septins into five distinct clades (Fig. 3.3A; Fig. 3.2): Groups 1 and 2 include septins from both animals and fungi, while Groups 3, 4, and 5 represent fungi-specific clades. Although limited sampling of non-opisthokont septins has previously placed some of them in Group 5 (Onishi and Pringle, 2016; Shuman and Momany, 2021), it is now clear that Group 5 septins are distinct from non-opisthokont septins, consistent with the proposal by Yamazaki et al. (2013).

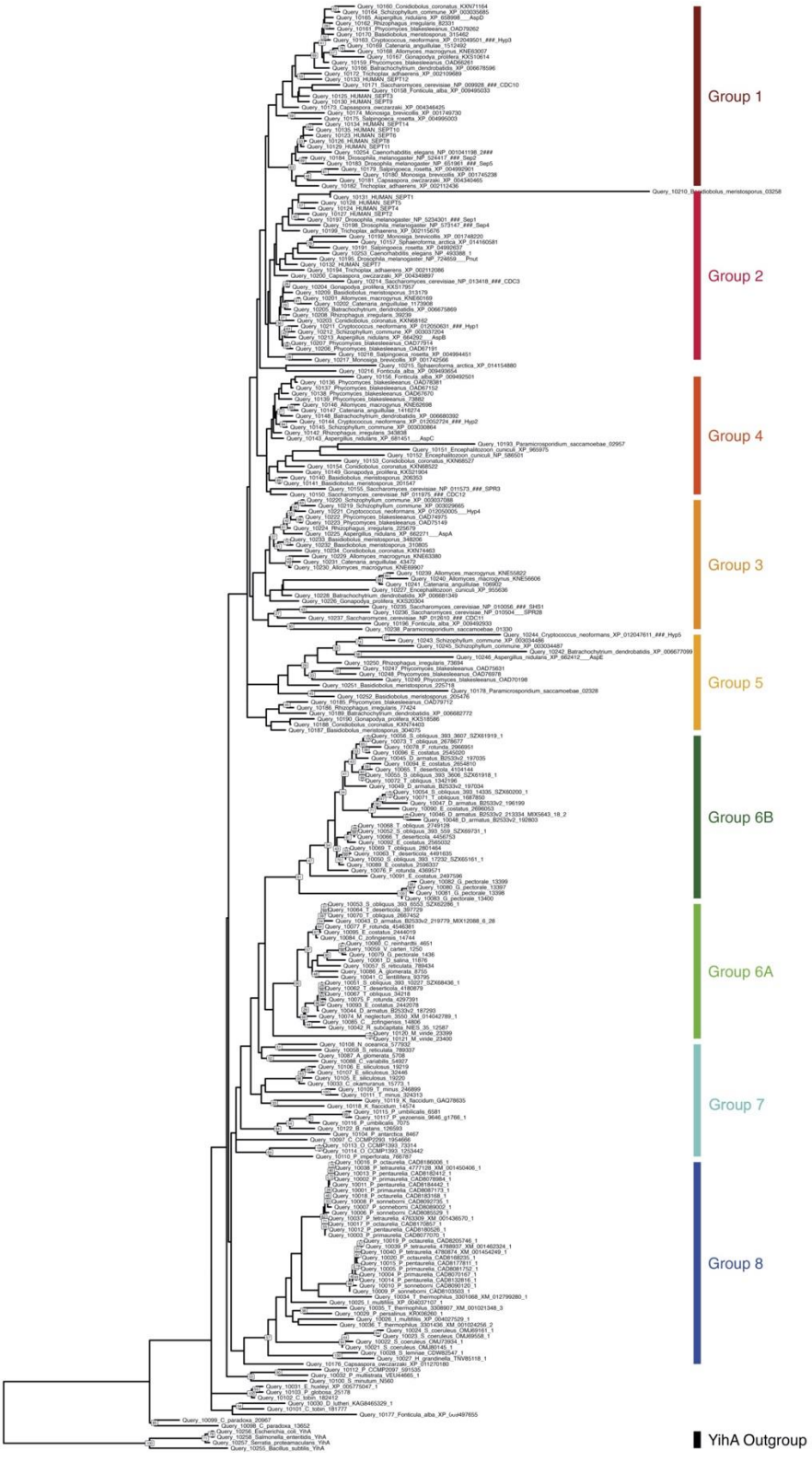

Figure 3.2. RAxML tree of all eukaryotic septins with 1000 bootstraps and YihA family as outgroup. Bootstrap values <25 are not shown. Defined phylogenetic groups are colored and displayed adjacent to tree tips.

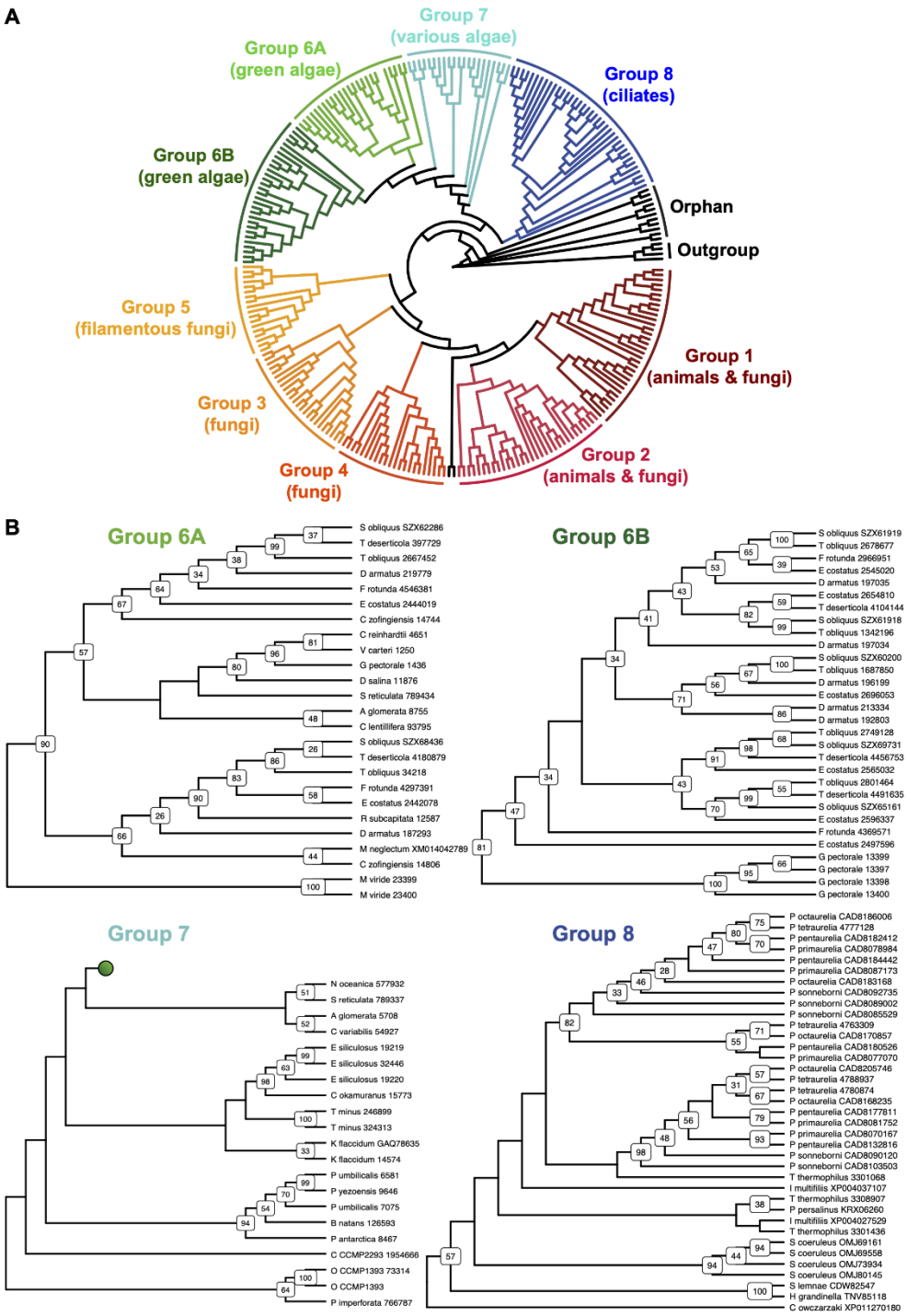

Figure 3.3. Identification of new septin groups in non-opisthokonts. (**A**) A simplified cladogram representation of a RAxML tree (Fig. 3.2) of 254 extant septin sequences across eukaryotic lineages. Individual septin phylogenetic clades are color-coded and labeled. The tree is rooted using four prokaryotic YihA proteins as an outgroup. (**B**) Magnified views of the four new phylogenetic clades. See Fig. 3.4 for the original RAxML trees. Bootstrap values greater than 25 are displayed at nodes.

The non-opisthokont septins themselves form three new groups (Groups 6-8) (Fig. 3.3B; Fig. 3.4). Group 6 is a monophyletic group of green algal species divided into two subgroups: Group 6A includes some septins that are encoded as a single gene in the genome, in species such as *C. reinhardtii* and *N. bacillaris* (Versele and Thorner, 2005; Yamazaki et al., 2013). Group 6B, in contrast, exclusively represents septins that appear to have emerged through gene duplication. For example, of five septins in the green alga *Gonium pectorale,* only one belongs to Group 6A while the remaining four belong to Group 6B (Fig. 3.3B; Fig. 3.4). The genes for these four septins form a cluster in the assembled *G. pectorale* genome. (Scaffold_65:140824 - 165695), suggesting a very recent gene duplication event. Similarly, of the seven septins in *Desmodesmus armatus,* five belong to Group 6B (Fig. 3.3B; Fig. 3.4). Group 7 is a paraphyletic group composed of septins from various groups of algae, such as additional green algae (e.g., *Symbiochloris reticulata*), heterokonts (*Ectocarpus siliculosus*), haptophytes (*Chrysochromulina tobinii*), cryptophytes (*Crytophyceae sp. CCMP2293*), chlorarachniophytes (*Bigelowiella natans*), and rhodophytes (*P. umbilicalis*) (Fig. 3.3B; Fig. 3.4). Finally, Group 8 is a monophyletic group comprised exclusively of septins from ciliates, except for one highly divergent sequence from the unicellular opisthokont *Capsaspora owczarzaki*. Within Group 8, septins from *Paramecium* and *Stentor coeruleus* formed genus-specific clades, suggesting recent expansion events of septin genes within their lineages (Fig. 3.3B; Fig. 3.4).

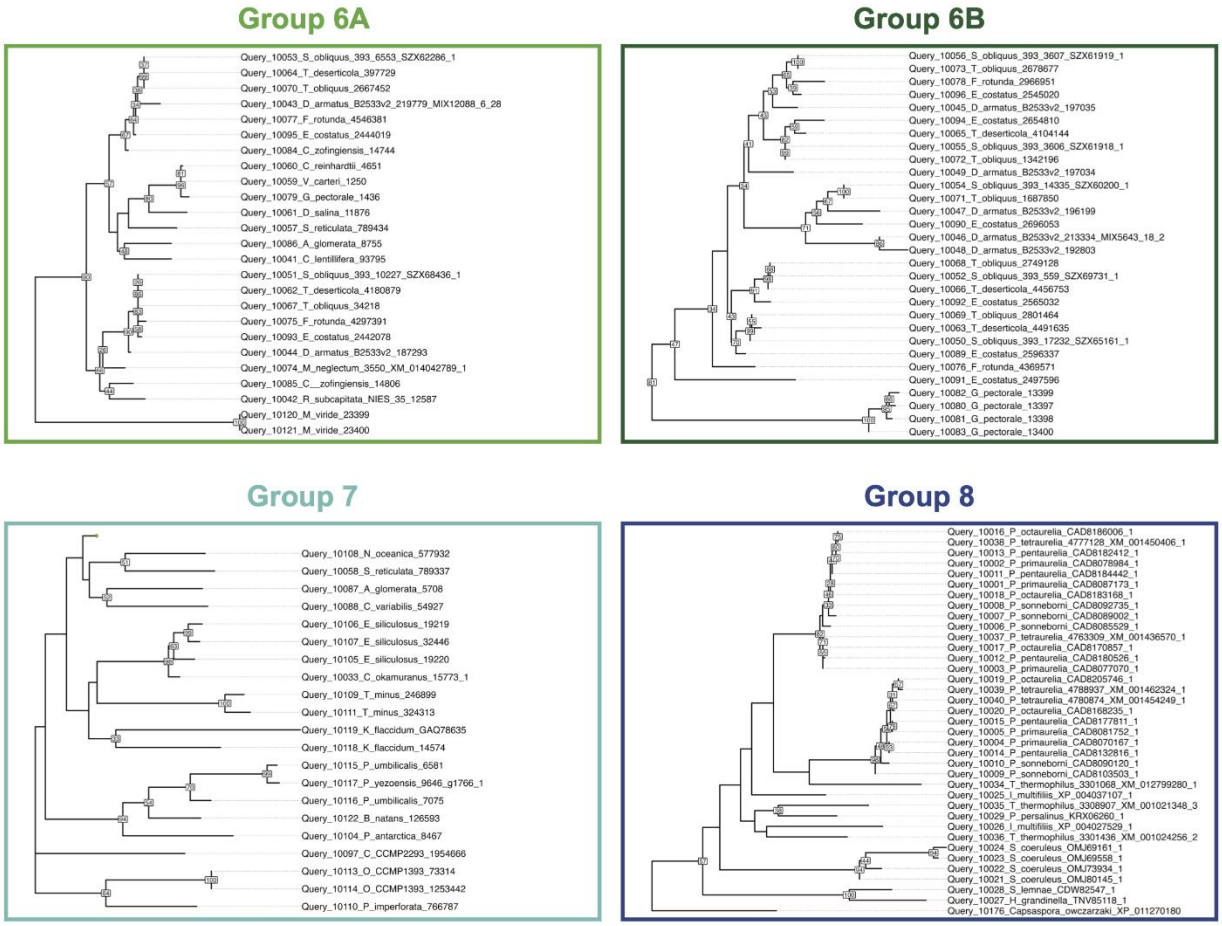

Figure 3.4. Magnified views of septin groups 6-8. The green dot in Group 7 indicates the node for Group 6.

Several non-opisthokont sequences are currently not classified in Groups 6-8 because their phylogenetic positioning was sensitive to the programs and parameters used (Fig. 3.3A; Fig. 3.2). These include sequences from glaucophytes (*C. paradoxa*), dinoflagellates (*S. minutum*, *Pseudonitzschia multistrata*), and coccolithophores and related haptophytes (*Emiliania huxleyi*, *Phaeocystis globosa*, *C. tobinii*, *Diacronema lutheri*). Curiously, a septin from *Fonticula alba*, an opisthokont cellular slime mold, also belonged to this orphan group. Additional sampling of sequences from these and related species will likely help improve the confidence in their phylogenetic positioning.

Conservation of G-interface residues in non-opisthokont septins

In previous studies, septins from Groups 1-5 were found to have several highly conserved regions in their GTPase domains (Fig. 3.5A) that participate in inter-subunit contacts across the G- and NC-interfaces (Fig. 3.5BC; Pan et al., 2007; (Pan et al., 2007; Auxier et al., 2019; Rosa et al., 2020; Shuman and Momany, 2021). To gain insights into the evolution of these interfaces in septins across the eukaryotic tree, we expanded the alignment to all 254 septins and generated a Weblogo representation for each septin group (Fig. 3.5D). In general, the GTPase-specific motifs (G1, G3, G4), septin-specific motifs (S2, S3, S4) except for the S1 motif (Pan et al., 2007; Auxier et al., 2019; Nishihama et al., 2011; Onishi & Pringle, 2016), and some key residues in the septin-unique element are all well conserved. More specifically, most of the key residues in the five G-interfaces (Gig1-Gig5) are all conserved, except for Gig2 which appears to be variable in Group 8 (Fig. 3.5D). In contrast, key residues in the four NC-interfaces (NCig1-4) are poorly conserved in Groups 6b, 7, and 8. These results suggest that non-opisthokont septins may primarily form homo- or hetero-dimers through the G-interface, and further addition of subunits through NC-interfaces may be limited to Group 6a. In support of this speculation, we found a unique arginine residue that is highly conserved in many Group 6-8 septins but not in Groups 1-5 (Fig. 3.5D); similar "arginine (R-) fingers" are found in other GTPases that form G-dimers (Koenig et al., 2008; Schwefel et al., 2013; see below).

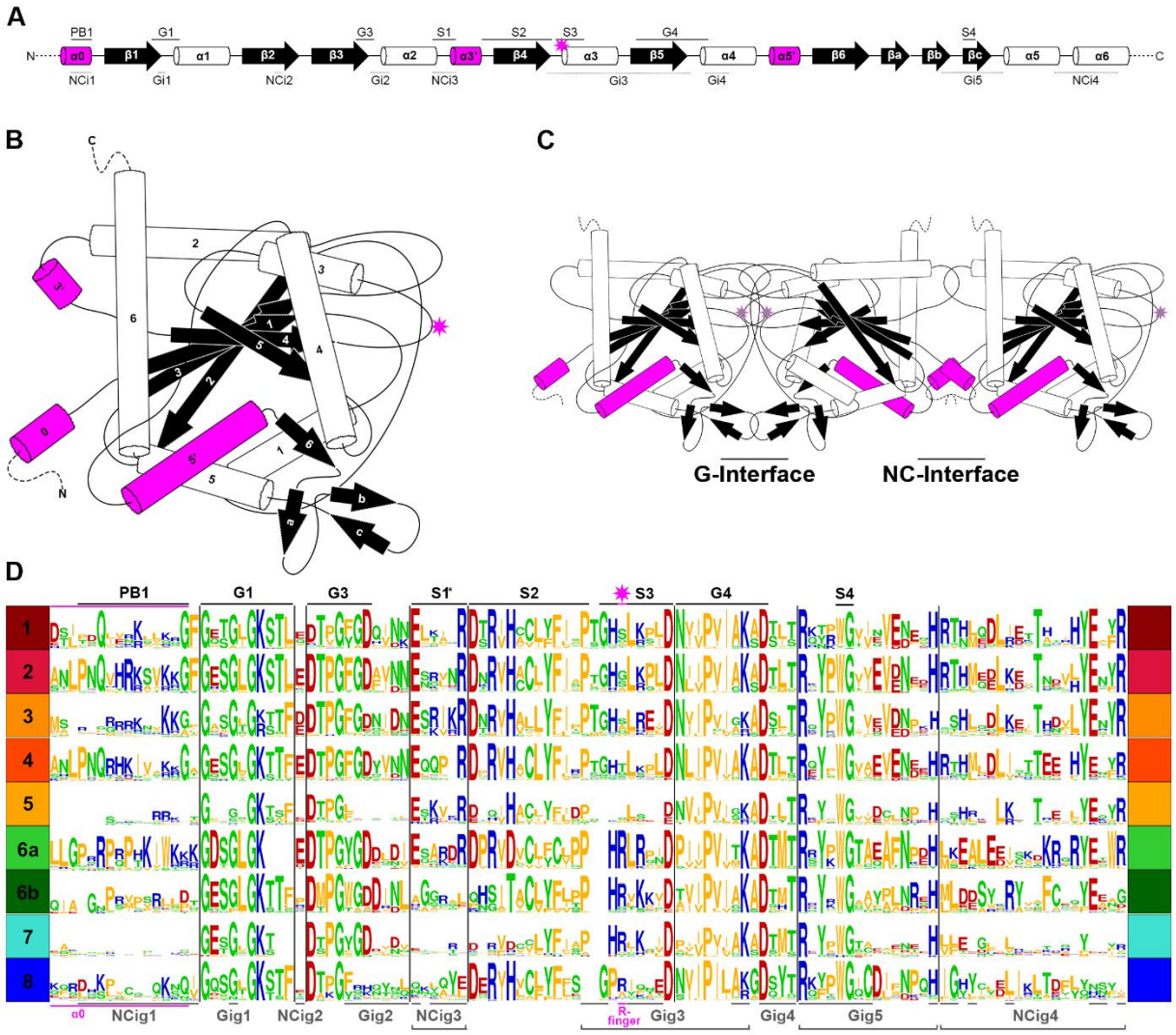

Figure 3.5. Patterns of conservation and diversity of interface motifs across septin Groups. (**A**) Topology diagram of the GTPase domain secondary structures from N to C-terminus. Conserved GTPase motifs and septin motifs are noted above by black lines (based on Grupp and Gronemeyer, 2023) and the NC and G-interacting group regions are noted below by dashed lines (based on Auxier et al., 2019). The typical position of the R-finger (when present) is indicated by the pink star. (**B**) A folded septin monomer. This aggregate depiction includes all predicted domains across eukaryote septins. Relative positions of secondary structures are based on PDB structures 7M6J and 8FWP (Mendonça et al., 2021; Grupp and Gronemeyer, 2023; Marques da Silva et al., 2023). (**C**) A septin trimer approximating interactions through their G- and NC-interfaces, based on PDB structure 7M6J. Grey stars with pink outline indicate the predicted positions of R-fingers if they are present in the subunits forming an interface. (**D**) Weblogo representation of select septin motifs, interacting groups, and structural elements across the eukaryotic septin groups. GTPase motifs and septin motifs are depicted above in black, and NC and G-interacting group regions are depicted below in grey. *Note, the location of S1 in groups 6a-8 was determined by relative position in the alignment to the beginning of S2. This loop region which resides between α2 and β4 has considerable sequence length variability and also includes a region where the α3' helix is predicted.

Reconstituted ancestral septins suggest that the arginine finger in the G-interface is an ancestral feature

To delve deeper into the evolution of the structural motifs within the septin GTPase domain, we used ancestral sequence reconstruction (ASR) (Ashkenazy et al., 2012) to resurrect ancestral septins. Due to the limitations of the program used, we reconstructed an IQTree of 200 of the 254 septins (Fig. 3.6A; Fig. 3.7). The grouping of septin clades and the overall topology of the tree were largely consistent with the RAxML tree (Fig. 3.3). Using this IQTree, ASR prediction was made for several key nodes representing Groups 1-8 and their parental nodes, and then AlphaFold2 (Jumper et al., 2021) was used to predict their 3D structures for the GTPase domain and the C-terminal extension (see Materials and Methods). Perhaps unsurprisingly given the conservation of the extant sequences (Fig. 3.5D), the tertiary structures of the ancestral sequences all appeared similar among themselves and with experimentally determined septin structures (Fig. 3.8).

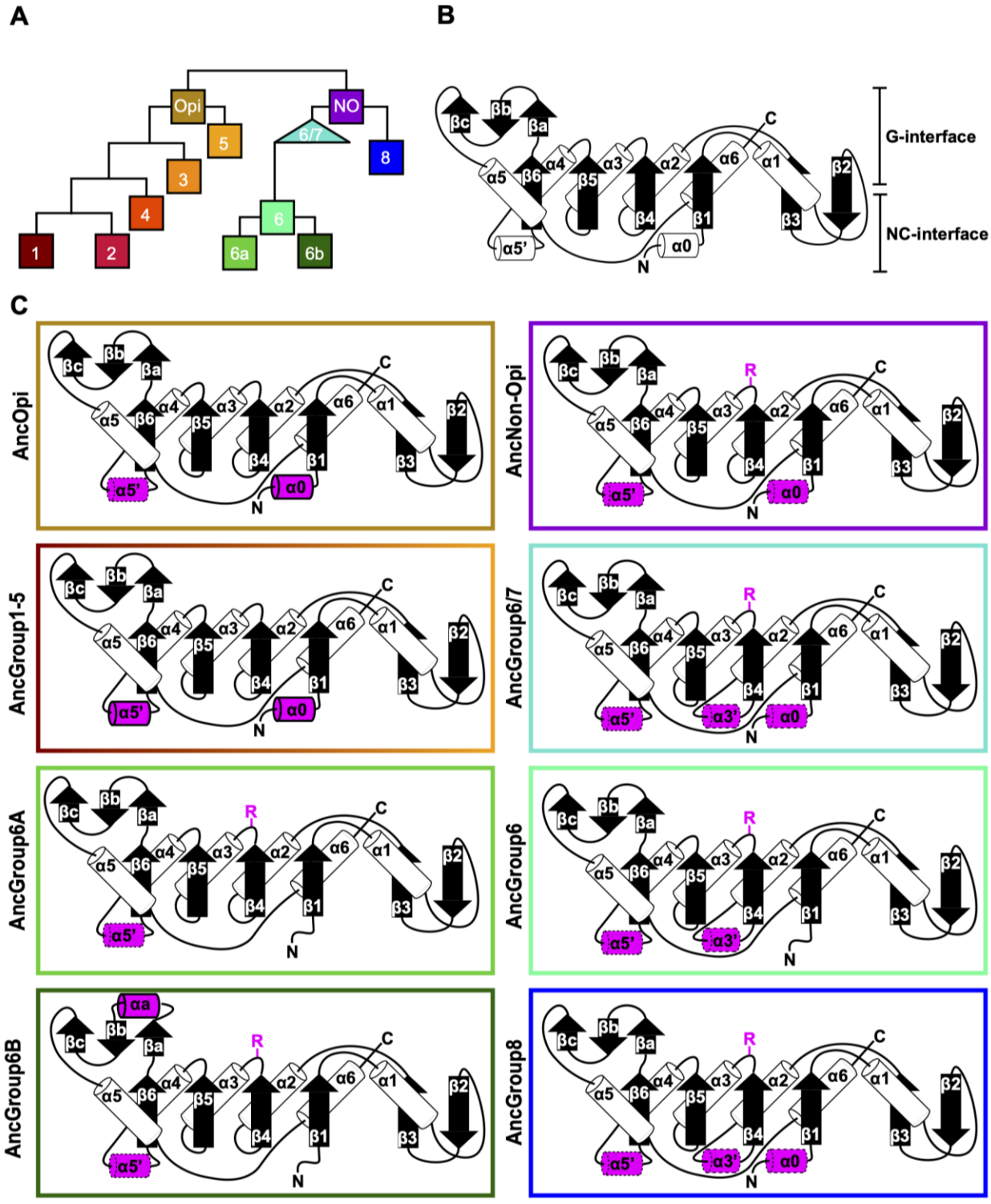

Figure 3.6. Ancestral sequence reconstruction of key evolutionary nodes throughout septin evolution. (**A**) Simplified tree diagram displaying the shape of the IQTree (Fig. 3.7) used in ancestral sequence reconstruction. Squares and triangle, key nodes with ancestral septins corresponding to interpretive diagrams shown in panel C. (**B**) Representative topology diagram of septin GTPase domain indicating both the G-interface and NC-interface. N and C represent

the N-terminal and C-terminal end of the protein. α helices and β sheets are each numbered sequentially from the N- to C-termini, except for those in the SUE (βa-βc). (**C**) Interpretive topology diagrams of the reconstructed ancestral septins at the nodes labeled in panel A. See Fig. 3.8 for the original AlphaFold2 predictions. Ancestral septins for Groups 1-5 are represented by a single diagram because their AlphaFold2 predictions appear largely identical. Structural motifs relevant to this study are highlighted in magenta. Secondary structures outlined in bold solid lines and dotted lines represent motifs with higher (pLDDT >70) and lower (pLDDT <70) AlphaFold confidence scores, respectively. R, arginine finger.

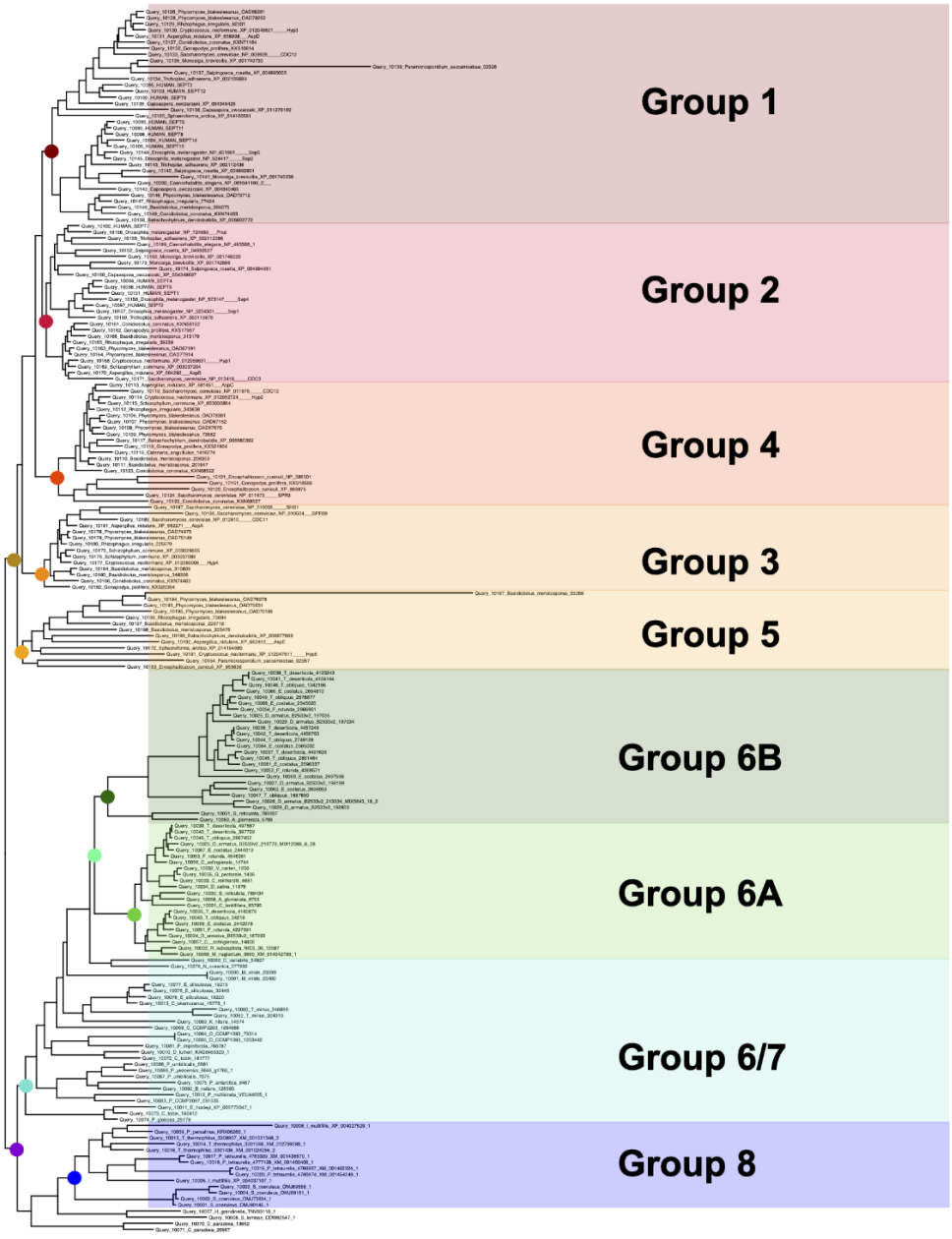

Figure 3.7. IQTree tree 200 septin sequences used in ancestral sequence reconstitution. Groups as defined in Figure 3.3 are redefined adjacent to branch tips. Colored nodes represent select ancestral sequences used for AlphaFold prediction.

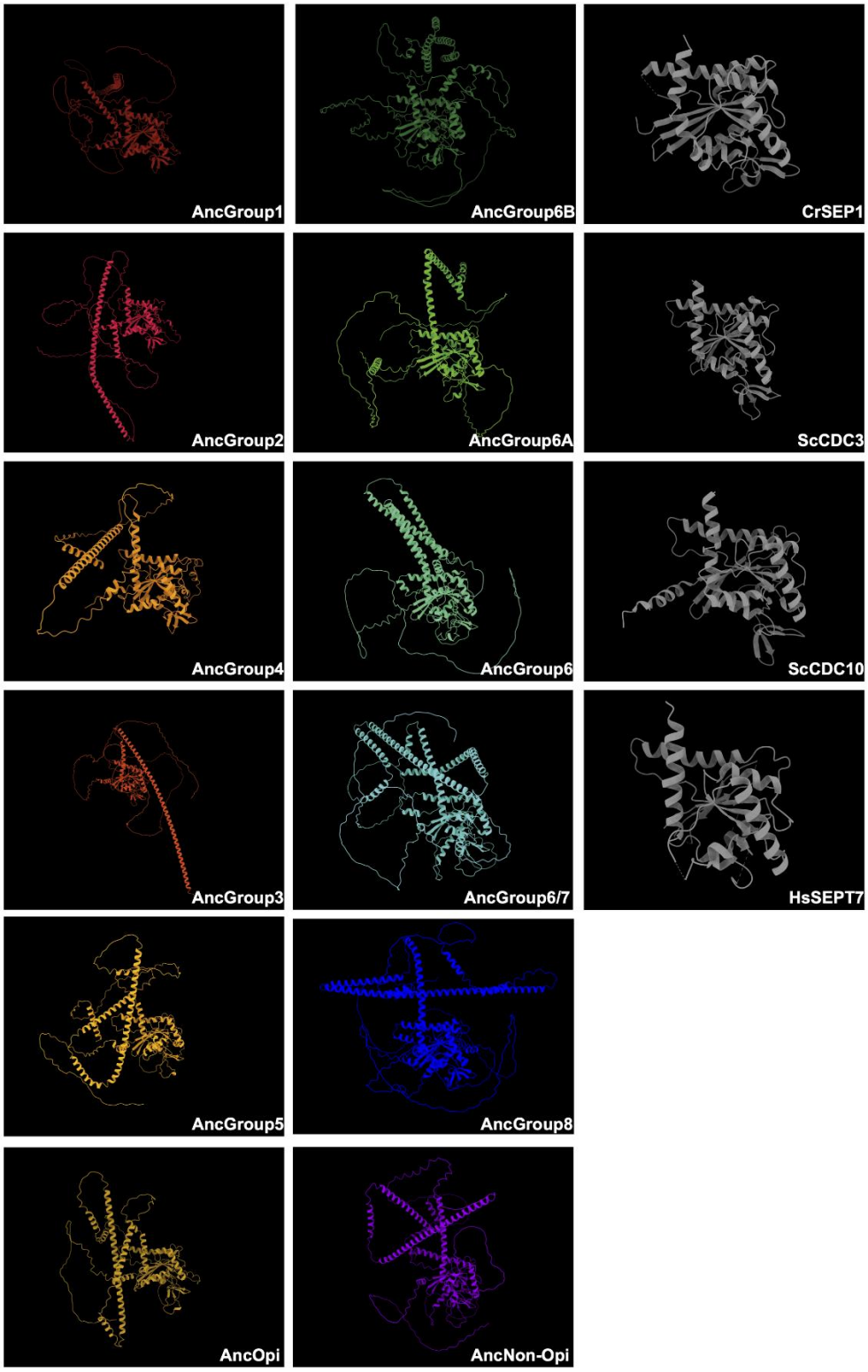

Figure 3.8. AlphaFold-predicted 3D structures of ancestral septins. Structures in grey are experimentally determined Protein Data Bank (PDB) files of septin GTPase domains, included here as references: CrSEP1 (PDB: 5IRR), ScCDC3 (PDB:8SGD), ScCDC10 (PDB:8SGD), HsSEPT7 (PDB:3TW4). Structures are orientated such that the NC-interface is towards the left of the monomer and the G-interface is towards the right.

To highlight gains and losses of sub-domain motifs during the evolution of ancestral septins, interpretive topology diagrams of the GTPase domains were generated based on the AlphaFold predictions (Fig. BC). This analysis revealed a largely consistent core structure of the GTPase domains consisting of six α-helices (α1-α6) and nine β-sheets (β1-β6 and βa-βc), as well as a few variable α-helices that emerged or were lost at specific ancestral nodes (see below). In addition to the helices and sheets, we identified an arginine residue positioned in the S3 motif of AncGroup 6-8 and LECA septins (Figs. 3.5BD and 3.6C). Although this residue is not found in the reconstructed AncGroup 1-5 septins (Fig. 3.6C), some extant Group 5 septins, such as *A. nidulans* AspE, appear to have it (see below). Thus, this "R-finger" arginine is an ancestral feature of septin family proteins that has been lost in most opisthokonts. Intriguingly, it has been reported that this R-finger in the single septin of *C. reinhardtii* is required for its homo-dimerization across the G-interface (Pinto et al., 2017), where it reaches into the GTP-binding pocket of the opposite subunit to accelerate GTP hydrolysis (see Fig. 3.5C, G-interface). Thus, we suspected that the R-finger would invariably be conserved in single septins found in other species. This prediction was partially confirmed: 20 of the 23 single septins that were included in our analysis have an R-finger at the expected position (Fig. 3.9A), suggesting that the dimerization mechanism observed in *C. reinhardtii* may be ancestral and conserved in many algae and protists. Of the other three that lacked an R-finger, the sequence from the dinoflagellate *S. minutum* is an extremely large 4484-aa protein, with a septin-like domain near the N-terminus and some additional domains (e.g., SMC domain, HSP70) that are not found in other septins. The other two (from the ciliates *Halteria grandinella* and *Stylonychia lemnae*) have the arginine replaced by a histidine residue. It is unknown whether these single septins still

form a G-dimer without an R-finger or have taken unique evolutionary paths to function without dimerizing through the G-interface.

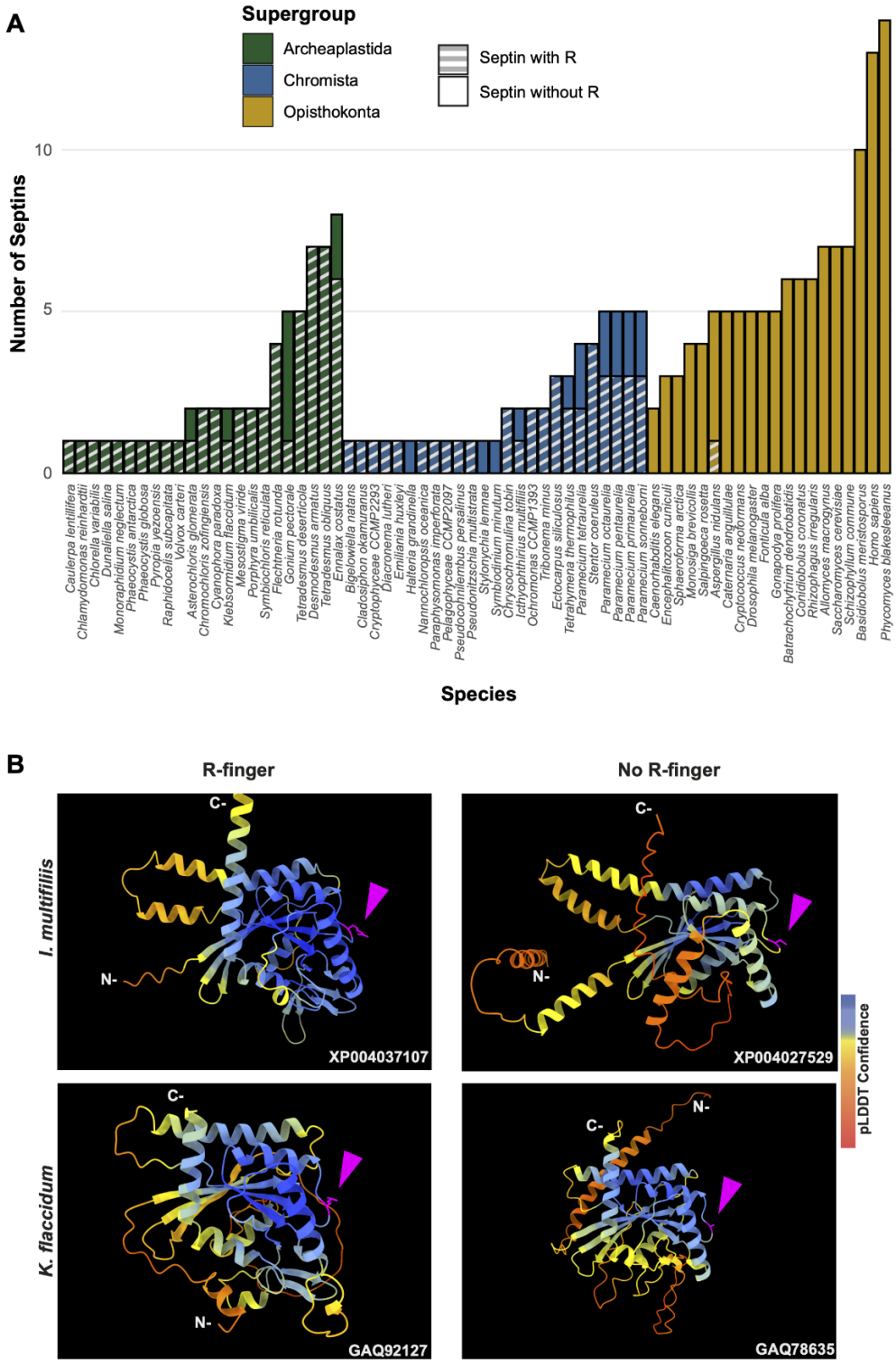

Figure 3.9. GAP-like R-finger is widely conserved in single septins. (**A**) Numbers of septins with and without R-finger in 68 species representing the three septin-harboring eukaryotic

supergroups. (**B**) AlphaFold predictions of septins with and without R-finger in the species *I. multifiliis* (top row) and *K. flaccidum* (bottom row). N- and C-, amino-terminus and carbonyl-terminus, respectively. Magenta arrowheads indicate the positions with the presence or absence of R-finger. Structures are colored according to the AlphaFold pLDDT confidence scores.

Interestingly, in many algae and protists with multiple septin genes, a loss of the R-finger is observed in some of the duplicated genes (Fig. 3.9A). For example, the ciliate *Ichthyophthirius multifiliis* possesses two septins: XP004037107 with an R-finger and XP004027529 without (Fig. 3.9B). Similarly, the filamentous charophyte green alga *Klebsormidium flaccidum* has two proteins with and without an R-finger (GAQ92127 and GAQ78635, respectively; Fig. 3.9B). Given the apparent selective pressure against the loss of R-finger in single septins as well as the loss of R-finger in most opisthokont septins that are invariably encoded as multiple copies in a genome (see below), it is tempting to speculate that these septins may have lost their R-finger because of evolution to form hetero-oligomers. Biochemical characterization of these septins is needed to address this possibility.

Unlike the non-opisthokont counterparts, the vast majority of opisthokont septins do not possess an R-finger between the S2-S3 motifs (Figs. 3.5D and 3.6C). In Group 1-4 septins, the arginine residue is replaced by small uncharged amino acids such as serine, glycine, or alanine. Although there is an invariant histidine residue in the adjacent position (Fig. 3.5D) that could potentially be involved in GTP hydrolysis (Weirich et al., 2008), a mutation to this amino acid in human SEPT2 did not affect its GTPase activity (Sirajuddin et al., 2009). Thus, it is unlikely that the Group 1-4 opisthokont septins employ an R-finger-like molecular mechanism to interact through their G-interfaces. The R-finger is also absent in most filamentous-fungus-specific Group 5 septins (Figs. 3.5D and 3.6C), consistent with the previous observation that the S1-S4

motifs in septins in these groups are highly variable (Shuman and Momany, 2021). However, some septins, such as *Aspergillus nidulans* AspE (Fig. 3.9A), have an arginine residue located between the divergent S2-S3 motifs. Available data suggest that AspE is not incorporated into canonical septin complexes, although it interacts with them in a developmental-stage-specific manner (Hernandez-Rodriguez et al., 2014). It is interesting to speculate that AspE-type Group 5 septins have retained the ancestral trait to form a homomeric G-dimer using their R-fingers.

Some opisthokont septins that lack the R-finger have lost their activity to hydrolyze GTP by losing a catalytically active threonine (or serine) within the switch I region, making them GTP-bound subunits (Fig. 3.10AB, yeast Cdc3, Cdc11, and human SEPT6; Rosa et al., 2020). We examined some representative non-opisthokont septins to ask if this residue is conserved. In septins with R-finger from *C. reinhardtii*, *V. carteri*, and *G. pectorale* (all Group 6A), this threonine is invariably conserved, consistent with the idea that these septins are active GTPases (Fig. 3.10A; Pinto et al., 2017). Interestingly, the other four septins from *G. pectorale* (Group 6B) all lack both the R-finger and the threonine (Fig. 3.10A), suggesting that they may have lost their GTPase activity. Similar concomitant loss of R-finger and catalytic threonine was observed in pairs of septins from *K. flaccidum* (Group 7) and *I. multifiliis* (Group 8). In contrast, all septins in *P. tetraurelia* (Group 8) contain the catalytic threonine, regardless of the presence or absence of their R-finger. These results suggest that sequential loss of R-finger (reduction of GTPase activity) and catalytic threonine in switch I (loss of GTPase activity) may have occurred independently in many (but not all) lineages during septin evolution, and that some non-opisthokonot species may form a septin complex consisting of a mix of GTP- and GDP-bound subunits, like their animal and fungal counterparts. We also observed that in some cases, such as *K. flaccidum* and *I. multfiliis*, where the catalytic threonine was lost, AlphaFold prediction

positions the Switch-I loop away from the G-interface (Fig. 3.10B). It is interesting to speculate that a potential rearrangement of Switch-I may have destabilized G-interface interactions in support of an emergent NC-interface interaction motif.

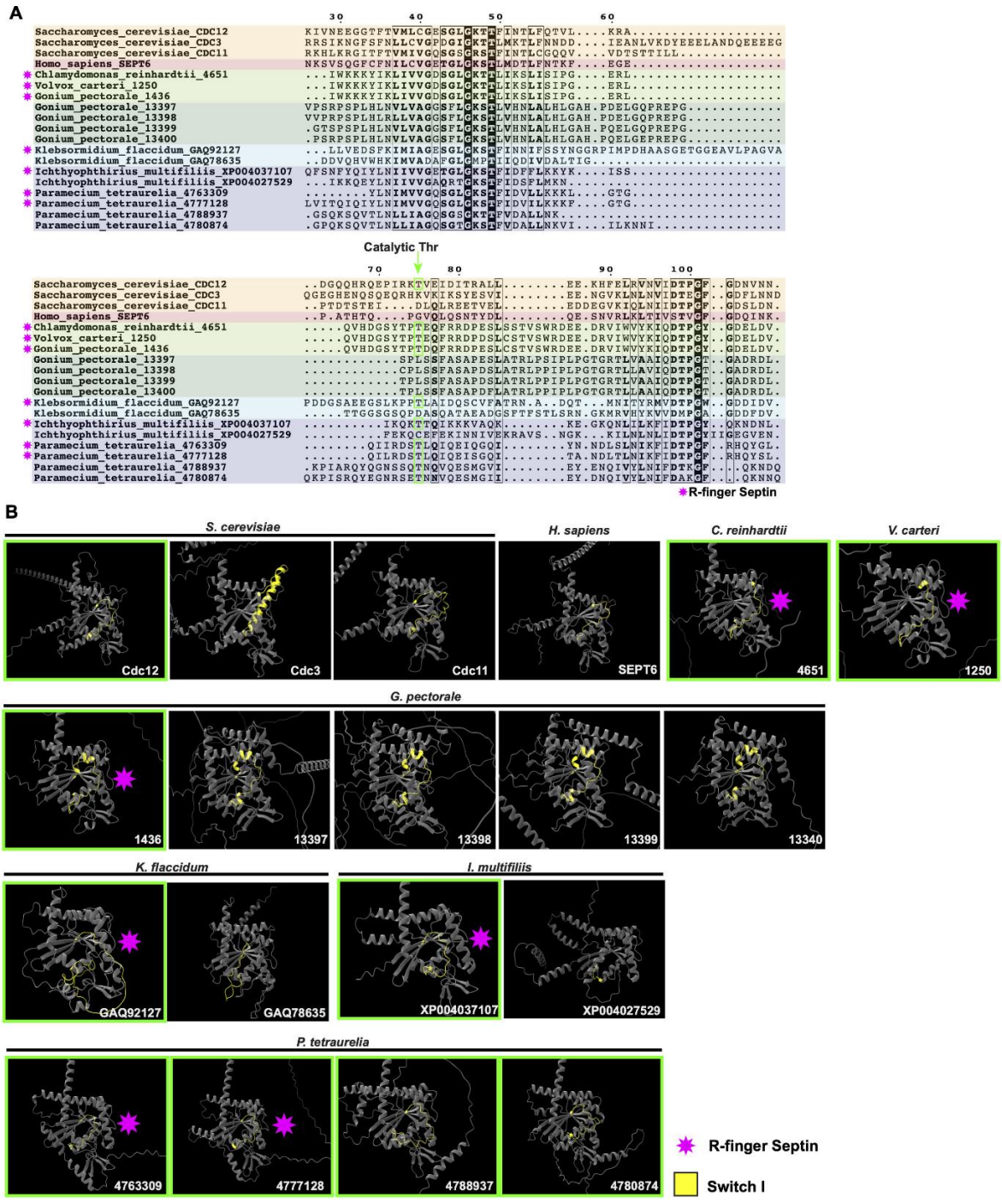

Figure 3.10. Apparent correlation between catalytic threonine in the Switch-I loop and R-finger septins. (A) Multiple-sequence alignment of a subset of representative septins across

eukaryotes. Sequences are colored according to their phylogenetic groupings. Regions between between G1 and G3 motifs are shown. Catalytic threonine residues conserved in some septins are indicated by green boxes and with an arrow. R-finger septins are indicated with a magenta star. (B) AlphaFold predicted structures of the septins displayed in (A). Structures are focused on the GTPase domain, with NC-interface towards the left and G-interface towards the right. Switch I region is highlighted in yellow. Green outline, septins with catalytic threonine; magenta star, septins with R-finger.

Conservation of α0 and α5' helices in opisthokont septins

In addition to the core helices and sheets, AlphaFold predictions of AncGroup 1-5 (opisthokont) septins displayed two additional invariant α-helices, both positioned in the NC-interface: α0 at the junction between the N-terminal extension and the GTPase domain, and α5' that is positioned in-between α4 and β6 (Fig. 3.6C). Interestingly, however, these helices are not predicted by AlphaFold in AncGroup 6-8 septins (Fig. 3.6C). In the human SEPT2/6/7 complex (and plausibly in many other opisthokont septins complexes), the α0 helix is an integral part of the NC interface where it forms an electrostatic inter-subunit interaction (Cavini et al., 2021). In addition, the α5'-helix contains a polyacidic region that is known to interact with the polybasic region 1 (PB1) within the α0 helix of a neighboring subunit across the NC interface (Fig. 3.5C; Cavini et al., 2021). Thus, it is conceivable that the α0 and α5' helices evolved together in the opisthokont lineage as the positioning of PB1 was fixed in the former (see below).

The PB1 domain in α0 helix binds to phospholipids such as phosphatidylinositol 4-phosphate, 4,5-bisphosphate, and 3,4,5-triphosphate (Zhang et al., 1999; Casamayor and Snyder, 2003; Bertin et al., 2010; Onishi et al., 2010; Krokowski et al., 2018). The PB1 domain has been observed in some septins in non-opisthokont species such as in *C. reinhardtii* (Wloga et al., 2008; Nishihama et al., 2011; Pinto et al., 2017) despite the lack of α0 in the same proteins (Figs. 3.5D and 3.6B), raising the possibility that the emergence of PB1 precedes that of α0. To test this, we examined the NTEs of the reconstructed ASR sequences for the presence of PB1 by

developing a Python script that calculates the isoelectric point of a 10 amino-acid window moving along protein sequences. We observed a basic region proximal to the beginning of the GTPase domain in AncGroup 1-5 septins (including in the very short NTE of AncGroup3 septin) (Fig. 3.11AB), consistent with the presence of PB1 in the majority of extant opisthokont septins (Nishihama et al., 2011; Shuman et al., 2021). Similarly, the regions immediately upstream of the G1 motif in AncGroup 6 and 6/7 septins are also highly basic (Fig. 3.11A). In contrast, the NTE of AncGroup8 is overall acidic (Fig. 3.11A), and a few basic residues found in this region are interdigitated by acidic residues (Fig. 3.11B), consistent with the reported ambiguity about the presence of polybasic regions in septins in *T. thermophila* and *P. tetraurelia* (Wloga et al., 2008). Interestingly, CLUSTALω alignment identified additional polybasic domains in AncGroup 6B and 6/7 septins at positions 339 and 214 aa upstream of the G1 motif, respectively, which exhibited greater similarity to the proximal PB1 observed in AncGroup 1-5 septins (Fig. 3.11B), and the G1-proximal sequences (PB1') are non-opisthokont-specific (Fig. 3.11B). Given the low overall sequence conservation of these regions in AncGroup 8 (Fig. 3.11B), it is not clear whether PB1' is an ancestral feature that has been lost in opisthokont septins, or it was newly inserted adjacent to the G1 motif in the lineage leading to Group 6 and 7 septins. Overall, however, the presence of a polybasic region in the NTE appears to be an ancestral feature that predates the emergence of opisthokont-specific α0.

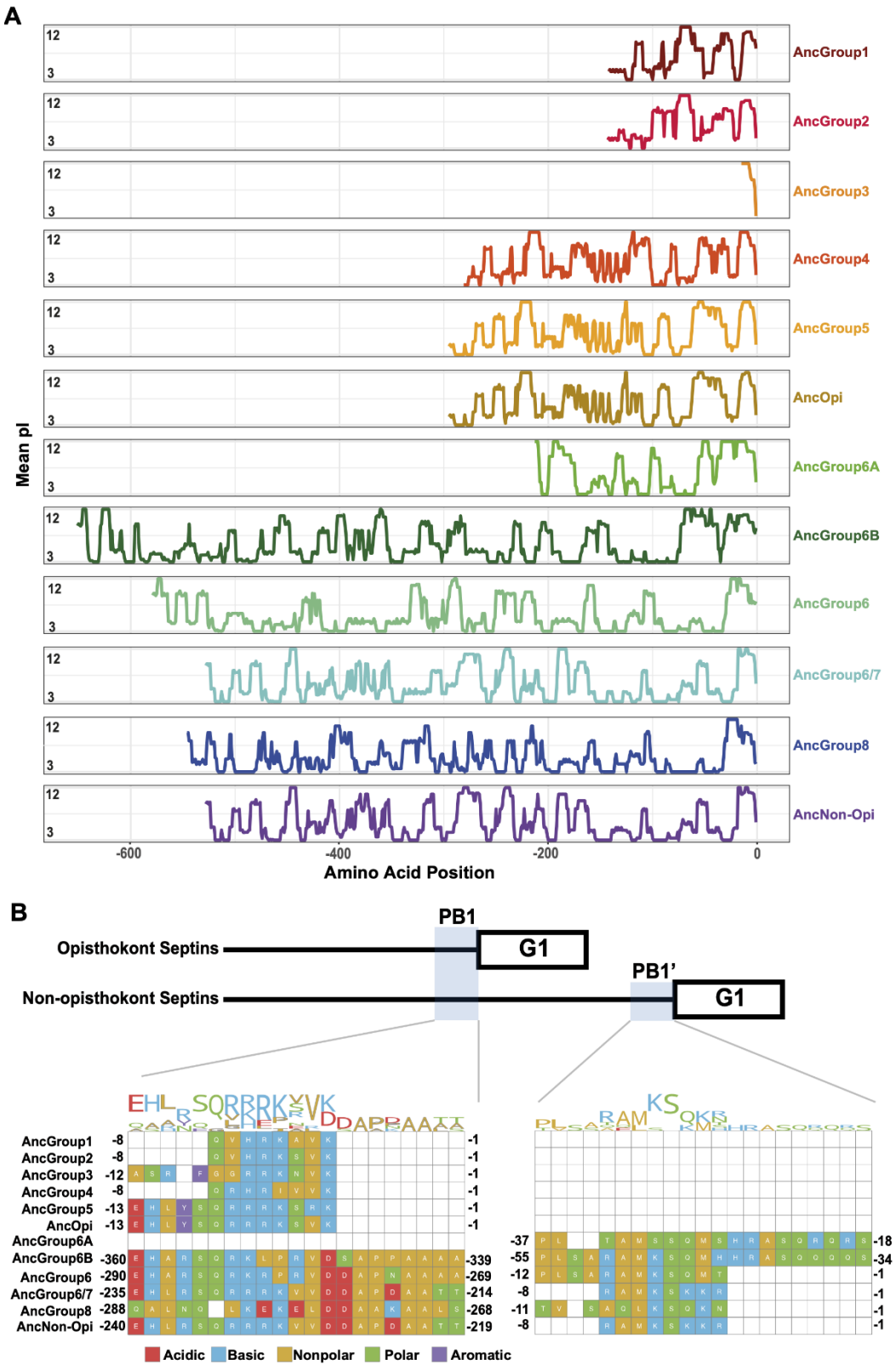

Figure 3.11. N-terminal polybasic domains across septins. (**A**) Calculation of isoelectric point windows across the NTE of reconstructed ancestral sequences. The average isoelectric point of a sliding 10 amino acid window is calculated across the NTE of reconstructed ancestral sequences.

X=0 represents the start of the GTPase domain. (**B**) CLUSTALw multiple sequence alignment of reconstructed ancestral sequences displaying two polybasic domains in non-opisthokont lineages. Numbers indicate the amino acid positions from the start of the GTPase domain.

Amphipathic helices are an ancestral feature of septins

Some opisthokont septins have the ability to recognize micron-scale membrane curvature through an amphipathic helix (AH) (Bridges et al., 2016; Cannon et al., 2019). Perturbation of these AHs can lead to abnormal subcellular localization of septin proteins (Cannon et al., 2019). To ask if putative membrane-binding AHs are found outside of opisthokonts and therefore can be an ancestral feature of septins, we developed a high-throughput pipeline to identify AH domains in a large number of polypeptide sequences by predicting alpha helices and then calculating their amphipathicity (see Materials & Methods), and applied it to the NTE and C-terminal extension (CTE) of our eukaryotic septin collection. This pipeline precisely identified previously reported AH domains in fungal and animal septins (Cannon et al., 2019; Lobato-Márquez et al., 2021; Woods et al., 2021), such as Cdc12 and Shs1 in *S. cerevisiae* and *Ashbya gossypii*, human SEPT6, *Caenorhabditis elegans* UNC-61, and *Drosophila melanogaster* Sep1 (Fig. 3.12). In some cases, multiple AHs were found in a single septin. These additional AHs could potentially be a result of the inherent amphipathicity of coiled-coil domains (where many AH domains reside). We limited our downstream analysis to hits with the largest calculated D-factor to focus on putative membrane-binding AHs. Our analysis revealed the presence of predicted AHs in septin sequences spanning all Groups (Fig. 3.13A; Table 3.2) with varying levels of conservation. In opisthokonts, for instance, predicted AHs were detected in 68% of Group 2 and Group 4 sequences, while only 13% of Group 3 sequences exhibited AHs. In Group 1, there is a striking difference between the two subclades: a predicted AH is completely absent in 1A

(animals and fungi), while it is found in 75% of septins in 1B (animal-specific). This suggests a potential connection between the evolution of AHs and the positioning of subunits within a canonical octameric protomer, in which 1A subunits occupy the central dimer. Like Group 3, only a small fraction of Group 5 septins (22%) have predicted AHs; unlike Group 1, there is no specific subgroup in which AHs are conserved, suggesting sporadic loss/gain of the domain within this group (Fig. 3.13A; Table 3.2). In general, the AHs in Groups 1-5 displayed features consistent with stereotypical amphipathicity, with a large hydrophobic window and a hydrophilic face composed of both positively and negatively charged residues (Fig. 3.12).

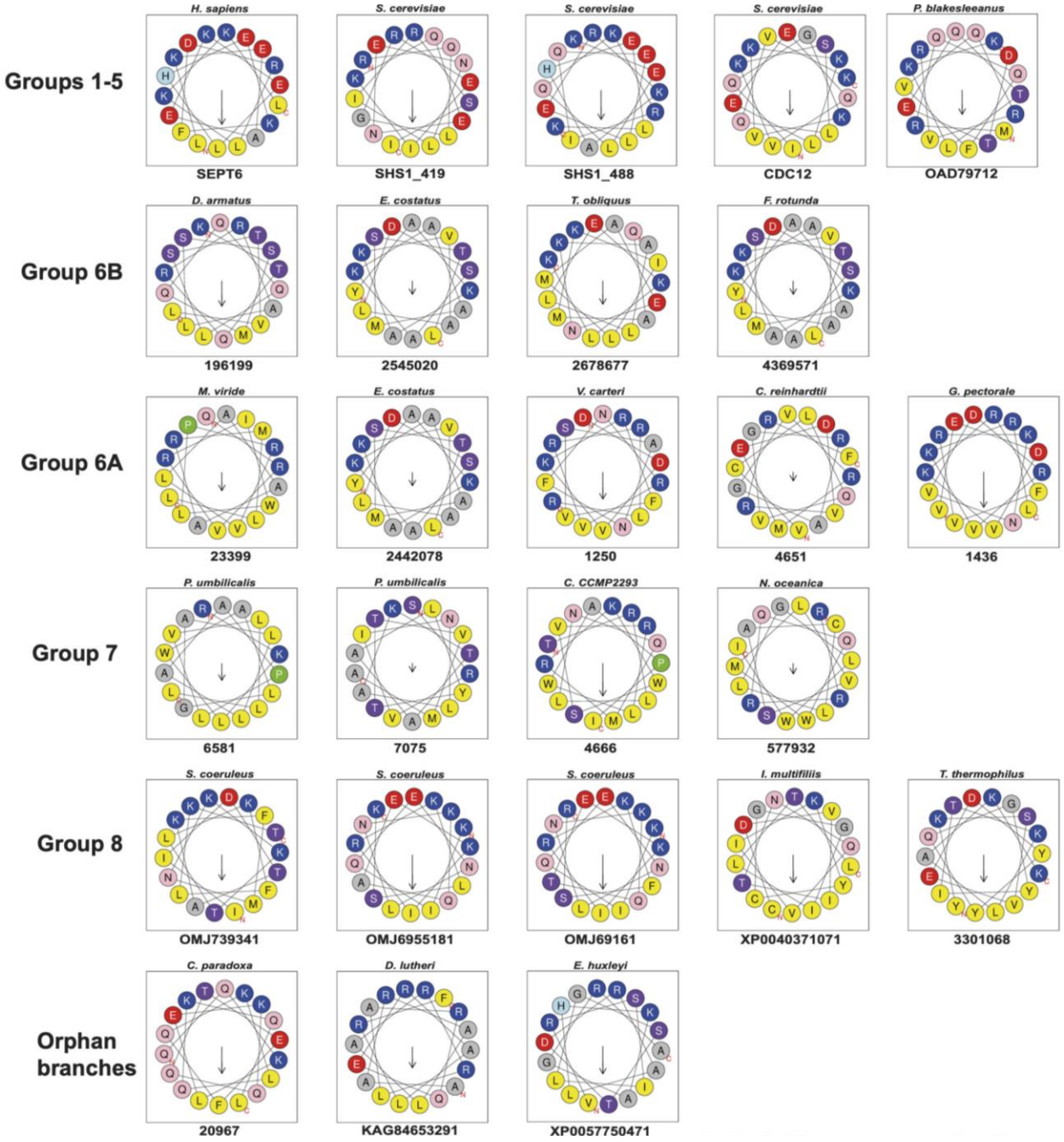

Figure 3.12. Representative helical wheel diagrams of predicted AHs across septin phylogenetic groups. Arrow represents the hydrophobic moment vector. Amino acids are colored according to their chemistry: yellow, hydrophobic; purple, Ser/Thr residues; grey, Gly/Ala residues; blue, basic residues; red, acidic residues; pink, Asp; green, Pro.

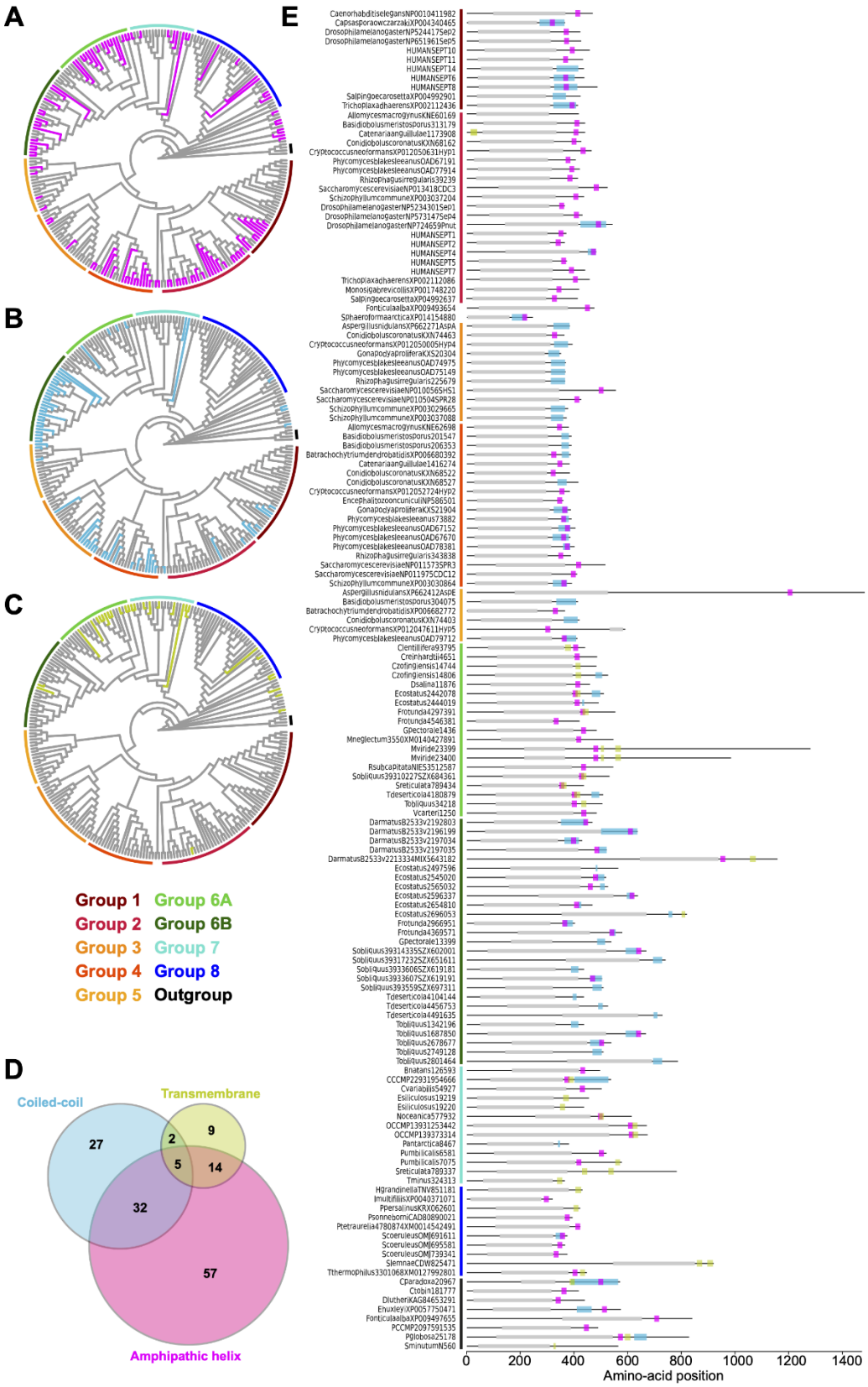
A
B
C
Group 1
Group 2
Group 3
Group 4
Group 5
Group 6A
Group 6B
Group 7
Group 8
Outgroup
D
Coiled-coil
Transmembrane
Amphipathic helix
27
9
2
5
14
32
57
E
0
200
400
600
800
1000
1200
1400
Amino-acid position

Figure 3.13. Distribution of AH, coiled-coil, and transmembrane domains across septin groups. (**A-C**) Simplified cladograms of the RAxML tree of 254 septins (see Fig. 3.3A), with individual sequences with AH (A, magenta), coiled-coil (B, blue), and transmembrane (C, green) domains highlighted. (**D**) Venn diagram showing the numbers of septins with AH, coiled-coil, and/or transmembrane domains. (**E**) Protein domain diagrams of septins with AH, coiled-coil, and/or transmembrane domains. Grey box, septin GTPase domain; magenta box, AH domain; blue box, coiled-coil domain; green box, transmembrane domain.

The wide distribution of AHs is also observed in all non-opisthokont groups (Fig. 3.13A; Table 3.2). Group 6A, consisting largely of single septins, has the highest rate of AH domains at 68%. In Group 6B, septins with predicted AHs were found in most subclades, with a total preservation rate of 50%. In Groups 7 and 8, septins with predicted AHs were found in at 38% and 19%, respectively. In the Heliquest visualization, both AHs present in Group 6B and Group 7 exhibited hydrophilic faces primarily composed of positively charged residues interspersed with small polar residues such as serines and threonines (Fig. 3.12). In some instances, weaker amphipathic helices were observed, as exemplified by *P. umbilicalis* 6581, which lacked a strongly pronounced hydrophilic face but still fulfilled the criteria of our search because of their high net charges that raised the *D*-factor (Fig. 3.13E; Fig. 3.12). Some Group 6A and Group 8 septins have predicted AHs similar to those observed in Groups 1-5 with a large hydrophobic window opposite the cluster of both positively and negatively charged residues.

Selective distribution of coiled-coil and transmembrane domains in specific septin groups

Many animal and fungal septins contain a coiled-coil (CC) motif in the CTE which is thought to be involved in polymer stabilization and the formation of bundles and filament pairs (Sirajuddin et al., 2007; Bertin et al., 2010; Cavini et al., 2021). We utilized the existing annotation of coiled-coil domains in the Uniprot database to identify them in our list of 254 extant septins. Interestingly, we observed the presence of CCs in Groups 1B, 3, 4, and 6B (Fig

3.13B; Table 3.2). The majority of these sequences were also positive for AH domains (Fig. 3.13D), with AH domains residing within CC domains in many cases, such as in *S. cerevisiae* Cdc12 (Fig 3.13E; Cannon et al., 2019). Interestingly, CC domains were almost entirely excluded from non-opisthokont Groups 6A, 7, and 8 (Fig. 3.13B; Table 3.2), suggesting that the CC domains observed in Group 6B were a result of convergent molecular evolution. It is interesting to speculate that septin gene duplication in some green algae (Fig. 3.9A) and the formation of heterooligomeric complexes may have led to the emergence of lateral pairing between septin subunits.

Lastly, it has previously been reported that some non-opisthokont septins possess putative transmembrane (TM) domains or short hydrophobic patches (Wloga et al., 2008; Nishihama et al., 2011). Thus, we searched for the presence of potential TM domains in our list of 254 extant septin sequences. Except for one sequence from the parasitic fungus *Catenaria anguillulae* (A0A1Y2I4M7, Group 2A, 46% identical to *S. cerevisiae* Cdc3) that has a unique N-terminal TM domain, all septins with a TM domain were found in the non-opisthokont lineages, with notable enrichment in Groups 6A and 7 (Fig. 3.13CE; Table 3.2). This distribution of TM domains in our dataset seems to suggest that they emerged early in the non-opisthokont branch after its split with opisthokonts and were subsequently lost in many species in Group 6B and 8. [See, however, Discussion for another possibility given a recent report by (Perry et al., 2023).] It is interesting to note that there is little overlap between the distributions of CC and TM domains in Group 6 septins (Fig. 3.13D), perhaps suggesting that the evolution of septin-septin interactions through CC domains necessitated a concomitant loss of TM that would otherwise restrict the accessibility of CTE.

In summary, our searches for α-helix-based structures that are often associated with septin CTE suggest that the AH and TM domains may have ancient origins in septin evolution, while the CC domain may have evolved independently in multiple lineages.

Table 3.2. Conservation of various features in septin groups.

Group	Phylum	R-finger (%)	α0 [a]	PB1 [a]	PB1' [a]	AH (%) [b]	CC (%) [b]	TM (%) [b]
1	Animals/fungi	0	Strong	Yes	No	27 [c]	18	0
2	Animals/fungi	0	Strong	Yes	No	**68**	6.5	3.2
3	Fungi	0	Strong	Yes	No	13	**33**	0
4	Fungi	0	Strong	Yes	No	**68**	**46**	0
5	Filamentous fungi	5.6	Strong	Yes	No	22	17	0
6A	Green algae	100	None	No	Yes	**68**	16	**44**
6B	Green algae	80	None	Yes	Yes	**50**	**87**	6.7
7	Various algae	91	Weak	Yes? [d]	Yes? [d]	**38**	9.5	**43**
8	Ciliates	60	Weak	No	No	18.9	2.7	11

[a] Based on AlphaFold predictions of ancestral protein structures.
[b] Based on analyses of extant sequences. Values greater than 30 are bold-faced. See Supplementary File 8 for details.
[c] 0% in 1A, 75% in 1B.
[d] Because Group 7 is paraphyletic, we could not confidently infer the conservation of PB domains based on AngGroup6/7.

Discussion

Septins have been reported in a variety of eukaryotic lineages outside of opisthokonts (Versele and Thorner, 2005; Wloga et al., 2008; Nishihama et al., 2011; Yamazaki et al., 2013; Onishi and Pringle, 2016; Shuman and Momany, 2021), although their phylogenetic relationships have not been fully explored. Here, we performed an updated search for septins in non-opisthokont lineages and found that septins are widely spread in two distinct non-opisthokont eukaryotic supergroups: Archaeplastida and Chromista. Because these two supergroups and opisthokonts share the ancestry only at the LECA level, our results strongly

support the idea that the first septin appeared in an early eukaryotic ancestor. We inferred structural features related to septin-septin interactions, membrane binding, and curvature sensing across eukaryotic evolution, and hypothesized functions related to ancestral septins.

Septins in Archaeplastida and Chromista form new phylogenetic clades outside of the previously defined Groups 1-5, herein named Groups 6A, 6B, 7, and 8. Group 6A and 6B are composed exclusively of septins from various green algae, while septins in Groups 7 and 8 belong to other various algae (some other green algae, red algae, heterokonts, haptophytes, cryptophytes, chlorarachniophytes) and ciliates, respectively. It is peculiar that these septins in algae from diverse groups formed a single clade separate from the ciliate septins, which is inconsistent with the general taxonomical classification of these species (compare Fig. 3.1 and Fig. 3.3A). It is tempting to speculate that these algal septins may have spread through horizontal transfer of nuclear genes, when ancestral red and green algae were taken up by other eukaryotes to form secondary and tertiary endosymbiosis (Keeling and Palmer, 2008; Archibald, 2012).

In this study, we found that the majority (but not all) of non-opisthokont septins have a conserved arginine residue within the G-interface. This arginine is predicted to act similarly to other R-fingers in GTPase-activating proteins (GAPs). Because R-fingers are also found in other "paraseptin" GTPases such as TOC34/TOC159 and AIG1/GIMAP (Leipe et al., 2002; Weirich et al., 2008), it is likely an ancestral feature that has been lost in some lineages. Biochemical and structural studies on the single Group 6A septin from *C. reinhardtii* have shown that this arginine is critical for the very high GTPase activity of this septin (40 times higher than human SEPT9, the most active septin GTPase in opisthokonts) and its homo-dimerization through the G-interface (Pinto et al., 2017). Interestingly, while Group 6A septins invariably have an R-finger, some Group 6B septins have lost this residue. It appears that the loss of R-finger is a crucial

evolutionary step associated with septin gene duplication in many eukaryotic lineages, including Group 6 (green algae), Group 8 (ciliates), and the transition from ancestral septin to opisthokonts.

Suppose we imagine an ancestral septin dimer with subunits possessing two potential interaction interfaces (G and NC). In that case, we predict that the presence of an R-finger strongly biases the interaction to the G-interface, suggesting that most ancestral septins formed a dimer across their G-interface. Upon gene duplication, some septins lost the R-finger and gained the NC-interface interaction motif, α0. These evolutionary events then would shift the equilibrium to favor the NC-interface, allowing for the formation of septin heterocomplex protomers. In some cases, evolution of non-opisthokont septin complexes may have involved further mutations in the GTP-binding pocket and the G-interface, causing some septins to be locked in apo-nucleotide or GTP-bound state, as seen in some opisthokont septins (Hussain et al., 2023).

When hypothesizing about the potential ancestral functions of septins, we sought to identify motifs that are crucial for septin function. We observe the presence of a polybasic domain immediately preceding the GTPase domain in all septins except for Group 8. Previous studies have implicated this domain to be important for membrane recognition, as well as stabilizing an NC-interaction interface (Bertin et al., 2010; Cavini et al., 2021). The wide distribution of the polybasic domain, but not an α0 helix in which it is found in opisthokonts, suggests that the role of ancestral septins involved their binding to lipid bilayers. In support of this, we found that AH domains were also present across many of the septin phylogenetic groups, suggesting that they are also an ancestral septin feature. By comparing helical wheel diagrams of these AH domains across species, we begin to see some level of heterogeneity in the

amino acid composition. Models to distinguish curvature sensing peptides highlight the importance of specific amino acid composition in either being a membrane sensor versus a membrane binder (van Hilten et al., 2023). It could be that the variation in amino acid composition confers distinct membrane binding properties, such as curvature sensing or subcellular localization. Within Groups 1-5, AH domains often had large hydrophobic faces and a large hydrophobic moment due to the presence of acidic and basic residues along the hydrophilic face. In contrast, in some lineages, particularly in group 6B and group 7, we observe the reduction of charged residues and often find threonine and serine residues. These residues may act as potential phosphorylation sites to adaptively regulate the functional properties of these helices (Byeon et al., 2022). Future biochemical studies of the AH domains of diverse septins would provide additional context to the ancestral role of this domain and to the question of whether membrane binding and/or curvature sensing are ancestral properties of septins.

We identified the presence of CC and putative TM domains in the CTE of septins across various phylogenetic groups. In non-opisthokonts, we observed an almost exclusive and ubiquitous conservation of CC domains in Group 6B, while TM domains are highly enriched in Group 6A. Considering that Group 6B is composed of septins that have undergone recent gene duplication, it raises an interesting possibility that septins utilize CC to form interactions between subunits and filaments only after the emergence of heterocomplexes. In this scenario, gene duplication and subsequent diversification would be a prerequisite for this specialization of function among subunits. It is important to note that our classification of septin groups was based solely on the sequences of the GTPase domain, independently of the CTE sequence. Therefore, the strong correlation between Group 6A/TM and Group6B/CC suggests a co-evolution between the GTPase and CTE.

In addition to Group 6, TM domains were found sporadically in the CTE of some Group 7 and 8 septins but largely missing from the opisthokont sequences we used in our analysis. We initially interpreted this as evidence that the TM domain emerged after the opisthokont/non-opisthokont split and was subsequently lost in some lineages. However, a recent study by Perry et al. (2023) reported the presence of TM domains in a transcript isoform of *C. elegans UNC-61* (Group 1) as well as many other opisthokont proteins currently annotated as septins on the Uniprot database (but were not included in our list of 254 septins). Interestingly, many of these TM domains are found in the NTE, as seen in *C. anguillulae* A0A1Y2I4M7 (Fig. 3.13E). Thus, we provide two possible interpretations: The N- and C-terminal TM domains evolved independently in opisthokonts and non-opisthokonts, respectively. Alternatively, the LECA septin possessed a TM in the C-terminus, which was inherited by some progeny in all septin groups; in opisthokonts, domain movement within a gene (Furuta et al., 2011) shifted the position of TM from C- to N-terminus.

For future studies of septin evolution and general principles of evolutionary constraints, two approaches appear particularly appealing. First, a comparative approach using green algae with single vs. multiple septins seems to provide a unique opportunity to understand the evolution of septin duplication and the formation of heterocomplexes. For example, while *C. reinhardtii* possesses a single Group 6A septin with R-finger, PB1/PB1', AH, and possible TM ((Wloga et al., 2008; Nishihama et al., 2011); though it is not currently annotated as such on Uniprot), a related green alga in the same Chlamydomonadales order, *G. pectorale*, has a total of five septins (one Group 6A and four 6B) with various combinations of septin features (Supplementary File 8). The Kinoshita rule (Kinoshita, 2003a) of opisthokont septins highlights the modularity and redundancy of opisthokont septin subunits at each position of a canonical

protomer, where a septin from the same group can replace one another. Biochemical and cell biological experiments of Group 6A and Group 6B septins can shed light on whether this rule also applies to non-opisthokont septins.

Second, to understand how – parsimoniously – a single septin with R-finger evolved into a highly variable family of five septin groups in opisthokonts, some filamentous fungi possessing Group 5 with putative R-fingers seem to be an ideal model. One such protein, AspE in *A. nidulans*, has been shown to be excluded from the heterooligomeric complex formed by other subunits (Hernández-Rodríguez et al., 2014). Perhaps this septin has an extremely high GTPase activity, forms a G-dimer, and works independently of canonical filaments or binds to filaments in a substoichiometric fashion.

Finally, although our study provided a general overview of septin evolution, it is important to consider these evolutionary events in the context of the cellular processes the ancestral septins were involved in. Given the near-universal role of animal and fungal septins in cytokinesis, it is tempting to speculate that ancestral septins had similar roles. In support of this, the single septin in the green alga *N. bacillaris* showed its localization at the division site (Yamazaki et al., 2013). However, the two and only other reports on non-opisthokont septins did not show division-site localization: in another green alga *C. reinhardtii,* a septin was found at the flagella-base region, and in the ciliate *T. thermophila,* septins were found associated with mitochondria (Wloga et al., 2008; Pinto et al., 2017). Further functional studies of septins in non-opisthokonts are necessary to reveal the ancestral and fundamental functions of septins.

Acknowledgments

We thank Jenna Perry and Amy Maddox for sharing unpublished information, and Holly Goodson, Brae Bigge, Natsumi Tajima-Shirasaki, Rossie Clark-Cotton, and Grace Hamilton for

their feedback on this manuscript. We also thank the two anonymous reviewers whose comments helped improve the manuscript. This work was supported by National Institutes of Health Grant R01 GM131004 (to S.B.), National Science Foundation Grants MCB 1818383 (to M.O. and John R. Pringle) and MCB CAREER 2337141 (to M.O.), Duke University Department of Biology, and The Franklin College of Arts and Sciences at the University of Georgia.

Special note

At a late stage of the review process, after the work was completed, it was brought to our attention that the NCBI Reference Sequence for *A. nidulans* AspE used in this study contained an apparent error in the database. The original ID, XP_662412, was linked to the protein model used in this study, XP_662412.1, which appears to be produced by an erroneous fusion between *aspE* and its neighboring gene. This protein model has since been replaced by XP_662412.2. Although this database error does not affect our overall conclusions, we provided a summary of this change in Supplementary File 9.

References

Archibald, J. M. (2012). "Chapter Three - The Evolution of Algae by Secondary and Tertiary Endosymbiosis," in Advances in Botanical Research, ed. G. Piganeau (Academic Press), 87–118. doi: 10.1016/B978-0-12-391499-6.00003-7

Ashkenazy, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., et al. (2012). FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Research 40, W580–W584. doi: 10.1093/nar/gks498

Auxier, B., Dee, J., Berbee, M. L., and Momany, M. (2019). Diversity of opisthokont septin proteins reveals structural constraints and conserved motifs. BMC Evol Biol 19, 4. doi: 10.1186/s12862-018-1297-8

Bertin, A., McMurray, M. A., Thai, L., Garcia, G., Votin, V., Grob, P., et al. (2010). Phosphatidylinositol-4,5-bisphosphate Promotes Budding Yeast Septin Filament Assembly and Organization. Journal of Molecular Biology 404, 711–731. doi: 10.1016/j.jmb.2010.10.002

Brawley, S. H., Blouin, N. A., Ficko-Blean, E., Wheeler, G. L., Lohr, M., Goodson, H. V., et al. (2017). Insights into the red algae and eukaryotic evolution from the genome of Porphyra umbilicalis (Bangiophyceae, Rhodophyta). Proceedings of the National Academy of Sciences 114, E6361–E6370. doi: 10.1073/pnas.1703088114

Bridges, A. A., Jentzsch, M. S., Oakes, P. W., Occhipinti, P., and Gladfelter, A. S. (2016). Micron-scale plasma membrane curvature is recognized by the septin cytoskeleton. J Cell Biol 213, 23–32. doi: 10.1083/jcb.201512029

Byeon, S., Werner, B., Falter, R., Davidsen, K., Snyder, C., Ong, S.-E., et al. (2022). Proteomic Identification of Phosphorylation-Dependent Septin 7 Interactors that Drive Dendritic

Spine Formation. Frontiers in Cell and Developmental Biology 10. Available at: https://www.frontiersin.org/articles/10.3389/fcell.2022.836746 (Accessed December 28, 2023).

Byers, B., and Goetsch, L. (1976). A highly ordered ring of membrane-associated filaments in budding yeast. Journal of Cell Biology 69, 717–721. doi: 10.1083/jcb.69.3.717

Cannon, K. S., Woods, B. L., Crutchley, J. M., and Gladfelter, A. S. (2019). An amphipathic helix enables septins to sense micrometer-scale membrane curvature. Journal of Cell Biology 218, 1128–1137. doi: 10.1083/jcb.201807211

Casamayor, A., and Snyder, M. (2003). Molecular dissection of a yeast septin: distinct domains are required for septin interaction, localization, and function. Mol Cell Biol 23, 2762–2777. doi: 10.1128/MCB.23.8.2762-2777.2003

Cavalier-Smith, T. (2018). Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. Protoplasma 255, 297–357. doi: 10.1007/s00709-017-1147-3

Cavini, I. A., Leonardo, D. A., Rosa, H. V. D., Castro, D. K. S. V., D'Muniz Pereira, H., Valadares, N. F., et al. (2021). The Structural Biology of Septins and Their Filaments: An Update. Frontiers in Cell and Developmental Biology 9. Available at: https://www.frontiersin.org/articles/10.3389/fcell.2021.765085 (Accessed May 29, 2023).

Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1982). The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature 299, 371–374. doi: 10.1038/299371a0

Fauchere, J., and Pliska, V. (1983). Hydrophobic parameters II of amino acid side-chains from the partitioning of N-acetyl-amino acid amides. Eur. J. Med. Chem. 18.

Feng, S.-H., Xia, C.-Q., and Shen, H.-B. (2022). CoCoPRED: coiled-coil protein structural

feature prediction from amino acid sequence using deep neural networks. Bioinformatics 38, 720–729. doi: 10.1093/bioinformatics/btab744

Field, C. M., al-Awar, O., Rosenblatt, J., Wong, M. L., Alberts, B., and Mitchison, T. J. (1996). A purified Drosophila septin complex forms filaments and exhibits GTPase activity. The Journal of cell biology 133, 605–616. doi: 10.1083/jcb.133.3.605

Furuta, Y., Kawai, M., Uchiyama, I., and Kobayashi, I. (2011). Domain movement within a gene: a novel evolutionary mechanism for protein diversification. PLoS One 6, e18819. doi: 10.1371/journal.pone.0018819

Gautier, R., Douguet, D., Antonny, B., and Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific alpha-helical properties. Bioinformatics 24, 2101–2102. doi: 10.1093/bioinformatics/btn392

Goodson, H. V., Kelley, J. B., and Brawley, S. H. (2021). Cytoskeletal diversification across 1 billion years: What red algae can teach us about the cytoskeleton, and vice versa. BioEssays 43, 2000278. doi: 10.1002/bies.202000278

Grupp, B., and Gronemeyer, T. (2023). A biochemical view on the septins, a less known component of the cytoskeleton. Biological Chemistry 404, 1–13. doi: 10.1515/hsz-2022-0263

Hartwell, L. H. (1971). Genetic control of the cell division cycle in yeast: IV. Genes controlling bud emergence and cytokinesis. Experimental Cell Research 69, 265–276. doi: 10.1016/0014-4827(71)90223-0

Hartwell, L. H., Culotti, J., Pringle, J. R., and Reid, B. J. (1974). Genetic Control of the Cell Division Cycle in Yeast. Science 183, 46–51. doi: 10.1126/science.183.4120.46

Hernández-Rodríguez, Y., Masuo, S., Johnson, D., Orlando, R., Smith, A., Couto-Rodriguez, M.,

et al. (2014). Distinct Septin Heteropolymers Co-Exist during Multicellular Development in the Filamentous Fungus *Aspergillus nidulans*. PLOS ONE 9, e92819. doi: 10.1371/journal.pone.0092819

Hussain, A., Nguyen, V. T., Reigan, P., and McMurray, M. (2023). Evolutionary degeneration of septins into pseudoGTPases: impacts on a hetero-oligomeric assembly interface. Front Cell Dev Biol 11, 1296657. doi: 10.3389/fcell.2023.1296657

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. doi: 10.1038/s41586-021-03819-2

Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338, 1027–1036. doi: 10.1016/j.jmb.2004.03.016

Keeling, P. J., and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9, 605–618. doi: 10.1038/nrg2386

Kinoshita, M. (2003a). Assembly of mammalian septins. J Biochem 134, 491–496. doi: 10.1093/jb/mvg182

Kinoshita, M. (2003b). The septins. Genome Biol 4, 1–9. doi: 10.1186/gb-2003-4-11-236

Koenig, P., Oreb, M., Rippe, K., Muhle-Goll, C., Sinning, I., Schleiff, E., et al. (2008). On the Significance of Toc-GTPase Homodimers *. Journal of Biological Chemistry 283, 23104–23112. doi: 10.1074/jbc.M710576200

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305, 567–580. doi: 10.1006/jmbi.2000.4315

Krokowski, S., Lobato-Márquez, D., Chastanet, A., Pereira, P. M., Angelis, D., Galea, D., et al. (2018). Septins Recognize and Entrap Dividing Bacterial Cells for Delivery to Lysosomes. Cell Host & Microbe 24, 866-874.e4. doi: 10.1016/j.chom.2018.11.005

Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B. M., Wägele, J. W., et al. (2010). Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Frontiers in Zoology 7, 10. doi: 10.1186/1742-9994-7-10

Kueck, P. (2017). ALICUT: a Perlscript which cuts ALISCORE identified RSS. Available at: https://github.com/PatrickKueck/AliCUT/blob/master/ALICUT_V2.31.pl

Lee, S., Carrasquillo Rodrı Guez, J. W., Merta, H., and Bahmanyar, S. (2023). A membrane-sensing mechanism links lipid metabolism to protein degradation at the nuclear envelope. J Cell Biol 222, e202304026. doi: 10.1083/jcb.202304026

Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases11Edited by J. Thornton. Journal of Molecular Biology 317, 41–72. doi: 10.1006/jmbi.2001.5378

Leonardo, D. A., Cavini, I. A., Sala, F. A., Mendonça, D. C., Rosa, H. V. D., Kumagai, P. S., et al. (2021). Orientational Ambiguity in Septin Coiled Coils and its Structural Basis. Journal of Molecular Biology 433, 166889. doi: 10.1016/j.jmb.2021.166889

Lobato-Márquez, D., Xu, J., Güler, G. Ö., Ojiakor, A., Pilhofer, M., and Mostowy, S. (2021). Mechanistic insight into bacterial entrapment by septin cage reconstitution. Nat Commun 12, 4511. doi: 10.1038/s41467-021-24721-5

Longtine, M. S., DeMarini, D. J., Valencik, M. L., Al-Awar, O. S., Fares, H., De Virgilio, C., et

al. (1996). The septins: roles in cytokinesis and other processes. Current Opinion in Cell Biology 8, 106–119. doi: 10.1016/S0955-0674(96)80054-8

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting Coiled Coils from Protein Sequences. Science 252, 1162–1164. doi: 10.1126/science.252.5009.1162

Marques da Silva, R., Christe dos Reis Saladino, G., Antonio Leonardo, D., D'Muniz Pereira, H., Andréa Sculaccio, S., Paula Ulian Araujo, A., et al. (2023). A key piece of the puzzle: The central tetramer of the *Saccharomyces cerevisiae* septin protofilament and its implications for self-assembly. Journal of Structural Biology 215, 107983. doi: 10.1016/j.jsb.2023.107983

McMurray, M. A., and Thorner, J. (2008). "Biochemical Properties and Supramolecular Architecture of Septin Hetero-Oligomers and Septin Filaments," in The Septins, (John Wiley & Sons, Ltd), 47–100. doi: 10.1002/9780470779705.ch3

Mendonça, D. C., Guimarães, S. L., Pereira, H. D., Pinto, A. A., de Farias, M. A., de Godoy, A. S., et al. (2021). An atomic model for the human septin hexamer by cryo-EM. Journal of Molecular Biology 433, 167096. doi: 10.1016/j.jmb.2021.167096

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees., in 2010 Gateway Computing Environments Workshop (GCE), 1–8. doi: 10.1109/GCE.2010.5676129

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37, 1530–1534. doi: 10.1093/molbev/msaa015

Misof, B., and Misof, K. (2009). A Monte Carlo Approach Successfully Identifies Randomness

in Multiple Sequence Alignments : A More Objective Means of Data Exclusion. Systematic Biology 58, 21–34. doi: 10.1093/sysbio/syp006

Moffat, L., and Jones, D. T. (2021). Increasing the accuracy of single sequence prediction methods using a deep semi-supervised learning framework. Bioinformatics 37, 3744–3751. doi: 10.1093/bioinformatics/btab491

Momany, M., Zhao, J., Lindsey, R., and Westfall, P. J. (2001). Characterization of the *Aspergillus nidulans* septin (asp) gene family. Genetics 157, 969–977. doi: 10.1093/genetics/157.3.969

Nishihama, R., Onishi, M., and Pringle, J. R. (2011). New insights into the phylogenetic distribution and evolutionary origins of the septins. Biol Chem 392, 681–687. doi: 10.1515/BC.2011.086

Omrane, M., Camara, A. S., Taveneau, C., Benzoubir, N., Tubiana, T., Yu, J., et al. (2019). Septin 9 has Two Polybasic Domains Critical to Septin Filament Assembly and Golgi Integrity. iScience 13, 138–153. doi: 10.1016/j.isci.2019.02.015

Onishi, M., Koga, T., Hirata, A., Nakamura, T., Asakawa, H., Shimoda, C., et al. (2010). Role of septins in the orientation of forespore membrane extension during sporulation in fission yeast. Mol Cell Biol 30, 2057–2074. doi: 10.1128/MCB.01529-09

Onishi, M., and Pringle, J. R. (2016). The nonopisthokont septins: How many there are, how little we know about them, and how we might learn more. Methods Cell Biol 136, 1–19. doi: 10.1016/bs.mcb.2016.04.003

Pan, F., Malmberg, R. L., and Momany, M. (2007). Analysis of septins across kingdoms reveals orthology and new motifs. BMC Evol Biol 7, 103. doi: 10.1186/1471-2148-7-103

Papadopoulos, J. S., and Agarwala, R. (2007). COBALT: constraint-based alignment tool for

multiple protein sequences. Bioinformatics 23, 1073–1079. doi: 10.1093/bioinformatics/btm076

Perry, J. A., Werner, M. E., Heck, B. W., Maddox, P. S., and Maddox, A. S. (2023). Septins throughout phylogeny are predicted to have a transmembrane domain, which in *Caenorhabditis elegans* is functionally important. 2023.11.20.567915. doi: 10.1101/2023.11.20.567915

Pinto, A. P. A., Pereira, H. M., Zeraik, A. E., Ciol, H., Ferreira, F. M., Brandão-Neto, J., et al. (2017). Filaments and fingers: Novel structural aspects of the single septin from *Chlamydomonas reinhardtii*. J Biol Chem 292, 10899–10911. doi: 10.1074/jbc.M116.762229

Rosa, H. V. D., Leonardo, D. A., Brognara, G., Brandão-Neto, J., D'Muniz Pereira, H., Araújo, A. P. U., et al. (2020). Molecular Recognition at Septin Interfaces: The Switches Hold the Key. Journal of Molecular Biology 432, 5784–5801. doi: 10.1016/j.jmb.2020.09.001

Schwefel, D., Arasu, B. S., Marino, S. F., Lamprecht, B., Köchert, K., Rosenbaum, E., et al. (2013). Structural Insights into the Mechanism of GTPase Activation in the GIMAP Family. Structure 21, 550–559. doi: 10.1016/j.str.2013.01.014

Shuman, B., and Momany, M. (2021). Septins From Protists to People. Front Cell Dev Biol 9, 824850. doi: 10.3389/fcell.2021.824850

Sirajuddin, M., Farkasovsky, M., Hauer, F., Kühlmann, D., Macara, I. G., Weyand, M., et al. (2007). Structural insight into filament formation by mammalian septins. Nature 449, 311–315. doi: 10.1038/nature06052

The UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research 51, D523–D531. doi: 10.1093/nar/gkac1052

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44, W232–W235. doi: 10.1093/nar/gkw256

van Hilten, N., Methorst, J., Verwei, N., and Risselada, H. J. (2023). Physics-based generative model of curvature sensing peptides; distinguishing sensors from binders. Science Advances 9, eade8839. doi: 10.1126/sciadv.ade8839

Versele, M., and Thorner, J. (2005). Some assembly required: yeast septins provide the instruction manual. Trends Cell Biol 15, 414–424. doi: 10.1016/j.tcb.2005.06.007

Weirich, C. S., Erzberger, J. P., and Barral, Y. (2008). The septin family of GTPases: architecture and dynamics. Nat Rev Mol Cell Biol 9, 478–489. doi: 10.1038/nrm2407

Wloga, D., Strzyzewska-Jówko, I., Gaertig, J., and Jerka-Dziadosz, M. (2008). Septins stabilize mitochondria in *Tetrahymena thermophila*. Eukaryot Cell 7, 1373–1386. doi: 10.1128/EC.00085-08

Woods, B. L., Cannon, K. S., Vogt, E. J. D., Crutchley, J. M., and Gladfelter, A. S. (2021). Interplay of septin amphipathic helices in sensing membrane-curvature and filament bundling. Mol Biol Cell 32, br5. doi: 10.1091/mbc.E20-05-0303

Yamazaki, T., Owari, S., Ota, S., Sumiya, N., Yamamoto, M., Watanabe, K., et al. (2013). Localization and evolution of septins in algae. The Plant Journal 74, 605–614. doi: 10.1111/tpj.12147

Zhang, J., Kong, C., Xie, H., McPherson, P. S., Grinstein, S., and Trimble, W. S. (1999). Phosphatidylinositol polyphosphate binding to the mammalian septin H5 is modulated by GTP. Current Biology 9, 1458–1467. doi: 10.1016/S0960-9822(00)80115-3

CHAPTER 4

DETECTION AND CONSEQUENCES OF INTRA-SPECIES CONTAMINATION IN WHOLE GENOME SEQUENCES[6]

[6] Scopel, E., Shuman, B., Ward, A., Momany, M., and Bensasson, B. To be submitted to G3.

Abstract

Phylogenetic and population genomic inferences made using whole genome sequences are dependent on pure sequences. Genome sequence contamination can come from a variety of causes. Contaminant sequences can come from within the same species or from others and previous research has focused primarily on cross-species contamination detection. This paper visualizes B-allele frequency to qualitatively test for contamination, and measures the effects of intra-species contamination on phylogenetic and admixture analysis in two fungal species. As little as 5-10% contamination was enough to change whole-genome tree topologies and make strains of distinct ancestries appear as hybrids. We suggest researchers consider using B-allele frequency plots to screen their isolates for contamination.

Introduction

The strength of genomic analyses is dependent on the quality of the genomes used. While genome assembly quality continues to increase, contamination in whole genome sequences is a problem that will continue to persist. Contamination of short-read sequences can be biological, experimental, and/or computational (as reviewed by Cornet and Baurain, 2022). Multiplex sequencing opens users to the possibility of *in silico* contamination through barcoding problems (Clark et al., 2019) or cross-contamination (Ballenghien et al., 2017) while *in vitro* contamination can arise from multiple sources including but not limited to accidental co-culture and DNA-contamination (Cornet and Baurain, 2022).

Contamination of short-read sequences can lead researchers to inaccurate conclusions. Interspecies contamination can suggest cross-species horizontal gene transfer (Merchant et al., 2014, Goig et al, 2020), while intra-species contamination can suggest sexual or parasexual reproduction in obligate asexual species (Wilson et al., 2018). Successful detection of contamination in WGS is necessary to ensure accurate and robust genomic analyses.

It is becoming common practice to test for contamination including using machine learning algorithms (Fierst and Murdoch, 2017, Cornet and Baurain, 2022), however many of the tools available are designed for haploid prokaryotes or inter-species contamination. Previous work has shown intra-species contamination can affect common population genomic analyses (Ballenghien et al., 2017). Additional tools are needed for rapid detection of within-species contamination integrable into standard SNP-calling pipelines. In this paper, we plot B-allele frequency from variant call files (Bensasson, 2018) to visualize intra-species contamination, and we analyze the effects of intra-species contamination on phylogenetic and population genetic analyses.

Materials and Methods

Selecting *A. fumigatus* strains

52 *A. fumigatus* strains from the 168 in Kang et al. (2020) were selected based on their genetic distances to outgroup strain CF098. To do this, pairwise genetic distances for all strains were calculated using the dnadist function of PHYLIP (v.3.697; Felsenstein 1989). A bespoke python script (getGenDist.py; available on request) was used to categorize all strains into 52 clusters, based on their genetic distances to CF098, using one-dimensional k-means clustering (Jenks 1967). Finally, one strain was randomly sampled from each cluster.

Selecting *S. cerevisiae* strains

52 *S. cerevisiae* strains were selected similarly as with *A. fumigatus*. A genetic distance matrix was calculated for 280 European Wine, and 8 Mediterranean Oak strains from the whole-genome genealogy in Scopel et al (2021) using PHYLIP. getGenDist.py was used to categorize all 280 European Wine strains into 26 clusters based on their genetic distances to the outgroup. Two strains were then randomly sampled from each cluster, resulting in 52 *S. cerevisiae* strains.

Building and analyzing *in silico* mixtures

Strains of known ploidy and heterozygosity were selected to be in-silico contaminated (recipient) by the contaminator strains (donor). Each recipient was mixed with different proportions of the donor (0, 1, 5, 10, 20, 30, 40, and 50%) by randomly sampling sequences without replacement (seqtk sample, version 1.2, Li, 2012 for *S. cerevisiae* and version 1.3 for *A. fumigatus*) and merging them, resulting in fastq.gz files that result in an average coverage of 50. Reads for each mixture were mapped to respective reference genomes (SacCer_Apr2011/sacCer3 from UCSC, and GCF_000002655.1_ ASM265v1 from Nierman et al., 2005) with Burrows-Wheeler Aligner (bwa mem, version 0.7.17; Li and Durbin, 2009).

Consensus sequences were generated for each mixture using Samtools mpileup (version 1.6; Li et al., 2009) and BCFtools call -c (version 1.6; Li et al., 2009) with the flag to remove indels on (-I), and limiting read depth to a maximum of 100,000 reads. Levels of heterozygosity were estimated from the consensus sequences with vcf2allelplot.pl (Bensasson et al 2018) using the default options. Consensus genome sequences in vcf were converted to fasta format using the vcfutils.pl vcf2fq function (available with Samtools) and the seqtk seq function, so that each mixture's consensus bases with a Phred-scaled quality score below 40 were counted as missing data (converted to N).

Phylogenetics

For each level of contamination, contaminated strains were aligned together with strains previously selected. Ambiguous and low-quality bases were converted to 'N' to conform with requirements of the software RAxML. Lengths of chromosome sequences were checked and filled in with Ns at the ends if they were not the same length using bespoke BioPerl scripts (v.1.7.2). Maximum likelihood (ML) trees were created using PHYLIP (v3.697) and RAxML (v8.2.11 for *S. cerevisiae* and 8.2.12 for *A. fumigatus*); Stamatakis 2014) with 100 bootstrap replicates. Neighbor-joining (NJ) trees were created using MEGA (v10.0.5; Stecher et al. 2020) with the Tamura-Nei model (Tamura & Nei 1993) 100 bootstrap replicates. Trees were visualized using the Interactive Tree of Life (Letunic and Bork, 2024).

Admixture analysis

Allele frequency plots showing percent shared ancestry between strains were constructed for previously-selected *S. cerevisiae* sequences using the methods described below. Variant call format (VCF) files were zipped, indexed, and merged using BCFtools (v1.6; SAMtools v.1.6; Li et al., 2009). Chromosome names were replaced, and mitochondrial sequences were removed in

merged VCF files. Variant sites were retrieved using BCFtools and thinned to 1 single nucleotide polymorphism (SNP) for every 2000 nucleotides using VCFtools (v0.1.16-Gcc-8.3.0-Perl-5.30.0). SNPs were filtered with VCFtools and PLINK (v1.9b_5-x86_64; Purcell et al., 2007) so that only those with a minimum Phred-scale quality score of 40 were included. Statistics for each file were coded using BCFtools, and files were converted to Browser Extensible Data format for use in ADMIXTURE (v.1.3.0; Alexander et al. 2009). Analyses were run using subpopulation (K) values of K=2 through K=20 in replicates of 5. Log-likelihood for each replicate within each K value was taken from output files and was plotted using R (+ RStudio). Allele frequencies for selected runs were clustered using the program CLUMPAK (v1.1; Kopelman et al. 2015). Plots were visualized using R.

Results and Discussion

B-allele frequency plots can be used to detect contamination

Plotting B-allele frequency can be used to visualize ploidy in fungi (Bensasson 2018). We plotted haploid isolates of *S. cerevisiae* (Figure 4.1A, 0% column) and *A. fumigatus* (Figure 4.1E, 0% column) and heterozygous polyploids of *S. cerevisiae* (Figure 4.1B-D, 0% column). As expected, all haploids showed BAF values of 1.0 and polyploids showed values between 0 and 1.0. Though only one chromosome is shown for each, the allele frequency patterns seen in Figure 4.1 were consistent across all chromosomes.

We wanted to understand the effects of contamination on phylogenetic and population genomic analyses in organisms of varying ploidies, so we first created *in silico* contaminated strains in both species. To do so, subsampled fastq sequences from strains eAF163 (*A. fumigatus*) and CLIB219.2b (*S. cerevisiae*), subsequently referred to as donors, were used to contaminate subsampled haploid strains eAF749 (*A. fumigatus*) and CBS1479 (*S. cerevisiae*), and heterozygous diploid, triploid, and tetraploid *S. cerevisiae* strains (DBVPG1074, NPA05a1, UCD_06-645, respectively), subsequently referred to as recipients, at various percentages to a read depth of approximately 50. For example, to make a 5% contaminated *A. fumigatus* strain with a read depth of 50 the total number of reads needs to be 6 million. 300,000 reads were subsampled from eAF163 (5% donor reads), 5,700,000 reads were subsampled from eAF749 (95% recipient reads), and all reads were concatenated into one strain file.

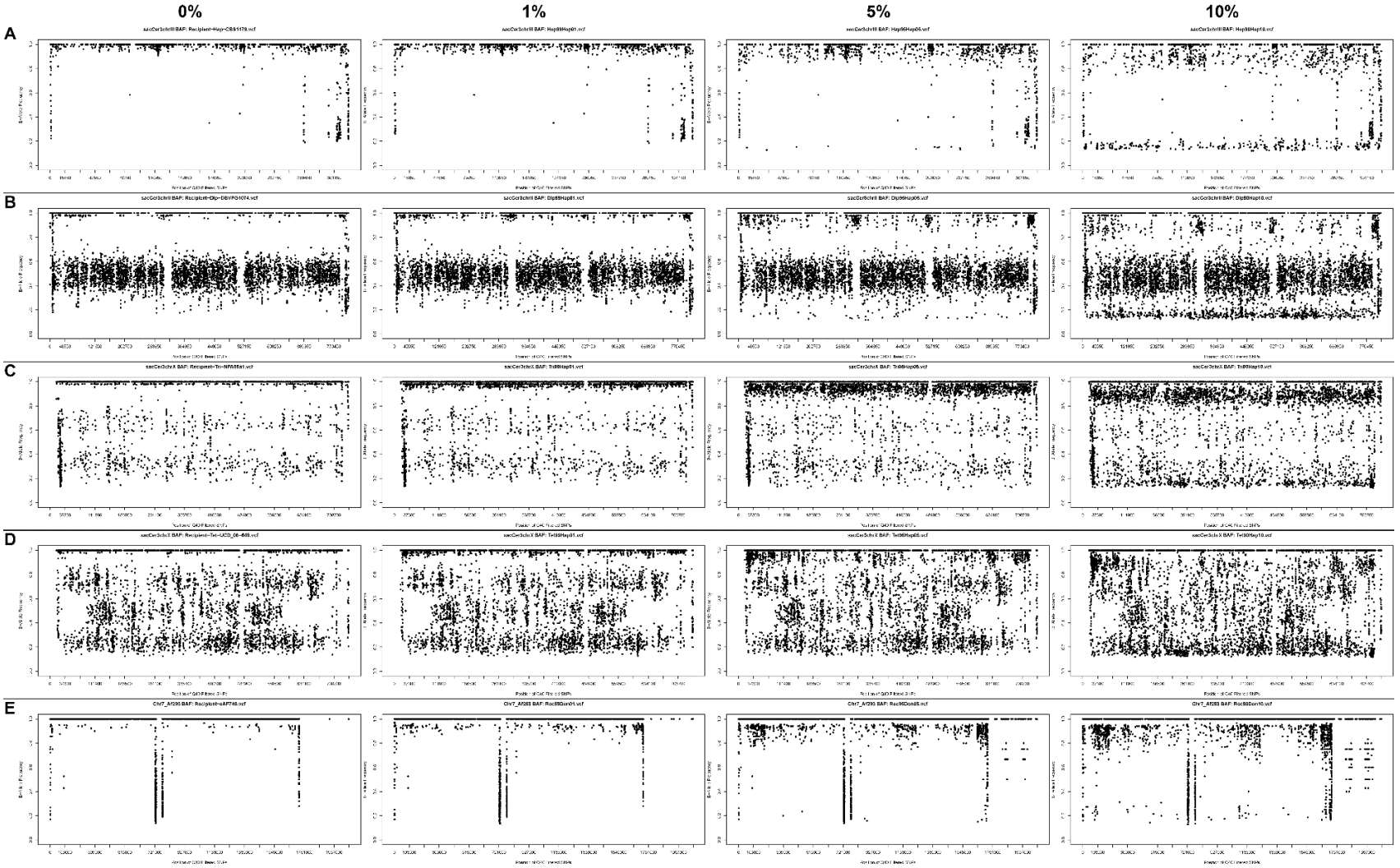

Figure 4.1. BAF plots can be used to identify contamination. BAF plots of 0%, 1%, 5% and 10% contamination of (A) chromosome 3 of the haploid recipient, (B) chromosome 2 of the diploid recipient, (C) chromosome 10 of the triploid recipient, and (D) chromosome 10 of the tetraploid recipient of *S. cerevisiae* strain mixtures and (E) chromosome 7 of the *A. fumigatus* mixture at 0, 1, 5, and 10% contamination.

In the *S. cerevisiae in silico* mixtures, the BAF plots are visibly different from the pure recipient strains at 5% or higher contamination (Figure 4.1). At 10% contamination the ploidy states of the haploid and diploid recipients remain resolvable however the triploid recipient looks tetraploid, and the ploidy state of the tetraploid would be difficult to discern. Consistent with *S. cerevisiae*, the BAF plots of contaminated *A. fumigatus* appear different from the 0% contaminated recipient at 5%. Across all mixtures for both species, the signature of contamination in BAF plots is a band of allele frequencies slightly less than 1.0. By recapitulating the contamination likely present in some publicly available whole genome sequences, we can test for what level of contamination affects phylogenetic and population genomic analyses.

Contamination alters whole-genome tree topology

Since both *S. cerevisiae* and *A. fumigatus* have well-defined lineages (Peter et al., 2018, and Celia-Sanchez et al., 2024), we wanted to understand how contamination might impact phylogenetic analysis. To determine the effect of contamination on phylogenetic tree building, 52 strains from both species were chosen to create whole genome species trees using both maximum-likelihood and neighbor-joining methods and a single strain from that group was contaminated *in silico* for both species. Maximum likelihood tree-building was less sensitive to contamination than neighbor-joining. 10% contamination was enough to change tree topologies for both species using a maximum likelihood approach (Figures 4.3 and 4.4). Using neighbor-joining methods, 1% contamination was enough to alter bootstrap support in the *A. fumigatus* whole-genome tree (Figure 4.6). 5% contamination altered which isolates were sister to the recipient strain, and 10% contamination and higher drastically altered tree topology for both species using neighbor-joining methods (Figures 4.5 and 4.6). In the event of contamination of

one strain by another within the same study, the two strains may appear more closely related than they should. Depending on the method used between 1% and 10% contamination in one strain could muddy inferred phylogenetic relationships within a species. Further research will compare the effects of keeping the donor strain within the whole genome trees in *A. fumigatus* and the effects, if any, of using the same software versions for both species.

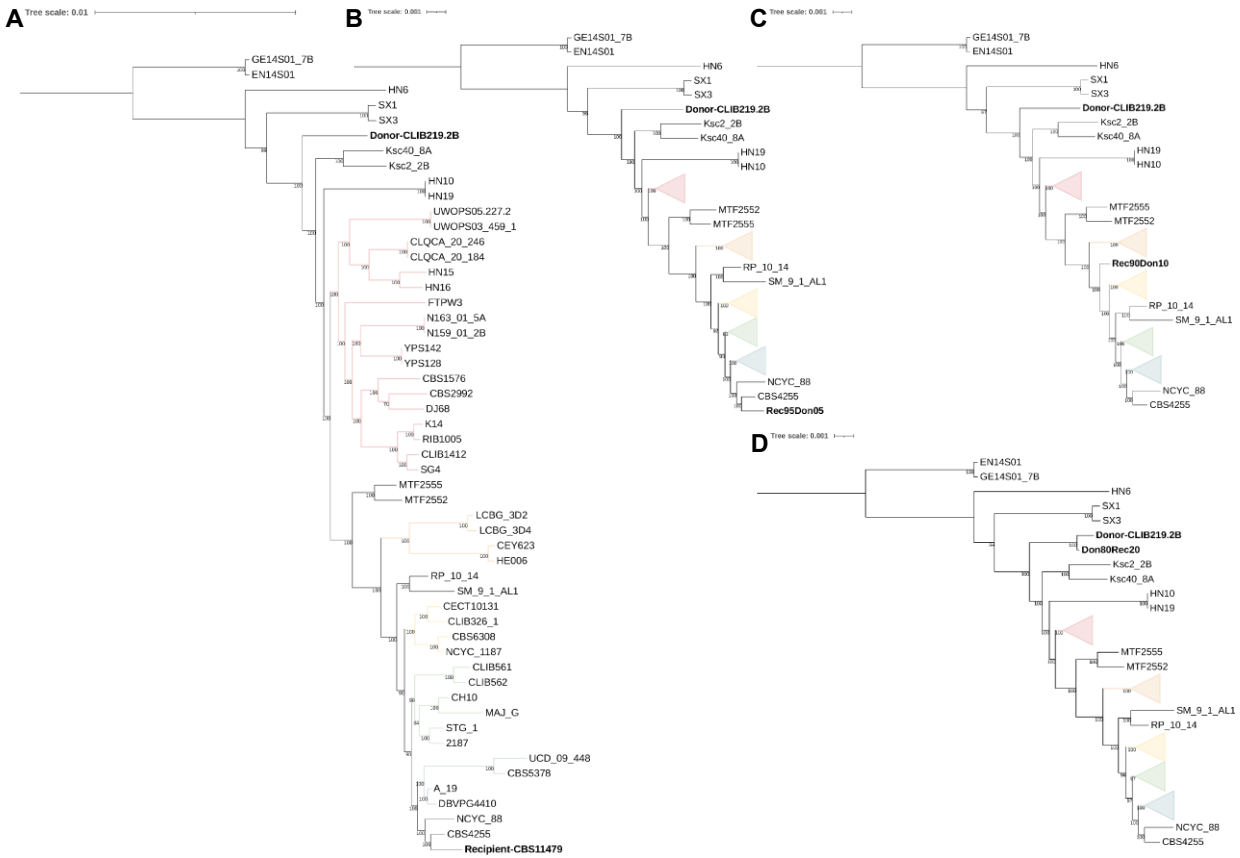

Figure 4.2. Contamination alters maximum likelihood trees of *S. cerevisiae*. The haploid *S. cerevisiae* recipient strain (CBS11479) was contaminated with the haploid donor strain (CLIB219.2B) at varying percentages. (A) No contamination. (B) 5% contamination. (C) 10% contamination shifts the position of the mixed recipient strain, suggesting a different ancestry than in A. (D) 20% contamination. The recipient and donor strains are shown in bold. Members of the collapsed nodes in B-D are represented by the colored branches in A.

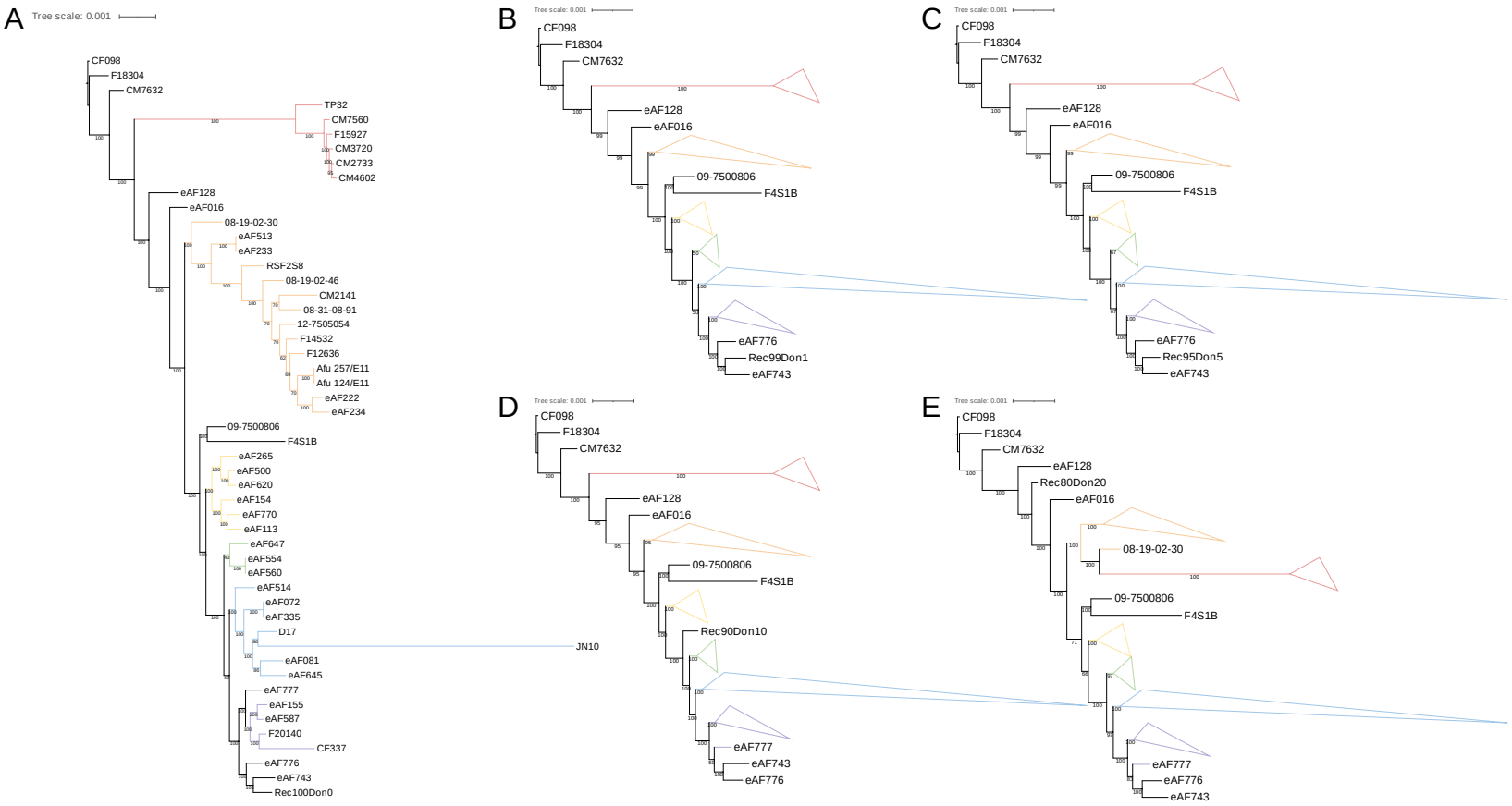

Figure 4.3. Contamination alters maximum likelihood trees of *A. fumigatus*. The recipient strain eAF749 was contaminated by the donor strain eAF163 at varying percentages. (A) 0% contamination. (B) 1% contamination. (C) 5% contamination. (D) 10% contamination. (E) 20% contamination. Arrows indicate the location of the mixture strain in the tree. Members of the collapsed nodes in B-E are represented by the colored branches in A.

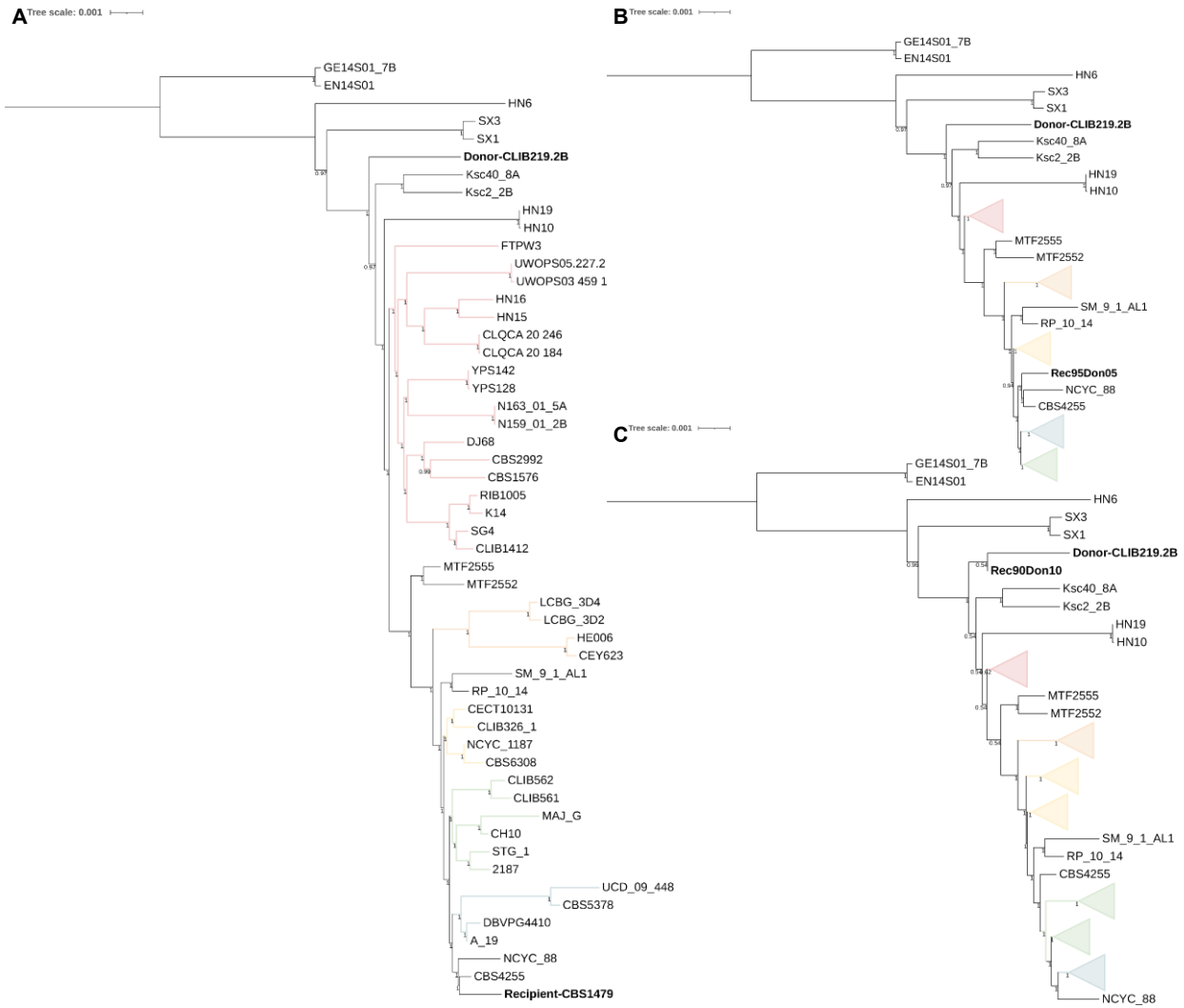

Figure 4.4. Contamination alters neighbor-joining trees of *S. cerevisiae*. The haploid *S. cerevisiae* recipient strain (CBS11479) was contaminated with the haploid donor strain (CLIB219.2B) at varying percentages. (A) 0% contamination. (B) 5% contamination. (C) 10% contamination. Arrows indicate the location of the mixture strain in the tree. Members of the collapsed nodes in B-C are represented by the colored branches in A.

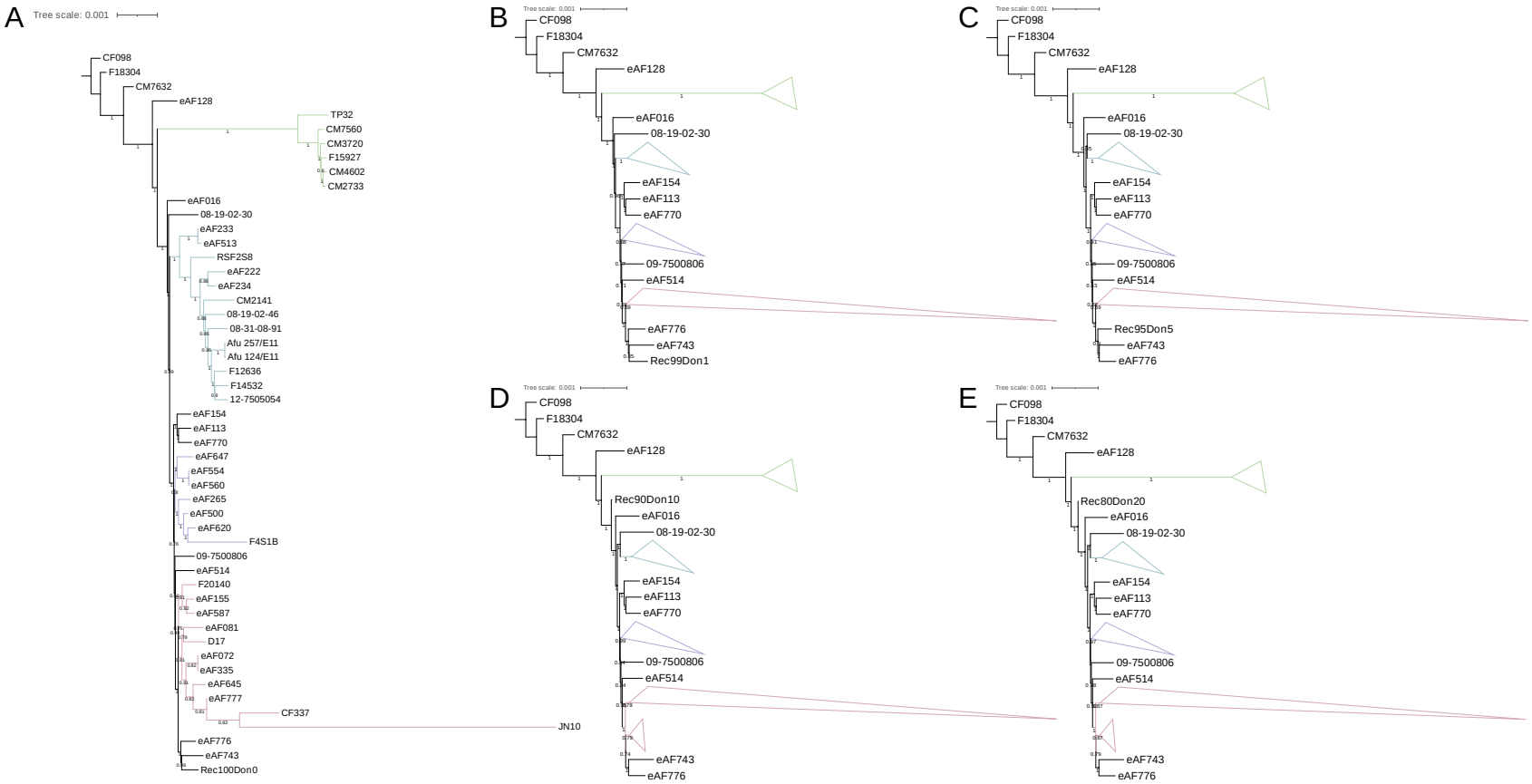

Figure 4.5. Contamination alters neighbor-joining trees of *A. fumigatus*. The recipient strain eAF749 was contaminated by the donor strain eAF163 at varying percentages. (A) 0% contamination. (B) 1% contamination. (C) 5% contamination. (D) 10% contamination. (E) 20% contamination. Arrows indicate the location of the mixture strain in the tree. Members of the collapsed nodes in B-D are represented by the colored branches in A.

Contaminated strains appear as hybrids in admixture analysis

S. cerevisiae is well known to have distinct lineages of related populations and gene flow between these populations can be detected by ADMIXTURE analysis (Purcell et al., 2007). To determine the effect of contamination on ADMIXTURE analysis, 52 strains from *S. cerevisiae* and *in silico* contaminated strains ranging from 1-10% were chosen for ADMIXTURE analysis at K=2 through K =20. We found, once again, between 5 and 10% contamination alters results (Figure 4.8). 10% contamination led to the recipient appearing as a hybrid between the original donor and recipient strains. This is true for all values of k tested (k=2 through k=20).

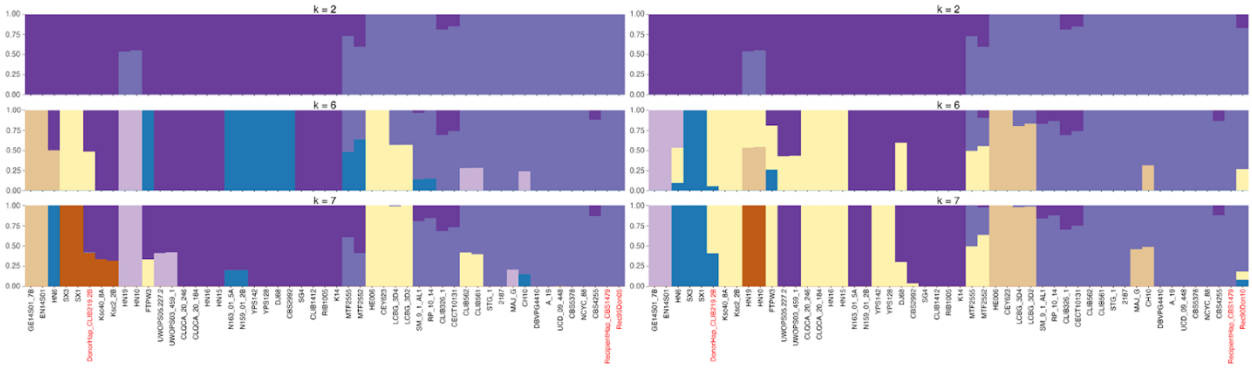

Figure 4.6. 10% contamination disrupts admixture analyses in *S.cerevisiae*. Regardless of k value (k=2-20), 10% contamination defines the recipient into at least two populations instead of the original population. Left: k=2,6&7 for the 5% contamination. Right: k=2,6&7 for the 10% contamination. Red print denotes the donor, recipient, and contaminated mixture strains.

We used BAF plots to screen isolates of *A. fumigatus* for contamination. We found two strains that had non-haploid B-allele frequencies (greater than 0.0 and less than 1.0) (Figure 4.7). Given that the BAFs in Figures 4.7A and 4.7B suggest decaploidy and pentaploidy, respectively, and since no polyploid *A. fumigatus* isolates have been reported, we assumed these strains were contaminated and omitted them from our downstream analysis.

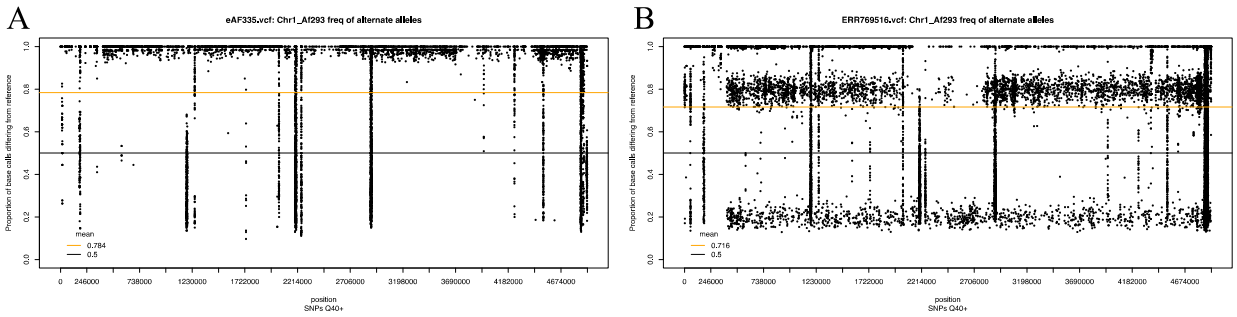

Figure 4.7. Plotting B-allele frequency suggests contamination in publicly available haploid genomes. BAF plots of *A. fumigatus* chromosome one from (A) eAF355 [PRJNA742769] and (B) Afu 343/P11 [PRJEB8623].

We suggest other researchers consider incorporating a qualitative check for contamination using BAF plots because it uses already existent variant call files generated in SNP calling, and takes only a few minutes to run. This is especially important for fungal researchers as yeasts can easily stick to working surfaces and filamentous fungal spores easily aerosolize.

We show here that BAF visualization is a viable method for detection of intra-species contamination and that contamination can affect population genetic and phylogenetic analyses. After 5% contamination BAF plots reliably show a preponderance of contaminated allele frequencies, the same is enough to alter tree topologies, and 10% contamination makes recipient strains appear as hybrids between the donor and recipient. Together this shows low levels of contamination increase the likelihood of placing a strain in an incorrect evolutionary niche across two vastly different fungal genomes.

References

Alexander, D.H., Novembre, J. and Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome research, 19(9), pp.1655-1664.

Ballenghien, M., Faivre, N. & Galtier, N. Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. *BMC Biol* 15, 25 (2017). https://doi.org/10.1186/s12915-017-0366-6

Bensasson, D., J. Dicks, J. M. Ludwig, C. J. Bond, A. Elliston et al., 2019 Diverse Lineages of Candida albicans Live on Old Oaks. *Genetics* 211: 277-288.

Bensasson, D. 2018. bensassonlab/scripts: Code for yeast population genomic analyses. DOI: 10.5281/zenodo.1488147. https://doi.org/10.5281/zenodo.1488147

Celia-Sanchez, B., Mangum, B., Gomez, L.L., Wang, C., Shuman, B., Brewer, M.T. and Momany, M., 2023. Pan-azole-and multi-fungicide-resistant Aspergillus fumigatus is widespread in the United States. bioRxiv, pp.2023-12.

Clark, M.A., Stankiewicz, S.H., Barronette, V. and Ricke, D.O. 2019, Letter to the Editor – Detecting HTS Barcode Contamination. *J Forensic Sci*, 64: 961-962. https://doi.org/10.1111/1556-4029.14027

Cornet, L., Baurain, D. (2022). Contamination detection in genomic data: more is not enough. *Genome Biol* 23, 60. https://doi.org/10.1186/s13059-022-02619-9

Felsenstein J. 1989. Phylogeny inference package (version 3.2). *Cladistics*. 5:164–166.

Fierst, J.L. and Murdock, D.A., 2017. Decontaminating eukaryotic genome assemblies with machine learning. *BMC bioinformatics*, *18*, pp.1-16.

Goig, G.A., Blanco, S., Garcia-Basteiro, A.L., Comas, I. 2020. Contaminant DNA in bacterial sequencing experiments is a major source of false genetic variability. *BMC Biol* 18, 24.

https://doi.org/10.1186/s12915-020-0748-z

Jenks, G.F. 1967. The data model concept in statistical mapping. *Int Year of Cart* , 7, 186–190.

Kang, S.E., Sumabat, L.G., Melie, T., Mangum, B., Momany, M., Brewer, M.T. (2022) Evidence for the agricultural origin of resistance to multiple antimicrobials in Aspergillus fumigatus, a fungal pathogen of humans. *G3*. 12(2). https://doi.org/10.1093/g3journal/jkab427

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. and Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular ecology resources, 15(5), pp.1179-1191.

Letunic, I. and Bork, P., 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Research, p.gkae268.

Li, H. 2012. seqtk Toolkit for processing sequences in FASTA/Q formats. GitHub, 767, p.69.

Li, H., and R. Durbin, 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinf* 25: 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinf* 25: 2078-2079.

Merchant S., Wood D.E., Salzberg S.L. 2014. Unexpected cross-species contamination in genome sequencing projects. *PeerJ* 2:e675 https://doi.org/10.7717/peerj.675

Nierman, W.C., Pain, A., Anderson, M.J., Wortman, J.R., Kim, H.S., Arroyo, J., Berriman, M., Abe, K., Archer, D.B., Bermejo, C. and Bennett, J., 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. *Nature*, 438(7071), pp.1151-1156.

Peter, J., De Chiara, M., Friedrich, A., Yue, J.X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A. and Cruaud, C., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature*, 556(7701), pp.339-344.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. and Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The Am Journ of Hum Gene*, 81(3), pp.559-575.

Scopel, E.F.C., Hose, J., Bensasson, D., Gasch, AP. (2021). Genetic variation in aneuploidy prevalence and tolerance across Saccharomyces cerevisiae lineages. *Genetics*. 217(4). https://doi.org/10.1093/genetics/iyab015

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinf*, 30(9), pp.1312-1313.

Stecher, G., Tamura, K. and Kumar, S., 2020. Molecular evolutionary genetics analysis (MEGA) for macOS. Molecular biology and evolution, 37(4), pp.1237-1239.

Tamura, K. and Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Bio and Evo*, 10(3), pp.512-526.

Wilson, G.C., Nowell, R.W., and Barraclough, T.G. (2018). Cross-Contamination Explains "Inter and Intraspecific Horizontal Genetic Transfers" between Asexual Bdelloid Rotifers. *Curr Biol*. 28(15), 2436-2444. https://doi.org/10.1016/j.cub.2018.05.070

CHAPTER 5

CONCLUSION

Chapter 2 outlined the history of septin classification and revisions to our understanding of septin group origins. Additionally, this chapter compared the septin-conserved sequences from the model organisms with the most septin research. Finally, this chapter posited the concept of positional orthology as it related to septin heteropolymers.

Septins from Momany groups 1-4 form heteropolymers. One member from each group occupies a position within a hetero-hexamer or hetero-octamer. It is intriguing that some members of larger septin clades occupy the same relative position within the heteropolymer. Group 1a septins (cdc10 in *S. cerevisiae* and Sept3 in *H. sapiens*) form the central dimer of the octamer and lack coiled coils. Members of Group 2 form the central dimer of the hexamer (Group 2a, cdc3, and Group 2b-7, Sept-7). Septins that occupy the remaining two positions within the heteropolymer are not monophyletic. Past and current trees do not place Group 1b with Group 4, nor do they place Group 2b-2 with Group 3.

In the ancestor of animals, the Group 2 septin ancestor appears to have duplicated creating both Sept2 and Sept7 orthologs, and the Group 1 ancestor duplicated creating both Sept6 and Sept3 orthologs. The origins of Groups 3 and 4 remain unclear.

Septin group member copy numbers hint at future research. In organisms that have more than one member in each group, the septin groups that have the most members are the groups that occupy the two positions furthest away from the homodimers (Groups 1b and 2b-2 in animals and Groups 3 and 4 in fungi). Further research into the origins of septins in the

opisthokont ancestor may reveal whether the interface residues of the outer two monomers have been under less strict selection and may reveal the mechanism by which septins have duplicated and diverged in this lineage.

Chapter 2 provides a "Rosetta Stone" for septin researchers across model organisms to use when comparing research on individual monomers. This review arose after numerous discussions with fellow researchers where we would go back and forth checking to make sure we were referencing the same septin homolog within a heteropolymer. Copious copies of a printed and laminated Figures 2.1 and 2.2 were brought to a septin research conference which allowed attendees to quickly note a septin being described and reference back to their organism of interest. I believe this chapter serves the septin field by clarifying cross-species protein comparisons especially since ancestry does not always match up with the functional equivalent in the heteropolymers.

During my defense, it was brought to my attention that the term "positional orthology" we coined can be misleading. "Positional orthology" was originally coined to remind the field that septins closest in ancestry do not always equate to the positions they occupy in the heteropolymer. Orthology refers to genes which are direct descendants from a single gene in their common ancestor, whereas parlogy refers to genes that share a common ancestor but have undergone duplication (Koonin, 2005). It seems neither of these fully explain the evolutionary relationships of the septins, and that across the septin gene family there are orthologs, co-orthologs, in-paralogs, and out-paralogs. Should a descriptor be necessary for future comparisons between kingdoms, perhaps the term positional homology would better encapsulate the variety of evolutionary relationships within the septin gene family. Further characterization of the major

septin gene duplications and deletions across eukaryotes may give insight into how the different septin monomer orders within heteropolymers arose.

While demonstrating the utility of Momany septin Groups 1-4, this Chapter 2 overfit the model of septin groupings to include non-opisthokont septins with those from opisthokonts, namely in Group 5. Chapter 3 identified new non-opisthokont septins, reclassified non-opisthokont septins into distinct phylogenetic groups, and used ancestral sequence reconstruction to predict the structure and function of septins at historic expansion nodes.

Expansion of the list of known septins has further suggested their presence in the last eukaryotic common ancestor. The new collection of septins has allowed for the expansion of Momany Groups 1-5 to include Onishi Groups 6A, 6B, 7, and 8. Group 8 is a monophyletic group separate from the monophyletic group containing 6A, 6B, and 7, and is unique in that it is only comprised of sequences from ciliates. Groups 6A and 6B are each monophyletic groups present within the 6A-6B-7 clade, and species may have multiple septins within each of these subgroups similar to *H. sapiens* and *S. cerevisiae*. Group 7 is the grade including all members not in Group 6. Whether these septins form heteropolymers and in what order, if any, remains to be determined.

When septins were first classified in mammals by Kinoshita, only four homology groups were created until more sequences were incorporated (Kinoshita, M. 2003). In Onishi groups 6A and 6B, multiple species have at least two septins in each group suggesting the 6A and 6B septin ancestors were each duplicated before speciation events leading to numerous lineage-specific septins. Take *S. obliquus* for example, this alga has two Group 6A septins and five Group 6B septins. In Chapter 2, I modified the Momany Group 2B to include the Kinoshita Sept2/Sept7 distinction. This is crucial because it combines the functional distinction of the Kinoshita

groupings, with the larger opisthokont septin evolutionary context. This highlights the necessity of future septin groupings to incorporate evolutionary history and function. Since biochemical and genetic evidence is lacking in these algae, I will list three possibilities for the implications of these duplications in reverse order of likelihood: (1) All of these septins act as monomers like in *C. reinhardtii*. (2) The ancestors of Groups 6A and 6B formed a dimer (which may or may not repeat into larger heteromers) and the duplicated members of their respective groups interchange depending on the specific function. In opisthokonts this is seen in the vast expansion of *H. sapiens* Groups 1A, 1B, and 2B-2 and in the meiotic septins in *S. cerevisiae*. (3) The duplications in Groups 6A and 6B have allowed distinct functional groups to diverge and heteropolymers similar to those in opisthokonts form. I predict that as more protist genomes are sequenced, septins are identified, and functional analyses are conducted the Onishi Groups will expand to show yet more diversity in the non-opisthokont septins.

Many non-opisthokont septins were found to have an arginine finger that appears crucial for G-interface homodimerization in *C. reinhardtii*. Loss of the R-finger was only seen in species that contained other septins retaining the R-finger. We hypothesized that septin duplication followed by the loss of the R-finger from the G-interface of the duplicate and gain of the α-0 helix made available the NC-interface for heteropolymerization. Algae with one septin in either Group 6A or 6B might be best suited for studying the establishment of the NC-interface to create homopolymers and after synthetic duplications, possibly even heteropolymers. Genetic and biochemical assays generating and testing mutants in these regions will be essential to understand the origins of heteropolymerization.

Chapter 4 uses an available B-allele frequency visualization pipeline to identify intra-species contamination. This compensates for a gap left by previous contamination detectors

which could identify inter-species contamination mostly in prokaryotes, and provides a fast method to qualitatively check for contamination. We found between 5 and 10% contamination was enough to disrupt phylogenetic and population genetic analysis in two species by making the contaminated strains appear as hybrids of the two mixed individuals. Detecting contamination is especially important for fungal and other microbial researchers as intra-species contamination can happen when working with more than one strain at a time. Further research is needed to examine how intra-species contamination affects other analyses.

We have provided fungal researchers with a simple, visual method to check their genomes for contamination. Creating b-allele frequency plots that use variant call files as their input streamlines this analysis as vcf files are generated when calling snps. Generating these plots for researchers working with fungal genomes especially serves as a quality control measure as these microbes are more easily contaminated than larger plants or animals.

REFERENCES

Kinoshita, M. 2003. Assembly of Mammalian Septins. *J. Biochem*. 134 (4), 491–496. doi:10.1093/jb/mvg182

Koonin, E. 2005. Orthologs, Paralogs, and Evolutionary Genomics. *Annu Rev Gen*. 39:309–38. doi: 10.1146/annurev.genet.39.073003.114725