

ABSTRACT

PHYLOGENETIC ANALYSIS, MODELING AND EXPERIMENTAL
STUDIES OF THE *SACCHAROMYCES CEREVISIAE*
PALMITOYLATED PROTEIN KINASE
GENE, *ENV7*

By

Stephanie M. Cocca

May 2014

Env7 is a vacuole membrane-localized protein kinase that is orthologous to the human serine/threonine protein kinase, STK16. It is evolutionarily well-conserved throughout Eukarya, and it has one ortholog in Bacteria. Phylogenetic analyses of sequences homologous to *Env7* revealed clades that are inconsistent with established eukaryotic phylogeny, suggesting that both horizontal and vertical gene transmission are responsible for their conservation. Conserved amino acid residues and motifs that are potentially important to *Env7*'s catalytic activity, localization, and interactions with other proteins were also identified and assessed. Additionally, one such conserved motif—the glycine-rich loop—was mutated in an effort to affect ATP binding in *Env7*. The phenotype resulting from this mutation was a slightly increased number of mutant cells exhibiting multi-lobed vacuoles under normal conditions.

PHYLOGENETIC ANALYSIS, MODELING AND EXPERIMENTAL
STUDIES OF THE *SACCHAROMYCES CEREVISIAE*
PALMITOYLATED PROTEIN KINASE
GENE, *ENV7*

A THESIS

Presented to the Department of Biological Sciences
California State University, Long Beach

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Microbiology

Committee Members:

Editte Gharakhanian, Ph.D. (Chair)
Judy Brusslan, Ph.D.
Simon Malcomber, Ph.D

College Designee:

Brian T. Livingston, Ph.D.

By Stephanie M. Cocca

B.S., 2008, California State University, Long Beach

May 2014

UMI Number: 1527907

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 1527907

Published by ProQuest LLC (2014). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Copyright 2014

Stephanie M. Cocca

ALL RIGHTS RESERVED

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. E. Gharakhanian and Dr. S. Manandhar for their excellent guidance and endless patience. This thesis would have been impossible if it were not for their unwavering support. I would also like to thank my committee members: Dr. S. Malcomber for his help with the phylogenetic sections of this thesis, and Dr. J. Brusslan for her astute observations and suggestions. Additionally, I would like to thank Dr. E. Sorin for sharing his knowledge about protein structure and modeling, and Dr. N. Nikolaidis for his assistance with the selection analyses.

Thank you to the former and current G-lab members—Maribeth, Florante, Ikha, Surya C., Daniele, Erika, Priscilla, Fiona, Kris, Teli, Lisa, Julie, Jonathan, Venessa, and, most definitely, Chris Frost—who lent me media in times of desperation, advice in times of confusion, and a sympathetic ear when I really needed one.

Finally, I am profoundly grateful to my parents, my fiancée, Ramin, and the rest of my family. Their selfless support, encouragement, and reassurance saw me through this degree process.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	x
CHAPTER	
1. INTRODUCTION	1
Why Study Cellular Trafficking? Human Lysosomal Storage and Neurodegenerative Diseases	1
<i>Saccharomyces cerevisiae</i> : A Model Organism for the Study of Human Disease	3
Discovery of the Late Endosome and Vacuole Interface Genes.....	4
<i>ENV7</i>	6
<i>STK16</i>	7
Eukaryotic Protein Kinases: Divisions and Subdomains.....	8
The Current Placement of <i>ENV7</i> in the Eukaryotic Protein Kinase Superfamily.....	10
Major Aims of this Study.....	11
2. MATERIALS AND METHODS.....	12
Protein Sequence Selection for Phylogenetic Analyses.....	12
Multiple Sequence Alignment	13
Determination of Evolutionary Models	13
Maximum Likelihood Tree Estimation.....	14
Bayesian Tree Estimation	14
Synteny Analysis	15
Tests for Evolutionary Selection.....	15
Generation and Manipulation of Tertiary Protein Structures	16
Phosphorylation Site Prediction.....	16
Yeast and Bacterial Strains	17
Competent Yeast Cells.....	17
Media and Reagents	17

CHAPTER	Page
Plasmids and Primers.....	19
Site-Directed Mutagenesis.....	19
Gel Purification of PCR I, II and III Products from Agarose Gel.....	20
Restriction Enzyme Site Analysis.....	20
Transformation of <i>env7Δ</i> Cells.....	22
Colony PCR.....	23
Yeast DNA Extraction.....	24
Amplification of Mutagenized Plasmid in DH5α <i>E. coli</i> Cells.....	24
Sequencing of Point-Mutated Plasmids.....	25
FM4-64 Staining and Microscopy.....	25
Liquid Growth Assays.....	26
Solid Growth Assays.....	26
Subcellular Fractionation.....	27
Protein Quantification.....	28
SDS-PAGE and Western Blotting.....	28
Phosphorylation Upshift Assay.....	30
 3. RESULTS.....	 32
Evolutionary Conservation of <i>ENV7</i>	32
The Conservation Pattern of the Env7 Orthologs Contradicts Established Monophyly.....	32
The N-Terminal Consecutive Cysteine Motif is Well-Conserved in Most Fungal and Metazoan Orthologs of Env7.....	35
Many Closely-Related Orthologs of Env7 Do Not Share the Disordered Glycines in the Glycine-Rich Loop that are Characteristic of NAK Kinases.....	38
No Discernable Pattern of Divergent Activation Segment APE Motifs Exists Among Env7 and its Orthologs.....	38
The Activation Segment Anchor DLG Motif is Conserved in Nearly All Saccharomycotina, Pezizomycotina, Hexapoda and Vertebrata Env7 Orthologs.....	39
The Canonical HxN Motif is Absent in Env7 and Most of its Fungal Orthologs.....	39
No Detectable Syntenic Conservation Exists Near <i>ENV7</i> Beyond its Very Close Fungal Orthologs.....	40
Individual Kinase Subdomain Analyses of Env7 Revealed Conserved Features and Evidence of Purifying Selection.....	43
The Relationship of <i>ENV7</i> to the NAK Kinase Family Members.....	44
Important Sequence and Structural Aspects of Env7.....	48
Well-Conserved Residues in Env7 are Appropriate to Form Salt Bridges Characteristic of Active Eukaryotic Protein Kinases.....	48

CHAPTER	Page
Env7's Catalytic Spine, Regulatory Spine, Magnesium-Positioning Loop, and Activation Segment are Composed of Conserved Residues	51
The Unique Subdomain VIA Insert in Env7.....	54
Proposed Phosphorylation Sites in Env7	56
Env7's Disordered Glycine-Rich Loop	58
Experimental Assessment of the Significance of Glycine 40 in the Glycine-Rich Loop.....	60
Generation of the Env7G ₄₀ V-HA Mutant	63
Analysis of Vacuole Morphology in Env7-HA, Env7G ₄₀ V-HA, and Env7C ₁₃₋₁₅ S-HA-Expressing Cells.....	64
 4. <i>ENV9</i> , <i>ENV10</i> , AND <i>ENV11</i> ANALYSES	 69
Phylogenetic Analysis of <i>ENV9</i>	69
Characteristics of <i>ENV9</i> and <i>RDH12</i>	69
Conservation of Residues Associated with Human Disease.....	69
Env9 is Evolutionarily Conserved in Every Domain of Life.....	71
Phylogenetic Analysis of <i>ENV10</i>	73
Phylogenetic Analysis of <i>ENV11</i>	77
 5. DISCUSSION	 82
A Combination of Vertical and Horizontal Gene Transfer is Likely Responsible for the Conservation of <i>ENV7</i> Orthologs in Eukarya and Bacteria	82
The NAK Kinase Family is Still Not Well-Defined.....	84
<i>ENV7</i> Orthologs May Rely on Different Forms of Activation Segment Regulation	85
The Subdomain VIA Insert in Env7 May Play a Role in its Localization and/or Interaction with Other Proteins.....	86
The Vacuolar Morphology of the Env7G ₄₀ V-HA Mutant is Slightly Aberrant	89
Promising Target Residues for Further Point Mutation in <i>ENV7</i>	90
 APPENDICES	 92
A. DETAILS REGARDING PHYLOGENETIC AND BIOINFORMATIC METHODS	93
B. DETAILED LISTS OF MEDIA AND REAGENTS USED IN THIS STUDY	97
C. ADDITIONAL FIGURES	103

APPENDIX	Page
D. POTENTIAL TARGETS FOR MUTATION IN <i>ENV7</i>	115
REFERENCES	122

LIST OF TABLES

TABLE	Page
1. Summary of Eukaryotic Protein Kinase Subdomains.....	9
2. Microbiological Strains Used in this Study	18
3. Primers Used for Site-Directed Mutagenesis.....	18
4. Thermocycler Settings for PCR I, II and III	21
5. Thermocycler Settings for Colony PCR	21
6. Individual Vacuole Morphology Counts for Cells Grown in YPD	68
7. Individual Vacuole Morphology Counts for Cells Grown in 0.5M NaCl YPD	68
8. Programs, Settings, and References for Multiple Sequence Alignments	94
9. Evolutionary Model Selection for Multiple Sequence Alignments.....	95
10. Settings and Average Standard Deviation of Split Frequencies (ASDSF) for Bayesian Analyses	95
11. Programs Used to Determine Possible Phosphorylation Sites in Env7	96
12. <i>Saccharomyces cerevisiae</i> Media Used throughout this Study.	98
13. <i>Escherichia coli</i> Media Used for Site-Directed Mutagenesis and Plasmid Preps	99
14. Subcellular Fractionation Buffers and Solutions.	99
15. Preparation of BSA Standards for Protein Quantification	99
16. Preparation of Large and Small 7.5% SDS-PAGE Gels	100
17. 6x SDS-PAGE Sample Buffer (6x Laemmli Buffer)	100

TABLE	Page
18. 5x SDS-PAGE Running Buffer	100
19. Western Transfer Buffer	101
20. Tris-Buffered Saline (TBS).....	101
21. TWEEN [®] 20 + Tris-Buffered Saline (TTBS).....	101
22. TTBS + Milk (5% TTBSM)	101
23. 5x Tris-Borate-EDTA	101
24. Poly-Lithium-Acetate-Tris-EDTA (PLATE) Mixture.....	102
25. Phosphorylation Upshift Assay Reagents.....	102
26. Potential Target Amino Acid Residues in Env7 for Further Analysis.....	116

LIST OF FIGURES

FIGURE	Page
1. Five major cellular trafficking pathways in <i>S. cerevisiae</i>	5
2. A Bayesian phylogenetic tree depicting Env7 and its fungal homologs	34
3. A Bayesian phylogenetic tree depicting homologs of Env7 from Eukarya and Bacteria	37
4. Genomic synteny is limited to <i>ENV7</i> and its close fungal orthologs.....	42
5. Comparisons of human and <i>S. cerevisiae</i> NAK Kinase protein sequences	45
6. An unrooted Bayesian phylogenetic tree of select NAK family protein sequences	47
7. An overview of the Env7 protein and its conserved residues and features	49
8. Three-dimensional rendering of conserved features in Env7	50
9. Three-dimensional rendering of Env7 featuring the putative conserved salt bridges.....	52
10. Three-dimensional rendering of Env7 featuring the catalytic and regulatory spines.....	52
11. Three-dimensional rendering of Env7 featuring the subdomain VIA insert ...	55
12. Prediction of putatively phosphorylated residues in Env7	57
13. Three-dimensional rendering of Env7 featuring the glycine-rich loop	59
14. Site-directed mutagenesis protocol used in this study	61
15. The 2 μ vector, pSMG17 + <i>env7G₄₀V-HA</i> , used in this study	62
16. Discovery and confirmation of mutant <i>S. cerevisiae</i> cell line used in this study.....	62

FIGURE	Page
17. FM4-64 staining and microscopy of Env7-HA, Env7G ₄₀ V-HA, and Env7C ₁₃₋₁₅ S-HA cells.....	67
18. Bar graphs displaying the percentages of the three vacuole morphologies observed in the Env7-HA, Env7G ₄₀ V-HA, and Env7C ₁₃₋₁₅ S-HA-expressing cells of <i>S. cerevisiae</i> incubated in 30°C.....	68
19. A pairwise sequence alignment of <i>H. sapiens</i> RDH12 and <i>S. cerevisiae</i> Env9 reveals conserved residues associated with human disease	70
20. A Bayesian phylogenetic tree of Env9 and 77 of its orthologs.....	72
21. Hidden Markov Model (HMM) Logo illustration of conserved amino acid .. residues in the DUF788 domain	75
22. An unrooted Bayesian phylogenetic tree of Env10 and 31 of its orthologs	76
23. An unrooted Bayesian phylogenetic tree of Env11 and Vid22 and their respective orthologs	81
24. An expanded view of the Bayesian phylogenetic tree depicting Env7 and its fungal orthologs	104
25. Individual kinase subdomain analyses of Env7 reveals conserved elements and evidence of negative (purifying) selection.	106
26. Conserved residues in Env7 based on an alignment with 151 of its orthologs from Eukarya and Bacteria.	110
27. Comparison of phylogenetic trees generated using complete and truncated NAK kinase sequences shows nearly the same topology	111
28. Preliminary growth curves for Env7-HA, Env7G ₄₀ V-HA, and Env7C ₁₃₋₁₅ S-HA –expressing strains in various liquid media reveal minor differences in growth pattern.....	112
29. Preliminary fitness analysis of WT-HA, Env7G ₄₀ V-HA, and Env7C ₁₃₋₁₅ S-HA expressing cells on various solid media does not show major growth differences	113
30. Preliminary analysis of subcellular localization and phosphorylation state of the Env7G ₄₀ V-HA mutant protein.....	114

CHAPTER 1

INTRODUCTION

Why Study Cellular Trafficking? Human Lysosomal Storage and Neurodegenerative Diseases

Nearly one in every 4000-7500 newborns worldwide is diagnosed with a lysosomal storage disease, often accompanied by meager odds for survival past childhood and a high likelihood of pain and suffering (Meikle et al., 1999; Poorthuis et al., 1999; Pinto et al., 2004; Poupetová et al., 2010). The lysosome is an acidic organelle in animal cells, and one of its major functions is to degrade cellular debris and other waste materials in order to keep the cell functioning properly. Lysosomal storage diseases arise from genetic mutations that either result in lysosomal malfunction or that cause materials typically destined for the lysosome to never reach the correct target (Ballabio and Gieselmann, 2009). As a result, cellular materials and wastes that are normally broken down in the lysosome accumulate in the cell, leading to tissue damage and even organ failure.

Lysosome-related diseases are certainly not limited to the young; many adult-onset neurodegenerative diseases have also been linked to lysosomal trafficking dysfunction. Recent research has revealed that autophagy—the lysosomal degradation of cellular contents in living cells—is impaired in the neuronal cells of individuals suffering from Huntington’s disease, Alzheimer’s disease, amyotrophic lateral sclerosis (ALS), and Parkinson’s disease (reviewed in Wong and Cuervo, 2010). For example, in

Alzheimer's-afflicted neurons, autophagic vesicles containing β -amyloid peptides fail to fuse with the lysosome, and this malfunction leads to the toxic buildup and aggregation of the peptides outside of the neuron (reviewed in Nixon et al., 2008; reviewed in Ułamek-Kozioł et al., 2013). Many other neurodegenerative lysosomal storage diseases follow a similar pattern of unsuccessful heterotypic fusion at the lysosome followed by harmful—and often lethal—accumulation of cellular materials in the brain (reviewed in Platt et al., 2012).

In addition to errors in cellular trafficking involving the lysosome, neurodegenerative diseases have been associated with abnormal fragmentation of the Golgi apparatus in mammalian neuronal cells. These dynamic organelles, composed of stacked membrane-bound sacs, are responsible for modifying, sorting, and preparing proteins for secretion (Cooper and Hausman, 2004). During mitosis, the Golgi disassemble and migrate into the daughter cell, but if this fragmentation fails to occur mitosis does not proceed (Robbins and Gonatas, 1964; reviewed in Warren, 1993; Sutterlin et al., 2002). The Golgi seem to play a particularly important role in neurons; when they fragment in these non-replicating cells, apoptosis soon follows (Nakagomi et al., 2008). This relationship between Golgi fragmentation and neuronal cell death has been observed in several neurodegenerative diseases, including ALS, Alzheimer's and Parkinson's (Gonatas et al., 2006; Fujita et al., 2006).

Lysosomal storage diseases and neurodegenerative diseases are serious problems worldwide, and the cellular origins of these illnesses have proven difficult to study in humans directly. The dawn of molecular genetics in the mid-20th century offered innumerable new directions in study of such diseases, but scientists had to find a model

organism that was genetically suitable, inexpensive, and ethically acceptable for their research. A popular microscopic organism soon became a useful tool for this research: *Saccharomyces cerevisiae*.

Saccharomyces cerevisiae: A Model Organism for the Study of Human Disease

S. cerevisiae is a unicellular budding yeast—commonly called baker’s yeast—that emerged as a species no more than 20 million years ago (Taylor and Berbee, 2006; Scannell et al., 2007; Caudy et al., 2013). Found ubiquitously in the environment, *S. cerevisiae* is a facultative anaerobe that humans have been utilizing for food production for more than 5000 years (Landry et al., 2006). Despite its evolutionary antiquity, *S. cerevisiae* has only been associated with the study of human disease since the latter part of the 20th century. In addition to its quick reproductive cycle, highly efficient mitotic recombination, and easily-controlled growth and proliferative processes, *S. cerevisiae* is an economic alternative to more-complicated eukaryotic models (Mager and Winderickx, 2005). Its genome, the first of the eukaryotes to be fully sequenced, was published in 1996 (Goffeau et al., 1996). About 40% of *S. cerevisiae* genes have orthologs—genes found in different species as a result of vertical transmission from a common ancestor—in the *Homo sapiens* genome (Botstein et al., 1997; Karathia et al. 2011). This suggests that the conserved genes may perform similar functions in the two species. These characteristics make *S. cerevisiae* an indispensable tool for lysosomal storage and neurodegenerative disease research.

S. cerevisiae’s vacuole is functionally analogous to the mammalian lysosome, and it exhibits cell cycle-associated fragmentation similar to the mitotic disassembly of the Golgi in mammalian cells (Wiemken et al., 1979; reviewed in Scott and Klionsky, 1998;

Shorter and Warren, 2002; Gieselmann and Braulke, 2009; reviewed in Li and Kane, 2009). Like the lysosome, the vacuole is also the terminus for many cellular trafficking pathways (Figure 1) (Conibear and Stevens, 1998; Luzio et al., 2007). Volumes of research since the 1970s have been testament to *S. cerevisiae*'s applicability to human medical research, particularly in the area of cellular trafficking disorders (Wiemken et al., 1979; Bryant and Stevens, 1998; Kucharczyk and Rytka, 2001; Gieselmann and Braulke, 2009; reviewed in Li and Kane, 2009).

Discovery of the Late Endosome and Vacuole Interface Genes

Research of human lysosomal storage and neurodegenerative diseases has demonstrated the necessity of proper cellular trafficking to the lysosome and the importance of the Golgi apparatus. Scientists have utilized *S. cerevisiae* to detect possible genetic factors for impaired cellular trafficking, and many genetic and mutagenic analyses aimed at isolating yeast mutants defective in various aspects of cellular trafficking have been successfully carried out (Jones, 1977; Banta et al., 1988; Robinson et al., 1988; Raymond et al., 1992). Two such studies were recently completed by our laboratory. Using novel immunodetection screens, *S. cerevisiae* mutants defective in processing a late endosome-dependent pathway cargo called pro-carboxypeptidase Y (proCPY) were isolated (Takahashi et al., 2008; Ricarte et al., 2011). In one set of deletion mutants—named late endosome and vacuole interface (*env*) mutants—proCPY accumulated inside the cells (Ricarte et al., 2011). Because no proCPY was found outside of unlysed cells, this phenotype suggested that the cellular trafficking defect must have occurred between the late endosome and the vacuole. Any problem at an earlier point in the endosome-dependent pathway would cause proCPY to be secreted as a

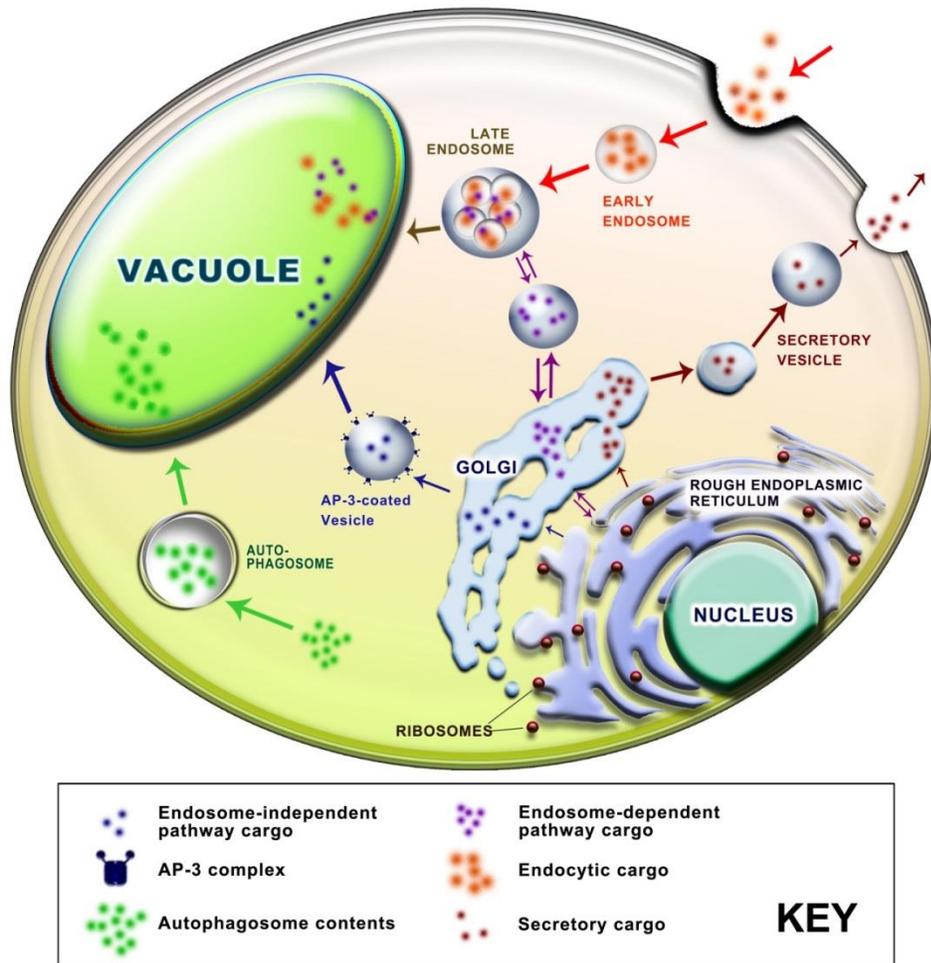


FIGURE 1. Five major cellular trafficking pathways in *S. cerevisiae*. The endosome-independent pathway (blue arrows and cargo) takes cargo via AP-3-coated vesicles directly from the Golgi to the vacuole. The endocytic pathway (orange arrows and cargo) begins with the binding of extracellular molecules to the cell membrane's surface. The cargo is then endocytosed into vesicles called early endosomes, which merge with vesicles from the endosome-dependent pathway to form the late endosomes. The late endosomes ultimately bring their cargo to the vacuole. The endosome-dependent pathway (purple arrows and cargo) begins with the translation/folding of proteins in the rough endoplasmic reticulum (ER). These proteins are budded off from the rough ER in vesicles that travel through the Golgi forming new vesicles. These Golgi vesicles fuse with the endocytic vesicles to form the late endosomes. Again, these vesicles ultimately deliver their cargo to the vacuole. The secretory pathway (dark red arrows and cargo) begins similarly to endosome-dependent pathway; however the Golgi vesicles form secretory vesicles which are targeted to the plasma membrane. There they fuse with the cell membrane and release their contents into the extracellular environment. Finally, the autophagic pathway (green arrows and cargo) is a way for the cell to recycle its own material during times of stress. Cellular contents are enveloped in the autophagosome, which brings its cargo to the vacuole for degradation. (Conibear and Stevens, 1998)

default cellular mechanism. Four of the *ENV* genes uncovered—*ENV7*, *ENV9*, *ENV10*, and *ENV11*—were previously uncharacterized, so our lab has focused on these genes. Along with traditional laboratory exploration, further understanding of the *ENV* genes will require the study of similar, evolutionarily-related genes—referred to generally as homologs.

ENV7

ENV7 (Gene locus: YPL236C) is primary focus of this study. It encodes an active 364-amino acid protein kinase (Env7) that localizes to the vacuolar membrane (Huh et al., 2003; Manandhar et al., 2013). Palmitoylation of its N-terminal triple cysteine motif (C₁₃C₁₄C₁₅) is required for proper function and localization of Env7 to the vacuolar membrane, and divalent magnesium (Mg²⁺) is essential for its catalytic activity (Manandhar et al., 2013). Under non-stressful conditions *env7Δ* cells do not exhibit a defect in vacuole morphology, but when *env7Δ* cells are exposed to a hypertonic environment the vacuoles do not stay fragmented after one hour as those in wild-type cells do (Ricarte et al., 2011; Manandhar et al., 2013). A similar phenotype was observed in budding cells; vacuoles in *env7Δ* cells remained prominent during the budding process, unlike those of wild-type cells which fragmented (Manandhar et al., 2013). This suggests that Env7 is involved with the negative regulation of vacuole membrane fusion during hyperosmotic stress and vacuole biogenesis. Although it can autophosphorylate, *in vivo* phosphorylation of Env7 is dependent on Yeast Casein Kinase 3 (Yck3), a vacuole membrane kinase also involved in vacuole size regulation (LaGrassa and Ungermann, 2005; Manandhar and Gharakhanian, 2013). No phylogenetic analysis concentrated on *ENV7* has been conducted before this study, but Serine/Threonine Kinase 16 (*STK16*)

was suggested as a human homolog for *ENV7* in multiple instances (Ligos et al., 1998; Stairs et al., 1998; Eswaran et al., 2008).

STK16

Human *STK16*, a 305-amino acid serine/threonine protein kinase homologous to *ENV7* in *S. cerevisiae*, has had an interesting research history. Between 1998 and 2000, five different groups spanning the globe converged on the gene nearly simultaneously, with each unfortunately giving it a different name in their own publication. *STK16* ended up with five aliases: Kinase Related to Cerevisiae and Thaliana (*KRCT*), Protein Kinase expressed in day 12 fetal Liver (*PKL12*), Embryo-Derived Protein Kinase (*EDPK*), Myristoylated and Palmitoylated Serine/threonine Kinase (*MPSK1*), and Transforming growth factor-beta (TGF-) Stimulated Factor (*TSF-1*) (Stairs et al., 1998; Ligos et al., 1998; Kurioka et al., 1998; Berson et al., 1999; Ohta et al., 2000). Eventually, to eliminate the ambiguity, the HUGO Gene Nomenclature Committee bestowed a single preferred name –*STK16*– upon the gene.

Although research of *STK16* intensified at the turn of the millennium, publications focusing specifically on the gene have recently dwindled because of the gene product's elusive function(s). Explorations into *STK16*'s cellular role have proposed that it is a serine/threonine kinase that localizes to the Golgi and is N-terminally myristoylated and palmitoylated, but studies regarding the function of the *Stk16* protein have not yielded repeatable results (Berson et al., 1999; Guinea et al., 2006; Eswaran et al., 2008). *STK16* was even suggested to play a role in cancer angiogenesis through involvement in transcriptional upregulation of Vascular Endothelial Growth Factor (*VEGF*), but this supposed function has not been addressed since a single publication in 2006 (Guinea et

al., 2006). The only two studied *STK16* homologs in fungi –*ENV7* in *S. cerevisiae* and *ppk3* in *Schizosaccharomyces pombe*– are nonessential (Bimbó et al., 2005; Ricarte et al., 2011). Our lab has undertaken an extensive characterization of *ENV7* in order to uncover more about its role in *S. cerevisiae* and potential functions for *STK16* in humans.

Eukaryotic Protein Kinases: Divisions and Subdomains

Almost two decades ago, pioneering protein kinase researchers Steven Hanks and Tony Hunter aligned eukaryotic protein kinase domains and identified sequence regions that were more than 95% conserved (Hanks and Hunter, 1995). They divided these well-aligned portions into two divisions and twelve subdomains, creating a framework of conserved primary sequence elements in the eukaryotic protein kinase domain. Hanks and Hunter reasoned that these conserved regions must be important to kinase activity, and extensive study since then has confirmed their supposition.

The first major division of the eukaryotic protein kinase domain separates the protein into two lobes: the N-lobe and the C-lobe. The N-lobe (NH₂-terminal lobe) primarily functions to position and bind the nucleotide, which in most cases is adenosine triphosphate (ATP). The larger C-lobe (COOH-terminal lobe) is involved with substrate-binding and transfer of the γ -phosphate from ATP. Structurally, the N-lobe is composed of mostly antiparallel β -sheets, and the C-lobe is mainly comprised of α -helices (Bossemeyer et al., 1994; Hanks and Hunter, 1995). The twelve eukaryotic protein kinase subdomains span the N- and C-lobes; subdomains I-V are part of the N-lobe, and subdomains VIA-XI are located in the C-lobe (Hanks and Hunter, 1995). Their functions and important residues are summarized in Table 1.

TABLE 1. Summary of Eukaryotic Protein Kinase Subdomains

Lobe	Subdomain	Important Residues	Function(s)
N	I	GxGxxG – glycine-rich loop	Forms flexible lid that folds over ATP, creating a hydrophobic binding pocket; interacts with α - and β -phosphates of ATP
	II	K – invariant lysine	Secures/positions ATP; forms salt bridge with conserved glutamic acid in subdomain III; interacts with α - and β -phosphates of ATP
	III	E – invariant glutamic acid	Forms α C-helix; forms salt bridge with conserved lysine in subdomain II; contributes to kinase stability during ATP binding
	IV	HxN – nearly invariant residues	Anchors α C-helix to C-lobe; stabilizes protein kinase in active configuration
	V	No observed invariant residues	Connects N- and C-lobes
C	VIA	No observed invariant residues	Support of active kinase configuration
	VIB	HRDxxxxN – catalytic loop	Conserved aspartic acid is the catalytic base; chelation of Mg^{2+} ion by conserved asparagine, allowing correct positioning of ATP γ -phosphate for transfer; asparagine also stabilizes catalytic loop
	VII	DFG – Mg^{2+} -positioning loop; activation N-terminal anchor	Chelation of primary activating Mg^{2+} ion, allowing correct positioning of ATP γ -phosphate for transfer
	VIII	APE – nearly invariant alanine-proline-glutamic acid; activation segment C-terminal anchor	Peptide substrate recognition; stabilization of C-lobe through glutamic acid salt bridge formation with arginine in subdomain XI; target for phosphorylation for kinase activation
	IX	D – invariant aspartic acid	Stabilization of catalytic loop; F-helix stabilization of entire C-lobe
	X	No observed invariant residues	Least conserved of the subdomains
	XI	R – nearly invariant arginine	Forms stabilizing salt bridge with conserved glutamic acid in subdomain VIII

Note: Information in this table was taken from: Hanks and Hunter, 1995; Taylor et al., 2004; Kannan et al., 2008.

The Current Placement of *ENV7* in the Eukaryotic Protein Kinase Superfamily

STK16 and *ENV7* are both classified as members of the NAK serine/threonine kinase family—a family that does not fit into a typical eukaryotic protein kinase group. The NAK kinase family, named after Numb-Associated Kinase in *Drosophila melanogaster*, is a poorly-defined collection of eukaryotic protein kinases. Its origins can be traced to a nearly twenty-year-old publication that briefly mentions the abnormal patterning of glycines in the glycine-rich loop of *D. melanogaster Nak* and three other similar kinases from *Caenorhabditis elegans* and *S. cerevisiae* (Chien et al., 1998). The glycine-rich loop in subdomain I is essential for catalysis in eukaryotic serine/threonine kinases. Through steric interactions and hydrogen bonds it enfolds the ATP molecule and buffers it from exposure to solvent (Taylor et al., 1993; reviewed in Bossemeyer, 1994). The glycine-rich loop has a consensus sequence of GxGxxG, where “x” denotes any amino acid (Hanks and Hunter, 1995). The first two glycines in most glycine-rich loops participate in phosphoryl transfer, and the third glycine contributes to the conformational flexibility of the loop, making it the most spatially variable part of the catalytic core (Hemmer et al., 1997; Grant et al., 1998). This NAK disordered glycine-rich loop pattern—xxxGGx instead of GxGxxG—is the only identical sequence feature shared by all of the known NAK kinases (Figure 5b); even so, the Polo-like Kinases also have two consecutive glycines in their glycine-rich loop (Hanks and Hunter, 1995).

The NAK kinase family has grown to include four well-known, albeit diverse, sequences from *H. sapiens*: AP2-Associated Kinase 1 (*AAK1*), Cyclin-G Associated Kinase (*GAK*), BMP-2 inducible Kinase (*BMP2K*), and Serine/Threonine Kinase 16 (*STK16*) (Manning et al., 2002). Additionally, four kinases in *S. cerevisiae*—Actin-

Regulating Kinase 1 (*ARK1*), p53-Regulating Kinase 1 (*PRK1*), Ark family Kinase-Like protein 1 (*AKLI*), and *ENV7*– are also thought to belong to this family (Cope et al., 1999). All of the human and yeast NAK kinases except for Env7 and STK16 have been shown to play a role in endocytosis, and their individual functions, for the most part, are not redundant (Cope et al., 1999; Greener et al., 2000; Zeng et al., 2001; Conner and Schmid, 2002; Henry et al., 2003; Smythe and Ayscough, 2003; Takahashi et al., 2008).

Major Aims of this Study

The three core goals of this report concentrate chiefly on *ENV7*. First, it was imperative to determine/augment information about *ENV7*'s evolutionarily relationships with other protein kinases and to evaluate its position as a NAK kinase. Moreover, this study sought to identify evolutionarily conserved and/or potentially important amino acid residues in Env7 in order to guide future creation of *env7* mutants. Finally, phenotypes of cells expressing an HA-tagged *env7* mutant harboring valine in place of a conserved glycine residue in the kinase's glycine-rich loop (Env7G₄₀V-HA) were explored experimentally. It was hypothesized that the glycine-to-valine substitution would alter the binding of ATP in the kinase cleft, leading to either increased or decreased kinase function. Although this study focused predominantly on phylogenetic and bioinformatical analyses of *ENV7*, some of these analyses were also performed on *ENV9*, *ENV10*, and *ENV11* to expand upon the already growing body of knowledge about these genes. The information presented in this work may be incorporated into future experimentation and discovery in the field of cellular trafficking and human disease.

CHAPTER 2

MATERIALS AND METHODS

Protein Sequence Selection for Phylogenetic Analyses

The Basic Local Alignment Search Tool (BLAST), sponsored by the National Center for Biotechnology Information (NCBI) was used to find similar protein sequences to the query polypeptides: Env7, Env9, Env10, and Env11 (Altschul et al., 1990). The default database, which contains non-redundant sequences from GenBank coding DNA sequence (CDS) translations, the Protein Data Bank (PDB), SwissProt, the Protein Information Resource (PIR), and the Protein Research Foundation (PRF) was chosen for each standard protein BLAST (blastp). A BLOSUM62 scoring matrix was used with compositional score matrix adjustment, a word size of 3, a gap existence cost of 11, and a gap extension cost of 1 (Yu et al., 2003; Yu and Altschul, 2005; Altschul et al., 2005). Sequence filtering and masking were not selected.

Results obtained from each blastp search were filtered by e-value and percent query coverage; sequences with expectation values (e-values) above 1.0×10.0^{-10} or query coverage below 60% were discarded. The remaining protein sequences were subjected to a reciprocal blastp to find potential orthologs of the original query sequence. This procedure was executed by using each protein sequence found by the initial blastp as a new individual query, with *S. cerevisiae* chosen as the target organism, for the reciprocal blastp. Sequences that showed the original *S. cerevisiae* query sequence (Env7, Env9, Env10, or Env11) as the top result of the reciprocal blastp, within the e-

value and query coverage parameters used for the first blastp, were selected as potential orthologs. If more than one sequence was found in the initial blastp query, both were subjected to reciprocal blastp. If both sequences differed in sequence identity to the reciprocal blastp query by less than one percent (while adhering to aforementioned the e-value and query coverage standards), they were considered possible paralogs.

Multiple Sequence Alignment

Multiple sequence alignments were generated using various algorithms in the Geneious[®] Pro 5.6.4 interface (Biomatters Ltd.). Each set of protein sequences was aligned many times with different multiple sequence alignment algorithms, and the algorithm that produced the best alignment of conserved elements (e.g. kinase domains, active sites, hydrophobic regions, etc.) was selected for manual editing. Manual editing of the multiple sequence alignments entailed visually scanning the alignments and moving residues in such a way that they were better aligned using the Geneious[®] Pro alignment viewer/editor. A detailed list of the algorithm, settings, and reference for each selected multiple sequence alignment included in this study can be found in Table 8 in Appendix A.

Determination of Evolutionary Models

Appropriate evolutionary models were determined for each aligned dataset using the maximum likelihood model selection function of the Molecular Evolutionary Genetics Analysis 5 (MEGA5) software suite (Tamura et al., 2011). A MEGA-formatted multiple protein sequence alignment was used as the input for each evolutionary model selection process. Neighbor-joining guide trees were constructed with a moderate branch swap filter and inclusion of all sites in the alignment. The model

with the lowest Bayesian Information Criterion (BIC) score was selected as the best to describe the amino acid substitution patterns present in the input multiple sequence alignment. The number of parameters was also taken into consideration; so models – even those with the lowest BIC score– that required amino acid frequencies to be specified were not considered for selection. The top evolutionary models for the multiple sequence alignments in this study are listed in Table 9 in Appendix A.

Maximum Likelihood Tree Estimation

Maximum likelihood analyses were performed on protein multiple sequence alignments using the PhyML 2.1.0 program in the Geneious Program interface (Biomatters Ltd.) (Guindon and Gascuel, 2003). Parameters were set based on the appropriate evolutionary model determined by the MEGA5 model selection process, as seen in Table 9 Appendix A. Nearest Neighbor Interchange (NNI) with no optimization was used for the tree topology search. The final maximum likelihood trees were viewed with the Geneious[®] Pro 5.6.4 interface (Biomatters Ltd.).

Bayesian Tree Estimation

Each Bayesian analysis was performed using the MrBayes program in the Geneious[®] Pro 5.6.4 interface (Biomatters Ltd.) (Huelsenbeck and Ronquist, 2001). The maximum likelihood trees generated by PhyML were used as the inputs for each analysis, and the settings chosen for each run are detailed in Table 10 in Appendix A. The average standard deviation of split frequencies (ASDSF) for each Bayesian analysis can also be found in Table 10 in Appendix A. None of the tree generated by Bayesian analysis had an ASDSF greater than 0.04, so it is reasonable to assume that the clade support values between the multiple chains in each analysis are fairly similar. The trees in this study

display clade credibility values of less than 90% as dotted lines and those with 90% or greater as solid lines.

Synteny Analysis

The Genome Evolution Analysis (GEvo) online server program (<http://genomeevolution.org/CoGe//GEvo.pl>) was used to determine syntenic regions around genes of interest (Lyons and Freeling, 2008). Accession numbers for each gene and its homologs were inputted into the sequence submission page, and the blastz (large regions) algorithm, with default parameters, was used to search for homologous regions between the chromosomal locations of each sequence. Sequences were inputted in order of species relationship to *Saccharomyces cerevisiae*; therefore the most closely related species (e.g. *Candida glabrata*, *Zygosaccharomyces rouxii*, *Naumovozyma sp.*) were analyzed first. Screenshots of the results were manipulated with Adobe Photoshop CS5.

Tests for Evolutionary Selection

Four programs from the Datamonkey webserver (<http://www.datamonkey.org/>), which is hosted by the University of California, San Diego, were utilized for selection analyses (Kosakovsky et al., 2006; Delpont et al., 2010). Single Likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), and Fast Unbiased Bayesian AppRoximation (FUBAR) were used for determination of diversifying or purifying selection (Kosakovsky et al., 2006; Murrell et al., 2013). To see if and which amino acid residues in the alignment were subject to change or conservation, PRoperty Informed Model of Evolution (PRIME) was used (Kosakovsky et al., 2006; Murrell et al., 2013). Inputs consisted of the translation-aligned nucleotide sequences for SLAC, FEL, and FUBAR; for the PRIME analysis protein sequence multiple sequence alignments were

inputted. Significance levels for all the likelihood ratio tests (LRTs) for the analyses were set at $p < 0.05$. Conant-Stadler amino acids properties –chemical composition, polarity, volume, isoelectric point, and hydrophathy– were chosen to be addressed by the PRIME program (Conant et al., 2007).

Generation and Manipulation of Tertiary Protein Structures

Three-dimensional renderings of tertiary protein structures were generated *in silico* by the iTASSER server provided by the Zhang Lab at the University of Michigan (Zhang, 2008; Roy et al., 2010). Fasta-formatted protein sequences were submitted to the online server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>), and the resulting Protein DataBase (PDB) files were viewed and manipulated (i.e. color adjustments, labeling, etc.) for presentation using ICM Browser Pro 3.7-2b student software (MolSoft L.L.C.). The estimated accuracy for the Env7 tertiary protein model was determined by three scores: a root-mean-square deviation (RMSD), a template-modeling score (TM-score), and a confidence score (C-score). The RMSD for Env7 was $8.4 \pm 4.5 \text{ \AA}$, and the TM-score was 0.61 ± 0.14 . The C-score, which can range from -5 to 2, was -0.82. Higher numbers for TM-scores and C-scores are desirable, while a lower RMSD is better (Zhang and Skolnick, 2004; Zhang, 2008; Roy et al., 2010).

Phosphorylation Site Prediction

Nine different methods were used to find amino acid residues in Env7 that possess good potential for phosphorylation. These methods and their respective references/sources are listed in Table 11 in Appendix A. Default settings were chosen for each prediction method, and, when possible, *S. cerevisiae* was specified as the target organism. If a putative phosphorylation site was only predicted by a single program, it

was precluded from further analysis. The resulting proposed phosphorylation sites were fashioned into a table and visually manipulated with Adobe Photoshop CS5 ver. 12.1 x64.

Yeast and Bacterial Strains

The *Saccharomyces cerevisiae* BY4742 background strain and *env7Δ* null mutant strain were obtained from a MAT- α haploid deletion library donated by Dr. Gregory Payne (UCLA). Chemically competent DH5 α *Escherichia coli* cells for site-directed mutagenesis were purchased from Invitrogen (Life Technologies; Carlsbad, CA). Table 2 defines the genotypes of these yeast and bacterial strains.

Competent Yeast Cells

Competent yeast cells for transfection were prepared using the Zymo Frozen EZ Yeast Transformation II Kit (Zymo Research; Irvine, CA). BY4742 *env7Δ* cells in YPD were grown in a 30°C shaker until mid-log phase (O.D.₆₀₀ = 0.8-1.0). The cells were then pelleted by centrifugation at 500 x g for 4 minutes and resuspended in 10ml EZ1 solution. The centrifugation was repeated, and the cells were ultimately resuspended in 1ml EZ2 solution. The competent *env7Δ* cells were stored at -80°C.

Media and Reagents

Yeast strains were mainly grown in two types of medium: yeast extract-peptone-dextrose (YPD) and synthetic minimal medium without uracil (SM-ura). However, many more types of media were used for growth experiments. Appendix B provides a full list of the media and reagents used in this study.

TABLE 2. Microbiological Strains Used in this Study

<i>Saccharomyces cerevisiae</i>		
Strain	Genotype	Source
<i>ENV7-HA</i> (Env7-HA)	<i>MATα his3Δ leu2Δ ura3Δ lys2Δ ENV7 3xHA</i>	Dr. Surya Manandhar
<i>env7g38cg41cg44c</i> (Env7C ₁₃₋₁₅ S-HA)	<i>MATα his3Δ leu2Δ ura3Δ lys2Δ env7g38cg41cg44c 3xHA</i>	Dr. Surya Manandhar
<i>env7g119t</i> (Env7G ₄₀ V-HA)	<i>MATα his3Δ leu2Δ ura3Δ lys2Δ env7g119t 3xHA</i>	This study
<i>env7Δ</i>	<i>MATα his3Δ leu2Δ ura3Δ lys2Δ env7Δ::KANMX4</i>	Dr. Gregory Payne, UCLA
<i>Escherichia coli</i>		
Strain	Genotype	Source
DH5 α	F- Φ 80 <i>lacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>) U169 <i>recA1 endA1 hsdR17</i> (rK-, mK+) <i>phoA</i> <i>supE44</i> λ - <i>thi-1 gyrA96 relA1</i>	Invitrogen (Life Technologies; Carlsbad, CA)

Note: All *S. cerevisiae* strains used were from the BY4742 background. The titles listed below the first three *S. cerevisiae* strain names are the corresponding gene product names.

TABLE 3. Primers Used for Site-Directed Mutagenesis

Mutant Primers		
Primer Name	Sequence (5' \rightarrow 3')	Source
<i>env7G40V</i> Forward	CGACTACTTGGAGAAGGTG <u>T</u> AATGTCCT TTGTGTATTTG	This study
<i>env7G40V</i> Reverse	CAAATACACAAAGGACATT <u>A</u> CACCTTCT CCAAGTAGTCG	This study
Non-Mutant Primers		
Primer Name	Sequence (5' \rightarrow 3')	Source
Primer 1 Forward	AGTTCTTAGATGCTTTCTTTTCT	Dr. Surya Manandhar
Primer 2 Reverse	GAACGTCATATGGATAGGATCCTG	Dr. Surya Manandhar
Primer 15 Forward	GGAGTTTACAAGACTCTATTAATC	Dr. Surya Manandhar

Plasmids and Primers

Dr. Surya Manandhar supplied the pSMG17 2-micron plasmid vector containing C-terminally HA-tagged *ENV7* that was used as a template for site-directed mutagenesis. Figure 15 shows this plasmid with the mutagenized *env7g119t* gene. Another plasmid, pWS479 + *RCE1-HA*, was used as the recipient 2 μ vector of the mutagenized *env7g119t* gene (Plummer et al., 2006). Mutant primers for site-directed mutagenesis were created using the PrimerX program (http://www.bioinformatics.org/primerx/cgi-bin/DNA_1.cgi) and then ordered from Eurofins MWG Operon (Huntsville, AL). Primers specific to the pSMG17 + *ENV7-HA* vector for the site-directed mutagenesis were also kindly provided by Dr. Surya Manandhar. A detailed list of all primers used in this study is displayed in Table 3. Point mutations are underlined and italicized in the mutant primer sequences.

Site-Directed Mutagenesis

Three separate PCR reactions were performed in order to carry out the desired site-directed mutagenesis of *ENV7* to *env7g119t*. The entire process is outlined in Figure 14. The Eppendorf Gradient Mastercycler settings are described in Table 4. The reaction mixtures for PCR I and II consisted of 2.0 μ l DNA template (pSMG17+*ENV7-HA*), 1.25 μ l of each appropriate primer, 1.0 μ l dNTPs, 10.0 μ l 5x Phusion polymerase buffer, 34.0 μ l sterile water, and 0.5 μ l Phusion polymerase. The reaction mixture for PCR III included 2.0 μ l gel-purified PCR I DNA product, 2.0 μ l gel-purified PCR II DNA product, 1.0 μ l dNTPs, 10.0 μ l 5x Phusion polymerase buffer, 32.0 μ l sterile water, and 0.5 μ l Phusion polymerase. After 8 repetitions of cycles 2-4, the machine was stopped and primers #1 and #2 were added to the mix. The machine was restarted at step 1 and cycles 2-4 were

then repeated 27 times. All reagents and enzymes were purchased from Sigma-Aldrich (St. Louis, MO).

Gel Purification of PCR I, II and III Products from Agarose Gel

To isolate PCR I and II products for use in PCR III, the results of the first two PCR rounds were run on a 1% agarose gel, identified by size, and removed from the gel. The same process was repeated for the PCR III DNA product, the mutagenized *env7g119t* fragment.

36.0ml deionized water, 4.0ml 5x TBE, 0.4g type II agarose, and 4.0µl ethidium bromide were combined and cast to make the 1% agarose gel. Once the gel solidified, 50µl of each PCR sample (run in triplicate) was mixed with 8µl 6x blue loading dye and pipetted into the wells. 0.3µl of 1Kb Plus DNA Ladder (Life Technologies; Carlsbad, CA) was also loaded as a size reference. The gel was run at 100 volts for 30 minutes.

Bands of appropriate size were cut from the gel with a razor blade. The PCR I band showed around 1Kb; the PCR II band was near 0.2Kb, and the PCR III band was expected at 1.2Kb (Figure 14). The Zymoclean™ Gel DNA Recovery Kit (Zymo Research; Irvine, CA) was used to extract the DNA fragments from the gel polymer and purify them. The purified DNA fragments were each resuspended in 15µl of sterile water and stored in -20°C.

Restriction Enzyme Site Analysis

To linearize the pWS479 + *RCEI-HA* recipient vector and allow *env7g119t* PCR III product to homologously recombine in the place of *RCEI*, thereby gaining the triple HA tag at its C-terminus, the plasmid was cut with a restriction enzyme. The restriction digest consisted of 1.0µl SphI-HF restriction enzyme, 2.0µl pWS479 + *RCEI-HA*, 4.0µl

TABLE 4. Thermocycler Settings for PCR I, II and III

PCR I and II		
Step	Temperature (°C)	Time (seconds)
1	98.0	300
2	98.0	50
3	68.0	30
4	72.0	60
-----Repeat steps 2-4 35x.-----		
5	72.0	600
6	4.0	HOLD
PCR III		
Step	Temperature (°C)	Time (seconds)
1	98.0	120
2	98.0	50
3	62.5	30
4	72.0	60
Repeat steps 2-4 8x.		
-----Then stop the machine and add 1.25µl of each appropriate primer.-----		
Start the machine over at step 1, then run steps 2-4 27x.		
5	72.0	600
6	4.0	HOLD

TABLE 5. Thermocycler Settings for Colony PCR

Step	Temperature (°C)	Time (seconds)
1	98.0	120
2	98.0	50
3	72.0	30
4	72.0	60
-----Repeat steps 2-4 35x.-----		
5	72.0	600
6	4.0	HOLD

NEB 10x Reaction Buffer #4, and 33.0µl sterile water. The restriction enzyme and NEB 10x Reaction Buffer #4 were acquired from New England Biolabs, Inc. (Ipswich, MA). The digest incubated overnight in 37°C.

To check the band sizes of the cut pWS479 + *RCE1-HA* plasmid, all 40µl of the digest results (plus 8µl loading dye) were run on a 1% agarose gel at 100 volts for 30 minutes. 0.3µl of 1Kb Plus DNA Ladder (Life Technologies; Carlsbad, CA) was again run as a size reference. The appropriate band (~8 Kb), representing the linearized vector with most of *RCE1* excised, was cut from the gel and purified with the Zymoclean™ Gel DNA Recovery Kit (Zymo Research; Irvine, CA).

Transformation of *env7Δ* Cells

In order for point-mutated *env7g119t* to homologously recombine in the place of *RCE1* in the pWS479 vector, *env7Δ* cells were transformed with the *env7g119t* DNA fragments from PCR III and the linearized pWS479 plasmid. 1.5ml of mid-log phase liquid YPD *env7Δ* culture was centrifuged at 15,000 x g for 15 seconds. The pellet was resuspended in 100.0µl YPD, and 4.0µl 10mg/ml salmon sperm DNA –carrier ssDNA– (Life Technologies; Carlsbad, CA) was added along with 2.0µl linearized pWS479 and 6.0µl PCR III product. The mixture was vortexed well, and then 500.0µl PLATE mixture was added. After more thorough vortexing, 20µl of 1.0M DTT was added. Again, the mixture was vortexed well. The transformation mixture was left to incubate at room temperature overnight.

After the overnight incubation, the cells were heat-shocked at 37°C for 15 minutes in a heating block. Then the cells were pelleted at 15,000 x g for 30 seconds, and all but 100.0µl of supernatant was aspirated. The transformed cells were plated on SM-ura and

incubated for 2 days in 30°C. Only cells containing circular (closed) plasmids were expected to grow on the SM-ura medium.

Colony PCR

To quickly determine which transformed colonies had plasmids with the *env7* insert (though not necessarily the g119t point mutation), colony PCR was performed on select colonies from the transformation. 24 discrete colonies from the transformation plates were chosen and placed into microfuge tubes containing 5µl of 0.02M NaOH. The cells were resuspended in each tube and boiled (100°C) for 10 minutes in a heating block. The tubes were briefly centrifuged at 15,000 x g to bring down any condensation.

PCR was used to amplify any DNA specific to the *env7* insert present in the cells. Each PCR reaction tube contained 0.625µl each of primers 1 and 2, 2.0µl DNA from the transformed cells (from a single colony), 0.25µl Phusion Polymerase, 5.0µl 5x Phusion Buffer GC, 0.5µl dNTPs, and 16.0µl sterile water. The detailed PCR settings for this analysis can be seen in Table 5.

If the *env7* insert in the pWS479 plasmid was present in a colony of cells, the PCR reaction should yield a ~1.19Kb fragment specific to the insert. To check this, all 25.0µl of each PCR reaction was run on a 1% agarose gel as described previously for the site-directed mutagenesis. Colonies that showed the correct band size after the 30 minute run were selected for further analysis, and the 1.19Kb bands were removed from the gel and purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research; Irvine, CA).

Yeast DNA Extraction

The pSMG17+*env7G40V-HA* plasmids must be isolated in order to determine if the correct point mutation is in the gene. To obtain these vectors, the Zymoprep™ Yeast Plasmid Miniprep II Kit (Zymo Research; Irvine, CA) was used on the colonies that exhibited the correct band size after colony PCR. Purified pSMG17+*env7G40V-HA* plasmid was eluted to a final volume of 10.0µl with the kit.

Amplification of Mutagenized Plasmid in DH5α *E. coli* Cells

A relatively high concentration of DNA is necessary for accurate nucleotide sequencing of a sample, so DH5α *E. coli* cells were transformed with the DNA extracted from the positive colony PCR cells to amplify the pSMG17+*env7G40V-HA* plasmids. 3.0µl of purified plasmid from the yeast cells were added to 250.0µl competent DH5α *E. coli* cells. Sterile water was used in place of the plasmid for the negative control, and pUC18 (Agilent; Santa Clara, CA) was used for the positive control. The cells were then incubated on ice for 10 minutes followed by heat shock in a 42°C heating block for 30 seconds. After a 2-minute incubation on ice, 250.0µl of room temperature liquid TYE was added to the cells. The cells were incubated for 1 hour in 37°C on a rotating platform. After the incubation, 100.0µl of each transformation mixture was plated onto pre-warmed TYE+ampicillin plates. The plates were incubated in 37°C overnight.

6.0ml of TYE+ampicillin was inoculated a single colony that grew on the TYE+ampicillin transformation plates from each representative yeast colony. These cultures were grown to log phase in the 37°C shaker incubator. The Zyppy™ Plasmid Midiprep Kit (Zymo Research; Irvine, CA) was then used to extract and purify the pSMG17+*env7G40V-HA* plasmid DNA from the bacterial cells. The protocol from the kit

was followed in every respect except for the final elution. 150.0µl sterile water was used to elute the DNA instead of 150.0µl of Zyppy Elution Buffer.

Sequencing of Point-Mutated Plasmids

Purified, mutagenized pSMG17+*env7G₄₀V-HA* plasmid DNA samples suspended in sterile water and primers 1, 2 and 15 were sent to the Macrogen Corporation (Seoul, South Korea) for DNA sequencing. Results returned from Macrogen were aligned and visually scanned for the g119t point mutation or any abnormalities in the DNA sequence.

FM4-64 Staining and Microscopy

Cells were grown to log phase (O.D.₆₀₀ = 0.8-1.0) in SM-ura, and 2.0ml of each culture was centrifuged at 5000 x g for 5 minutes. The supernatants were aspirated, and the pellets were resuspended in 50.0µl YPD and 15.0µl FM4-64 stock solution. The samples were then incubated on a rotating platform in 30°C for 90 minutes. After the incubation, the cells were centrifuged at 5000 x g for 5 minutes. Each sample was washed 4 times with 1.0ml fresh YPD. On the fourth wash, each 1.0ml sample was split into 2 tubes of 500.0µl. The cells were centrifuged again at 5000 x g for 5 minutes, and the pellets were resuspended in either 1.0ml YPD or 1.0ml 0.5M NaCl YPD. The samples were incubated in 30°C for 1 hour. When the incubation was complete, the samples were centrifuged at 5000 x g for 5 minutes, and the pellets were suspended in 2 parts fresh YPD.

4µl of each sample were pipetted onto concanavalin A-coated glass microscope slides and viewed at 1000x using differential interference (DIC) microscopy and fluorescent microscopy via a green excitation filter. At least 400 non-budding cells were

scored based on their prominent or multi-lobed vacuolar morphology, as illustrated in Figure 17b. Images were manipulated using Adobe Photoshop CS5 ver. 12.1 x64.

Liquid Growth Assays

To see any growth defects or advantages were produced by the G₄₀V mutation or the C₁₃₋₁₅S mutation, liquid growth tests were performed in triplicate using the Synergy H1 plate reader (BioTek, Winooski, VT). The O.D.₆₀₀ of log phase liquid cell cultures (Env7-HA, Env7G₄₀V-HA, and Env7C₁₃₋₁₅S-HA) were standardized to 0.02 in 200µl of appropriate medium in a sterile Falcon Microtest™ 96-well, flat-bottom, polystyrene microplate with a lid (BD Biosciences, San Jose, CA). The cultures were grown for 30 hours in the plate reader at 30°C with orbital agitation every 15 minutes. O.D.₆₀₀ readings were taken automatically every 30 minutes by the Gen5 data analysis program (BioTek, Winooski, VT). A triplicate blank was also run for each medium to ensure that there was no contamination of the plate or medium.

Statistical analyses were performed to determine if differences in growth in liquid culture were significant. The point during the growth of the cultures that was used as the milestone for the hypotheses was mid-log phase (the time at which the culture had the fastest rate of growth). A one-sided t-test, calculated by hand, was used to determine if differences were significant for each set of hypotheses. The 90%, 95%, or 98% confidence interval with 2 degrees of freedom was used when referencing the t-critical values table.

Solid Growth Assays

To if see any fitness advantages or defects were affected by the G₄₀V mutation or the C₁₃₋₁₅S mutation, solid growth tests were performed using dilutions of YPD liquid

culture. Liquid cultures of Env7-HA, Env7G₄₀V-HA, and Env7C₁₃₋₁₅S-HA cells were grown in SM-ura to O.D.₆₀₀ ≈ 1.0. These cultures were then diluted to O.D.₆₀₀ = 0.1 in 500μl fresh YPD. The cells were pelleted at low speed and resuspended in 500μl YPD. To obtain a serial dilution of cells, 20μl of this initial culture were removed and added to 180μl YPD. These steps were repeated twice more to yield 1:1, 1:10, 1:100, and 1:1000 diluted cultures containing about 15,000, 1,500, 150, and 15 cells, respectively. 5μl of each strain's culture for each dilution were spotted onto a solid medium and left to dry on a room temperature countertop for 1 hour. Once the liquid culture spots had dried onto the media, the plates were inverted and incubated for up to 12 days in either 30°C or 37°C.

Subcellular Fractionation

Liquid cell cultures were grown to O.D.₆₀₀ ≈ 1.0 in SM-ura and pelleted. The supernatant was discarded, and the pellet was resuspended in 1ml of DTT solution. The cells were incubated for 10 minutes in the 30°C water bath and then centrifuged at 4600xg for 4 minutes at 4°C. The supernatant was discarded, and the pellet was resuspended in 500μl of spheroplasting buffer and 10μl of Zymolyase (Zymogen). The cells were incubated in the 30°C water bath for 20 minutes, followed by centrifugation at 1500xg for 3 minutes in 4°C. The supernatant was again discarded, and the pellet was resuspended in 1ml of lysis buffer and 4μl of DEAE-Dextran solution. Next, the cells were incubated on ice for 5 minutes, then in the 30°C water bath for 2 minutes. The cells were centrifuged at 400xg for 10 minutes in 4°C, and the supernatant collected was labeled the "S0.4" fraction. 50μl of this fraction were set aside, and the rest of the supernatant was then centrifuged at 13,000xg for 15 minutes in 4°C. The resulting

supernatant was labeled the “S13” fraction, and the pellet was labeled “P13”. The P13 fraction was resuspended in 100µl lysis buffer.

Protein Quantification

The DC™ Protein Assay kit (Bio-Rad, Hercules, CA) was used for protein quantification. 100µl of each bovine serum albumin (BSA) standard was prepared as shown in Table 15 in Appendix B. 5µl of each sample or standard were pipetted into appropriate wells of a sterile Falcon Microtest™ 96-well microplate (BD Biosciences, San Jose, CA). Then 200µl of Bio-Rad DC™ Protein Assay Reagent B were pipetted into each well containing sample or standard, followed by 25µl of Bio-Rad DC™ Protein Assay Reagent A. The solutions were incubated at room temperature for 10 minutes to allow any color change to occur. The 96-well plate was then transferred to the SynergyH¹ microplate reader (BioTek, Winooski, VT), and the absorbance at 750 nm for each well was measured and recorded using Gen5 data analysis software (BioTek, Winooski, VT). Standard concentrations versus their respective absorbances were graphed in Microsoft Excel, and linear regression was conducted to generate a function describing the relationship between absorbance and protein concentration. This function was applied to the sample absorbances to obtain the total protein concentration for each subcellular fraction. Standard and sample readings were performed in triplicate.

SDS-PAGE and Western Blotting

A large, 1.5mm thick 7.5% SDS-PAGE gel was cast (see Table 16 in Appendix B for gel preparation) and loaded onto a vertical gel rig to set for 1.5 hours. 5µl of 6x SDS-PAGE sample buffer were combined with 25µl of each S0.4, S13, and P13 subcellular fraction to make a final protein concentration of 1.667µg/µl. The samples in buffer were

boiled for 10 minutes and briefly centrifuged to collect any condensation. The entire 30 μ l of sample + buffer were loaded into the wells of the large gel. 10 μ l of Precision Plus WesternC Protein Ladder (Bio-Rad, Hercules, CA) were also added to a lane to provide reference protein sizes. The gel was run for 15 minutes at 170 volts so the samples could clear the stacking portion of the gel, and then the voltage was increased to 190 volts for an additional 3 hours and 45 minutes while the samples traveled through the resolving part of the gel.

After the SDS-PAGE run was complete, the gel was removed from the glass plates and soaked in Western blot transfer buffer. A nitrocellulose membrane and filter papers were also soaked in the transfer buffer. The wetted components of the blot were stacked with the filter papers sandwiching the gel (closer to cathode) and the nitrocellulose (closer to anode) on the Bio-Rad TransBlot Semidry transfer cell (Bio-Rad, Hercules, CA). After thoroughly soaking the stack with transfer buffer and using a test tube to smooth out any bubbles, the transfer cell was run at a constant amperage of 0.1 for 60 minutes.

To develop the Western blot, the nitrocellulose was blocked for 30 minutes in 30ml 5% TTBSM on a rocking platform at room temperature. Then 5 μ l of primary antibody (anti-HA mouse monoclonal antibody; Cell Signaling Technologies, Beverly, MA) were added to the nitrocellulose in 5% TTBSM. This was incubated overnight (~18 hours) on the rocking platform at 4°C. After the overnight incubation, unbound primary antibody was washed off by incubating the nitrocellulose with 20ml TBS for 10 minutes, then 20ml TTBS for 10 minutes, and finally 20ml more of TBS for 10 minutes – all on the rocking platform at room temperature. 4 μ l of secondary antibody (immunopure goat

anti-mouse IgG peroxidase conjugate, prepared by M. Seranilla) was thoroughly mixed with 30ml 5% TTBSM and added to the nitrocellulose membrane. This was incubated for 2 hours on the rocking platform at room temperature. Excess secondary antibody was removed by the same washing procedure described for the primary antibody.

Chemiluminescent reagents (15ml SuperSignal West Pico luminol/enhancer and 15ml SuperSignal West Pico stable peroxide – Thermo Fisher Scientific, Rockford, IL) were added to the washed nitrocellulose membrane, and this was covered in foil to protect from light and incubated for 15 minutes on the rocking platform at room temperature. The nitrocellulose membrane was developed onto Kodak film and visualized using a Kodak M35A X-OMAT film processor.

The SDS-PAGE and Western blotting protocol for the phosphorylation upshift assay was essentially the same as the one for the subcellular fractionation samples described above. The major differences arise in the casting/running of the gel and the amount of sample added to the gel. A small, 0.5mm thick 7.5% SDS-PAGE gel was cast in an upright Bio-Rad gel rig and set for 1 hour. Samples from the phosphorylation upshift assay were combined with 6x SDS sample buffer to a final protein concentration of 2.0µg/µl. 18µl of sample + buffer were loaded into each well, along with 6µl of Precision Plus WesternC Protein Ladder (Bio-Rad, Hercules, CA). The gel was run at 120 volts for 10 minutes, then 140 volts for 1 hour. The protein transfer from gel to nitrocellulose was run at 0.06 amps for 1 hour.

Phosphorylation Upshift Assay

The appropriate cell fraction was suspended in 100.0µl fresh lysis buffer. Then 15.0µl ATP-regenerating system, 15.0µl 10x fusion reaction buffer, 0.66µl 2.5mM

coenzyme A, and 19.34 μ l 0% ficoll were added to the sample. For the negative control, 19.34 μ l more ficoll were added in the place of the ATP-regenerating system to another 100.0 μ l sample in lysis buffer. These solutions were incubated in 30°C for 1 hour. After the incubation period, the samples were centrifuged at 14,000 x g in 4°C for 15 minutes. The supernatants were carefully aspirated, and the pellets were resuspended in 30.00 μ l 1x SDS-PAGE sample buffer and vortexed for a short time to mix thoroughly. The samples were boiled for 10 minutes in a heating block and briefly centrifuged at 500 x g to collect condensation.

CHAPTER 3

RESULTS

Evolutionary Conservation of *ENV7*

The Conservation Pattern of Env7 Orthologs Contradicts Established Monophyly

The *ENV7* gene is conserved throughout Eukarya and in one bacterial species (*Methylotenera mobilis*). Reciprocal blastp, multiple sequence alignments, and phylogenetic tree generation revealed that STK16, along with many other eukaryotic sequences and the bacterial sequence, is an ortholog of Env7 (Figures 2 and 3).

However, results from these analyses indicate that the pattern of conservation of Env7 and its orthologs is not consistent with established relationships among organisms.

Env7 orthologs were found in four major fungal phyla: Chytridiomycota, Zygomycota, Basidiomycota, and Ascomycota. The phylogenetic tree shown in Figure 2 was rooted on the sequence from *Batrachochytrium dendrobatidis*

(BATDEDRAFT_33346), a chytrid fungus that is basal in the fungi phylogenomic tree (Förster et al., 1990; Hedges, 2002; Hibbett et al., 2007; Ebersberger et al., 2012).

Surprisingly, Env7 orthologs representing the largest phylum, Ascomycota, did not form a clade (posterior probability = 0.99) (Figure 2). Using a variety of genetic elements for phylogenetic analyses, Ascomycota has been repeatedly shown to be a monophyletic group of organisms (Lutzoni et al., 2004; James et al., 2006; Ebersberger et al., 2012).

For Env7 and its fungal orthologs to deviate from this cladistic grouping, other evolutionary methods besides vertical gene transmission likely occurred. Lateral gene

transfer could explain why the phylogenetic tree in Figure 2 does not reflect Ascomycota as a single clade. Additionally, the positioning of the sequence from Saccharomycetaceae family member *Komagataella pastoris* (XP_002491616) could be a result of incomplete lineage sorting.

The same departure from recognized relationships was apparent in the phylogenetic tree of Env7 and its orthologs from Bacteria and Eukaryota. Fungi and metazoans are each monophyletic, but the only true clade reflected among Env7 and its orthologs was that of the land plants and green algae (posterior probability = 0.96; Figure 3) (Wainright et al., 1993; Hibbett et al., 2007; Dunn et al., 2008). One unexpected result of the Bayesian analysis was the proximity of the *Methylotenera mobilis* bacterial sequence (Mmol_0176) to Env7. Although the branch length of Mmol_0176 suggests that it is very divergent from Env7, the topology of the tree shows that it is closely related to Env7 among the sequences chosen for the tree (product of posterior probabilities = 0.92; Figure 3). Orthologs for Env7 could not be found in any other sequenced bacterial species. *M. mobilis* is only distantly related to a core group of Betaproteobacteria and has an average of only 32.95% of its proteins in common with its closest Methylophilaceae relatives (Kalyuzhnaya et al., 2006; Lapidus et al., 2011). More importantly, it has many distinctive gene islands in its genome that are products of horizontal gene transfer (Lapidus et al., 2011). Although the Mmol_0176 gene is not part of these gene islands, it is located in a very small section of genome between two islands predicted to be the result of DNA transfer from a phage and a plasmid insert (Lapidus et al., 2011). However, there are no conserved syntenic blocks between *ENV7* and Mmol_0176 regions in the two genomes (Results not shown). *M. mobilis* is a motile, aerobic freshwater

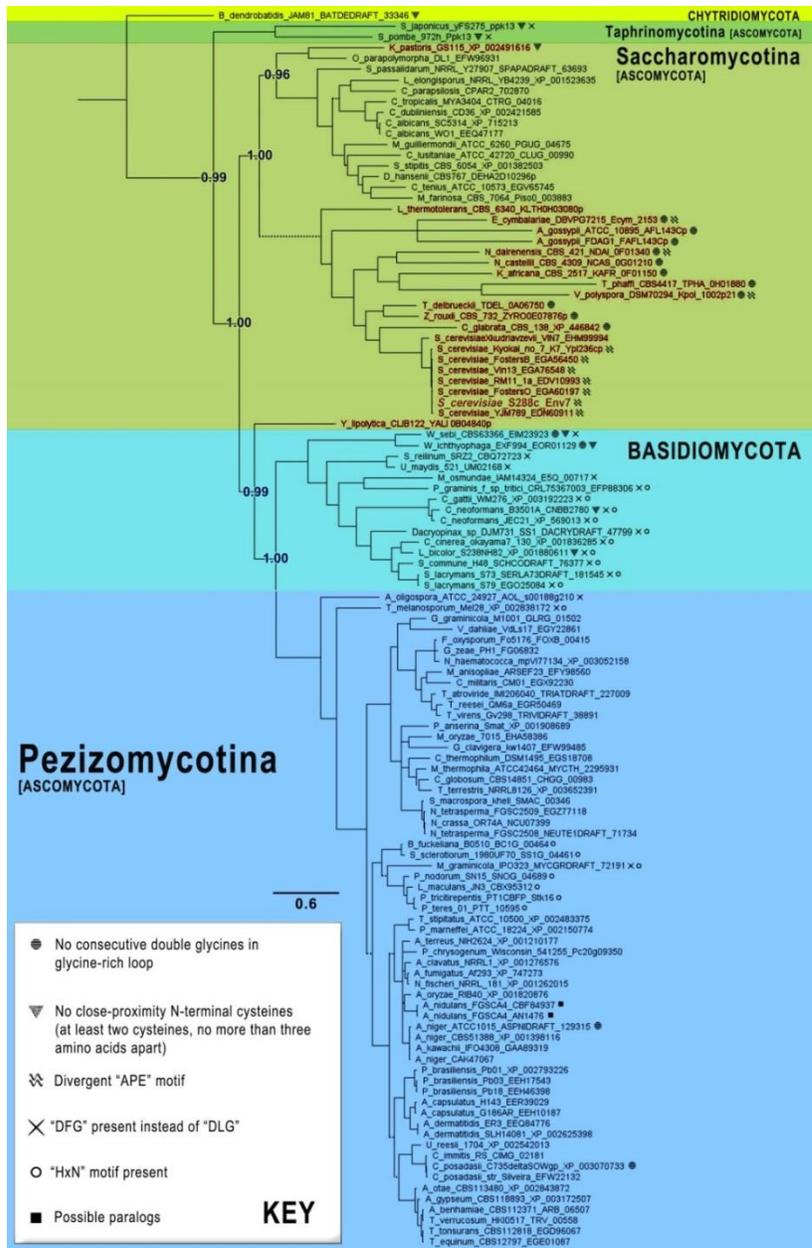


FIGURE 2. A Bayesian phylogenetic tree depicting Env7 and its fungal homologs. A MAFFT multiple sequence alignment of the Env7 sequence and 114 of its representative fungal homologs (orthologs and paralogs) was subjected to Bayesian analysis to produce a phylogenetic tree. The tree was rooted on the basal chytrid fungus, *Batrachochytrium dendrobatidis* (Förster et al., 1990). Dotted branches represent clades with less than 90% posterior probability. Some posterior probability values are shown for specific nodes. The scale bar depicts the branch length on the tree that signifies 0.6 amino acid substitutions per site. Symbols after sequence names signify a particular characteristic of the protein sequence, described in the key. Clades are color-coded by the phylum (all capital letters) or subphylum of its constituent species. Sequences belonging to Saccharomycetaceae family organisms are highlighted in red-orange.

sediment dweller that prefers a temperature of 30°C and a pH of 7.5, so it is possible that it and *S. cerevisiae* could have shared the same environment at some point in their evolutionary histories, allowing horizontal gene transfer to occur.

The presence of three Env7 paralogs in an early-diverging fungus, *Mucor circinelloides* (HMPREF1544_04705, HMPREF1544_01328, and HMPREF1544_05109), was another interesting discovery. These sequences are clearly not positioned basally to the fungal sequences on the tree (Figure 3). As no evidence of genome duplication for *M. circinelloides* exists, these paralogs might have been duplicated specifically within this species, acquired via lateral gene transfer, or a combination of both (Hedges, 2002).

The N-Terminal Consecutive Cysteine Motif is Well-Conserved in Most Fungal and Metazoan Orthologs of Env7

The short, N-terminal pre-kinase domain portion of the multiple sequence alignment of Env7 and its orthologs was poorly aligned (results not shown). Generally, only orthologs within the same family had very similar N-terminal elements upstream of their kinase domains, suggesting that this region confers organism-specific functions. For this reason, the codons in this part of the protein were not used in the selection analyses described later in this chapter and illustrated in Figure 25 of Appendix C. The most apparent similarities in this region of the aligned sequences were multiple cysteine residues. These conserved cysteine residues are significant because studies have shown that they are necessary for palmitoylation of Env7 and its human ortholog, STK16 (Berson et al., 1999; Roth et al., 2006; Manandhar et al., 2013). The addition of palmitic acid to a terminal region of a protein can anchor the protein in a membrane (Linder, 2007). Palmitoylation is the only known reversible fatty acid modification of proteins,

and its regulatable dynamic nature plays a large role in intracellular trafficking and signaling (Dunphy and Linder, 1998; Smotryst and Linder, 2004; Rocks et al., 2005).

Studies from our lab have indicated that Env7 is a vacuolar membrane protein, and when the N-terminal palmitoylation of the cysteine residues is compromised the kinase becomes cytosolic and non-functional (Manandhar et al., 2013).

At least two cysteine residues, no more than three amino acids apart, were conserved in most of the fungal and metazoan sequences (Figures 2 and 3). The absence of these residues in the bacterial ortholog is understandable due to the lack of endomembranes in prokaryotes, but the dearth of these palmitoylatable cysteine residues in plants is not so easily explained. Recent studies into the specificities of plant palmitoylation have uncovered novel sequence and structural signatures necessary for the addition of palmitate molecules to proteins (Hemsley et al., 2013); however, none of these features are present in the viridiplantae orthologs of Env7 (Figure 3). All but one (*Micromonas pusilla* MICPUCDRAFT_34378) of these plant orthologs contains a glycine directly after the initiator methionine residue, so it is probable that they are myristoylated. Addition of a myristoyl group would tether them to membranes, though not in a reversible fashion such as in palmitoylation. It is also conceivable that the viridiplantae orthologs of Env7 are less dynamically regulated by fatty acid addition than homologs in fungi, metazoan, and protists.

The chordates had either CVC or CIC conserved on their N-termini, while the fungal sequences tended to have CxxC. Env7 and its orthologs in *Torulasporea delbrueckii*, *Z. rouxii*, and *Naumovozyma dairenensis* were the only sequences with three or more consecutive cysteines in this region. Not enough is currently known about the

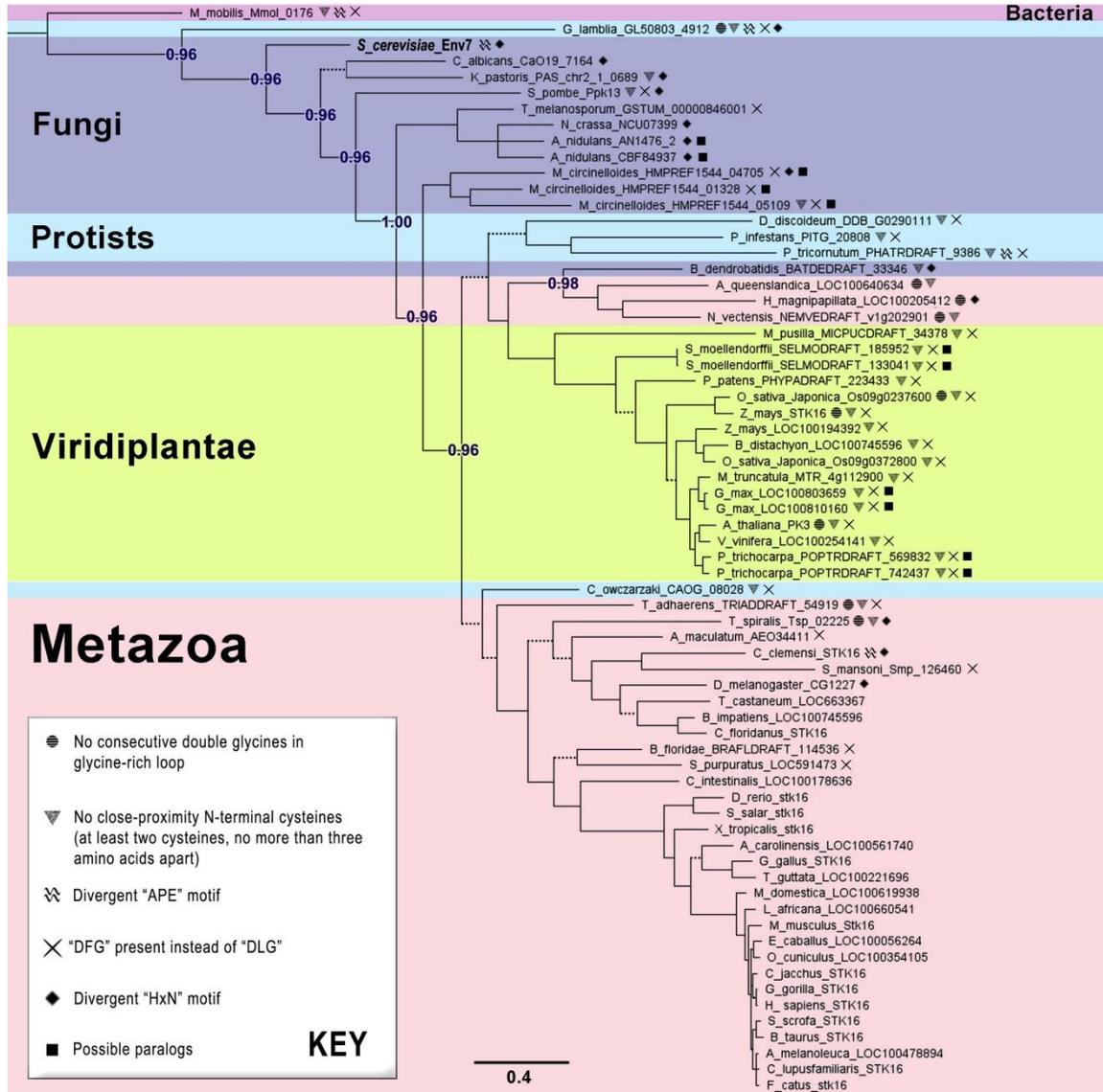


FIGURE 3. A Bayesian phylogenetic tree depicting homologs of Env7 from Eukarya and Bacteria. A multiple sequence alignment of the Env7 sequence and 67 of its representative homologs (orthologs and paralogs) was subjected to Bayesian analysis to produce a phylogenetic tree. The tree was rooted on the *Methylobacterium mobilis* bacterial sequence (Munoz et al., 2007). Dotted branches represent clades with less than 90% posterior probability. Some posterior probability values are shown for specific nodes. The scale bar depicts the branch length on the tree that signifies 0.4 amino acid substitutions per site. Symbols after sequence names signify a particular characteristic of the protein sequence, described in the key. Clades are color-coded depending on the phylum or group of its constituent species.

localization of the other Env7/STK16 orthologs, so whether or not the sequences that do not contain conserved N-terminal cysteines are membrane-associated has yet to be determined.

Many Closely-Related Orthologs of Env7 Do Not Share the Disordered Glycines in the Glycine-Rich Loop that are Characteristic of NAK Kinases

The conserved subdomain I glycine-rich loop pattern in the vast majority of eukaryotic protein kinases is GxGxxG; however, in Env7, STK16, and the established NAK kinases, the glycine-rich loop pattern is xxxGGx (Figure 6b) (Hanks and Hunter, 1995; Chien et al., 1998; Eswaran et al., 2008). The two consecutive glycine residues in STK16 were proposed to make the loop much more flexible, although the consequence of increased flexibility, be it greater reluctance to bind or release ATP or constitutive kinase activity, has not been demonstrated (Eswaran et al., 2008).

Surprisingly, almost all of the orthologous sequences from *S. cerevisiae*'s fellow Saccharomycetaceae family members do not have a double glycine motif in their glycine-rich loops (Figure 2). The longer branch lengths of the Saccharomycetaceae sequences suggest that their rates of evolution may be faster than those of other sequences in other groups. This observation could explain why the Saccharomycetaceae sequences exhibit varying conservation the motifs represented in Figures 2 and 3. Among the eukaryotic and bacterial orthologs, absence of the disordered glycine motif does not seem to follow any pattern (Figure 3).

No Discernable Pattern of Divergent Activation Segment APE Motifs Exists Among Env7 and its Orthologs

Env7 and a few of its orthologs do not contain the classic alanine-proline-glutamic acid (APE) motif at the C-terminal anchor of their activation segment; instead,

the alanine in the first position of the motif is substituted with proline, serine, threonine, or even cysteine (Figures 2 and 3; data not shown). There was no pattern of these alanine substitutions evident among the clades represented in the Bayesian trees, and, considering that the activation segment is one of the most structurally-variable parts of a eukaryotic protein kinase, this could reflect precise substrate binding specificities of the kinases at the organismal level (reviewed in Nolen et al., 2004).

The Activation Segment Anchor DLG Motif is Conserved in Nearly All Saccharomycotina, Pezizomycotina, Hexapoda and Vertebrata Env7 Orthologs

Most eukaryotic protein kinases have a DFG motif anchoring the N-terminus of their activation loop (Taylor and Kornev, 2011), but Env7 and almost all of its orthologs in Saccharomycotina, Pezizomycotina, Hexapoda and Vertebrata contain a leucine residue in place of the phenylalanine (Figures 2 and 3). The phenylalanine residue is thought to play a role in regulation of kinase activity, and its presence in Viridiplantae, Protist, and non-Hexapoda invertebrate sequences suggests that it might be part of their catalytic regulation as well.

The Canonical HxN Motif is Absent in Env7 and Most of its Fungal Orthologs

One feature that sets most eukaryotic protein kinases apart from eukaryotic-like protein kinases is a conserved HxN motif in the α C- β 4 loop that anchors the α C-helix to the C-lobe, stabilizing the active kinase (Kannan et al., 2006). However, some kinases, most notably the human AGC Serine/threonine Protein Kinase Family, lack this motif, and it is thought that their regulation of the α C-helix and conformational stability is achieved through interaction of the kinase core with the C-tail (Kannan et al., 2008). Env7 and most of its fungal homologs do not have the conserved HxN motif; instead many of them, including Env7, have an SxY motif (S94-Y96) (Figures 2 and 3). This is

not unheard of, as the human Map Kinase Kinase-1 (MEK1) also shares this residue pattern (Kannan et al., 2008). Env7 and other eukaryotic protein kinases that do not have the classic HxN motif may secure their α C-helix via intramolecular mechanisms that are still unknown.

No Detectable Syntenic Conservation Exists Near *ENV7* Beyond its Very Close Fungal Orthologs

The Comparative Genomics (CoGe) web-tool, Genome Evolution Analysis (GEvo), was used to look for synteny in the chromosomal regions around *ENV7* (Lyons et al., 2008). Default settings for the blastz search algorithm were selected, and the adjacent regions (100kb upstream and downstream) of *ENV7*'s orthologs, along with those of *ENV7*, were submitted for analysis (Schwartz et al., 2003). Results of the analyses revealed a lack of conserved gene contiguity outside of the subphylum, Saccharomycotina (Figure 4).

The strongest syntenic linkages between chromosomal regions around *ENV7* and its orthologs were found in members of the Saccharomycetaceae family. *Ashbya gossypii* ATCC 10895, *Ashbya gossypii* FDAG1, *Eremothecium cymbalariae* DBVPG#7215, *Lachancea thermotolerans* CBS 6340, and *N. dairenensis* CBS 421 shared the highest numbers of conserved chromosomal blocks with *S. cerevisiae*. *Zygosaccharomyces rouxii* CBS 732 also exhibited a large amount of synteny with *S. cerevisiae* around *ENV7*; however the conserved portions of the *Z. rouxii* chromosome were inverted in respect to those of *S. cerevisiae*. No major gene rearrangements (i.e. more than 2-3) or gene duplications were evident in the syntenic alignments of the chromosomal region regions of these species.

The lack of synteny in fungal clades outside of Saccharomycetaceae suggests two possible scenarios. The first is that recombination rates in those particular fungal organisms were high enough to eliminate conserved gene order and proximity around the *ENV7* ortholog. The second is that the immense amount of time that passed between speciation events of the available sequenced fungal genomes is adequate to eliminate synteny outside of the subphylum Saccharomycotina. The approximate time of Saccharomycotina's emergence as a clade is rough; estimations vary from 400-1000 million years ago (Taylor and Berbee, 2006). However, more specific divergences between organisms have been calculated. For example, approximately 500 million years have passed since *Schizosaccharomyces pombe*, a member of subphylum Taphrinomycotina, and *S. cerevisiae* diverged (Sipiczki, 2000; Rhind et al., 2011). And, in spite of the strong sequence similarity of *ENV7* to even mammalian orthologs, syntenic conservation is limited to the Saccharomycotina.

Additionally, the guanine and cytosine (GC) usage in the wobble codon was examined using GEvo, and it was found to vary quite a bit among *Env7* and its orthologs, especially the protists and single-celled eukaryotes (data not shown). Even closely related organisms, like *S. cerevisiae* and *T. delbrueckii* had different biases at the wobble codon position; *S. cerevisiae ENV7* favored adenine and thymine (AT: 64.95%; GC: 35.07%), while the *T. delbrueckii* ortholog used AT and GC almost equally in the wobble position (AT: 48.79%; GC: 51.21%). Wobble codon bias is important to analyze because it can provide a clue to the type of gene transmission that occurred for a set of homologs. If the wobble GC preference was found to be the same for two genes in distantly-related

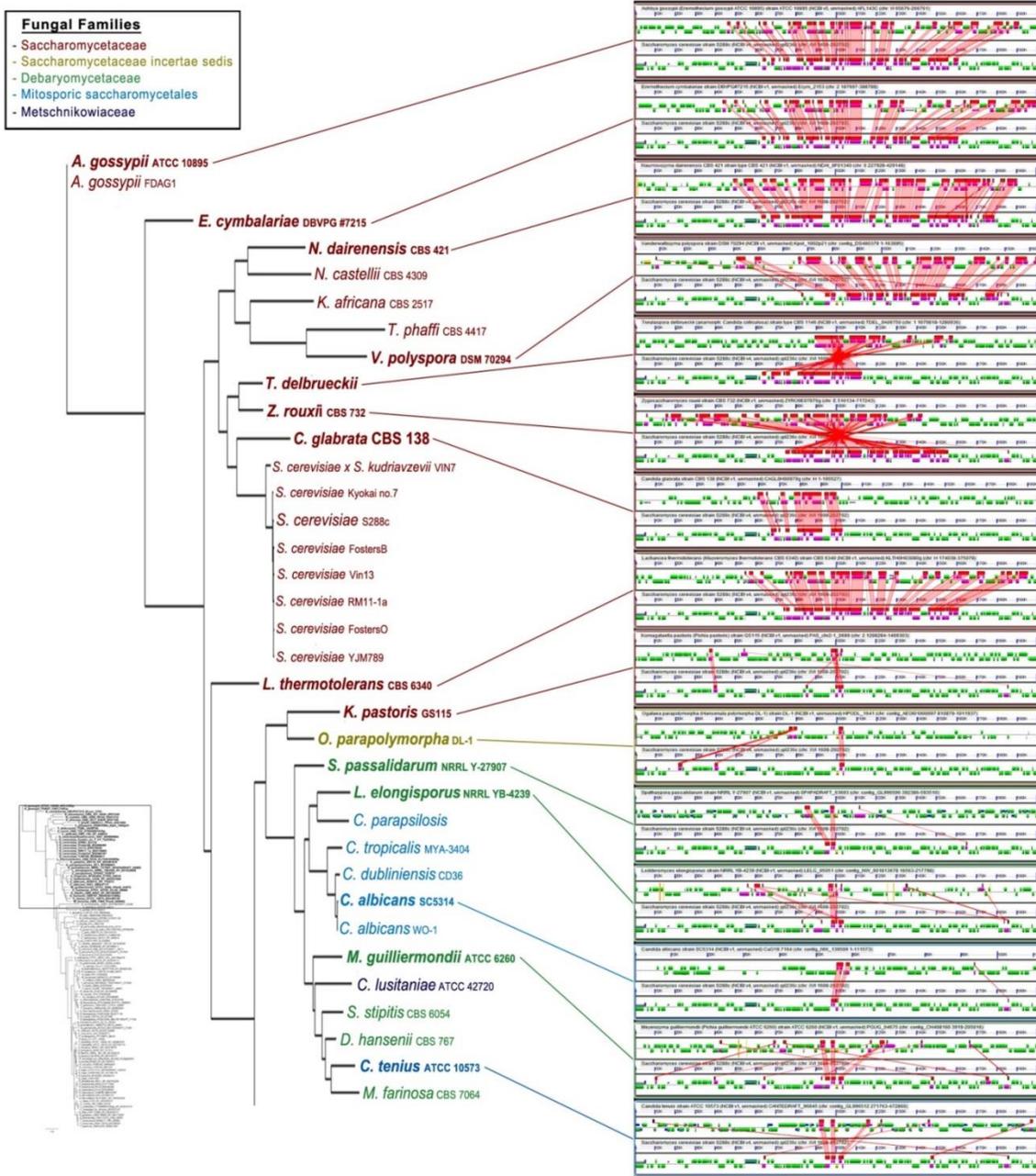


FIGURE 4. Genomic synteny is limited to *ENV7* and its close fungal orthologs. All orthologs of *ENV7* were subjected to syntenic analysis using the GEvo program, hosted on <http://genomeevolution.org/CoGe/>. The results are listed above on the right. Each rectangle section has a top and bottom panel; the bottom panel is *ENV7* and 100 kilobases (Kb) upstream/downstream, and the top panel is the ortholog and 100Kb upstream/downstream. The small green and purple blocks are coding sequences, with purple blocks indicating syntenically conserved genes. Pink wedges connecting the top and bottom panel genes represent conserved syntenic blocks between the genomes. On the right, a portion of a maximum likelihood phylogenetic tree (shown small, at the bottom left) is matched to the GEvo results.

organisms—like humans and yeast—while the rest of the flanking genes in one of the organisms showed a vastly different wobble codon bias, it would suggest that horizontal transfer had occurred. Unfortunately, there is no real pattern in wobble codon bias among *ENV7* and its orthologs, so no inferences about horizontal gene transfer can be made (results not shown).

Individual Kinase Subdomain Analyses of Env7 Revealed Conserved Features and Evidence of Purifying Selection

For the study of Env7's protein features, two multiple sequence alignments—the fungal alignment of 114 sequences orthologous to Env7 and the alignment of the kinase domains of 67 Env7 orthologs from the Bacteria and Eukaryota—were assessed. From these multiple sequence alignments, well-conserved residues were uncovered and evolutionary selection analyses were executed using the Datamonkey webserver and the HyPhy program package (Kosakovsky et al., 2005; Pond et al., 2005; Delport et al., 2010). The specific methods used for examining evolutionary selection were Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), Single Likelihood Ancestor Counting (SLAC), and Property Informed Models of Evolution (PRIME) (Kosakovsky et al. 2005; Murrell et al. 2013). FEL, FUBAR, and SLAC are each different methods for detecting positive, negative, or neutral evolution among aligned amino acid codons. PRIME, which employs the same algorithms as FEL and SLAC, specifically looks for conserved amino acid properties among codon inputs. Using the eukaryotic protein kinase subdomains as a framework, sections of aligned codon sequences from the two domains of life were inputted into the four programs. The results are shown in Figure 25 in Appendix C.

Overall, negative, or purifying selection was responsible for the conservation of many amino acids in the alignment of 67 orthologous sequences and Env7 (Figure 25 in Appendix C). This means that any deleterious nucleotide substitutions were most likely eliminated over the course of evolution of the sequences. No positive selection was detected by any of the programs.

The Relationship of *ENV7* to the NAK Kinase Family Members

The NAK Kinase Family's existence relies almost entirely on the kinase domain sequence similarity of its members. All of the NAK kinase sequences contain a well-conserved N-terminal serine/threonine kinase domain, sometimes followed by up to ~1000 amino acids; however, Env7, STK16, and their orthologs only have the kinase domain with relatively few flanking amino acids (Figure 5a). It is the conserved kinase domain that provides the sequence similarity among the members of the NAK Kinase family. The C-terminal portions of the NAK Kinase sequences vary significantly in length and composition. These areas generally have low sequence identity to the same regions of their respective orthologs. In fact, blastp searches for orthology analyses with a strict lower limit of 70% query coverage (instead of 60%) yielded interesting results. Orthologs for *S. cerevisiae* Prk1 were only found in the genera *Millerozyma* (*Pichia*), *Kazachstania*, *Kluyveromyces*, *Naumovozyma*, and *Saccharomyces*. Ark1-like sequences were only present in *Naumovozyma* and *Saccharomyces*, and Ak11 matches were found only in *Candida glabrata* and *Saccharomyces*. Taken together, Ark1 and Ak11 did not have any orthologs that aligned to their complete sequences –kinase domain and large downstream sequence region– outside of Saccharomycetaceae, but Prk1,

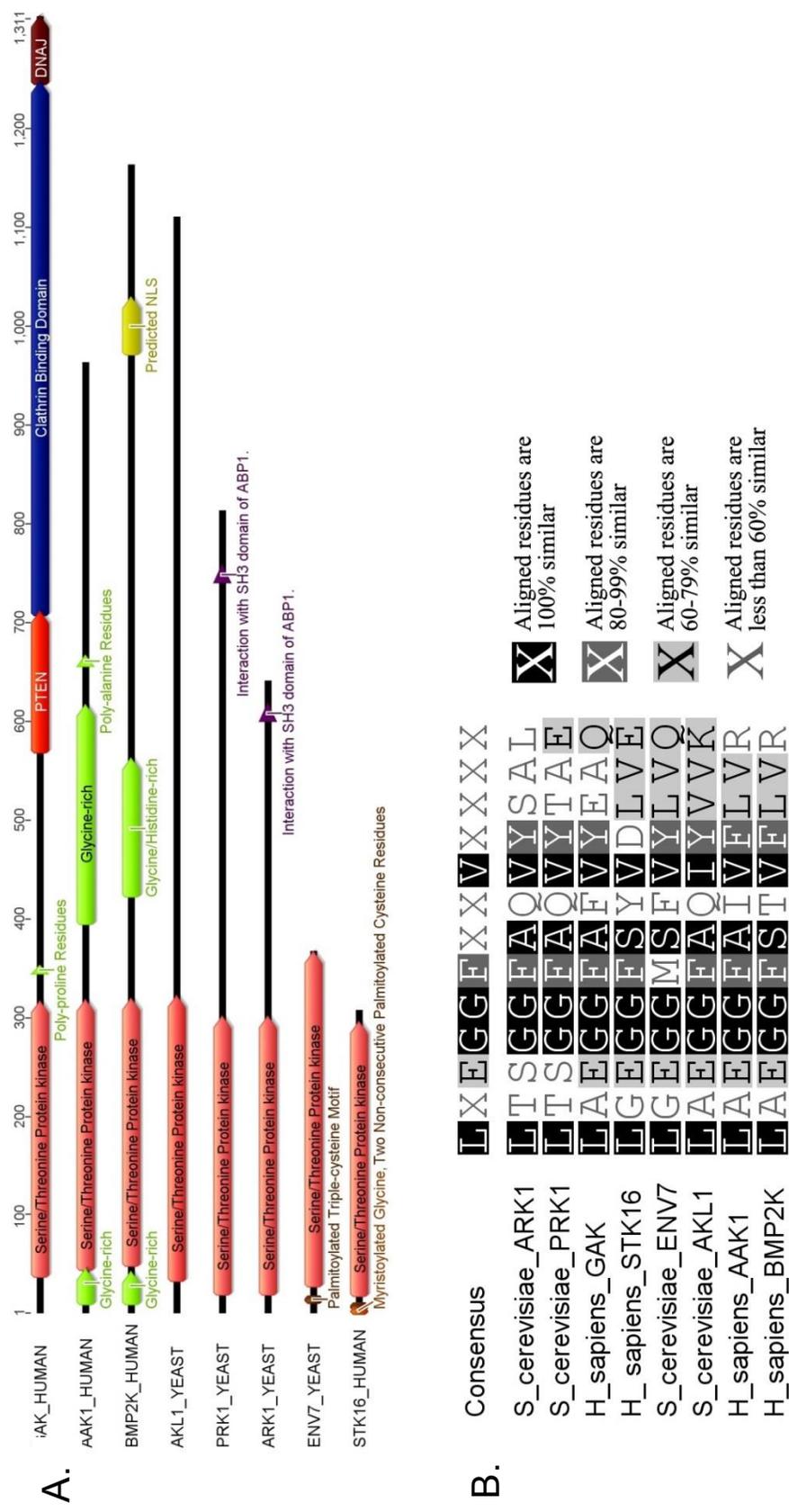


FIGURE 5. Comparisons of human and *S. cerevisiae* NAK Kinase protein sequences. A) A visual comparison of annotated NAK kinases shows the well-conserved N-terminal kinase domain. B) Multiple sequence alignment of the glycine-rich loops of the human and *S. cerevisiae* NAK kinases illustrates the conserved double-glycine motif.

though mostly conserved in Saccharomycetaceae, had one complete ortholog in Debaryomycetaceae. The human NAK kinases were a little more broadly distributed; GAK was limited to metazoans, and AAK1 was found in complete form in osteichthyes only. BMP2K was not conserved in full outside of Sarcopterygii (results not shown)

Despite the lack of complete sequence conservation, the NAK kinases do form a distinct clade, with specific subfamilies spanning Eukarya (Figure 6). The Ark/Prk Subfamily of sequences clustered most closely with AAK1 and BMP2K sequences, while GAK/Auxillin and STK16/Env7 orthologs remained as their own more evolutionarily separated groups (Figure 6). The STK16/Env7 clade was the only one that included plant, protist, and bacterial sequences, and the relatively long branch lengths of some of the sequences in the group suggest a rapid mutation rate and possible functional divergence (Figure 6).

It is difficult to trace the evolution of the separate NAK Kinase subfamilies with respect to each other because there is no immediate relative. No orthology or paralogy exists between the subfamilies; though the kinase domains by themselves are similar within the NAK Kinase family. Paralogy definitely exists within protein subfamilies, suggesting possible neo- or sub-functionalization (Figure 6) (Zeng and Cai 1999; Byrne and Wolfe, 2005). Even more peculiar, the non-kinase portions of the NAK Kinases display little resemblance to each other among subfamilies, so their C-terminal attachment to the conserved kinase domain is not the product of one or two gene fusion events.

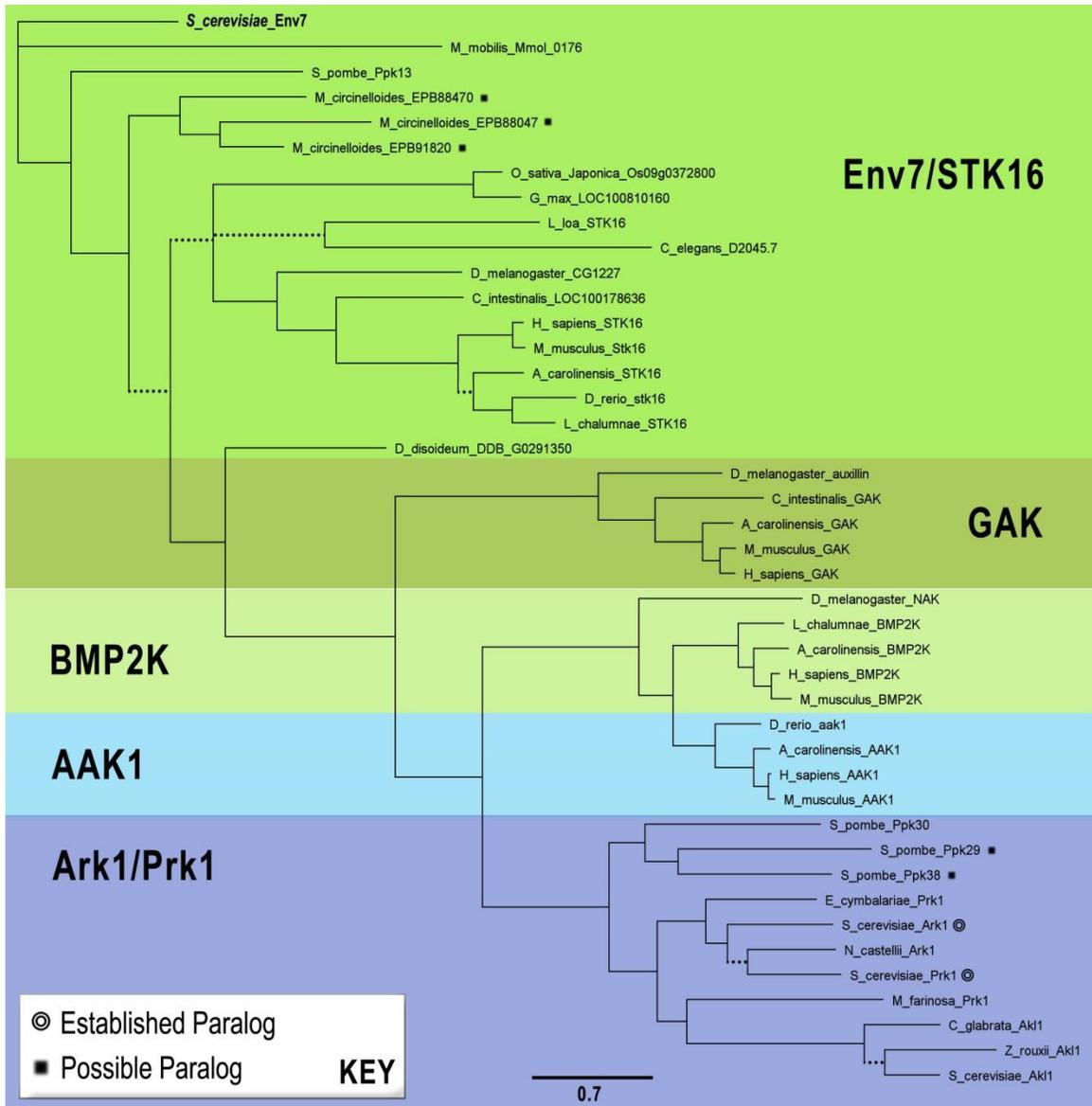


FIGURE 6. An unrooted Bayesian phylogenetic tree of select NAK family protein sequences. Sequence names consist of genus abbreviation, species, and gene name/locus. Dotted lines represent clades with less than 90% posterior probability, and the square symbol denotes sequences belonging to a single species, making them likely paralogs. The scale bar represents the number of amino acid changes per site.

Important Sequence and Structural Aspects of Env7

Well-Conserved Residues in Env7 are Appropriate to Form Salt Bridges Characteristic of Active Eukaryotic Protein Kinases

Intact salt bridges have been shown to be extremely important for protein kinase catalytic activity because they maintain protein kinase stability and protect parts of the catalytic loop from the destabilizing influence of solvents (Yang et al., 2012). Env7 contains four well-conserved residues that could form such salt bridges: lysine 69, glutamic acid 85, glutamic acid 269, and arginine 349 (Figure 25 in Appendix C). These residues are highly conserved via purifying selection in almost all of Env7's orthologs (Figure 25 in Appendix C); so mutations of these residues are likely to greatly affect kinase activity. A study from our lab found that when glutamic acid 269 is mutated to alanine, Env7 becomes completely unstable and lacks kinase activity; however, mutations substituting aspartic acid or serine for the glutamic acid do not render Env7 unstable or non-functional (Manandhar et al., 2013). Point mutations associated with the conserved glutamic acid 269-arginine 349 (Env7 numbering) salt bridge in human protein kinases result in disease (Torkamani et al., 2008).

Two extremely well-conserved STK16 residues, corresponding to lysine 69 and glutamic acid 85 in Env7, were shown to form a salt bridge in the crystalline structure (Eswaran et al., 2008). The residues are close enough on the *in silico* Env7 structure to form a bridge as well (Figure 9), and it has been shown that in cAMP-dependent protein kinase catalytic subunit alpha (PKAC α) and many other protein kinases, the existence of this salt bridge between the N- and C-lobes is critical for a catalytically active kinase (Kornev et al., 2006; Taylor and Kornev, 2011). Therefore, mutation of either of these

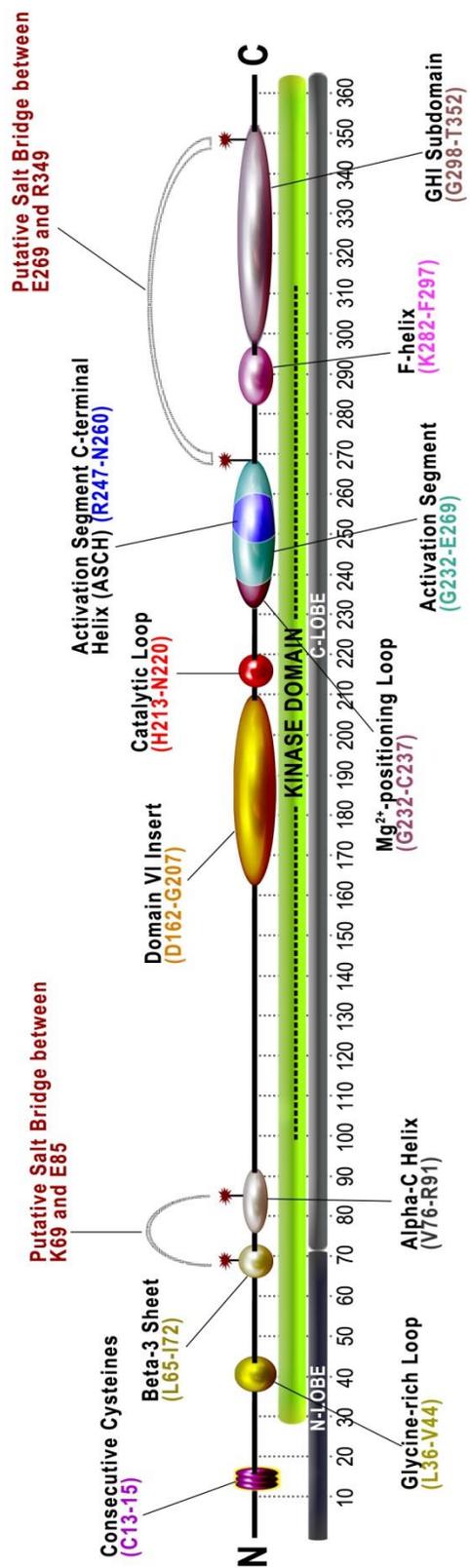


FIGURE 7. An overview of the Env7 protein and its conserved residues and features.

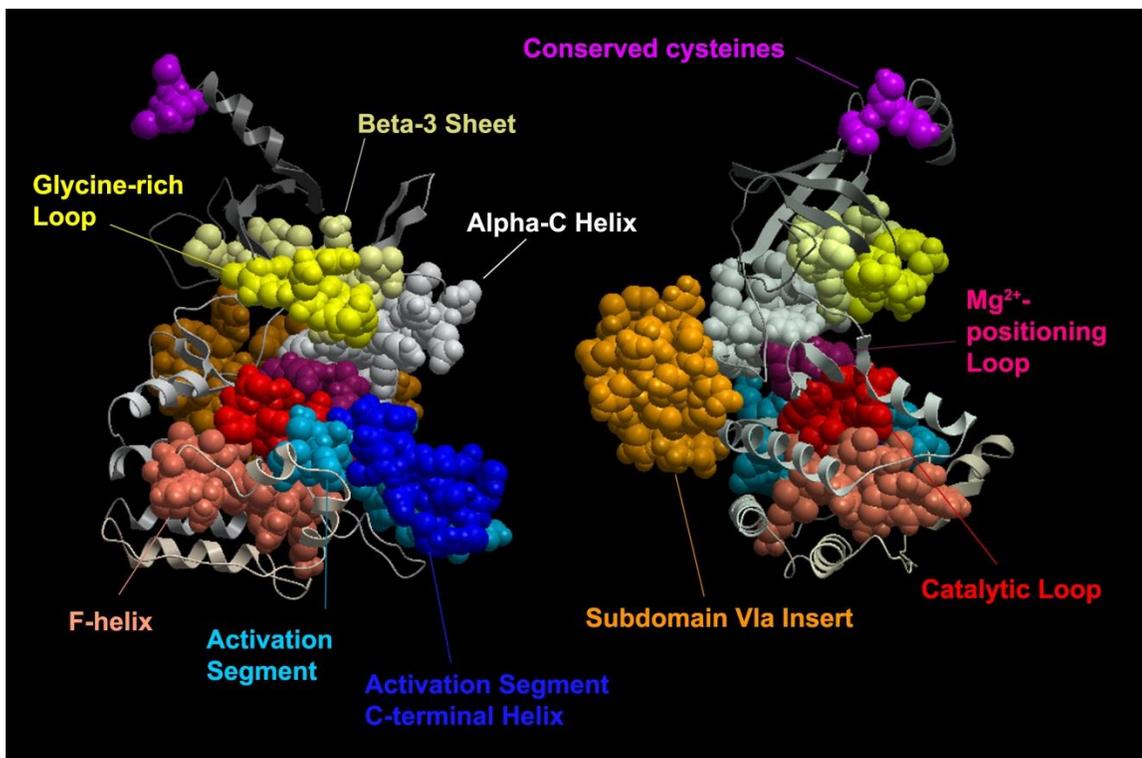


FIGURE 8. Three-dimensional rendering of conserved features in Env7. Nine important structural components in Env7p are illustrated above in CPK representation. Residues that are not part of these features are shown in the ribbon representation. Colors of each feature coordinate with those displayed in Figure 7.

residues to an uncharged and/or smaller amino acid could abolish kinase activity in Env7. Studies in our lab have shown that the Env7K₆₉M mutant protein is a stable, functional kinase, suggesting that this second salt bridge may not be essential Env7 function (Manandhar et al., 2013).

Env7's Catalytic Spine, Regulatory Spine, Magnesium-Positioning Loop, and Activation Segment are Composed of Conserved Residues

The catalytic spine (C-spine) is a spatially-conserved structure in protein kinases that facilitates docking of the adenine ring of ATP (Kornev, 2008). It does this by forming a backbone between the N and C lobes that is only completed by the direct binding of the adenine ring. Based on pairwise protein sequence alignment with human PKAC α , Env7p has eight residues—valine 44, alanine 67, leucine 127, isoleucine 221, leucine 222, phenylalanine 223, threonine 291, and leucine 295—that make up its catalytic spine (Figure 10). This suggests that valine 44 and alanine 67, both in the N-lobe, directly bind the adenine ring during catalysis. Both residues are highly conserved in fungi, and alanine 67 is particularly well-conserved throughout most of Env7's orthologs. Leucine 127 is well-conserved in most of Env7's orthologs (Figure 25 in Appendix C). Interestingly, the rest of the proposed C-spine residues in Env7 are not well-conserved in many of Env7's orthologs, but studies have shown that sequence position-specific conservation of an amino acid is not as important as the spatial, three-dimensional conservation of an amino acid in the C-spine (Kornev, 2008; Kornev and Taylor, 2010). The two N-lobe components of the C-spine (V44 and A67) must be conserved as specific amino acids most likely because they are the only ones that interact with the ATP molecule directly (Kornev, 2008). Though the PRIME analysis did not detect any

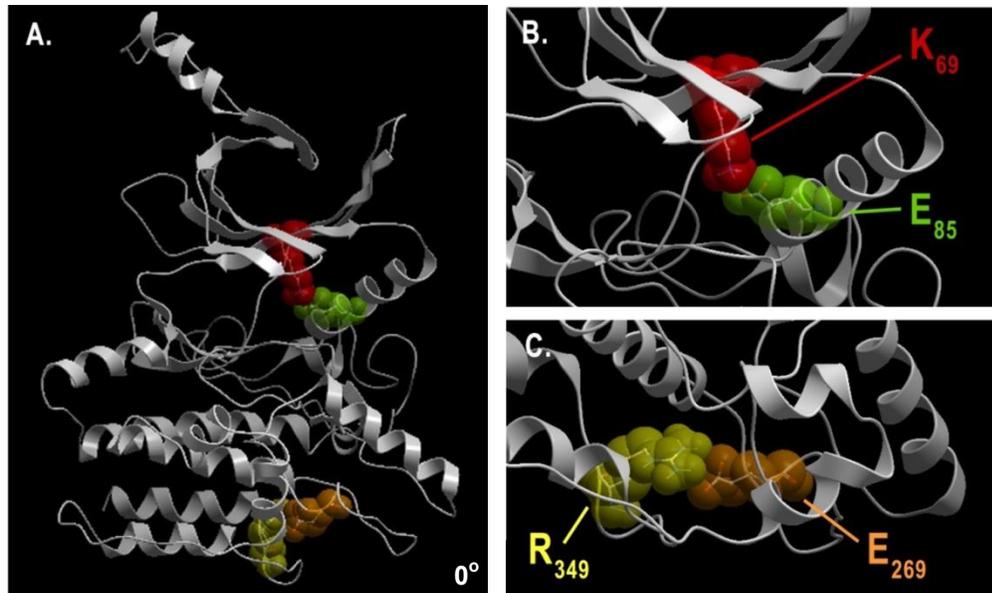


FIGURE 9. Three-dimensional rendering of Env7 featuring the putative conserved salt bridges. A) Full view of Env7, with conserved structurally-stabilizing salt bridges highlighted in translucent color space-filling model. B) The putative salt bridge between lysine 69 (red) and glutamic acid 85 (green) connects the N- and C-lobes in catalytically active protein kinases. C) The putative salt bridge between glutamic acid 269 (orange) and arginine 349 (yellow). The ball and stick models of the residues are visible through the space-filling representation.

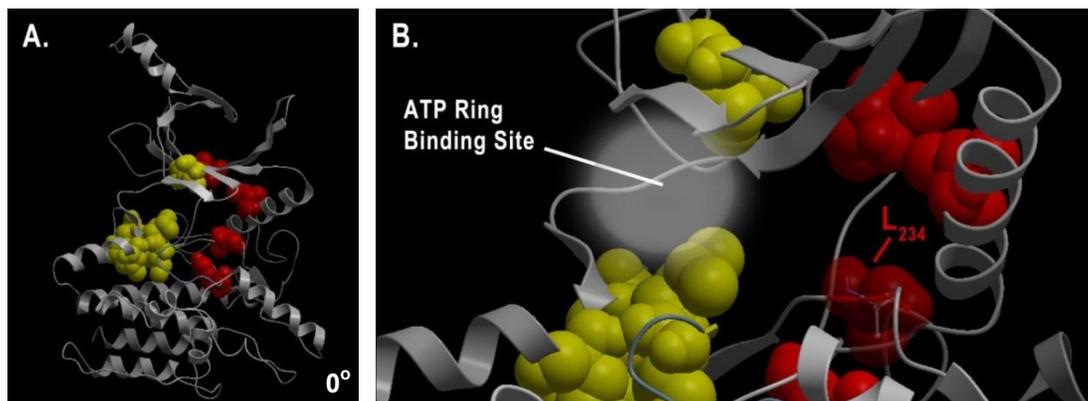


FIGURE 10. Three-dimensional rendering of Env7 featuring the catalytic and regulatory spines. A) Full views of Env7p with C-spine residues (yellow) and R-spine residues (red) highlighted by space-filling model representation. B) A closer view of the conserved C- and R-spine residues shows that they do form spine-like structures that connect different parts of Env7. The C-spine residues only form a complete spine when the ring of an ATP molecule binds the protein during catalysis. The R-spine controls the catalytic activity of the kinase, and most eukaryotic protein kinases contain a phenylalanine residue that acts like a switch to determine if the kinase will be active (Taylor et al., 2012). Env7 and its orthologs have a leucine (234 in Env7) in place of this phenylalanine.

conserved amino acid properties at these two positions—possibly because there was not enough variation between amino acids at these positions in the alignment—mutation of these residues to drastically different amino acids could alter the catalytic activity of Env7 by affecting the binding of the ring portion of ATP (Figure 25 in Appendix C).

Another important structure conserved in active protein kinases is called the regulatory spine (R-spine). This arrangement of five conserved hydrophobic amino acids creates a link between the N and C-lobes, and without this rigid constraint the kinase loses its catalytic activity (Johnson et al., 2001; Huse and Kuriyan, 2002; Taylor and Kornev, 2011). In Env7, the residues that make up the R-spine, again based on pairwise similarity to PKA, are: isoleucine 86, isoleucine 101, histidine 213 (part of the HRD motif), leucine 234, and aspartic acid 284 (Figure 10). Env7, as well as its many orthologs, has a distinct variation; leucine 234—in the DLG motif of the activation segment—is actually phenylalanine in almost all other protein kinases (Figure 25 in Appendix C) (Taylor and Kornev, 2011). The importance of this difference is that the phenylalanine has been proposed to be a catalytic switch for protein kinases (Taylor and Kornev, 2011).

Subdomains VIB and VII contain residues that chelate Mg^{2+} ions in order to properly position the γ -phosphate of ATP for phosphoryl transfer (Hanks and Hunter, 1995). Env7 has two well-conserved residues in these subdomains—aspartic acid 233 and asparagine 220—that align with the Mg^{2+} -coordinating residues on PKA (Figure 26 in Appendix C). As seen in Figure 25 in Appendix C, there is evidence of purifying selection at these residues as well.

Spanning subdomains VII and VIII, the activation segment is a major characteristic that distinguishes eukaryotic protein kinases from eukaryotic-like protein kinases, and it is controlled by phosphorylation of residues within it (Taylor, 2012). Only residues towards the N- and C-terminal anchors—DFG and PPE in Env7, respectively—are well conserved (Figure 25 in Appendix C). This paucity of strongly conserved residues may reflect a genus- or species-specific nature catalytic regulation via phosphorylation of the activation segment. Additionally, the activation segment of Env7 and most of its orthologs is also unusual in that, instead of a simple helical turn, their activation segments seem to contain a full alpha helix towards their C-terminus (Figure 8) (Scheeff and Bourne, 2005; Eswaran et al., 2008). The significance of this extended activation segment in Env7 orthologs may be associated with regulation of kinase activity.

The Unique Subdomain VIA Insert in Env7

A feature exclusive to Env7 and most of its fungal orthologs was revealed after evaluating the various multiple sequence alignments generated in this study. An insert from aspartic acid 162 through glycine 207 in Env7, predicted to consist of coils, turns, and a small helix, became apparent (Figure 11a). This insert is highly variable even among closely related Saccharomycetaceae Family orthologs (Figure 11b). Only glutamic acid 197 in Env7 is well-aligned with identical or similar residues in the other sequences, and no clear conservation of hydrophobic properties exists in the alignment of these inserts either (Figure 11b). Interestingly, there seems to be an acidic di-leucine sorting signal conserved in the unique subdomain VIA insert of Env7 and some of its Saccharomycetaceae orthologs (Figure 11b). This sorting signal, which has a consensus sequence of [D/E]XXXL[L/I/V], has been shown to interact with the delta subunit of the

AP-3 clathrin-associated protein complex (Vowels and Payne, 1998; Darsow et al., 1998).

Proposed Phosphorylation Sites in Env7

The activity of almost all eukaryotic protein kinases is regulated in some way by phosphorylation of serine, threonine, and/or tyrosine residues within the kinase itself (reviewed in Rubenstein and Schmidt, 2007). In general, if at least one residue in the activation segment is not phosphorylated, the kinase cannot complete the R-spine configuration and is not activated (Johnson et al., 1996; Yang et al., 2011; Taylor et al., 2012). However, other phosphorylatable residues outside of the activation may also provide multiple layers of regulation by phosphorylation (Prowse et al., 2001; reviewed in Nolen et al., 2004). Env7 has been shown to be phosphorylated *in vivo* and *in vitro* (Manandhar et al., 2013; Manandhar and Gharakhanian, 2013). Using a well-rounded group of computational methods (listed and described in in Table 11 of Appendix A), Env7 was analyzed *in silico* to determine which residues are most likely to be phosphorylated. When evaluating the most likely candidate for phosphorylation, residue conservation was also taken into consideration, even though some phosphorylated residues are expected to be unique based on the nature of substrate binding.

Figure 12 shows the positive hits for phosphorylation for each phosphorylatable residue in Env7 along with the percentage conservation of that residue in an alignment of 151 of Env7's orthologs. Serines dominate the list of potential phosphorylation targets in Env7. Six residues –serine 126, threonine 194, threonine 217, serine 238, serine 280, and serine 331– stand out with the most positive predictions by the algorithms. Among these residues, only serine 238 is within the activation segment. This residue is also not well-

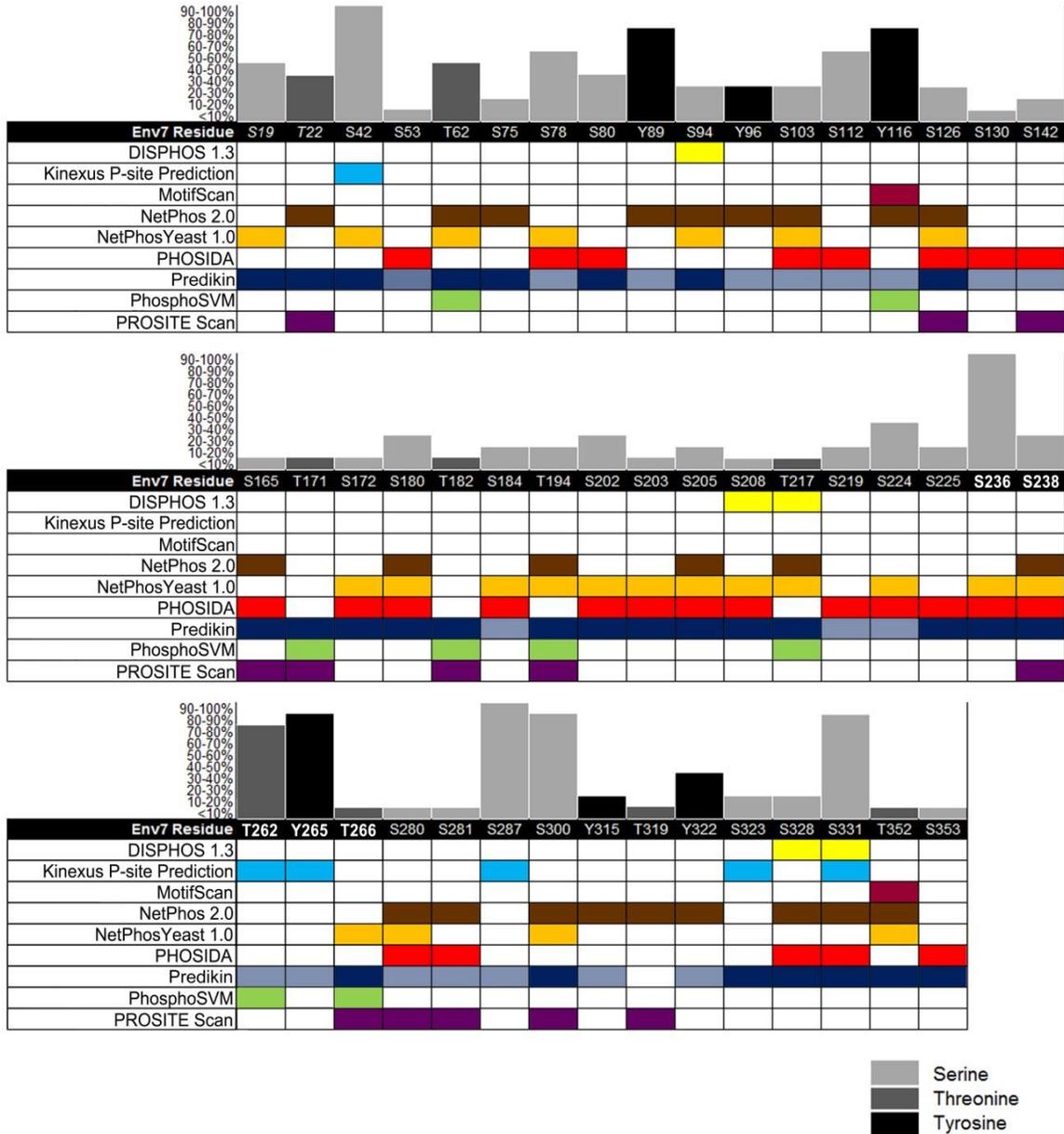


FIGURE 12. Prediction of putatively phosphorylated residues in Env7. Nine phosphorylation prediction programs, listed at the left, were used to determine possible phosphorylation sites on Env7p. Percent conservation, indicated by the bar graphs above the residues, is based on multiple sequence alignment of Env7p and 151 of its orthologs (eukaryotic and bacterial). For the Predikin program, darkly shaded blue boxes represent sites chosen by the SDR method, and lightly shaded blue boxes represent sites chosen by the KDS or PANTHER methods, which are less refined. Residues in the activation segment are in bold, and residues that are not part of the kinase domain are italicized.

conserved, even among close orthologs, but conservation of residues in the activation segment is not extremely important because is it very often unique in each organism (Taylor et al., 2013). Threonine 194 falls within the subdomain VIA insert region that is unique to Env7 and a few of its close orthologs in Saccharomycetaceae, so phosphorylation of this residue could be involved with protein-protein interactions or localization. Serine 217 is in the catalytic loop at the base of the ATP-binding cleft. Serine 331 is in the GHI subdomain, but it is unclear how phosphorylation of these residues could affect Env7 activation or function. These findings are consistent with experiments in our lab that have shown that at least the C-terminal fragment of Env7 is phosphorylated (S. Manandhar, unpublished results). The rest of the residues are not in any definable area of Env7, so as of now the role of their potential for phosphorylation is also unknown.

Env7's Disordered Glycine-rich Loop

The first and second glycines in Env7 (G37 and G39) are the most highly conserved in almost all eukaryotic protein kinases, and their mutation to different amino acid residues in other eukaryotic kinases alters catalytic activity (Hemmer et al., 1997; Davies et al., 2002). The third glycine (G40) is more variable (reviewed in Bossemeyer, 1994). This pattern of conservation was not observed in Env7 orthologs; seven of the nine residues in the glycine-rich loop are conserved in Env7 orthologs (Figure 25 in Appendix C). Negative selection seems to be responsible for the preservation of these glycine-rich loop residues, which form the consensus sequence LGEGGxSxV (Figure 25 in Appendix C). Conservation of amino acid volume is a major trend in the glycine-rich loop region—a possible consequence of its role in ATP-binding. The glycine-rich loop

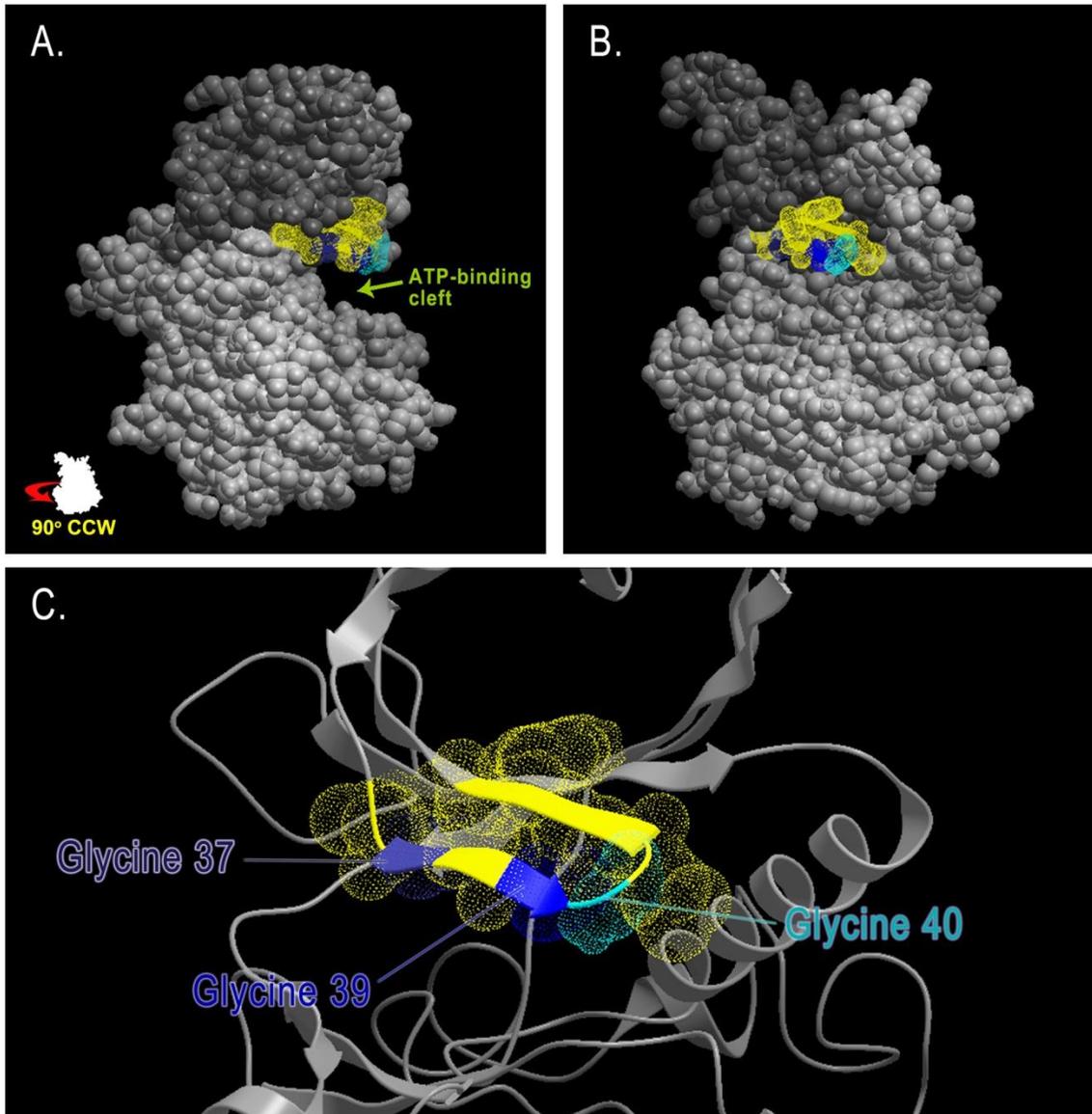


FIGURE 13. Three-dimensional rendering of Env7 featuring the glycine-rich loop. The loop is highlighted in color, with the three conserved glycines in various shades of blue and the rest of the residues in yellow. A) and B) A view of the whole protein as a space-filling model shows that the glycine-rich loop is positioned above the ATP-binding cleft, between the N- and C-lobes. C) A closer view of the glycine-rich loop, with the dot envelope representation emphasizing its specific residues.

must be flexible enough to fold over ATP as it binds in the catalytic cleft, and glycine is known to confer flexibility to protein structure (Figure 13) (Hanks and Hunter, 1995; Yan and Sun, 1997). Over the course of evolution, larger amino acid substitutions in this region may have been disadvantageous because they interfered with ATP binding in the catalytic cleft.

Experimental Assessment of the Significance of Glycine 40 in the Glycine-Rich Loop

According to the few publications that mention the NAK kinase family, *ENV7*'s membership is based upon the conservation of two consecutive glycines in its glycine-rich loop, but the significance of these glycines has not been determined for the NAK kinases (Chien et al., 1998; Eswaran et al., 2008). The glycine-rich loop in a eukaryotic protein kinase is responsible for positioning and binding the ATP molecule during catalysis, and many studies have shown that point mutations of residues, particularly the glycine triad, this loop can affect kinase activity and cause disease (Odawara et al., 1989; reviewed in Bossemeyer, et al., 1994; Hemmer et al., 1997; Chan et al., 2002; Davies et al., 2002; Rajagopalan et al., 2002; Yuen et al., 2002; Ikenoue et al., 2004). In *Env7*, glycine 40 corresponds to the last of the three conserved glycines and, more specifically, the second of the consecutive glycines which NAK kinases have in common (Figure 6b). To probe the significance of this last conserved glycine, an HA-tagged point mutant – *Env7*G₄₀V-HA– was created and expressed in the *env7Δ* background to ensure that no endogenous *Env7* would be present in the cells. Glycine 40 was chosen to be mutated specifically to valine because it is a bigger amino acid that could create steric interference with ATP binding. It is estimated that 99% of eukaryotic protein kinases have a small amino acid, such as glycine, alanine, or serine, in this position (Hemmer et

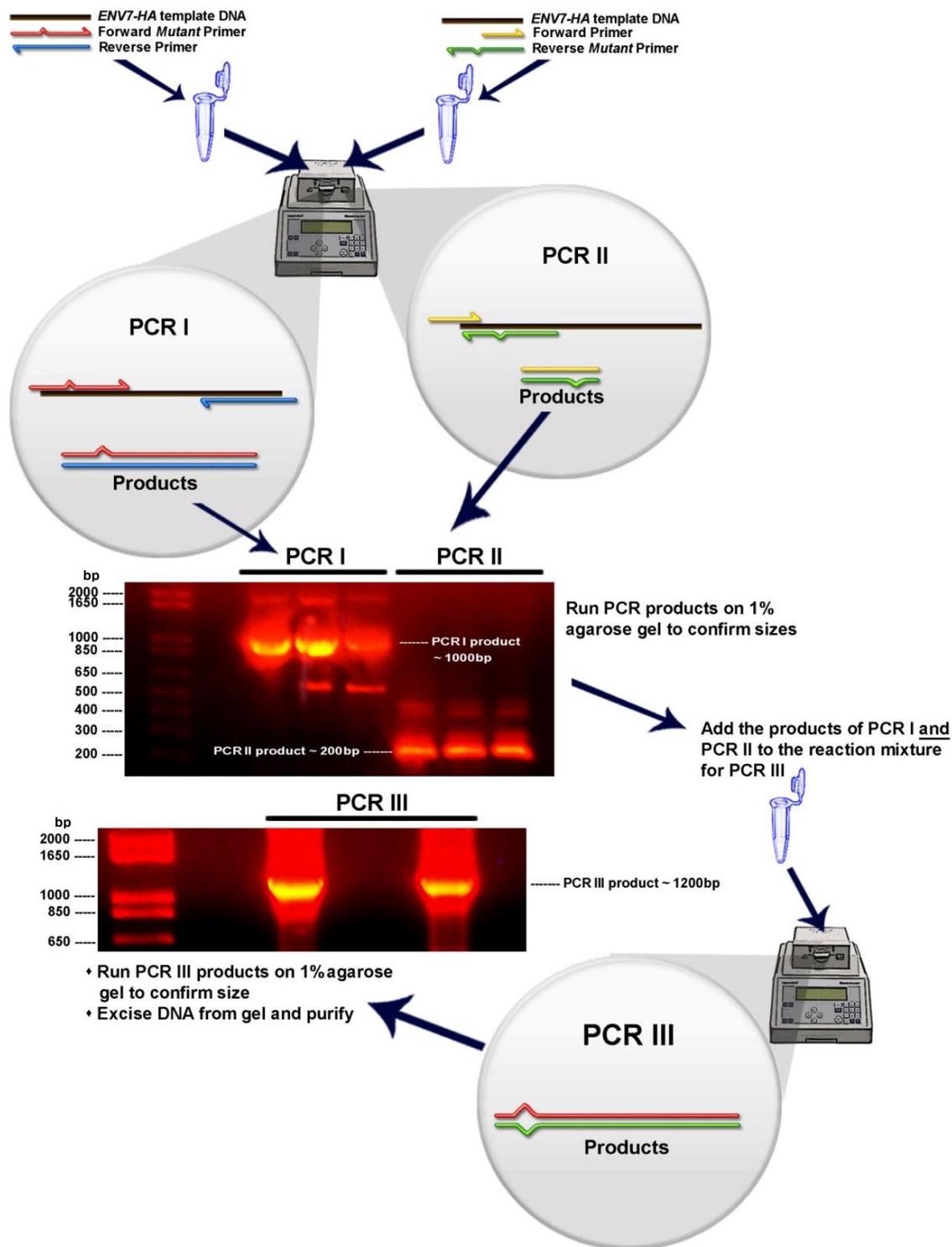


FIGURE 14. Site-directed mutagenesis protocol used in this study. Agarose gel checks of the three sets of PCR are shown above as well.

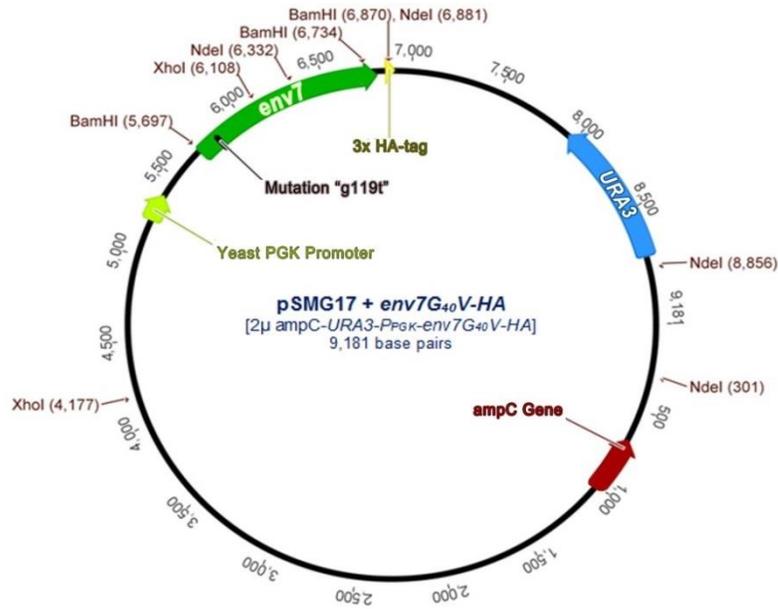


FIGURE 15. The 2μ vector, pSMG17 + *env7G₄₀V-HA*, generated by and used in this study. The *URA3* gene (encodes uracil) and the *ampC* gene (confers ampicillin resistance through β -lactamase production) were used as selective markers. The yeast *PGK* promoter sequence upstream of *env7G₄₀V-HA* controlled the transcription of the mutant gene, ensuring that multiple copies of the mutant product would be generated. Restriction digestion cut sites for restriction enzymes used in this study are indicated on the outside of the plasmid ring.

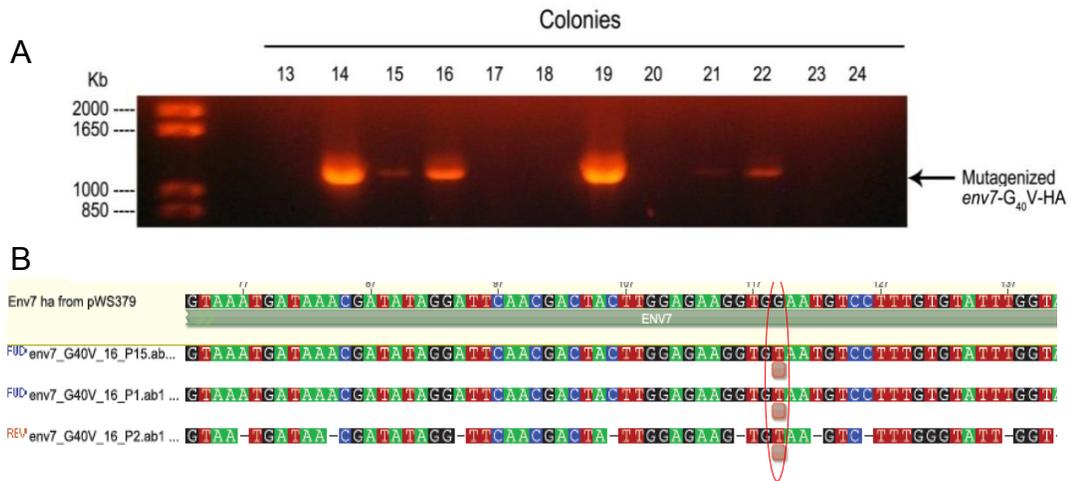


FIGURE 16. Discovery and confirmation of mutant *S. cerevisiae* cell line used in this study. A) 1% agarose gel of colony PCR products from yeast cells transfected with potentially point-mutated *ENV7* DNA showed some colonies containing appropriately-sized DNA fragments. Colonies 14, 16, 19 and 22 were chosen to be sequenced to determine if the g119t ($G_{40}V$) mutation was truly present in the DNA sequences. B) DNA sequencing (Macrogen, Korea) confirmed the presence of the correct mutation, shown circled in red on the alignment of DNA contigs provided by the sequencing company. Colony 16 cells were used for all subsequent experiments.

al., 1997). Additionally, a mutation of the third glycine of the ordered (non-NAK-like) glycine-rich loop to valine in human insulin receptor kinase was linked to insulin resistance due to reduced kinase activity (Odowara et al., 1989).

The two *S. cerevisiae* control strains expressing mutagenized and/or tagged versions of *ENV7* in the *env7Δ* background used in this study were Env7-HA and Env7C₁₃₋₁₅S-HA. Env7-HA was chosen as a positive control because it is simply wild-type *ENV7* with a triple HA tag fused to its C-terminus. Most importantly, Env7-HA was demonstrated to be a functionally active kinase *in vitro* (Manandhar et al., 2013). Also possessing a triple C-terminal HA-tag, the Env7C₁₃₋₁₅S-HA-expressing mutant was used as a negative control because the triple cysteine (C₁₃, C₁₄, and C₁₅) mutation to serines was found to inactivate the protein kinase (Manandhar et al., 2013).

Generation of the Env7G₄₀V-HA Mutant

The Env7G₄₀V-HA mutant was generated by site-directed mutagenesis. The 2μ pWS479+*RCE1-HA* plasmid, provided by Dr. Surya Manandhar, was used as the target vector for the mutagenized DNA (Plummer et al., 2006). Figure 14 details the site-directed mutagenesis process used to create a point mutation in *ENV7* rendering guanine 119 to thymine, effectively changing glycine 40 to valine in the gene product. Then, by transforming *env7Δ* cells with the site-directed mutagenesis product and linearized pWS479+*RCE1-HA* 2μ vector, the mutagenized *env7G₄₀V* gene was inserted into the vector in the place of HA-tagged *RCE1*, creating pSMG17+*env7G₄₀V-HA* as shown in Figure 15. The transformed cells were streaked onto SM-ura selective medium, and the colonies that grew in the absence of uracil were subjected to colony PCR. Using primers 1 and 2, which only bind to *ENV7* and generate a 1.19Kb product, yeast colonies with

pSMG17 containing some sort of *ENV7* gene (not necessarily mutagenized) were confirmed (Figure 16a). DNA from confirmed colonies was extracted, and the plasmids were amplified in *E. coli*. The mutagenesis was confirmed by DNA sequencing (Figure 16b). All three forms of *ENV7* used in this report –Env7-HA, Env7C₁₃₋₁₅S-HA, and Env7G₄₀V-HA– were overexpressed by a yeast PGK promoter in the 2 μ vector in the *env7 Δ* background. The two control plasmids were kindly provided by Dr. Surya Manandhar.

Analysis of Vacuole Morphology in Env7-HA, Env7G₄₀V-HA, and Env7C₁₃₋₁₅S-HA-Expressing Cells

Previous studies in our lab have shown that Env7 negatively regulates vacuole fusion (Manandhar et al., 2013). More specifically, when *env7 Δ* mutant cells are subjected to hypertonic conditions, their vacuoles initially fragment like those in WT cells; however, unlike WT cells, after 60 minutes of hypertonic treatment the vacuoles in the *env7 Δ* cells are no longer multi-lobed. Most of them contain 1-5 rounded vacuoles and are considered to have a prominent vacuole phenotype, as shown in Figure 17b. These results suggested that Env7 –though not involved with initial vacuole fragmentation– could be functioning to maintain the fragmentation of the vacuoles in hypertonically stressed cells (Manandhar et al., 2013).

To determine if the Env7G₄₀V-HA mutant cells exhibited deviant vacuole phenotypes compared to those of Env7-HA and Env7C₁₃₋₁₅S-HA cells, log phase cells in YPD were stained for 45 minutes with FM4-64. FM4-64 is a fluorescent, styryl dye commonly used for labeling membranes because its hydrocarbon tail intercalates between the lipids in the plasma membrane upon addition to liquid cell culture. This causes the lipophilic dye to enter the endocytic pathway, terminating at the vacuolar membrane

(Vida and Emr, 1995). After the staining period, the cells were washed four times with fresh YPD, and then chased for 90 minutes to allow the dye to fully migrate to the vacuole. The stained cells were pelleted and resuspended in either YPD or 0.5M NaCl YPD and allowed to incubate for 60 more minutes on a rotating platform at 30°C before viewing with the DIC/fluorescent microscope. The cells were scored as having either prominent or multi-lobed phenotypes (Figure 17).

After Chi-square analyses were performed on the results, it became evident that the vacuoles in Env7G₄₀V-HA mutant cells were significantly different from those of Env7-HA cells in YPD ($p < 0.0005$) (Table 6). The mutant cells displayed more multi-lobed vacuoles than the Env7-HA cells (Figure 18a). Additionally in YPD, the vacuoles of the Env7G₄₀V-HA mutant cells were significantly different from Env7C₁₃₋₁₅S-HA vacuoles ($p < 0.05$). The Env7G₄₀V-HA mutant cells again showed a greater percentage of multi-lobed vacuoles (Figure 18a).

In 0.5M NaCl YPD, the Env7G₄₀V-HA mutant cell vacuoles were not significantly different from those in Env7-HA cells. Both strains had a nearly even mix of prominent and multi-lobed vacuole phenotypes, although the Env7G₄₀V-HA cells showed about 10% more prominent vacuoles than Env7-HA (Figure 18b). The difference in phenotypes between Env7G₄₀V-HA and Env7C₁₃₋₁₅S-HA vacuole phenotypes in 0.5M NaCl YPD was also significant ($p < 0.0005$). The Env7C₁₃₋₁₅S-HA cells had nearly the same amount of prominent vacuoles as they did without additional salt in the medium (Figure 18a-b). This last result is important because it is supported by similar analyses done in our lab which showed that mutants lacking functional Env7 did not maintain many multi-lobed vacuoles 60 minutes after hypertonic treatment

(Manandhar et al., 2013; S. Manandhar and E. Calle, unpublished results). The Env7G₄₀V-HA cells seem to show a much less extreme vacuole phenotype than Env7C₁₃₋₁₅S-HA in high salt conditions, suggesting that the G₄₀V mutation does not abolish the normal function of Env7 in the cell. However, the subtle dissimilarities in vacuole phenotypes between this mutation and non-mutated Env7 in YPD may imply that the G₄₀V alteration has some effect on the protein and its function(s), perhaps different from a loss-of-function mutation.

Four other preliminary experiments using Env7G₄₀V-HA expressing cells were conducted: solid media growth analysis, liquid media growth analysis, subcellular localization assay, and phosphorylation assay. The results of these experiments can be found in Appendix C.

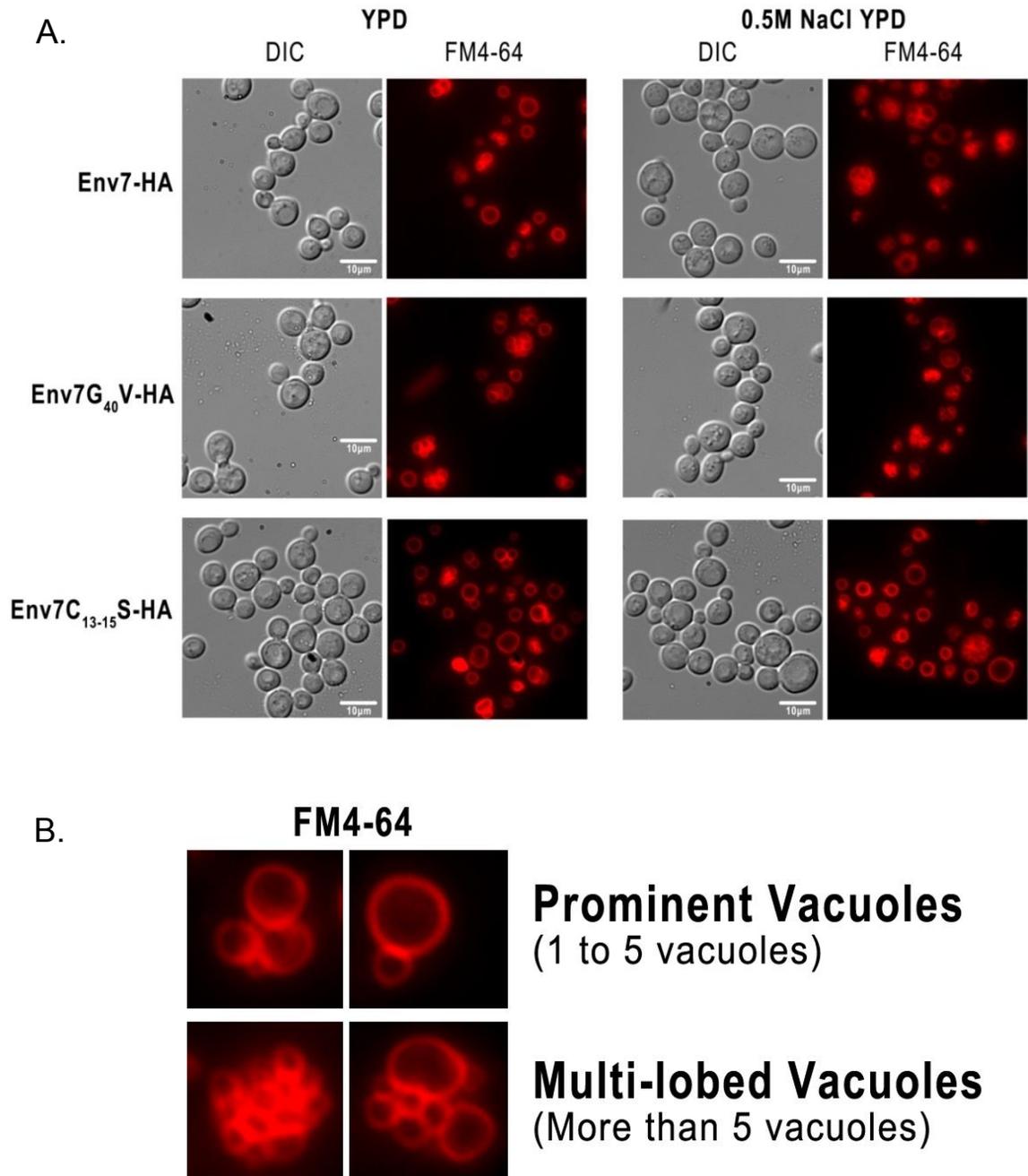


FIGURE 17. FM4-64 staining and microscopy of Env7-HA, Env7G₄₀V-HA, and Env7C₁₃₋₁₅S-HA cells. A) Log phase cells stained with FM4-64 were incubated in YPD or 0.5M NaCl YPD on a rotating platform (30°C) for 60 minutes then viewed under DIC/fluorescence microscopy at 1000x. B) Examples of vacuole phenotypes scored in Env7-HA, Env7G₄₀V-HA, and Env7C₁₃₋₁₅S-HA *S. cerevisiae* cells. Cells were considered to have prominent vacuoles if they exhibited 1-5 round FM4-64 staining patterns. Vacuoles counted as multi-lobed contained more than 5 rounded FM4-64 staining patterns.

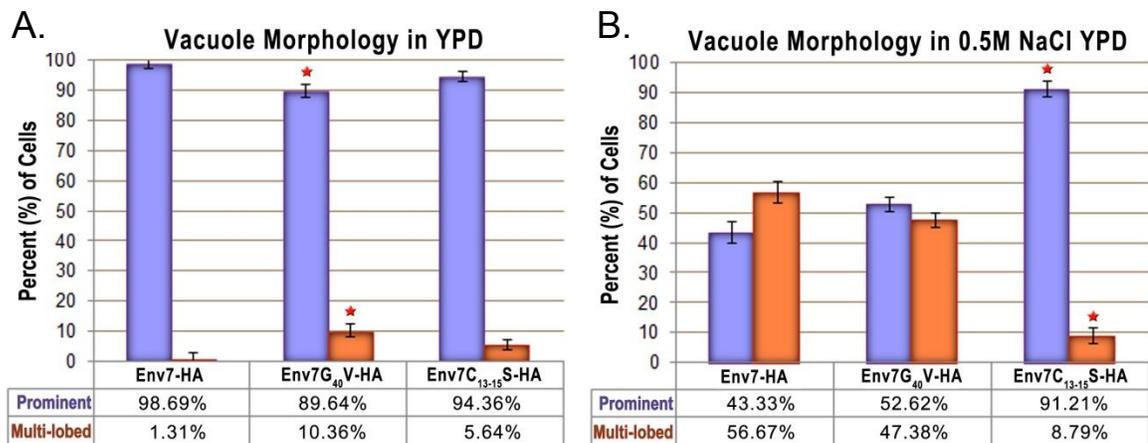


FIGURE 18. Bar graphs displaying the percentages of the three vacuole morphologies observed in the Env7-HA, Env7G₄₀V-HA, and Env7C₁₃₋₁₅S-HA-expressing cells of *S. cerevisiae* incubated in 30°C. Cells were grown in either A) YPD or B) 0.5M NaCl YPD. Vacuole morphology of the cells was determined by staining them with FM4-64 and then visualizing them under a fluorescent light microscope. Cells were scored as having prominent (1-5) vacuoles or multi-lobed (more than 5) vacuoles. Budding or dead yeast cells were not included in the analysis. Chi-square tests were used to determine significance. Bars with red stars above them represent significantly different percentages from the Env7-HA positive control.

TABLE 6. Individual Vacuole Morphology Counts for Cells Grown in YPD

Strain	Prominent	Multi-lobed	Total
Env7-HA	509.00	7.00	516.00
Env7G ₄₀ V-HA	422.00	48.00	470.00
Env7C ₁₃₋₁₅ S-HA	462.00	27.00	489.00

TABLE 7. Individual Vacuole Morphology Counts for Cells Grown in 0.5M NaCl YPD

Strain	Prominent	Multi-lobed	Total
Env7-HA	224.00	291.00	515.00
Env7G ₄₀ V-HA	241.00	217.00	458.00
Env7C ₁₃₋₁₅ S-HA	411.00	40.00	451.00

CHAPTER 4

ENV9, ENV10, AND ENV11 ANALYSES

Phylogenetic Analysis of *ENV9*

Characteristics of *ENV9* and *RDH12*

ENV9 (Gene locus: YOR246C) is a putative oxidoreductase studied by our lab. Experiments conducted by Teli Hsueh (T. Hsueh, Master's Thesis) showed that Env9 is localized to lipid droplets in the cell, and, although *env9Δ* cells internally accumulate proCPY like the other *env* null mutants do, they do not show altered vacuole acidification, lipid droplet morphology, or autophagy defects. However, *env9Δ* mutants have a somewhat unique, vesiculated vacuolar phenotype and tend to be caffeine sensitive. Additionally, topology predictions of Env9 suggested that it has one C-terminal transmembrane domain which might allow it to be a tail-anchored protein. Env9 is 28.0% identical to its human ortholog, Retinol Dehydrogenase 12 (RHD12), which is recognized as an integral ER membrane protein with most of its polypeptide chain exposed to the cytoplasm (Lee et al., 2010). Found mostly in retinal cells, it can oxidize/reduce retinoids and reduce toxic lipid peroxidation molecules to relatively harmless products (Belyaeva et al., 2005; Marchette et al., 2010).

Conservation of Residues Associated with Human Disease

RDH12 has been linked to human diseases, and a few studies in particular have elucidated the point mutations in the gene that are associated with Leber Congenital

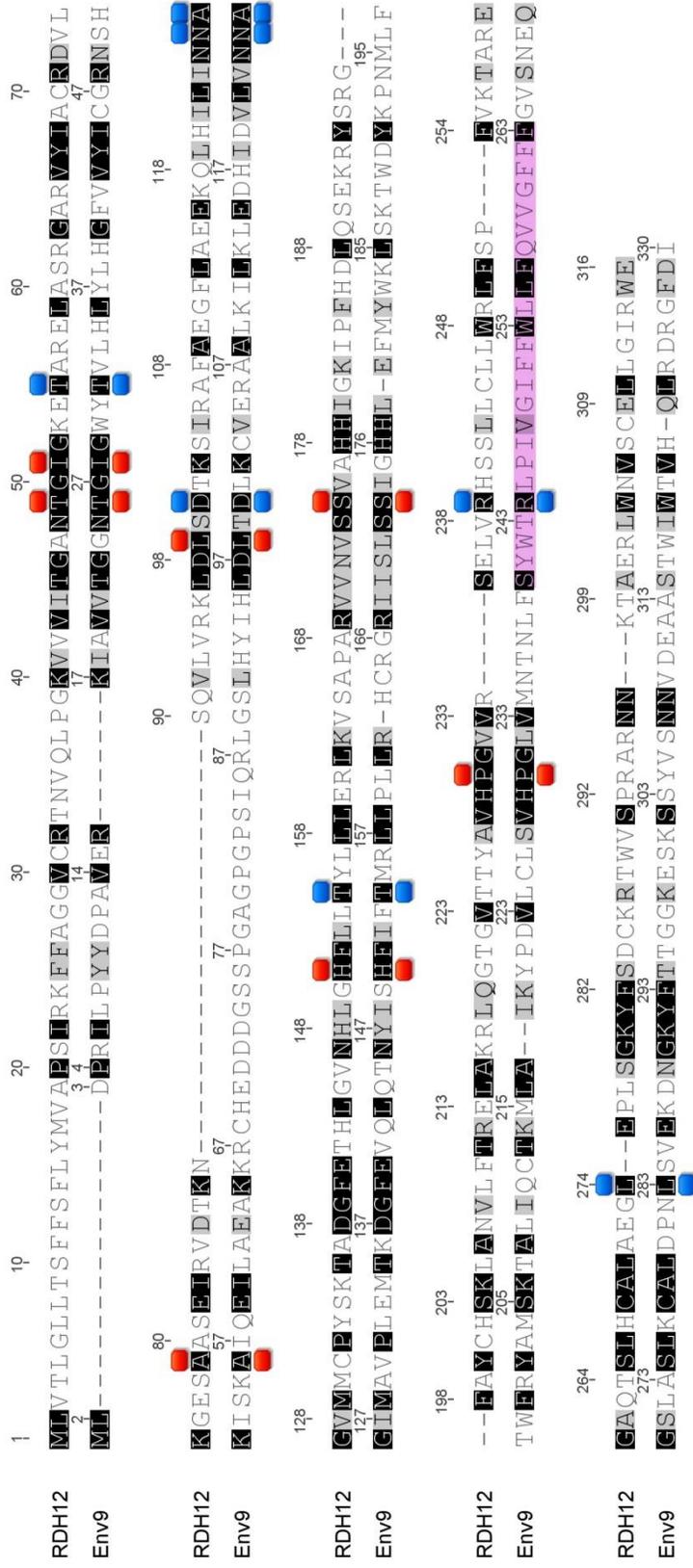


FIGURE 19. A pairwise sequence alignment of *H. sapiens* RDH12 and *S. cerevisiae* Env9 reveals conserved residues associated with human disease. Black shaded residues are identical, and gray shaded residues are similar to each other. Residues associated with Leber Congenital Amaurosis 13 are highlighted in red, and the residues connected to Retinitis Pigmentosa 53 is indicated in blue (Janecke et al., 2004; Perrault et al., 2004; Thompson et al., 2005; Benayoun et al., 2009). The membrane-spanning domain in Env9 is highlighted in purple.

Amaurosis 13 and Retinitis Pigmentosa 53, both of which can lead to blindness (Janecke et al., 2004; Perrault et al., 2004; Thompson et al., 2005; Benayoun et al., 2009). A pairwise alignment of Env9 and RDH12 revealed that many of the residues found mutated in diseased cells are conserved in Env9 (Figure 19). Also, many more residues are well-conserved between the two proteins, denoted by gray or black shaded boxes in Figure 19.

Env9 is Evolutionarily Conserved in Every Domain of Life

Reciprocal blastp searches using Env9 as the query revealed that the putative oxidoreductase has orthologs in Bacteria, Archaea, and Eukarya. This was not surprising given that oxidoreductase sequences are commonly found in clades throughout eukaryotes and prokaryotes (Gibney et al., 2013). After the Env9 orthologs were aligned and relationships were reconstructed, it became clear that Env9 was situated with the rest of the Ascomycetes, which formed a clade (Figure 20). Interestingly, the sequences belonging to Ascomycota showed the most intra-clade divergence, with branches adding up to nearly 1.5 amino acid substitutions per site between its most divergent sequences (*A. niger* and *C. lusitaniae*). Another unexpected result was that the bacterial and archaeal sequences seemed to be more closely related to the Metazoan sequences than the Fungi. Moreover, three mosquito sequences – *A. aegypti*, *A. gambiae*, and *C. quinquefasciatus* – were found to be more closely related to the human and chimp orthologs than those of the mouse and rat. This, along with the unanticipated prokaryotic portion of the tree topology, could suggest horizontal transfer events. However, several instances of homoplasy and splits with less than 90% clade credibility could undermine

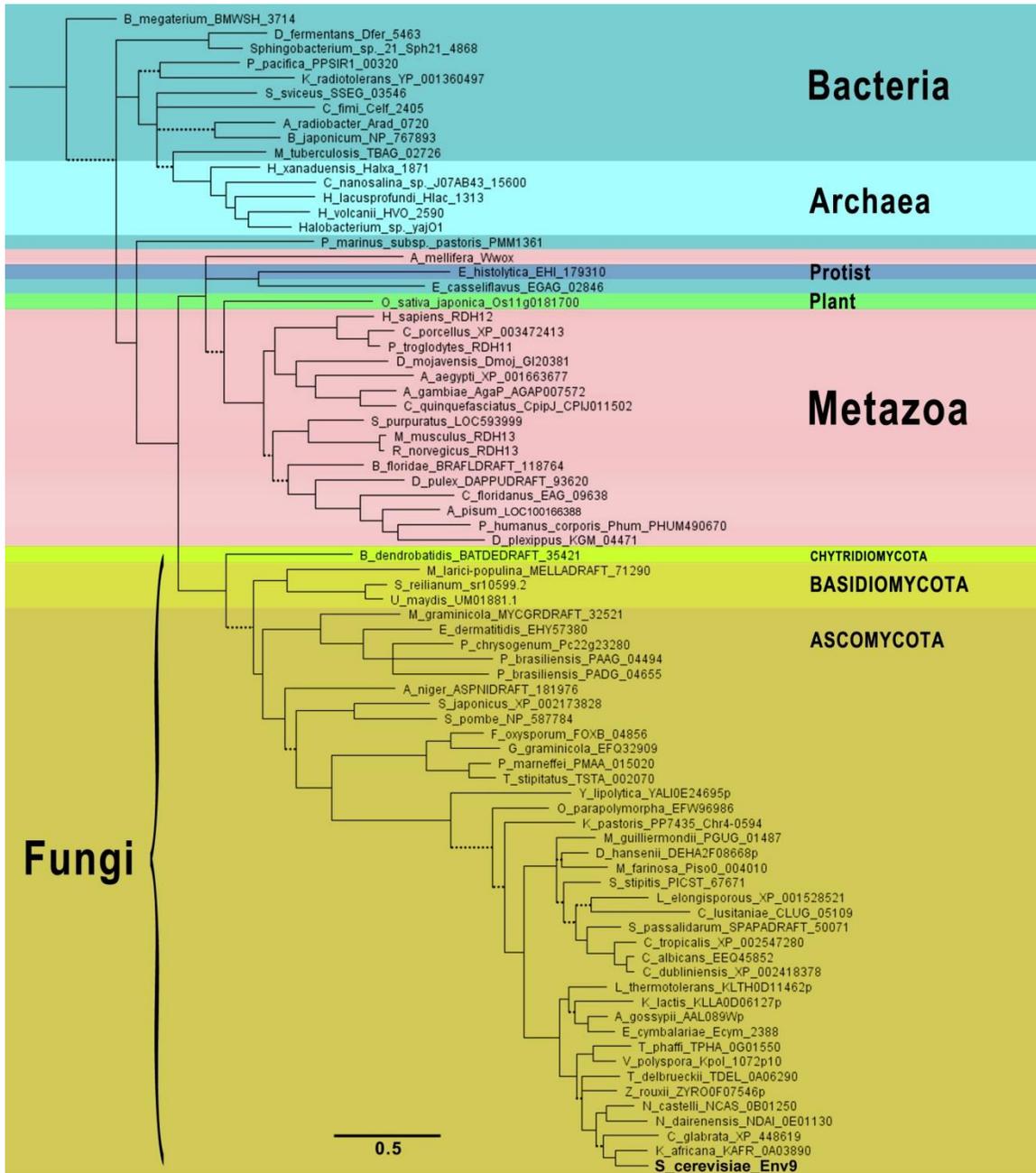


FIGURE 20. A Bayesian phylogenetic tree of Env9 and 77 of its orthologs. The tree is rooted on the *Bacillus megaterium* sequence because Firmicutes have been suggested as a basal bacterial group (Munoz et al., 2011). Sequence names consist of genus abbreviation, species, and gene name/locus. Dotted lines represent clades with less than 90% posterior probability, and the scale bar represents the number of amino acid changes per site. Clade shading reflects sequence membership in larger groups of organisms. Groups written in all capital letters are three of the fungal phyla.

these inferences. As additional data from orthologous genes become available, better supported relationships may be revealed.

Phylogenetic Analysis of *ENV10*

ENV10 (Gene locus: YLR065C) is a non-essential gene uncovered by our lab during a genome-wide immunodetection assay of *S. cerevisiae* null mutants exhibiting defective proCPY processing (Ricarte et al., 2011). Further studies in our lab by Lisa Oliveira have shown that Env10, predicted to be a transmembrane protein, localizes to the endoplasmic reticulum and that its null mutant is cold-sensitive. Unlike the single mutants, the *env9Δ/env10Δ* double mutant showed drastic decreases in fitness under a variety of growth conditions, characteristic of a negative genetic interaction (Oliveira, Master's Thesis). The double mutant was sensitive to alkaline conditions, very low pH, and high salt treatment. Although the function of Env10 is still unknown, our lab postulated that it might interact with the lipid droplet resident Env9 as the lipid droplets form from the endoplasmic reticulum.

Env10 has a domain of unknown function, denoted DUF788, spanning the majority of its 181 amino acid sequence (residues 1-172) that is conserved in many eukaryotes, including humans (InterPro entry: IPR008506). DUFs are a substantial group of protein domains that are conserved but have no known function (Bateman et al., 2010). They make up about 20% of the domains defined in Pfam (Sanger Institute), and their numbers are growing rapidly. The DUF788 domain has a set of specific residues that define it as a unique protein domain (Figure 21). 442 other proteins in Eukarya are known to possess the DUF788 domain, and only human Transmembrane Protein 208

(TMEM208) and *S. pombe* Meiotically Upregulated Gene 69 (Mug69) have been studied. *mug69Δ* vegetative cells in *S. pombe* are curved and inviable under standard growing conditions (Hayles et al., 2013). Mug69 is ER-localized and is an integral membrane protein (Matsuyama et al., 2006). A very recent study implicated human TMEM208 in autophagy and ER stress regulation (Zhao et al., 2013). This was the first publication to address the DUF788-containing protein, and it also confirmed the presence of 3 transmembrane domains in TMEM208. Upon heat treatment, TMEM208 protein also formed SDS-resistant aggregates. TMEM208 knockdown and silencing showed an increase in autophagy in human osteosarcoma and embryonic kidney cells, while overexpression of the gene resulted in decreased autophagy and ER stress (Zhao et al., 2013). While the ER localization of TMEM208 and Mug69 is related to what is known about Env10, it is important to remember that just because these proteins have a DUF788 domain does not mean that they are orthologous to Env10. Although TMEM208, for example, shares about 31% sequence identity with Env10, the identities only occur at the key residues that define the DUF788 domain, which comprises nearly 95% of each protein sequence. The rest of the amino acid residues in TMEM208, as well as in Mug69 (which shares only 22.3% sequence identity with Env10), do not align well with those in Env10 (data not shown). Still, because the DUF788 region spans most of the length of these proteins, they may be functional counterparts in the different species.

Based on reciprocal blastp searches, orthologs of Env10 are not found outside of the Ascomycete order, Saccharomycetales. All of the orthologs shown in Figure 22 contain a DUF788 domain like Env10 and are also predicted to have 2-4 transmembrane domains (TMHMM). However, the *Ashbya gossypii* and *Eremothecium cymbalarie*

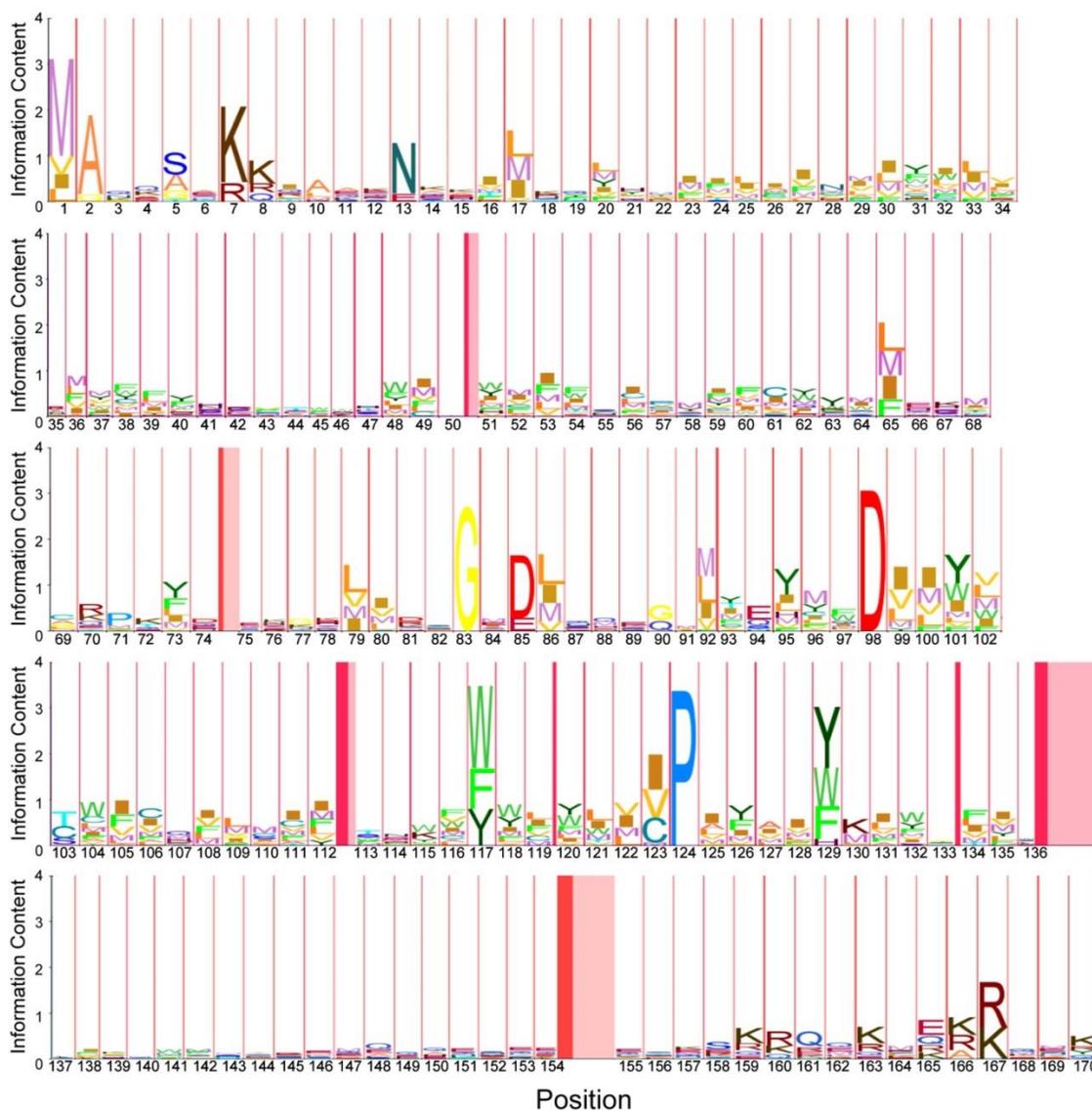


FIGURE 21 . Hidden Markov Model (HMM) Logo illustration of conserved amino acid residues in the DUF788 domain. The DUF788 HMM Logo profile (Schuster-Böckler et al., 2004) was generated and copied from Pfam (Sanger Institute, <http://pfam.sanger.ac.uk/family/PF05620#tabview=tab4>) and styled into the Figure above. Each position is divided by pink lines, which represent possible amino acid inserts. Dark pink represents the probability that at least one residue is inserted at the position, and light pink represents the probability that more than one amino acid is inserted there. The bigger the pink area is, the higher the likelihood of the respective insertion event. A larger letter indicates greater conservation of that particular amino acid residue in the DUF788 domain at that position. Taller stacks of letters signify greater information content, or certainty, of the residues (and their sizes) shown in that position.

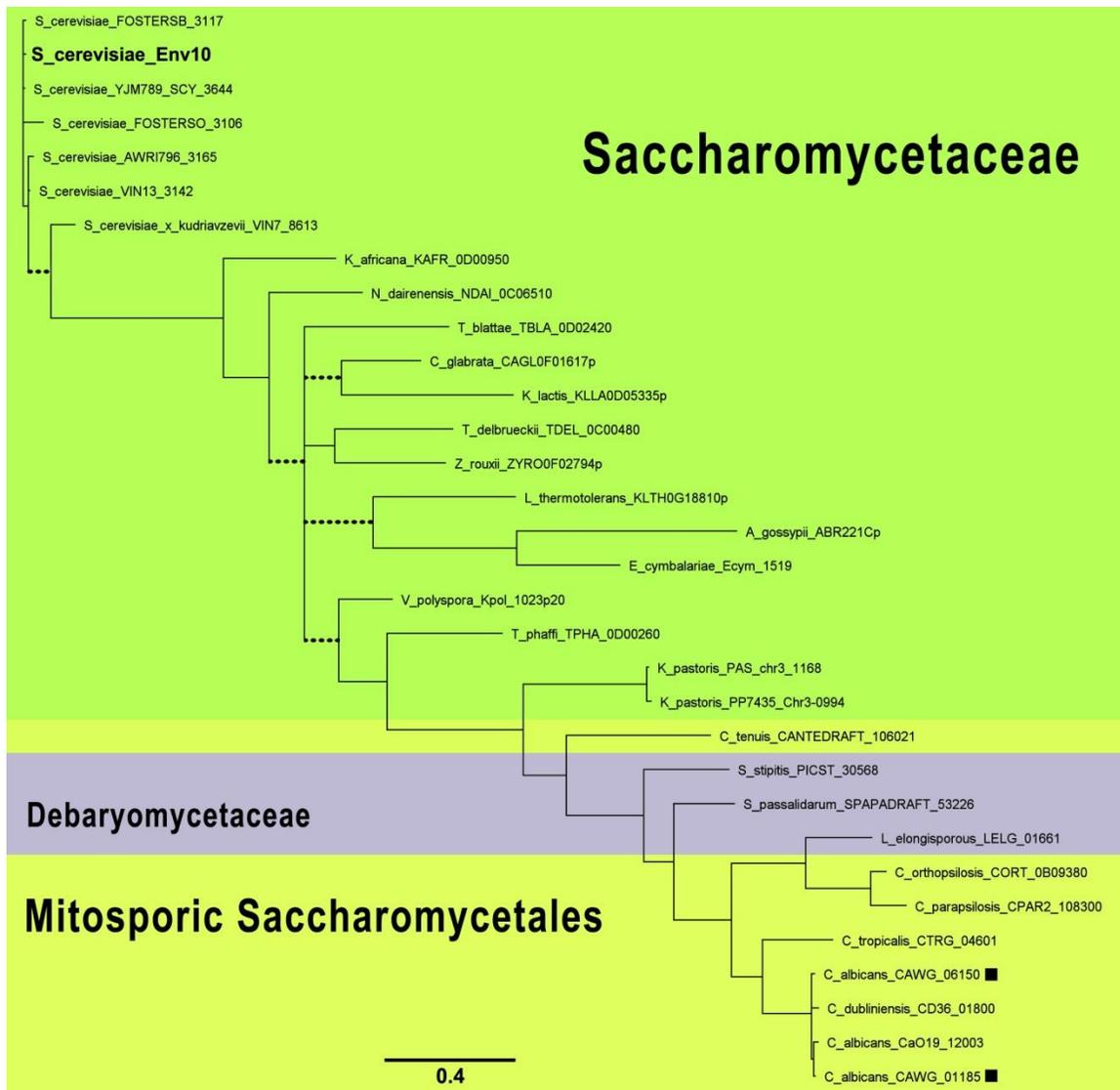


FIGURE 22. An unrooted Bayesian phylogenetic tree of Env10 and 31 of its homologs. Dotted branches represent clades with less than 90% posterior probability. The scale bar depicts the branch length on the tree that signifies 0.4 amino acid substitutions per site. Clades are color-coded depending on the family membership of its constituent species in the order, Saccharomycetales. Sequences marked with a black square are most likely paralogous.

sequences are missing the C-terminal quarter of their DUF788 domain, almost directly after the last predicted transmembrane domain in each sequence. The Saccharomycetaceae portion of the Bayesian phylogenetic tree is not well resolved, as is evident from the large polytomy present and clade credibility values below 90% (Figure 22). The aligned Saccharomycetaceae sequences have 39.9% average pairwise sequence identity, compared to 57.3% average sequence identity among aligned Debaryomycetaceae and Mitosporic Saccharomycetales sequences. The *Candida tenuis* sequence is grouped more closely with the *Scheffersomyces stipitis* and *Spathaspora passalidarum* sequences than those of its fellow Mitosporic Saccharomycetales members. The link between these three organisms is their ability to ferment pentose sugars, though it is unclear what correlation, if any, this physiological ability has with orthologs of Env10 (Wohlbach et al., 2011). Two possible paralogs in *Candida albicans* WO-1, CAWG 06150 and CAWG 01185, suggest an incidence of gene duplication and subsequent divergence, and this is not seen in any other organism represented in the tree (Figure 22). Env10, as expected, was grouped with the other *Saccharomyces* species, with its closest ortholog belonging to *Kazachstania africana*.

Phylogenetic Analysis of *ENV11*

ENV11 (Gene locus: YGR071C), the last novel late endosome and vacuole mutants identified, has been studied by our lab and others. While the *env11Δ* mutant is not sensitive to ionic stress or common drugs, like caffeine, it does accumulate higher levels of glycogen than wild-type cells and exhibits abnormally elongated buds during replication (Wilson et al., 2002; Watanabe et al., 2009; Ricarte et al., 2011). Additionally, the *env11Δ* cells contain fragmented vacuoles and exhibit internal

accumulation of pro-CPY (Ricarte et al., 2011). Env11 localizes to the nucleus in specific quantities by unknown nuclear import mechanisms.

ENV11 is an ohnolog (paralog that arose from whole genome duplication) of another *S. cerevisiae* gene, Vacuole Import and Degradation 22 (*VID22*) (Brown et al., 2002; Byrne et al., 2005). Both of their gene products contain a BED-type zinc finger domain that functions mainly to bind DNA (Aravind 2000). Much is known about Vid22; it is a glycosylated, integral plasma membrane protein, and it travels to the nucleus to interact with Telomere Binding Protein 1 (Tbf1) to repair DNA double-strand breaks (Brown et al., 2002; Bonetti et al., 2013). Vid22 and Tbf1 work with histone-modifying enzymes and chromatin remodeling proteins near the site of the DNA double-strand break, to which they are both recruited (Bonetti et al., 2013). Env11 becomes a player in the Vid22/Tbf1 relationship when the three proteins colocalize to promoter regions of endomembrane trafficking genes, including *ENV9*, among others (Krogan et al., 2006; Oliveira, Master's Thesis; Preti et al., 2010). Experiments with an *env11Δ/vid22Δ* double deletion mutant performed by our lab established the two genes' importance in endomembrane trafficking; cells from this double mutant strain were abnormally large with fragmented vacuoles.

Multiple sequence alignment of Env11 and Vid22 protein sequences highlights some important conserved regions and residues. As previously mentioned, both proteins share a conserved BED-type zinc finger domain towards their N-terminus. Additionally, a stretch of amino acids in Vid22 proposed to create a membrane-spanning helix shares many identical residues with the part of Env11 with which it is aligned. An interesting asparagine-rich region of Vid22 does not align with Env11 (results not shown). Env11 is

missing most of the residues that make up this region of Vid22. It has been shown that many yeast transcriptional regulators have asparagine-rich stretches, and these regions are essential for activation of transcription (Myers et al., 1998; Michelitsch and Weissman, 2000; Sim and Creamer, 2002; Titz et al., 2006).

Another feature of Vid22 that is not generally present in Env11 is the abundance of both putative and experimentally-determined glycosylation sites. Glycosylation is important for some proteins because the added carbohydrate moieties aid in protein folding and trafficking to the correct cellular membrane. As an integral plasma membrane protein, Vid22 may move through the secretory pathway to reach its cellular destination, but because Env11 shares very few of the conserved glycosylation sites, it may not rely on carbohydrate addition for localization and/or function.

Env11 and Vid22 are not orthologous to any protein outside of the Saccharomycetaceae family (Figure 23). This is consistent with their proposed functions as transcriptional regulators; groups limited to Fungi or subgroups of Fungi tend to be significantly enriched in nucleic acid binding and transcription factor activity (Gibney et al., 2013). Interestingly, only the Saccharomyces Vid22 orthologs contained the asparagine-rich region that could be important in transcriptional regulation, and the VIN7 8867 *S. cerevisiae* strain sequence did not even possess a complete zinc finger domain (Figure 23). Most of the sequences on the Env11/Vid22 phylogenetic tree are either orthologous to Env11 or Vid22. However, there are three exceptions: *Kazachstania africana*, *Eremothecium cymbalariae*, and *Ashbya gossypii*. *K. africana* has two sequences, paralogous to each other, that both are co-orthologous to *S. cerevisiae* Vid22. These two sequences are only 25.7% identical to each other, and the KAFR 0A06130

sequence contains acidic stretch of residues towards its N-terminus that the other sequence does not possess. This acidic stretch is unique to KAFR 0A06130, and it may impart differential regulation during transcription activation in *K. africana*. *E. cymbalariae* and *A. gossypii* belong to the same genus (Eremothecium), but neither of them have both an Env11 ortholog and a Vid22 ortholog. Using the reciprocal blastp technique, it was clear that the *E. cymbalariae* sequence was orthologous to Env11, and the *A. gossypii* sequence was orthologous to Vid22. Unfortunately, further BLAST searches with the genera featured in the tree in Figure 23 excluded, failed to show any significant hits, even in other Saccharomycetaceae family members (data not shown). This suggests that perhaps the whole genome duplication that occurred in yeast 100-150 million years ago could be responsible for the presence of the paralogs, Env11 and Vid22, in some extant yeast clades (Kellis et al., 2004; Capra et al., 2010). Genomic duplication events are usually followed by massive gene loss, so that might account for the many Saccharomycetaceae members that have neither Vid22 nor Env11. This idea is supported by the large relative divergence of the *E. cymbalariae* and *A. gossypii* sequences from the rest of the sequences on the tree (Figure 23).

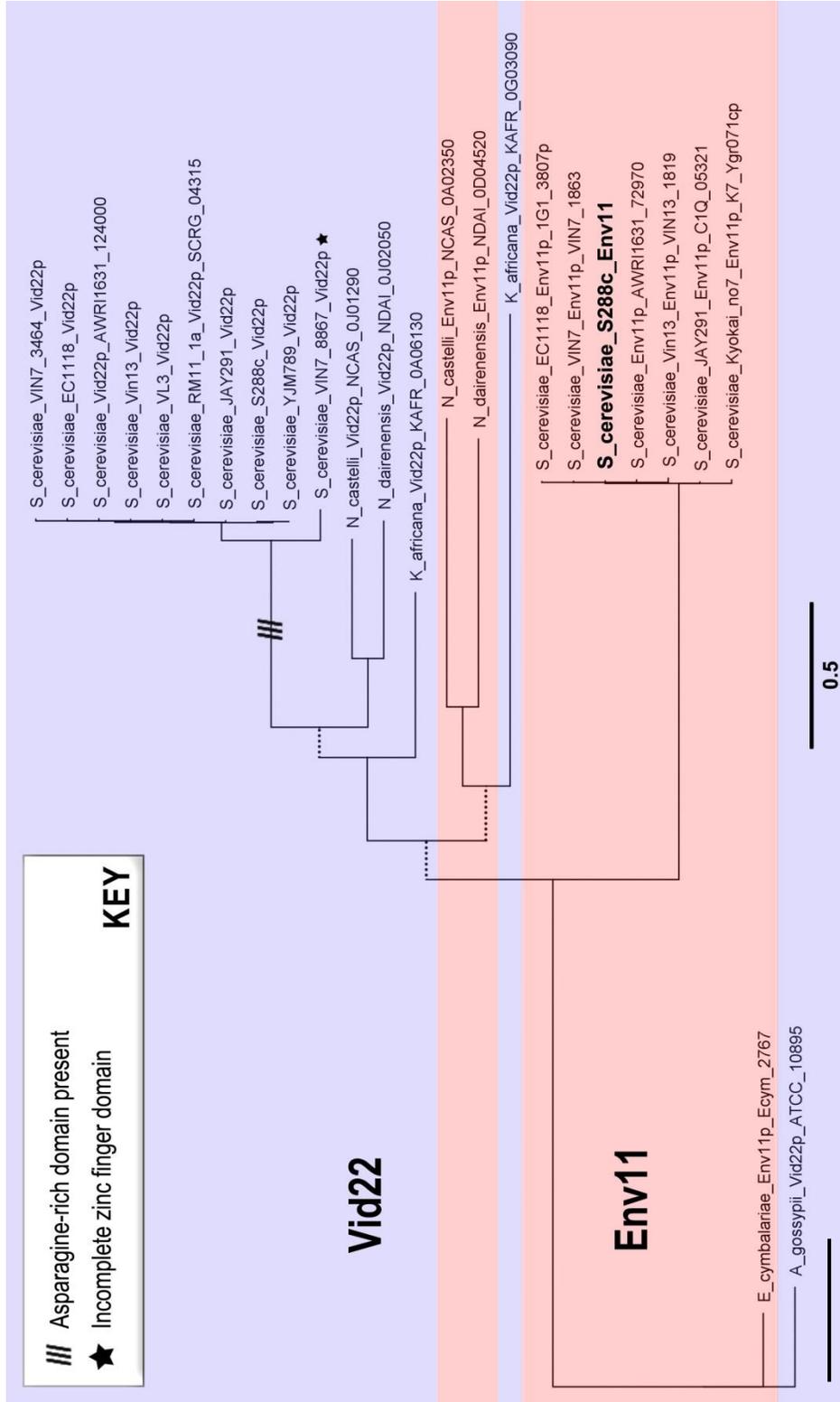


FIGURE 23. An unrooted Bayesian phylogenetic tree of Env11 and Vid22 and their respective homologs. Sequence names consist of genus abbreviation, species, strain (if applicable), NCBI designation as “Env11-like” or “Vid22-like” (if applicable), and gene name/locus. Sequences/groups of sequences are shaded based on NCBI annotation as Env11-like or Vid22-like. Dotted lines represent clades with less than 90% posterior probability. The scale bar represents the number of amino acid changes per site.

CHAPTER 5

DISCUSSION

A Combination of Vertical and Horizontal Gene Transfer is Likely Responsible for the Conservation of *ENV7* Orthologs in Eukarya and Bacteria

An aim of most phylogenetic studies is reconstructing relationships among related sequences; however, when the topology of the obtained phylogenetic tree violates established monophyletic groups, modes of gene transmission other than vertical must be considered. There are four major reasons why gene and species trees can differ (Galtier and Daubin, 2008). The first has to do with tree-building errors, including sequence selection, multiple sequence alignment, and deviations from the selected evolutionary model. Incomplete lineage sorting is another explanation for gene and species tree incongruence because neutral evolution can sometimes lead to gene divergence occurring before or after speciation events. A third cause of gene and species tree dissimilarity is hidden paralogy, which has a greater influence when sequences are taken from incompletely sequenced genomes. Horizontal gene transfer is the fourth source of gene and species tree topological discrepancies. Though this form of gene transmission is most common in prokaryotes, it has been detected in *S. cerevisiae* and a growing number of other eukarya (Hall et al., 2005).

Assuming there are no issues with the phylogenetic analyses in this study and considering that the sequences used were all from completely sequenced genomes, the two explanations remaining for the departure of the gene trees in Figures 2 and 3 from

accepted species trees are incomplete lineage sorting and horizontal gene transfer. Incomplete lineage sorting has been shown to be responsible for gene and species tree discrepancies in more closely-related organisms, mostly in phylogenomic trees consisting of long branches terminating in much shorter diverging branches (Degnan, 2006). This could explain why the *K. pastoris* sequence (XP_002491616) is separated from the rest of the Saccharomycetaceae family by sequences from other Saccharomycotina families (posterior probability = 1.0) (Figure 2). However, the placement of the Basidiomycotal sequences between two parts of the Ascomycota phylum might not be explained easily by incomplete lineage sorting.

Horizontal gene transfer, particularly among single-celled eukaryotes, is not as uncommon as once thought. Conjugation events between fungi and bacteria have been observed in laboratory settings, and more than 300 examples of horizontal gene transfer into fungi have been dually confirmed by individual research groups (Heinemann and Sprague, 1989; Inomata et al., 1994; Sawasaki et al., 1996; reviewed in Richards et al., 2011). In a few instances, horizontal gene transfer between different fungal species growing in proximity to each other has occurred through cytoplasmic interconnections (reviewed in Richards et al., 2011). The estimated divergence of Ascomycota and Basidiomycota predates the emergence of Pezizomycotina by about 135 million years, so the Pezizomycotina clade should be found within the Ascomycotal clade (Blair, 2009). The phylogenetic tree in Figure 2 does not reflect this topology, so it may be possible that the Env7 sequences present in Basidiomycota and Pezizomycotina were initially acquired horizontally from a divergent Saccharomycotina member, such as *Yarrowia lipolytica*, that shared the same environmental niche.

Env7 orthologs are almost exclusively found in Eukarya, and the phylogenetic tree in Figure 3 demonstrates that methods of gene transmission other than vertical are likely responsible for its topology. First, the lone bacterial ortholog (Mmol_0176) could either be the ancestor of all the Env7 orthologs or a result of horizontal gene transfer from a eukaryote. Because Env7 orthologs were not identified in a wide range of fully sequenced bacterial genomes, the horizontal gene transfer explanation seems to be the more likely one. If the bacterial sequence was truly ancestral to all of the eukaryotic ones, orthologs should have been found in more Methylothera species as well as in more of the early-emerging fungi. Additionally, a well-supported clade consisting of sequences from an early-diverging chytrid fungus (*B. dendrobatidis*), a sponge (*Amphimedon queenslandica*), a fresh-water polyp (*Hydra vulgaris*), and a sea anemone (*Nematostella vectensis*) suggest that horizontal gene transfer may have been at work (posterior probability = 0.98; Figure 3). Finally, the early-diverging fungal sequences in Figure 3, belonging to *Mucor circinelloides* and *B. dendrobatidis*, were not basally situated in relation to the sequences from Ascomycota and Basidiomycota when the tree was rooted on the bacterial sequence. This contradicts established species relationships and suggests that these organisms may have also acquired their sequences by means other than vertical gene transmission. Based on results of this study, the *ENV7* homologs may have been conserved via a combination vertical and horizontal gene transmission in Eukarya and Bacteria.

The NAK Kinase Family is Still Not Well-Defined

Since the tenuous establishment of the NAK kinase family of eukaryotic protein kinases a little more than a decade ago, scant evidence besides similarity among the

kinase domains of its members has surfaced to substantiate its existence (Chien et al., 1998; Eswaran et al., 2008). This study showed that the sequence similarity among the NAK kinase subfamilies does exist, but there is little else to support the membership of the *ENV7*/*STK16* clade as a part of the NAK kinase family. Some studies have suggested that the disordered glycines in the glycine-rich loop are the sole requirement for NAK membership, but the Polo-like Kinases also have this double glycine motif (Eswaran et al., 2008). There are no invariant residues unique to the NAK kinases, and even some of *Env7*'s orthologs do not possess the requisite disordered glycines (Figures 2 and 3). The other NAK subfamilies—GAK, BMP2K, AAK1, and Ark1/Prk1—share similarities in function in addition to kinase sequence similarity. However, *Env7*/*STK16* subfamily members do not likely interact with cytoskeletal elements like the rest of the NAK subfamily members because their sequences are hardly longer than a single eukaryotic protein kinase domain. Furthermore, alignment and maximum likelihood analysis of just the kinase domains of the NAKS still showed roughly the same topology as the tree of the full length sequences (Figure 6), suggesting that even the kinase domains of the *Env7*/*STK16* clade members are divergent (Figure 27 in Appendix C). Solved three-dimensional structures for more of NAK Family members could elucidate important tertiary structural similarities in future, but as of this time the *Env7*/*STK16* clade seems to be only loosely related to the rest of the NAK subfamilies.

ENV7 Orthologs May Rely on Different Forms of Activation Segment Regulation

The DFG motif in subdomain VII is the N-terminal anchor of the activation segment in nearly all eukaryotic protein kinases (Hanks and Hunter, 1995; Taylor and Kornev, 2011). The phenylalanine residue is particularly important because it can

regulate the activity of the kinase. When the R-group of the phenylalanine is facing towards the catalytic cleft, called “DFG-in”, the kinase is catalytically competent because the N- and C-lobes are securely positioned to bind the ATP molecule. When the phenylalanine R-group is facing away from the catalytic cleft and is “DFG-out”, the two lobes become unstable because the R-spine has been disrupted. This renders the kinase inactive (Taylor et al., 2012). If Env7 and its orthologs in Saccharomycotina, Pezizomycotina, Hexapoda, and Vertebrata have a leucine residue instead of the phenylalanine, can the R-spine still be intact when the leucine is “DLG-in”? Leucine and phenylalanine are both fairly large, with the major difference between them being that the phenylalanine has an aromatic benzyl group while leucine has an alkyl side chain. Studies from our lab have demonstrated that Env7 is a functional serine/threonine kinase, so even with leucine in place of phenylalanine the kinase can still be catalytically active (Manandhar et al., 2013). Similar studies have demonstrated that STK16 is an active kinase, even with a DLG motif (Kurioka et al., 1998; Berson et al., 1999; Eswaran et al., 2008). What remains unclear is whether or not the leucine can act as a switch like phenylalanine, and, if not, that could mean that Env7 and its orthologs that rely on other methods of functional regulation.

The Subdomain VIA Insert in Env7 May Play a Role in its Localization and/or Interaction with Other Proteins

The potential acidic di-leucine sorting signal in Env7's subdomain VIA insert may allow it to be trafficked to the vacuole by the ALP pathway. This sorting signal, which has the consensus sequence [D/E]xxxL[L/I/V], has been shown to interact with the delta subunit of the AP-3 clathrin-associated protein complex (Vowels and Payne, 1998; Darsow et al., 1998). Embedded in vesicle membranes, the AP-3 complex guides

proteins from the Golgi to the vacuole via the ALP pathway (Figure 1) (Cowles et al., 1997; Odorizzi et al., 1998). The three-dimensional position of the subdomain VIA insert in Env7, on the side opposite of the likely ATP and substrate binding sites, implies that it is not involved with catalysis directly (Figure 11a). STK16, a Golgi membrane protein, does not contain an exposed di-leucine signaling motif far from the catalytic site, which may be the reason why it is localized to the Golgi (Eswaran et al., 2008).

Env7's subdomain VIA insert sequence may also serve as an interaction interface with other proteins. Structurally, Env7 is little more than a kinase core, with relatively few flanking N- and C-terminal residues. This suggests that it could be part of a larger protein complex, perhaps as a regulatory element. The conservation of the glutamic acid in the insert sequence is consistent with findings which have shown that charged residues are generally conserved at protein interaction surfaces (Figure 11b) (Zhao et al., 2011). Plus, despite the fact that nearly all of Env7's fungal orthologs have an insert in this area of the protein, the residues that comprise the inserts vary greatly throughout the fungi (Figure 11b; results not shown). Such unique stretches of residues could mean that the inserts serve specialized functions that are conserved at the family or species level.

The specific protein subunit that the di-leucine sorting signal interacts with in *S. cerevisiae* is called AP-3 complex subunit delta (Apl5) (Cowles et al., 1997). When Apl5 is deleted, a defect in vacuolar fragmentation under salt stress similar to what our lab has observed in *env7Δ* cells is seen (Michaillat and Mayer, 2013; Manandhar et al., 2013). It was proposed that adaptor proteins, like Apl5, might even bind to the vacuolar membrane and affect its curvature, but this has yet to be experimentally verified (Michaillat and Mayer, 2013). This could explain why some experiments from our lab showed that, in

addition to normal localization to the vacuolar membrane, overexpressed GFP-tagged Env7 localized to punctate formations in the Golgi (Manandhar et al., 2013). If copious amounts of Env7 protein were to be trafficked to the vacuole by AP-3 coated vesicles, and assuming a limited number of Env7 molecules can interact with the AP-3 delta subunits on a single given AP-3 coated vesicle, then the leftover Env7 molecules would likely remain in the late Golgi until they could be ushered to the vacuole somehow. Additionally, many of Env7's orthologs that lack the subdomain VIA insert, including human STK16 and *S. pombe* Ppk3, localize to the Golgi (Ligos et al., 1998; Bimbó et al., 2005). This further supports the idea that Env7's subdomain VIA insert may be responsible for its vacuolar localization and protein interaction.

Many mammalian proteins related to lysosomal storage diseases, such as Niemann-Pick Type C disease, Batten Disease, and mucopolidosis type IV, are targeted to the lysosome by di-leucine signaling motifs, and disease occurs when these proteins are not properly trafficked to their destination (Watari et al., 1998; Storch et al., 2004; Berger et al., 2005; Vergarajauregui and Puertollano, 2006; Berger et al., 2007). These types of lysosomal storage diseases are characterized by accumulation of late endosomes and other vesicles in the cell and the subsequent damage caused by this buildup (reviewed in Platt et al., 2012). Studies from our lab have shown the proCPY accumulates in the *S. cerevisiae env7Δ* cells due to a trafficking defect between the late endosome and the vacuole (Ricarte et al., 2011), and it would be interesting to determine if deletion of the potential di-leucine signaling motif in subdomain VIA could cause a similar phenotype. Furthermore, some lysosomal storage disease-associated proteins have homologs in *S. cerevisiae*, and it may be worthwhile to compare the phenotypic effects of deletion or

mutation of *ENV7* to phenotypes observed when the lysosomal storage disease homologs are deleted (Pearce et al., 1999; Berger et al., 2005).

The Vacuolar Morphology of the Env7G₄₀V-HA Mutant is Slightly Aberrant

An unexpected finding of this study was that the Env7G₄₀V-HA cells exhibited slightly more multi-lobed vacuoles in YPD than Env7-HA and Env7C₁₃₋₁₅S-HA in YPD (Figure 18). However, when under hyperosmotic stress, the Env7G₄₀V-HA cells' vacuoles were not significantly different from those of Env7-HA. Because the G₄₀V mutation is in the ATP-binding region of the kinase, this phenotypic outcome could be the result of Env7G₄₀V-HA's constitutive activity or simple failure to phosphorylate its substrate(s). This mutant kinase form may bind ATP too tightly or fail to release ADP appropriately after the catalytic event. It is unlikely that the G₄₀V mutation prevents ATP binding altogether, because in many experimental instances in this study, Env7G₄₀V-HA and Env7-HA phenotypes are not different from one another (Figures 28-30 in Appendix C). Therefore the kinase should be active in some way.

It is possible that the G₄₀V mutation elicited multi-lobed vacuoles in the absence of salt stress due to constitutive Env7G₄₀V-HA activity. The kinase-dead Env7C₁₃₋₁₅S-HA mutant showed prominent vacuoles like Env7-HA in YPD, so that may mean that the Env7G₄₀V-HA mutant was not inactive but maybe inappropriately active. Instead of requiring salt treatment to induce the multi-lobed vacuoles through some unknown signaling cascade involving Env7, the G₄₀V mutation itself was sufficient to cause the a weaker version of the multi-lobed stress phenotype. Unfortunately, until Env7's *in vivo* substrates are identified, the precise effect of the G₄₀V mutation on vacuole morphology will remain undisclosed.

Clues to Env7's role in vacuole morphology and the Env7G₄₀V-HA mutant's multi-lobed vacuoles in standard YPD medium may lie once again in the relationships between *ENV7/Env7* and other genes/proteins. *ENV7* has a negative genetic interaction with Phosphatidic Acid phosphoHydrolase (*PAH1*), which is a yeast homolog of mammalian lipin (Santos-Rosa et al., 2005; Costanzo et al., 2010). Pah1 plays a role in vacuole homeostasis and vacuolar fusion events. In cells lacking Pah1, the vacuoles appear deformed and fragmented (Sasser et al., 2012). Additionally, deletion of both *ENV7* and a transcription factor called Mbf and Sbf Associated 2 (*MSA2*) causes decreased vegetative growth (Sharifpoor et al., 2012). *MSA2* also negatively genetically interacts with *YCK3*, which in turn, as our lab has discovered for *ENV7*, negatively regulates vacuolar fusion during hyperosmotic stress (Lagrassa and Ungermann 2005; Manandhar et al., 2013). Perhaps Yck3/Env7 and Pah1 could be cogs in compensatory pathways governing vacuole membrane equilibrium during salt stress, as is often the case with negatively genetically interacting genes.

Golgi fragmentation in mammalian neuronal cells has been observed in neurodegenerative diseases, and studies have suggested that this disassembly may be a stress response that triggers cell death (Gonatas et al., 2006; Nakagomi et al., 2008). Further research is needed to elucidate the mechanisms of Golgi fragmentation in disease, but perhaps Golgi-localized STK16, as a ortholog of Env7, plays a role in Golgi stress fragmentation, just as Env7 does in vacuolar membrane homeostasis during stress.

Promising Target Residues for Further Point Mutation in *ENV7*

A main goal of this study was to elucidate potential amino acid targets in Env7 for further experimentation. These residues are listed in Table 26 in Appendix D. Many of

the residues in the table may be involved in catalytic activity, substrate binding, and/or localization. Other residues are simply very well-conserved throughout Env7's orthologs, suggesting that they could be important. Although our lab has already initiated/completed experimentation on some of these residues, there are many questions about *ENV7* that are still unanswered. Hopefully, future manipulation and analysis of these residues will yield information about Env7's substrates, its precise role in the regulation of vacuolar fusion, and the cellular events/other proteins that dictate its expression and activity in yeast cells. The broader aim of discovering and organizing this type of information about Env7 is to connect it to lysosomal storage disorders in mammalian cells. If our lab is able to accomplish this, Env7 will become one more tool in the search for cures for these serious diseases.

APPENDICES

APPENDIX A

DETAILS REGARDING PHYLOGENETIC AND BIOINFORMATIC METHODS

TABLE 8. Programs, Settings, and References for Multiple Sequence Alignments

Alignment Name	Associated Figure(s)	MSA Program	Scoring Matrix	Settings	Reference
Env9 + 77 orthologs	20	MAFFT v6.814b	BLOSUM62	Algorithm: L-INS-i Gap open penalty: 1.53 Offset value: 0.123	Katoh et al., 2002; Katoh et al., 2005
Env10 + 31 homologs	22	MAFFT v6.814b	BLOSUM62	Algorithm: L-INS-i Gap open penalty: 1.53 Offset value: 0.123	Katoh et al., 2002; Katoh et al., 2005
Env11 and Vid22 + 23 orthologs	23	MAFFT v6.814b	BLOSUM62	Algorithm: L-INS-i Gap open penalty: 1.53 Offset value: 0.123	Katoh et al., 2002; Katoh et al., 2005
Env9 and RDH12	19	ClustalW	Identity	Gap open cost: 10 Gap extent cost: 1	Thompson et al., 1994
Env7 + 114 fungal homologs	2	MAFFT v6.814b	BLOSUM62	Algorithm: L-INS-i Gap open penalty: 1.53 Offset value: 0.123	Katoh et al., 2002; Katoh et al., 2005
Env7 + 67 eukaryotic and bacterial homologs	3	MAFFT v6.814b	BLOSUM62	Algorithm: L-INS-i Gap open penalty: 1.53 Offset value: 0.123	Katoh et al., 2002; Katoh et al., 2005
NAK Kinase Families	5, 27	MAFFT v6.814b	BLOSUM62	Algorithm: L-INS-i Gap open penalty: 1.53 Offset value: 0.123	Katoh et al., 2002; Katoh et al., 2005

Abbreviation: BLOSUM = Blocks Substitution Matrix (Henikoff and Henikoff, 1992); MSA = Multiple Sequence Alignment

TABLE 9. Evolutionary Model Selection for Multiple Sequence Alignments

Alignment Name	Model	# Parameters	BIC	AICc	lnL	Invariant	Gamma
Env9 + 77 orthologs	WAG+G+I	155	77045.0	75785.0	-37737.0	0.012256	1.07365
Env10 + 31 homologs	WAG+G	50	13960.0	13639.0	-6768.8	n/a	1.41934
Env11 and Vid22 + 23 orthologs	JTT+G	48	44765.0	44380.0	-22142.0	n/a	4.30154
Env7 + 114 fungal homologs	WAG+G	228	79070.0	77082.0	-38312.0	n/a	0.85072
Env7 + 67 eukaryotic and bacterial homologs	WAG+G+I	191	93402.0	91825.0	-45720.0	0.006768	1.28762
NAK Kinase Families	JTT+G	92	132730.0	131958.0	-65887.0	n/a	1.24040

Abbreviations: WAG = Whelan and Goldman (Whelan and Goldman, 2001); JTT = Jones, Taylor and Thornton (Jones et al., 1992); G = Discrete Gamma Distribution; I = Evolutionarily Invariant Sites; BIC = Bayesian Information Criterion; AICc = Akaike Information Criterion, corrected; lnL = Maximum Likelihood value; n/a = Not Applicable

TABLE 10. Settings and Average Standard Deviation of Split Frequencies (ASDSF) for Bayesian Analyses

Tree Name	Associated Figure(s)	Fixed Rate Matrix	Rate Variation	Number of Gamma Categories	Heated Chains	Heated Chain Temperature	Burn-in Length	Number of Generations	ASDSF
Env9 + 77 orthologs	20	WAG	gamma + invariant	5	4	0.20	250,000	1,000,000	0.030625
Env10 + 31 homologs	22	WAG	gamma	5	2	0.20	250,000	1,000,000	0.011059
Env11 and Vid22 + 23 orthologs	23	JTT	gamma	5	4	0.20	164,250	657,000	0.009932
Env7 + 114 fungal homologs	2	WAG	gamma	5	2	0.25	250,000	1,000,000	0.015174
Env7 + 67 eukaryotic and bacterial homologs	3	WAG	gamma + invariant	5	2	0.20	250,000	1,000,000	0.035982
NAK Kinase Families	5, 27*	JTT	gamma	5	2	0.20	111,500	446,000	0.008418

Notes: WAG = Whelan and Goldman (Whelan and Goldman, 2001); JTT = Jones, Taylor and Thornton (Jones et al., 1992)
* = Maximum likelihood, Bayesian settings not used

TABLE 11. Programs Used to Determine Possible Phosphorylation Sites in Env7

Program	Type of Analysis	Organism	Web Address	Source/Reference(s)
DISPHOS 1.3	Disorder-Enhanced phosphorylation sites prediction	<i>S. cerevisiae</i>	http://www.dabi.temple.edu/disphos/	Dunker et al., 2002; Diella et al., 2004; Iakoucheva et al., 2004
Kinexus P-site Prediction	Amino acid frequency scoring matrices based on 22,000+ kinase-substrate pairs; Experimental results mined from literature	Human	http://www.phosphonet.ca/	Kinexus Bioinformatics Corporation, Vancouver, Canada; Eswaran et al., 2008
Motifscan	Pattern, profile, and HMM multiple database search	Not specified	http://myhits.isb-sib.ch/cgi-bin/motif_scan	Artimo et al., 2012
NetPhos 2.0	Neural network prediction	Eukaryotes	http://www.cbs.dtu.dk/services/NetPhos/	Blom et al., 1999
NetPhosYeast	Neural network prediction	Yeast	http://www.cbs.dtu.dk/services/NetPhosYeast/	Ingrell et al., 2007
PHOSIDA	Amino acid frequency scoring matrices based on <i>in vivo</i> yeast phosphorylation sites	<i>S. cerevisiae</i>	http://www.phosida.com/	Gnad et al., 2007; Gnad et al., 2011
Predikin	HMM profile search for SDRs	Eukaryotes	http://predikin.biosci.uq.edu.au/	Ellis and Kobe, 2011
PhosphoSVM	Sequence-based features: SE, RE, SS, PD, ASA, OP, ACH, and KNN	Animal	http://sysbio.unl.edu/PhosphoSVM/prediction.php	System Biology Laboratory Of Chi Zhang, University Nebraska-Lincoln
PROSITE Scan	PROSITE database pattern/profile scan	Not specified	http://www.ebi.ac.uk/Tools/pfa/ps_scan/	Sigrist et al., 2002; Sigrist et al., 2013

Abbreviations: HMM = Hidden Markov Model; SDR = Substrate-Determining Residues; SE = Shannon entropy; RE = relative entropy; SS = predicted protein secondary structure; PD = predicted protein disorder; ASA = accessible surface area; OP = overlapping properties; ACH = averaged cumulative hydrophobicity; KNN = k-nearest neighbor

APPENDIX B

DETAILED LISTS OF MEDIA AND REAGENTS USED IN THIS STUDY

TABLE 12. *Saccharomyces cerevisiae* Media Used throughout this Study.

Medium	Constituent/Amount per 100ml	Sources
YPD	Yeast extract – 1.00g	Difco Laboratories, USA
	Bacto™-peptone – 2.00g	Difco Laboratories, USA
	40% dextrose solution – 5.00ml	Sigma-Aldrich, St. Louis, MO
SM-uracil	Yeast nitrogen base without amino acids – 0.67g	Difco Laboratories, USA
	20 mg/L Histidine solution – 200.00µl	Sigma-Aldrich, St. Louis, MO
	60 mg/L Leucine solution – 1.00ml	Sigma-Aldrich, St. Louis, MO
	30 mg/L Lysine solution – 300.00µl	Sigma-Aldrich, St. Louis, MO
	Sodium hydroxide – half a small pellet	Sigma-Aldrich, St. Louis, MO
	40% dextrose solution – 5.00ml	Sigma-Aldrich, St. Louis, MO
Nitrogen-limited	Yeast nitrogen base without amino acids and ammonium sulfate – 0.17g	Difco Laboratories, USA
	40% dextrose solution – 27.75µl	Sigma-Aldrich, St. Louis, MO
	Sodium sulfate – 0.50g	Sigma-Aldrich, St. Louis, MO
	Ammonium sulfate – 62.50mg	Sigma-Aldrich, St. Louis, MO
Nitrogen-starvation	Yeast nitrogen base without amino acids and ammonium sulfate – 0.17g	Difco Laboratories, USA
	40% dextrose solution – 27.75µl	Sigma-Aldrich, St. Louis, MO
	Sodium sulfate – 0.50g	Sigma-Aldrich, St. Louis, MO
YPD 0.5/1.5M NaCl	Same as YPD 2% dextrose except: Sodium chloride – 2.93g/8.78g	Sigma-Aldrich, St. Louis, MO
YPD 1M Sorbitol	Same as YPD 2% dextrose except: Sorbitol – 18.22g	Sigma-Aldrich, St. Louis, MO
YPD 10mM CuSO ₄	Same as YPD 2% dextrose except: Copper (II) sulfate pentahydrate – 1.25g	Sigma-Aldrich, St. Louis, MO
YPD 10mM MnCl ₂	Same as YPD 2% dextrose except: Manganese (II) chloride tetrahydrate – 1.00g	Sigma-Aldrich, St. Louis, MO
YPD 1%/10%/15% Dextrose	Same as YP (YPD without dextrose) except: 40% dextrose solution - 2.50ml/25.00ml/37.50ml	Sigma-Aldrich, St. Louis, MO
YPD 0.2% Dextrose	Same as YP (YPD without dextrose) except: 40% dextrose solution – 500.00µl	Sigma-Aldrich, St. Louis, MO
2% Glycerol	Glycerol – 2.00ml	Sigma-Aldrich, St. Louis, MO

Notes: After autoclavable components were combined, deionized water was used to fill to 100ml or the appropriate volume based on final dextrose concentration. The media listed above also contained agar (1.50g/100ml) when used as solid media plates (Difco Laboratories, USA).

TABLE 13. *Escherichia coli* Media Used for Site-Directed Mutagenesis and Plasmid preps.

Medium	Constituent/Amount per 100ml	Sources
TYE (LB)	Bacto™-tryptone – 1.00g	Difco Laboratories, USA
	Yeast extract – 0.50g	Difco Laboratories, USA
	Sodium chloride – 1.00g	Sigma-Aldrich, St. Louis, MO
TYE (LB) + Ampicillin	Same as TYE (LB) except: Ampicillin – 10.00mg	Sigma-Aldrich, St. Louis, MO

Notes: After autoclavable components were combined, deionized water was used to fill to 100ml. The media listed above also contained agar (1.50g/100ml) when used as solid media plates. (Difco Laboratories, USA)

TABLE 14. Subcellular Fractionation Buffers and Solutions.

Solution/Buffer	Component	Amount
Lysis Buffer	4M Sorbitol	2.50ml
	1M Potassium Acetate	2.50ml
	0.5M EDTA	500.00µl
	1M HEPES-Potassium hydroxide (pH 6.8)	1.00ml
	200 mM PMSF	50.00µl
	10 µg/ml Aprotinin in DMSO	5.00µl
	5 mg/ml Pepstatin A in ethanol	10.00µl
	1M DTT	5.00µl
DTT Solution	1M DTT	50.00µl
	1M Tris (pH 9.4)	500.00µl
	Sterile water	4.45ml
DEAE-Dextran Solution	10 mg/ml DEAE-Dextran	40.00µl
	0% Ficoll	960.00µl
Spheroplasting Buffer	0.2x YPD	8.00ml
	1M Monopotassium phosphate buffer (pH 7.5)	500.00µl
	4M Sorbitol	1.50ml

Note: All chemicals were acquired from Sigma-Aldrich (St. Louis, MO).

TABLE 15. Preparation of BSA Standards for Protein Quantification

Standard	Preparation	Final BSA Concentration
1	0.00µl 2 mg/ml BSA + 100.00µl sterile water	0.00 mg/ml
2	12.50µl 2 mg/ml BSA + 87.50µl sterile water	0.25 mg/ml
3	25.00µl 2 mg/ml BSA + 75.00µl sterile water	0.50 mg/ml
4	50.00µl 2 mg/ml BSA + 50.00µl sterile water	1.00 mg/ml
5	75.00µl 2 mg/ml BSA + 25.00µl sterile water	1.25 mg/ml
6	100.00µl 2 mg/ml BSA + 0.00µl sterile water	2.00 mg/ml

Note: Concentrated BSA was acquired from Sigma-Aldrich (St. Louis, MO).

TABLE 16. Preparation of Large and Small 7.5% SDS-PAGE Gels

Resolving Gel Component	Amount (Lg/Sm)
30% Protogel (37.5:1 Acrylamide to Bisacrylamide Stabilized Solution)	10.00ml/2.47ml
4x Protogel Resolving Buffer	10.00ml/2.47ml
Deionized water	20.00ml/4.95ml
10% Ammonium Persulfate	405.00µl/100.00µl
Tetramethylethylenediamine (TEMED)	45.00µl/10.00µl

Stacking Gel Component	Amount (Lg/Sm)
30% Protogel (37.5:1 Acrylamide to Bisacrylamide Stabilized Solution)	1.31ml/657.00µl
4x Protogel Stacking Buffer	2.52ml/1.26ml
Deionized water	6.16ml/3.08ml
10% Ammonium Persulfate	105.00µl/50.00µl
Tetramethylethylenediamine (TEMED)	12.00µl/5.00µl

Notes: Isopropanol was poured on top to even-out resolving gel. Gel set for 1 hour. Stacking gel was poured on top of solidified resolving gel. After insertion of the comb, the gel set for 30 minutes. Protogel materials were from National Diagnostics (Atlanta, GA); all other materials were purchased from Sigma-Aldrich (St. Louis, MO).

TABLE 17. 6x SDS-PAGE Sample Buffer (6x Laemmli Buffer)

Component	Amount Required for 10.00ml
Sodium dodecyl sulfate (SDS)	1.20g
Bromophenol Blue	6.00mg
100% Glycerol	4.70ml
0.5M Tris (pH 6.8)	1.20ml
Deionized water	2.10ml
Dithiothreitol (DTT)	0.93g

Note: All chemicals were acquired from Sigma-Aldrich (St. Louis, MO).

TABLE 18. 5x SDS-PAGE Running Buffer

Component	Amount Required for 500.00ml
Trizma base	7.20g
Glycine	34.56g
Sodium dodecyl sulfate (SDS)	2.40g
Deionized water	Fill to 500.00ml

Notes: 5x SDS-PAGE running buffer was diluted prior to the running of the gel to 1x with deionized water. All chemicals were acquired from Sigma-Aldrich (St. Louis, MO).

TABLE 19. Western Transfer Buffer

Component	Amount Required to Make 1.00L
Trizma base	5.82g
Glycine	2.93g
Methanol	200.00ml
Deionized water	Fill to 800.00ml

Note: All chemicals were acquired from Sigma-Aldrich (St. Louis, MO).

TABLE 20. Tris-Buffered Saline (TBS)

Component	Amount Required to Make 500.00ml
1.0M Tris-HCl	25.00ml
3.0M NaCl	25.00ml
Deionized water	450.00 μ l

Note: All chemicals were acquired from Sigma-Aldrich (St. Louis, MO).

TABLE 21. TWEEN[®]20 + Tris-Buffered Saline (TTBS)

Component	Amount Required to Make 500.00ml
1.0M Tris-HCl	25.00ml
3.0M NaCl	25.00ml
TWEEN [®] 20 (Polyoxyethylenesorbitan monolaurate)	250.00 μ l
Deionized water	449.75 μ l

Note: All chemicals were acquired from Sigma-Aldrich (St. Louis, MO).

TABLE 22. TTBS + Milk (5% TTBSM)

Component	Amount Required to Make 100.00ml
TTBS	100.00ml
Dry milk	5.00g

Note: All chemicals were acquired from Sigma-Aldrich (St. Louis, MO).

TABLE 23. 5x Tris-Borate-EDTA

Component	Amount Required to Make 500.00ml
Trizma base	27.00g
Boric acid	13.75g
0.5M EDTA	10.00ml
Deionized water	Fill to 500.00ml

Note: All chemicals were acquired from Sigma-Aldrich (St. Louis, MO).

TABLE 24. Poly-Lithium-Acetate-Tris-EDTA (PLATE) Mixture

Component	Amount Required to Make ~100.00ml
45% Polyethylene glycol (PEG) 8000	90.00ml
1M Lithium acetate	10.00ml
1M Tris (pH 7.5)	1.00ml
0.5M EDTA	0.20ml

Note: Mixture was filter-sterilized. All chemicals were acquired from Sigma-Aldrich (St. Louis, MO).

TABLE 25. Phosphorylation Upshift Assay Reagents

10x Reaction Buffer	
Component	Amount Required to Make 12.00ml
3M Potassium chloride	6.00ml
1M Manganese chloride	60.00 μ l
1M Magnesium chloride	60.00 μ l
4M Sorbitol	600.00 μ l
1M Pipes/Potassium hydroxide solution (pH 8.6)	120.00 μ l
Sterile water	5.16ml
10x ATP-Regenerating System	
Component	Total Concentration
ATP	5.00mM
Creatine kinase	1.00mg/ml
Creatine phosphate	400.00mM
Pipes (pH 6.8)	10.00mM
Sorbitol	200.00mM
0% Ficoll	
Component	Amount Required to Make 50.00ml
4M Sorbitol	2.50ml
1M Pipes/potassium hydroxide (pH 8.6)	0.50ml
Sterile water	47.00ml
200mM ATP Solution	
Component	Amount
ATP	1.65g
Sterile water	15.00ml
Potassium hydroxide (pH 7.0)	600.00 μ l

Note: 10x Reaction Buffer and 0% Ficoll were filter-sterilized. All chemicals were acquired from Sigma-Aldrich (St. Louis, MO).

APPENDIX C
ADDITIONAL FIGURES

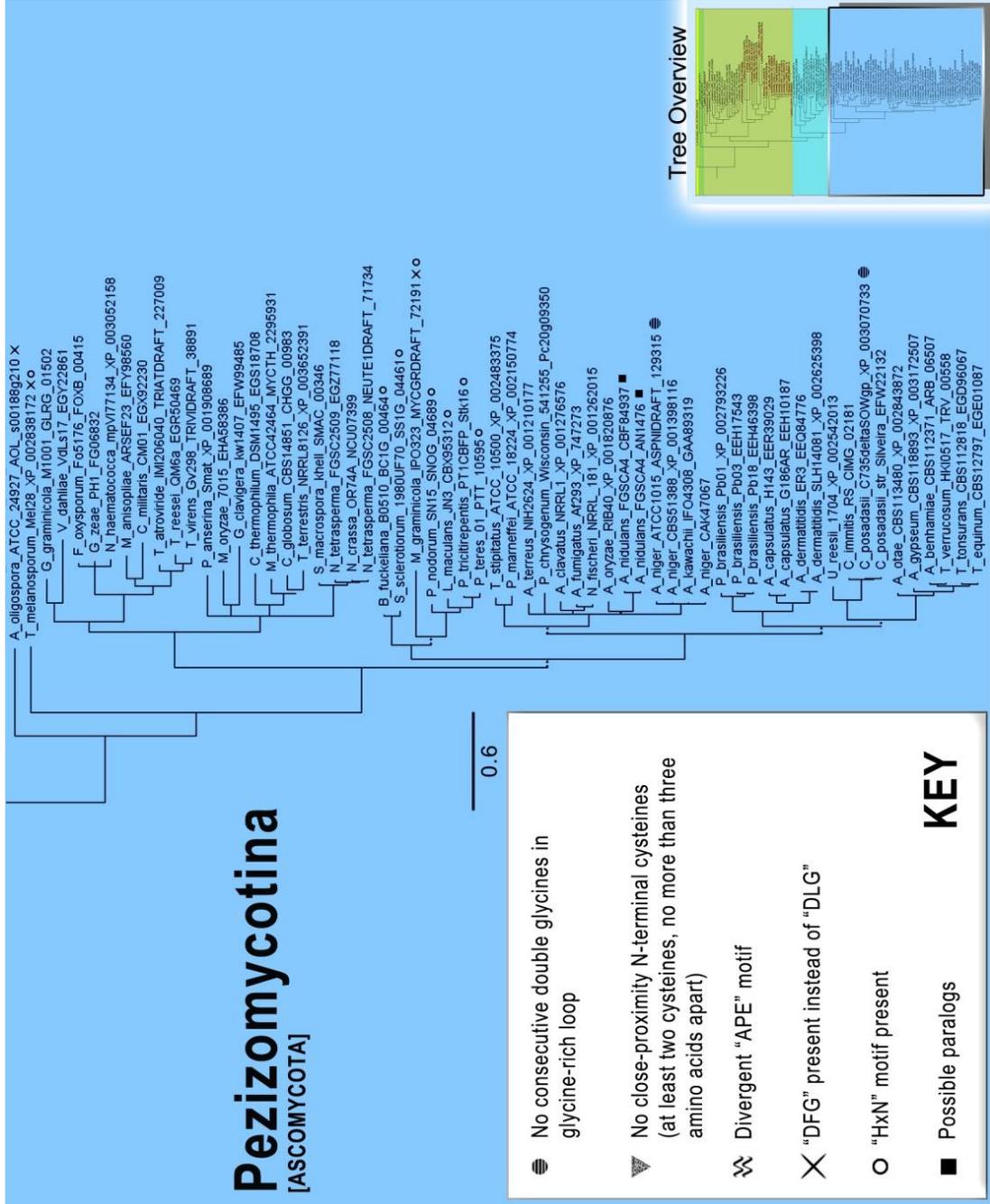


FIGURE 24. (Continued)
 Dotted branches represent clades with less than 90% posterior probability. Some posterior probability values are shown for specific nodes. The scale bar depicts the branch length on the tree that signifies 0.6 amino acid substitutions per site. Symbols after a species and protein name signify a particular characteristic of the protein sequence. Clades are color-coded by the phylum (all capital letters) or subphylum of its constituent species. Sequences belonging to Saccharomycetaceae family organisms are highlighted in red-orange.

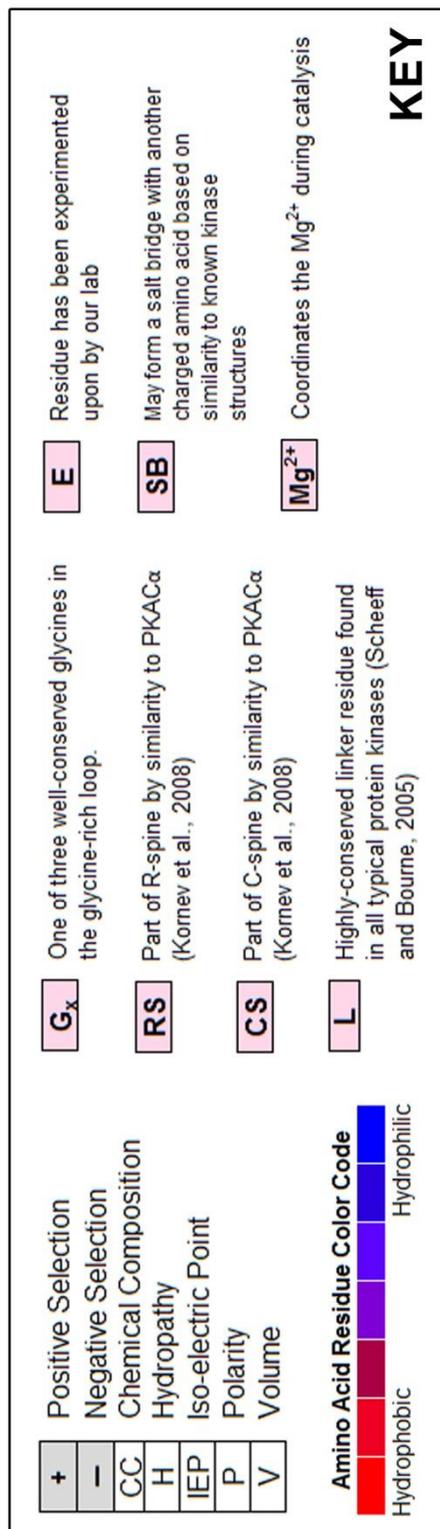
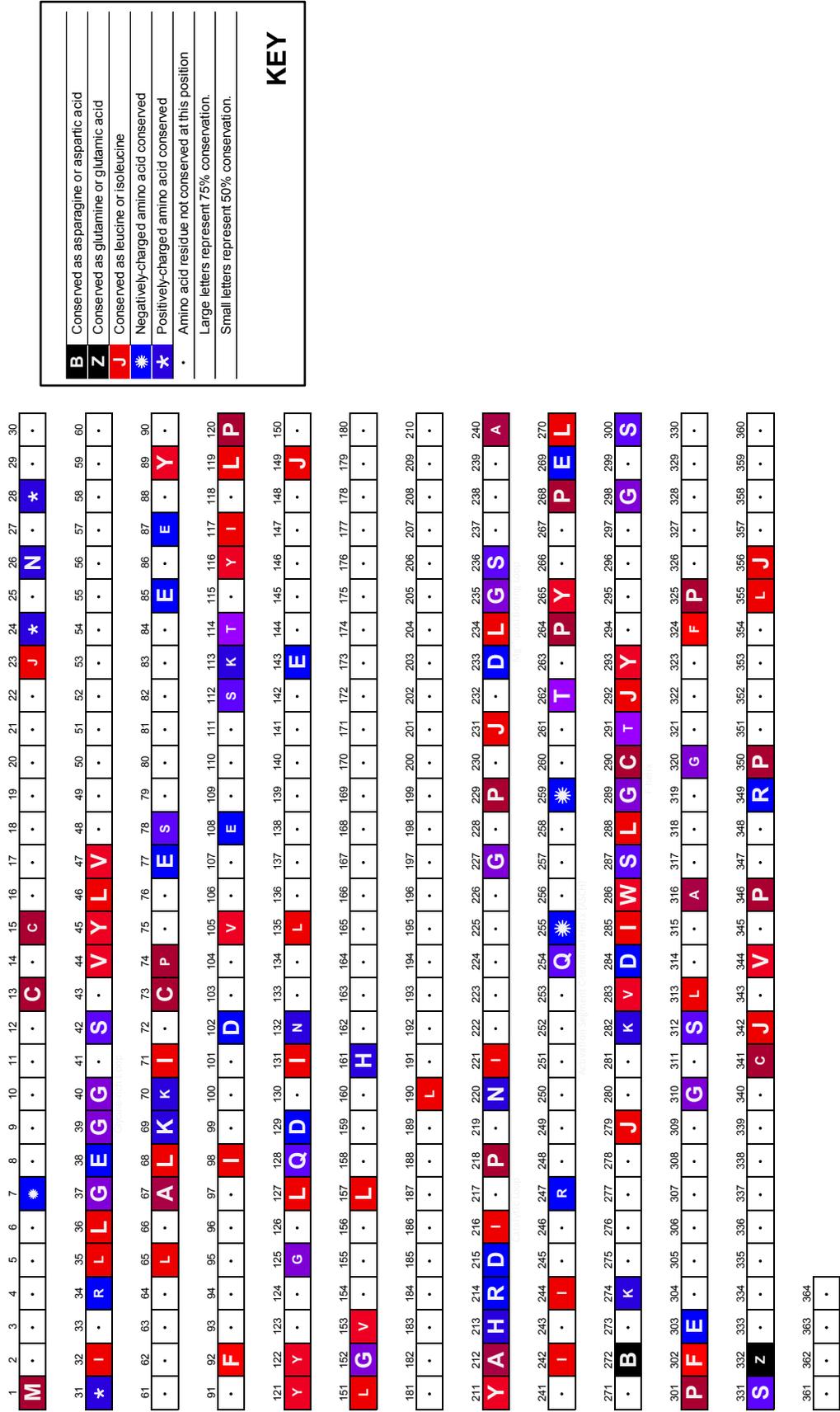
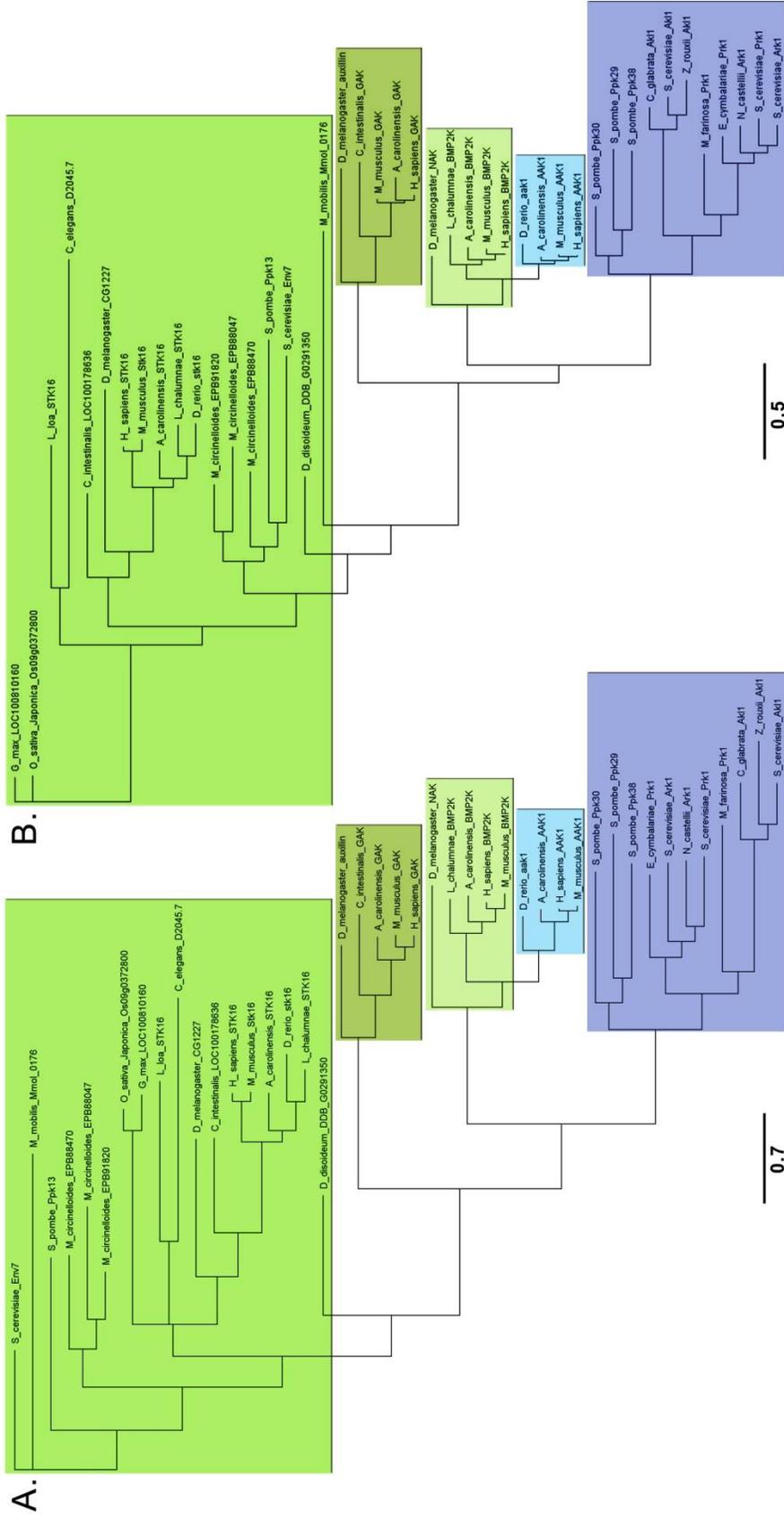


FIGURE 25. (Continued) These charts represent the 12 kinase subdomains of Env7. Amino acid residues are color-coded by their hydrophobicity/hydrophilicity. Conserved residues are marked with either a black box with a white check mark, or a grey box with a black check mark. Grey letters represent residues in the orthologs that do not match those in Env7 but are nonetheless well-conserved. Negative selection, relative to the program used, is seen as the dominant form of evolutionary selection in the light grey-shaded boxes with minus signs in them. Green and yellow boxes with letter abbreviations in them indicate that the amino acid properties at a residue site are changing or are being conserved, respectively. Below the chart, light pink boxes containing a letter abbreviation can be found for putatively important residues, and long navy blue bars denote specific three-dimensional structures present on Env7. The brown boxes in stretching from subdomains VII and VIII represent the activation segment. Descriptions of all abbreviations can be found in the Figure key.



B	Conserved as asparagine or aspartic acid
Z	Conserved as glutamine or glutamic acid
J	Conserved as leucine or isoleucine
*	Negatively-charged amino acid conserved
*	Positively-charged amino acid conserved
.	Amino acid residue not conserved at this position
Large letters represent 75% conservation.	
Small letters represent 50% conservation.	
KEY	

FIGURE 26. Conserved residues in Env7 based on an alignment with 151 of its orthologs from Eukarya and Bacteria.



Complete NAK Sequences

Truncated NAK Sequences

Figure 27. Comparison of phylogenetic trees generated using complete and truncated NAK kinase sequences shows nearly the same topology. A) Unrooted Bayesian tree of complete NAK kinase sequences (same as in Figure 6). B) Maximum likelihood tree of only the kinase domains of the NAK kinase sequences. Settings for the maximum likelihood analysis were the same as the Bayesian one listed in Table 10 in Appendix A. The scale bar indicated the number of amino acid changes per site.

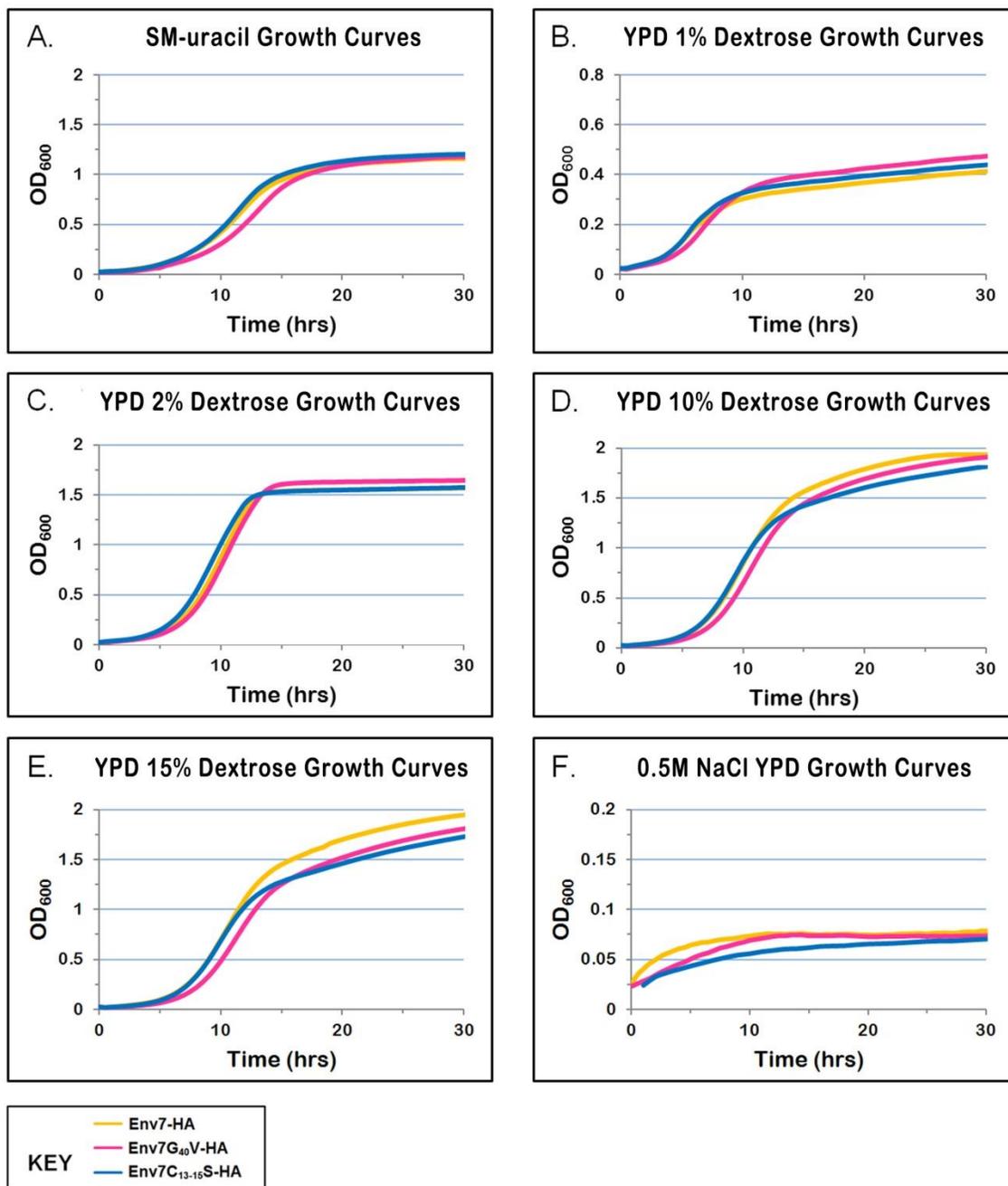


FIGURE 28. Preliminary growth curves for Env7-HA, Env7G₄₀V-HA, and Env7C₁₃₋₁₅S-HA –expressing strains in various liquid media reveal minor differences in growth pattern. Liquid cultures (200µl) of the three strains were grown for 30 hours in 30°C, with orbital agitation every 15 minutes. Growth (O.D.₆₀₀) was measured every 30 minutes. A) SM-ura. B) 1% dextrose YPD. C) 2% dextrose YPD. D) 10% dextrose YPD. E) 15% dextrose YPD. F) 0.5M NaCl YPD (2% dextrose).

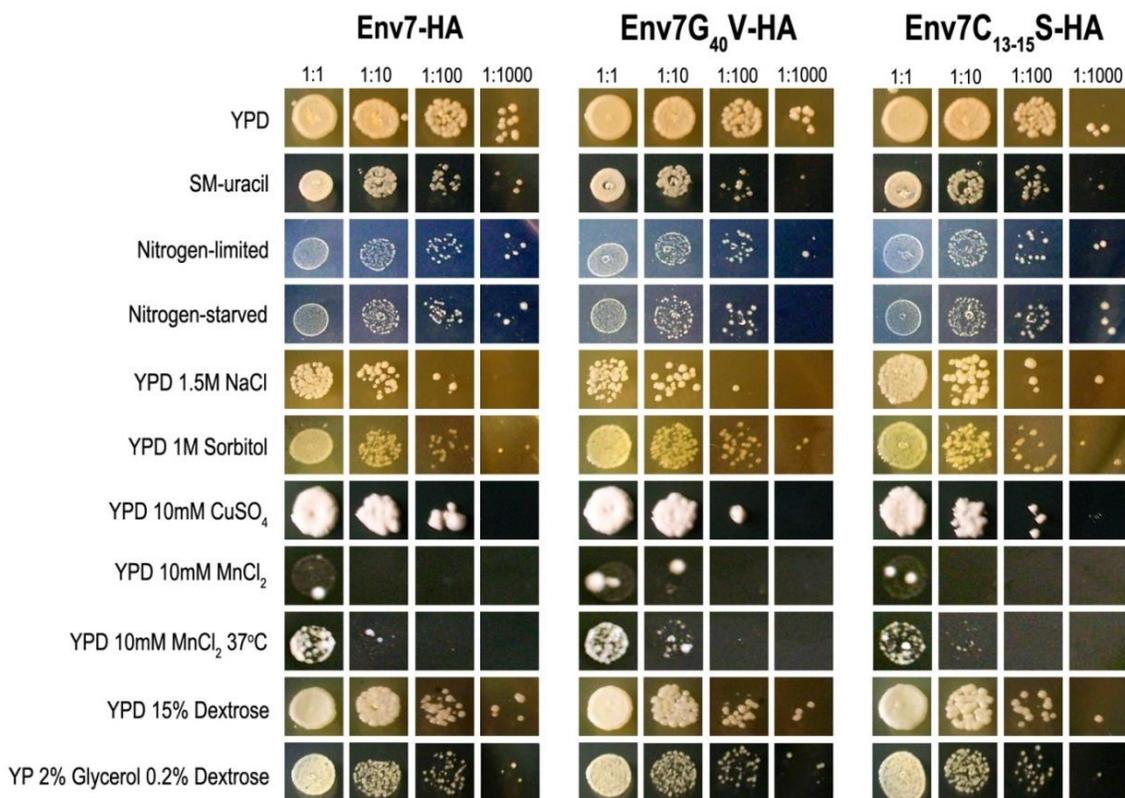


FIGURE 29. Preliminary fitness analysis of WT-HA, Env7G₄₀V-HA, and Env7C₁₃₋₁₅S-HA expressing cells on various solid media does not show major growth differences. Exactly 5 μ l of each dilution of liquid YPD culture were spotted onto each plate for each strain. These dilutions – 1:1, 1:10, 1:100, and 1:1000 – contained about 15,000, 1,500, 150, and 15 cells respectively. All plates shown, with the exception of the YPD 10mM MnCl₂ 37°C plate, were grown in 30°C. Incubation times varied from 2 days to 12 days. Unless otherwise specified, YPD media contained 2% dextrose.

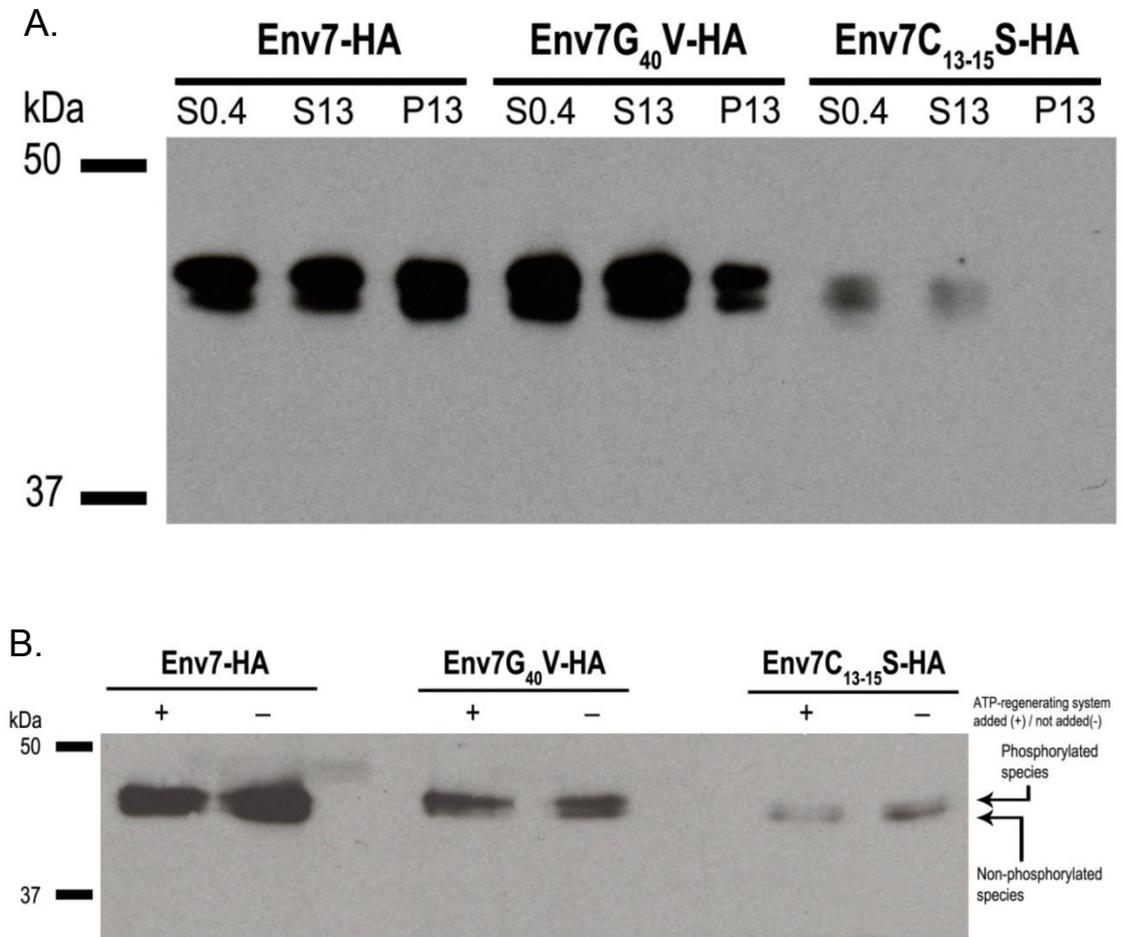


FIGURE 30. Preliminary analysis of subcellular localization and phosphorylation state of the Env7G₄₀V-HA mutant protein. A) Western blot of subcellular fractions probed for HA-tagged Env7. Cells were subjected to subcellular fractionation and run on a large 7.5% SDS-PAGE gel. B) Phosphorylation assay of the P13 fraction of subcellular lysates run on a 7.5% SDS-PAGE gel. The P13 (vacuoles/ER/plasma membrane/mitochondria) fractions of the subcellular lysates for each strain were subjected to a phosphorylation assay and then Western blotted. Like Env7-HA, the Env7G₄₀V-HA mutant is able to autophosphorylate with (+) and without (-) added ATP-regenerating system, as evidenced by the double band in both lanes for the mutant.

APPENDIX D

POTENTIAL TARGETS FOR MUTATION IN *ENV7*

TABLE 26. Potential Target Amino Acid Residues in Env7 for Further Analysis

Env7 Residue	Conservation	(Potential) Function	Notes
Cysteine 13	89.5%	Attachment to the vacuole	<ul style="list-style-type: none"> - Experimentally shown to be important for attachment to the vacuole membrane (Manandhar et al., 2013) - Env7C₁₃₋₁₅S = kinase-dead (Manandhar et al., 2013)
Cysteine 14	21.1%	Attachment to the vacuole	
Cysteine 15	86.2%	Attachment to the vacuole	
Leucine 36	94.7%	ATP-binding	<ul style="list-style-type: none"> - Homologous to PKA L₄₉, which interacts hydrophobically with the adenine base (reviewed in Bossemeyer 1994)
Glycine 37	86.8%	ATP-binding: Correct positioning and securing of ATP in the catalytic cleft	<ul style="list-style-type: none"> - Homologous to PKA G₅₀, which must be a small amino acid in order to make room for ribose ring in ATP (reviewed in Bossemeyer 1994) - G₅₀ mutation to A or S leads to decreased PKA catalytic activity (Hemmer et al., 1997) - Considered to be the most catalytically important of the three glycine residues in the glycine-rich loop (Grant et al., 1998) - B-raf G₄₆₃E mutation found in human cancer leads to increased catalytic activity (Davies et al., 2002) - B-raf G₄₆₃V mutation showed decreased catalytic activity (Ikenoue et al., 2004)
Glycine 39	90.8%	ATP-binding: Correct positioning and securing of ATP in the catalytic cleft	<ul style="list-style-type: none"> - Homologous to G₅₂ in PKA; G₅₂S mutation in PKA affected ATP-binding and lowered rate of phosphoryl transfer (Grant et al., 1998) - Thought to help position backbone of S₅₃ (possibly S₄₂ in Env7) in order for it to make contact with γ-phosphate of ATP (Grant et al., 1998)
Glycine 40	88.8%	ATP-binding: Correct positioning and securing of ATP in the catalytic cleft	<ul style="list-style-type: none"> - In serine/threonine kinases, this is the most variable of the three conserved glycines (reviewed in Bossemeyer 1994) - Mutation of glycine-rich loop third glycine to valine in insulin receptor tyrosine kinase linked to non-insulin dependent diabetes mellitus- most likely due to lowered catalytic activity (Odawara et al., 1989) - B-raf G₄₆₈A mutation found in human colorectal cancer leads to dramatically increased catalytic activity Ikenoue et al., 2004) - G₄₀V mutation in Env7 led to some phenotypic variances from WT (this study)

TABLE 26. (Continued)

Env7 Residue	Conservation	(Potential) Function	Notes
Serine 42	88.8%	ATP-binding	<ul style="list-style-type: none"> – Part of the glycine-rich loop that folds on top of gamma phosphate of ATP for catalysis (Grant et al., 1998) – PKA S₅₃P compromised catalytically (Aimes et al., 2000) – It is the backbone amide of S₅₃ on PKA that is necessary for catalysis, not the side-chain. (Aimes et al., 2000) – PKA S₅₃K was not expressed easily in <i>E. coli</i> (Pflug et al. 2011)
Valine 44	88.2%	ATP-binding C-spine Residue	<ul style="list-style-type: none"> – Homologous to PKA V₅₇, which is part of the C-spine that contacts the adenine ring of ATP (Kornev et al., 2008) – Partakes in essential hydrophobic interaction with adenosine (reviewed in Bossemeyer 1994)
Alanine 67	94.1%	ATP-binding C-spine Residue	<ul style="list-style-type: none"> – Homologous to PKA A₇₀, which is part of the C-spine that contacts the adenine ring of ATP (Kornev et al. 2008) – Contributes to hydrophobic adenine ring binding pocket (Hanks and Hunter, 1995)
Lysine 69	98.0%	Possible Salt Bridge-forming Residue	<ul style="list-style-type: none"> – K₆₉R and K₆₉M Env7 mutants both stable and able to autophosphorylate (Manandhar et al., 2013)
Lysine 70	73.7%	Unknown	<ul style="list-style-type: none"> – K₇₀R and K₇₀M Env7 mutants both stable and able to autophosphorylate (Manandhar et al., 2013)
Glutamic Acid 85	98.7%	Possible Salt Bridge-forming Residue	<ul style="list-style-type: none"> – Homologous to PKA E₉₁ – Known to form a salt bridge with PKA K₇₂ (Env7 K₆₉) when in the active conformation (Taylor et al., 2012)
Isoleucine 86	30.9%	R-spine Residue	<ul style="list-style-type: none"> – Conserved as isoleucine or valine in 87.5% of Env7 orthologs
Phenylalanine 92	92.8%	Unknown	<ul style="list-style-type: none"> – Well-conserved among Env7 orthologs
Isoleucine 101	44.7%	R-spine Residue	<ul style="list-style-type: none"> – Conserved as leucine, isoleucine, or valine in 86.8% of Env7 orthologs
Isoleucine 117	54.0%	ATP-binding	<ul style="list-style-type: none"> – Possible “gatekeeper” residue based on alignment with PKA (M₁₂₀) (Taylor and Kornev 2011); could control access of ATP to activation site; smaller gatekeeper residue = larger ATP analog can bind (Zuccotto et al., 2009) – 96.1% of residues at this position in the alignment are either isoleucine, leucine, or valine
Proline 120	96.7%	Unknown	<ul style="list-style-type: none"> – Well-conserved among Env7 orthologs
Glycine 125	90.1%	Unknown	<ul style="list-style-type: none"> – Well-conserved among Env7 orthologs
Serine 126	26.3%	Phosphorylation	<ul style="list-style-type: none"> – Strongly predicted to be phosphorylated
Leucine 127	92.1%	C-spine Residue	<ul style="list-style-type: none"> – Part of the C-spine

TABLE 26. (Continued)

Env7 Residue	Conservation	(Potential) Function	Notes
Histidine 161	97.4%	Possible Linker Residue	<ul style="list-style-type: none"> – Positioned immediately before the subdomain VIa insert that structurally sits at the back of the kinase – Could form an H-bond with D₂₈₄ which links together Helices E, F, and the crossing loops in the catalytic region (Scheeff & Bourne, 2005) – The H-bond between the two residues described above is very visible on the <i>in silico</i> 3D Env7 structure (results not shown) – Extremely well-conserved among Env7 orthologs
Aspartic Acid 185	20.4%	Localization	<ul style="list-style-type: none"> – Located in the subdomain VIA insert region – Could be part of an acidic di-leucine signaling motif that allows vacuolar localization via the ALP pathway (Vowels and Payne 1998; Darsow et al., 1998). – Conserved as D or E in 42.8% of orthologs
Glutamic Acid 186	45.4%	Localization	<ul style="list-style-type: none"> – Located in the subdomain VIA insert region – Could be part of an acidic di-leucine signaling motif that allows vacuolar localization via the ALP pathway (Vowels and Payne 1998; Darsow et al., 1998). – Conserved as D or E in 52.0% of orthologs
Leucine 190	67.1%	Localization	<ul style="list-style-type: none"> – Located in the subdomain VIA insert region – Could be part of an acidic di-leucine signaling motif that allows vacuolar localization via the ALP pathway (Vowels and Payne 1998; Darsow et al., 1998).
Threonine 194	10.5%	Phosphorylation	<ul style="list-style-type: none"> – Located in the subdomain VIA insert region – Strongly predicted to be phosphorylated
Histidine 213	95.4%	R-spine Residue	<ul style="list-style-type: none"> – Part of “HRD” motif and R-spine (Kornev et al. 2008); homologous to PKA Y₁₆₄
Arginine 214	82.2%	ATP-binding	<ul style="list-style-type: none"> – Homologous to PKA R₁₆₅; part of universally conserved “HRD” motif; anchors ATP γ-phosphate on PKA T₁₉₇ (Taylor et al. 2004)

TABLE 26. (Continued)

Env7 Residue	Conservation	(Potential) Function	Notes
Aspartic Acid 215	94.7%	Phosphorylation	<ul style="list-style-type: none"> – Homologous to human AAK1 D₁₇₆; AAK1 D₁₇₆A mutation did not phosphorylate AP2 substrate (Conner and Schmid, 2003) – Pkr1p D₁₅₈Y (homologous residue) = kinase dead (Zeng and Cai, 1999; Huang et al., 2003) – Homologous to PKA D₁₆₆; serves as a catalytic base (Taylor et al., 1995)
Threonine 217	87.5%	ATP-binding Phosphorylation	<ul style="list-style-type: none"> – Homologous to K₁₆₈ in PKA, which converges on γ-phosphate (along with S₄₂ and D₂₁₅ “HRD”) to phosphorylate substrate – Could H-bond with N₂₂₀ in Env7p 3d <i>in silico</i> structure (just as N₁₇₁ and K₁₆₈ do in PKA) (Taylor et al., 2012) – Strongly predicted to be phosphorylated – The side-chain of the lysine that is conserved in all the other sequences and in PKA serves to stabilize the γ-phosphate (Taylor et al., 2004)
Asparagine 220	96.1%	ATP-binding Mg ²⁺ -binding	<ul style="list-style-type: none"> – Part of catalytic loop; also coordinates with ATP and bound Mg²⁺; homologous residue N₂₁₀A in PLK2 = kinase dead (Cizmecioglu et al., 2008) – H-bonds with T₂₁₇ in Env7 3D <i>in silico</i> structure (just as N₁₇₁ and K₁₆₈ do in PKA) (Taylor et al., 2012)
Isoleucine 221	65.1%	C-spine Residue	<ul style="list-style-type: none"> – Conserved as either isoleucine or valine in 95.4% of Env7 orthologs
Leucine 222	31.6%	ATP-binding C-spine residue	<ul style="list-style-type: none"> – Corresponding residue on PKA docks directly onto adenine ring; when homologous residue on PKA (L₁₇₃) mutated to alanine catalytic activity is abolished – Conserved as a non-polar amino acid in 98.0% of Env7 orthologs
Phenylalanine 223	9.2%	C-spine Residue	<ul style="list-style-type: none"> – Conserved as a non-polar amino acid in 100.0% of Env7 orthologs
Aspartic Acid 233	96.7%	ATP-binding Mg ²⁺ -binding	<ul style="list-style-type: none"> – Homologous to PKA D₁₈₄, which is known to bind the catalytic Mg²⁺ ion via side-chain oxygens; this Mg²⁺ ion chelates both β- and γ-phosphates of ATP (Adams, 2001; Valiev et al. 2007)
Leucine 234	75.7%	R-spine Residue	
Serine 236	96.1%	Unknown	<ul style="list-style-type: none"> – Part of the activation segment in Env7 – Very well-conserved among Env7 orthologs – Possibly phosphorylated
Serine 238	21.1%	Phosphorylation	<ul style="list-style-type: none"> – Strongly predicted to be phosphorylated – Found in the activation segment

TABLE 26. (Continued)

Env7 Residue	Conservation	(Potential) Function	Notes
Threonine 262	79.6%	Phosphorylation	– Conserved as either a serine or a threonine in 96.7% of Env7 orthologs
Tyrosine 265	89.5%	Phosphorylation	– Homologous to STK16 Y ₁₉₈ , which is an autophosphorylation site (Eswaran et al., 2008) – Homologous to PKA Y ₂₀₄ , which anchors the activation segment by working with PKA T ₂₀₁ (also conserved as T ₂₆₂ in Env7) to orient K ₁₆₈ 's side chain (Moore et al., 2003; Yang et al. 2005; Masterson et al., 2008) – Env7 has a threonine (217) residue aligned with PKA K ₁₆₈ (STK16 has a lysine there, as well.) – All of the Env7 orthologs have an amino acid with an aromatic benzene ring (Y, W, or F) conserved at this position.
Proline 267	5.3%	Unknown	– Part of the “APE” motif – Conserved as alanine in 91.5% of Env7 orthologs; the rest have serine, proline or cysteine.
Proline 268	100.0%	Invariant Residue	– Part of the “APE” motif
Glutamic Acid 269	100.0%	Kinase activity Possible Salt Bridge-forming Residue Invariant Residue	– Mutation to alanine in PKA (E ₂₀₈) abolished kinase activity without disrupting structure (Yang et al., 2012) – E _{202A} in STK16 = kinase-dead (results not shown) (Eswaran et al., 2008) – E _{269S} and E _{269D} in Env7 = stable and able to autophosphorylate (Manandhar et al., 2013) – E _{269A} in Env7 completely unstable-proteasomally degraded, most likely because salt bridge could not be formed (Manandhar et al., 2013)
Leucine 270	96.7%	Unknown	– Well-conserved in Env7 orthologs
Serine 280	7.2%	Phosphorylation	– Strongly predicted to be phosphorylated
Serine 281	4.6%	Phosphorylation	– Strongly predicted to be phosphorylated
Aspartic Acid 284	100.0%	R-spine Residue Invariant Residue	
Tryptophan 286	99.3%	Unknown	– Extremely well-conserved among Env7 orthologs
Serine 287	98.0%	Unknown	– Extremely well-conserved among Env7 orthologs
Glycine 289	100.0%	Invariant Residue	
Cysteine 290	98.7%	Unknown	– Extremely well-conserved among Env7 orthologs
Threonine 291	65.8%	C-spine Residue	
Tyrosine 293	94.1%	Unknown	– Well-conserved among Env7 orthologs
Leucine 295	27.0%	C-spine Residue	
Serine 300	84.9%	Phosphorylation	– Strongly predicted to be phosphorylated

TABLE 26. (Continued)

Env7 Residue	Conservation	(Potential) Function	Notes
Proline 301	98.7%	Unknown	– Extremely well-conserved in Env7 orthologs
Glutamic Acid 303	81.6%	Unknown	– Conserved as D or E in 97.4% of Env7 orthologs
Glycine 310	96.1%	Unknown	– Extremely well-conserved in Env7 orthologs
Serine 328	15.1%	Phosphorylation	– Strongly predicted to be phosphorylated
Serine 331	83.6%	Phosphorylation	– Strongly predicted to be phosphorylated
Arginine 349	97.4%	Possible Salt Bridge-forming Residue	– Mutation to alanine in PDK (R ₂₈₀) abolished kinase activity without disrupting structure (Yang et al., 2012)
			– Interacts with universally conserved “APE” motif glutamic acid (Kornev et al., 2008)
Proline 350	95.4%	Unknown	– Extremely well-conserved in Env7 orthologs
Threonine 352	6.6%	Phosphorylation	– Strongly predicted to be phosphorylated

REFERENCES

REFERENCES

- Adams J. 2001. Kinetic and catalytic mechanisms of protein kinases. *Chem Rev* 101: 2271-2290.
- Aimes R, Hemmer W, Taylor S. 2000. Serine-53 at the tip of the glycine-rich loop of cAMP-dependent protein kinase: Role in catalysis, P-site specificity, and interaction with inhibitors. *Biochemistry* 39: 8325-8332.
- Altschul s, Wootton J, Gertz E, Agarwala R, Morgulis A, Schäffer A, Yu Y. 2005. Protein database searches using compositionally adjusted substitution matrices. *FEBS J* 272: 5101-5109.
- Altschul S, Gish W, Miller W, Myers E, Lipman D. 1990. Basic local alignment search tool. *J Mol Biol* 215: 403-410.
- Anand V, Daboussi L, Lorenz T, Payne G. 2009. Genome-wide analysis of AP-3-dependent protein transport in yeast. *Mol Biol Cell* 20: 1592-1604.
- Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier E, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechi R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* 40: 597-603.
- Balderhaar H, Ungermann C. 2013. CORVET and HOPS tethering complexes - coordinators of endosome and lysosome fusion. *J Cell Sci* 126: 1307-1316.
- Ballabio A, Gieselmann V. 2009. Lysosomal disorders: From storage to cellular damage. *Biochim Biophys Acta* 1793: 684-696.
- Banta L, Robinson J, Klionsky D, Emr S. 1988. Organelle assembly in yeast: Characterization of yeast mutants defective in vacuolar biogenesis and protein sorting. *J Cell Biol* 107: 1369-1383.
- Bartholomew C, Hardy C. 2009. p21-activated kinases Cla4 and Ste20 regulate vacuole inheritance in *Saccharomyces cerevisiae*. *Eukaryot Cell* 8: 560-572.
- Bateman A, Coggill P, Finn R. 2010. DUFs: Families in search of function. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 66: 1148-1152.

- Belyaeva O, Korkina O, Stetsenko A, Kim T, Nelson P, Kedishvili N. 2005. Biochemical properties of purified human retinol dehydrogenase 12 (RDH12): Catalytic efficiency toward retinoids and C9 aldehydes and effects of cellular retinol-binding protein type I (CRBPI) and cellular retinaldehyde-binding protein (CRALBP) on the oxidation and reduction of retinoids. *Biochemistry* 44: 7035-7047.
- Benayoun L, Spiegel R, Auslender N, Abbasi A, Rizel L, Hujeirat Y, Salama I, Garzozzi H, Allon S, Stavit B, Yosef T. 2009. Genetic heterogeneity in two consanguineous families segregating early onset retinal degeneration: The pitfalls of homozygosity mapping. *Am J Med Genet A* 149A: 650-656.
- Berger A, Hanson P, Wylie Nichols J, Corbett A. 2005. A yeast model system for functional analysis of the Niemann-Pick type C protein 1 homolog, Ncr1p. *Traffic* 6: 907-917.
- Berger A, Salazar G, Styers M, Newell-Litwa K, Werner E, Maue R, Corbett A, Faundez V. 2007. The subcellular localization of the Niemann-Pick Type C proteins depends on the adaptor complex AP-3. *J Cell Sci* 120: 3640-3652.
- Bernasconi P, Rausch T, Struve I, Morgan L, Taiz L. 1990. An mRNA from human brain encodes an isoform of the B subunit of the vacuolar H(+)-ATPase. *J Biol Chem* 265: 17428-17431.
- Berson A, Young C, Morrison S, Fujii G, Sheung J, Wu B, Bolen J, Burkhardt A. 1999. Identification and characterization of a myristylated and palmitylated serine/threonine protein kinase. *Biochem Biophys Res Commun* 259: 533-538.
- Bimbó A, Jia Y, Poh S, Karuturi R, den Elzen N, Peng X, Zheng L, O'Connell M, Liu E, Balasubramanian M, Liu J. 2005. Systematic deletion analysis of fission yeast protein kinases. *Eukaryot Cell* 4: 799-813.
- Biswas A, Mukherjee S, Das S, Shields D, Chow C, Maitra U. 2011. Opposing action of casein kinase 1 and calcineurin in nucleo-cytoplasmic shuttling of mammalian translation initiation factor eIF6. *J Biol Chem* 286: 3129-3138.
- Blair J. 2009. Fungi. In: Hedges S, Kumar S, editors. *The Timetree of Life*. New York: Oxford University Press. p. 215-219.
- Blom N, Gammeltoft S, Brunak S. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 294: 1351-1362.
- Bonetti D, Anbalagan S, Lucchini G, Clerici M, Longhese M. 2013. Tbf1 and Vid22 promote resection and non-homologous end joining of DNA double-strand break ends. *EMBO J* 32: 275-289.

- Booher R, Deshaies R, Kirschner M. 1993. Properties of *Saccharomyces cerevisiae* wee1 and its differential regulation of p34CDC28 in response to G1 and G2 cyclins. *EMBO J* 12: 3417-3426.
- Bossemeyer D. 1994. The glycine-rich sequence of protein kinases: A multifunctional element. *Trends Biochem Sci* 19: 201-205.
- Botstein D, Chervitz S, Cherry J. 1997. Yeast as a model organism. *Science* 277: 1259-1260.
- Brown R, McCann J, Hung G, Elco C, Chiang H. 2002. Vid22p, a novel plasma membrane protein, is required for the fructose-1,6-bisphosphatase degradation pathway. *J Cell Sci* 115: 655-666.
- Bryant N, Stevens T. 1998. Vacuole biogenesis in *Saccharomyces cerevisiae*: Protein transport pathways to the yeast vacuole. *Microbiol Mol Biol Rev* 62: 230-247.
- Byrne K, Kenneth K. 2005. The yeast gene order browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* 15: 1456-1461.
- Capra J, Pollard K, Singh M. 2010. Novel genes exhibit distinct patterns of function acquisition and network integration. *Genome Biol* 11: R127.
- Caudy A, Guan Y, Jia Y, Hansen C, DeSevo C, Hayes AP, Agee J, Alvarez-Dominguez J, Arellano H, Barrett D, Bauerle C, Bisaria N, Bradley P, Breunig J, Bush E, Cappel D, Capra E, Chen W, Clore J, Combs P, Doucette C, Demuren O, Fellowes P, Freeman S, Frenkel E, Gadala-Maria D, Gawande R, Glass D, Grossberg S, Gupta A, Hammonds-Odie L, Hoisos A, Hsi J, Hsu Y, Inukai S, Karczewski K, Ke X, Kojima M, Leachman S, Lieber D, Liebowitz A, Liu J, Liu Y, Martin T, Mena J, Mendoza R, Myhrvold C, Millian C, Pfau S, Raj S, Rich M, Rokicki J, Rounds W, Salazar M, Salesi M, Sharma R, Silverman S, Singer C, Sinha S, Staller M, Stern P, Tang H, Weeks S, Weidmann M, Wolf A, Young C, Yuan J, Crutchfield C, McClean M, Murphy C, Llinás M, Botstein D, Troyanskaya OG, Dunham M. 2013. A new system for comparative functional genomics of *Saccharomyces* yeasts. *Genetics* 195: 275-287.
- Chan E, Stang S, Bottorff D, Stone J. 2002. Mutations in conserved regions 1, 2, and 3 of raf-1 that activate transforming activity. *Mol Carcinog* 33: 189-197.
- Chien C, Wang S, Rothenberg M, Jan L, Jan Y. 1998. Numb-associated kinase interacts with the phosphotyrosine binding domain of numb and antagonizes the function of numb in vivo. *Mol Cell Biol* 18: 598-607.
- Conant G, Wagner G, Stadler P. 2007. Modeling amino acid substitution patterns in orthologous and paralogous genes. *Mol Phylogenet Evol* 42: 298-307.

- Conibear E, Stevens T. 1998. Multiple sorting pathways between the late golgi and the vacuole in yeast. *Biochim Biophys Acta* 1404: 211-230.
- Conner S, Schmidt S. 2002. Identification of an adaptor-associated kinase, AAK1, as a regulator of clathrin-mediated endocytosis. *J Cell Biol* 156: 921-929.
- Cooper G, Hausman R. 2004. The Golgi Apparatus. In: Cooper G, Hausman R, editors. *The Cell: A Molecular Approach*, 3rd edition. Sunderland: Sinauer Associates. p. 374-380.
- Cope M, Yang S, Shang C, Drubin D. 1999. Novel protein kinases Ark1p and Prk1p associate with and regulate the cortical actin cytoskeleton in budding yeast. *J Cell Biol* 144: 1203-1218.
- Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear E, Sevier C, Ding H, Koh J, Toufighi K, Mostafavi S, Prinz J, St. Onge R, VanderSluis B, Makhnevych T, Vizeacoumar F, Alizadeh S, Bahr S, Brost R, Chen Y, CokoL M, Deshpande R, Li Z, Lin Z, Liang W, Marback M, Paw J, San Luis B, Shuteriqi E, Tong A, Hin Y, van Dyk N, Wallace I, Whitney J, Weirauch M, Zhong G, Zhu H, Houry W, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth F, Giaever G, Nislow C, Troyanskaya O, Bussey H, Bader G, Gingras A, Morris Q, Kim P, Kaiser C, Myers C, Andrews B, Boone C. 2010. The genetic landscape of a cell. *Science* 327: 425-431.
- Cowles C, Odorizzi G, Payne G, Emr S. 1997. The AP-3 adaptor complex is essential for cargo-selective transport to the yeast vacuole. *Cell* 91: 109-118.
- Cvrcková F, De Virgilio C, Manser E, Pringle J, Nasmyth K. 1995. Ste20-like protein kinases are required for normal localization of cell growth and for cytokinesis in budding yeast. *Genes Dev* 9: 1817-1830.
- Darsow T, Burd C, Emr S. 1998. Acidic di-leucine motif essential for AP-3-dependent sorting and restriction of the functional specificity of the Vam3p vacuolar t-SNARE. *J Cell Biol* 142: 913-922.
- Davies H, Bignell G, Cox C, Stephens P, Edkins S, Cleggs, Teague J, Woffendin H, Garnett M, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson B, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins G, Bigner D, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho J, Leung S, Yuen S, Weber B, Seigler H, Darrow T, Paterson H, Marais R, Marshall C, Wooster R, Stratton M, Futreal P. 2002. Mutations of the BRAF gene in human cancer. *Nature* 417: 949-954.
- Delpont W, Poon A, Frost S, Kosakovsky Pond S. 2010. Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26: 2455-2457.

- Diella F, Cameron S, Gemünd C, Linding R, Via A, Kuster B, Sicheritz Pontén T, Blom N, Gibson T. 2004. Phospho.ELM: A database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* 5: 79.
- Drummond A, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A. 2010. Geneious 4.8.5.
- Dunker A, Brown C, Lawson J, Iakoucheva L, Obradović Z. 2002. Intrinsic disorder and protein function. *Biochemistry* 41: 6573-6582.
- Dunn C, Hejnal A, Matus D, Pang K, Browne W, Smith S, Seaver E, Rouse G, Obst M, Edgecombe G, Sørensen M, Haddock SHD, Schmidt Rhaesa A, Okusu A, Kristensen R, Wheeler W, Martindale M, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452: 745-749.
- Dunphy J, Linder M. 1998. Signalling functions of protein palmitoylation. *Biochim Biophys Acta* 1436: 245-261.
- Ebersberger I, de Matos Simoes R, Kupczok A, Gube M, Kothe E, Voigt K, von Haeseler A. 2012. A consistent phylogenetic backbone for the fungi. *Mol Biol Evol* 29: 1319-1334.
- Ellis J, Kobe B. 2011. Predicting protein kinase specificity: Predikin update and performance in the DREAM4 challenge. *PLoS ONE* 6: e21169-e21169.
- Eswaran J, Bernad A, Ligos J, Guinea B, Debreczeni J, Sobott F, Parker S, Najmanovich R, Turk B, Knapp S. 2008. Structure of the human protein kinase MPSK1 reveals an atypical activation loop architecture. *Structure* 16: 115-124.
- Förster H, Coffey M, Elwood H, Sogin M. 1990. Sequence analysis of the small subunit ribosomal RNAs of three zoosporic fungi and implications for fungal evolution. *Mycologia* 82: 306-312.
- Fujita Y, Ohama E, Takatama M, Al-Sarraj S, Okamoto K. 2006. Fragmentation of Golgi apparatus of nigral neurons with alpha-synuclein-positive inclusions in patients with Parkinson's disease. *Acta Neuropathol* 112: 261-265.
- Galtier N, Daubin V. 2008. Dealing with incongruence in phylogenomic analyses. *Philos Trans R Soc Lond B Biol Sci* 363: 4023-4029.
- Gibney P, Hickman M, Bradley P, Matese J, Botstein D. 2013. Phylogenetic portrait of the *Saccharomyces cerevisiae* functional genome. *G3 (Bethesda)* 3: 1335-1340.
- Gieselmann V, Bräulke T. 2009. Lysosomes. *Biochim Biophys Acta* 1793: 603-604.

- Gladfelter A, Kozubowski L, Zyla T, Lew D. 2005. Interplay between septin organization, cell cycle and cell shape in yeast. *J Cell Sci* 118: 1617-1628.
- Gnad F, Gunawardena J, Mann M. 2011. PHOSIDA 2011: The posttranslational modification database. *Nucleic Acids Res* 39: D253-D260.
- Gnad F, Ren S, Cox J, Olsen J, Macek B, Oroshi M, Mann M. 2007. PHOSIDA (phosphorylation site database): Management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol* 8: R250-R250.
- Goffeau A, Barrell B, Bussey H, Davis R, Dujon B, Feldmann H, Galibert F, Hoheisel J, Jacq C, Johnston M, Louis E, Mewes H, Murakami Y, Philippsen P, Tettelin H, Oliver S. 1996. Life with 6000 genes. *Science* 274: 546-563.
- Gonatas N, Stieber A, Gonatas J. 2006. Fragmentation of the Golgi apparatus in neurodegenerative diseases and cell death. *J Neurol Sci* 246: 21-30.
- Grant B, Hemmer W, Tsigelny I, Adams J, Taylor S. 1998. Kinetic analyses of mutations in the glycine-rich loop of cAMP-dependent protein kinase. *Biochemistry* 37: 7708-7715.
- Greener T, Zhao X, Nojima H, Eisenberg E, Greene L. 2000. Role of cyclin G-associated kinase in uncoating clathrin-coated vesicles from non-neuronal cells. *J Biol Chem* 275: 1365-1370.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696-704.
- Guinea B, Ligos J, de Lera T, Martn-Caballero J, Flores J, de la Pea M, Garca-Castro J, Bernad A. 2006. Nucleocytoplasmic shuttling of STK16 (PKL12), a Golgi-resident serine/threonine kinase involved in VEGF expression regulation. *Exp Cell Res* 312: 135-144.
- Hall C, Brachat S, Dietrich F. 2005. Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. *Eukaryot Cell* 4: 1102-1115.
- Hanks S, Hunter T. 1995. Protein kinases 6. the eukaryotic protein kinase superfamily: Kinase (catalytic) domain structure and classification. *FASEB J* 9: 576-596.
- Harvey S, Charlet A, Haas W, Gygi S, Kellogg D. 2005. Cdk1-dependent regulation of the mitotic inhibitor Wee1. *Cell* 122: 407-420.
- Hayles J, Wood V, Jeffery L, Hoe K, Kim D, Park H, Salas-Pino S, Heichinger C, Nurse P. 2013. A genome-wide resource of cell cycle and cell shape genes of fission yeast. *Open Biol* 3:130053.

- Hedges S. 2002. The origin and evolution of model organisms. *Nat Rev Genet* 3: 838-849.
- Heinemann J, Sprague G. 1989. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* 340: 205-209.
- Hemmer W, McGlone M, Tsigelny I, Taylor S. 1997. Role of the glycine triad in the ATP-binding site of cAMP-dependent protein kinase. *J Biol Chem* 272: 16946-16954.
- Hemsley P, Weimar T, Lilley K, Dupree P, Grierson C. 2013. A proteomic approach identifies many novel palmitoylated proteins in arabidopsis. *New Phytol* 197: 805-814.
- Henikoff S, Henikoff J. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* 89: 10915-10919.
- Henry K, D'Hondt K, Chang J, Nix D, Cope M, Chan C, Drubin D, Lemmon S. 2003. The actin-regulating kinase Prk1p negatively regulates Scd5p, a suppressor of clathrin deficiency, in actin organization and endocytosis. *Curr Biol* 13: 1564-1569.
- Hibbett D, Binder M, Bischoff J, Blackwell M, Cannon P, Eriksson O, Huhndorf S, James T, Kirk P, Lücking R, Thorsten Lumbsch H, Lutzoni F, Matheny P, McLaughlin D, Powell M, Redhead S, Schoch C, Spatafora J, Stalpers J, Vilgalys R, Aime M, Aptroot A, Bauer R, Begerow D, Benny G, Castlebury L, Crous P, Dai Y, Gams W, Geiser D, Griffith G, Gueidan C, Hawksworth D, Hestmark G, Hosaka K, Humber R, Hyde K, Ironside J, Kõljalg U, Kurtzman C, Larsson K, Lichtwardt R, Longcore J, Miądlikowska J, Miller A, Moncalvo J, Mozley-Standridge S, Oberwinkler F, Parmasto E, Reeb V, Rogers J, Roux C, Ryvarden L, Sampaio J, Schüssler A, Sugiyama J, Thorn R, Tibell L, Untereiner W, Walker C, Wang Z, Weir A, Weiss M, White M, Winka K, Yao Y, Zhang N. 2007. A higher-level phylogenetic classification of the fungi. *Mycol Res* 111: 509-547.
- Hsueh T. 2012. *ENV9*, a novel gene in *Saccharomyces cerevisiae*, is responsible for the pleiotropic phenotype of *env9Delta* and encodes a membrane protein that is localized to lipid droplets.
- <http://genomeevolution.org/CoGe//GEvo.pl> (last modified: February 27, 2013; retrieved: March 7, 2014)
- http://myhits.isb-sib.ch/cgi-bin/motif_scan (retrieved: March 7, 2014)
- <http://pfam.sanger.ac.uk/family/PF05620#tabview=tab4> (last modified: June 2012; retrieved: March 7, 2014; Rob Finn, Jody Clements and Goran Ceric)
- <http://predikin.biosci.uq.edu.au/> (retrieved: March 7, 2014; Bostjan Kobe)

- <http://sysbio.unl.edu/PhosphoSVM/prediction.php> (retrieved: March 7, 2014)
- http://www.bioinformatics.org/primerx/cgi-bin/DNA_1.cgi (last modified: August 6, 2006; retrieved March 7, 2014; Carlo Lapid)
- <http://www.cbs.dtu.dk/services/NetPhos/> (last modified: April 24, 2013; retrieved: March 7, 2014; Nikolaj Blom)
- <http://www.cbs.dtu.dk/services/NetPhosYeast/> (last modified: October 12, 2012; retrieved: March 7, 2014; Nikolaj Blom)
- <http://www.dabi.temple.edu/disphos/> (last modified: July 29, 2010; retrieved: March 7, 2014; Chad Haynes)
- <http://www.datamonkey.org/> (retrieved: March 7, 2014)
- http://www.ebi.ac.uk/Tools/pfa/ps_scan/ (last modified: March 5, 2014; retrieved: March 7, 2014)
- <http://www.phosida.com/> (retrieved: March 7, 2014)
- <http://www.phosphonet.ca/> (last modified: March 1, 2013; retrieved: March 7, 2014)
- <http://zhanglab.ccmb.med.umich.edu/I-TASSER/> (last modified: March 6, 2014; retrieved: March 7, 2014)
- Huang B, Zeng G, Ng A, Cai M. 2003. Identification of novel recognition motifs and regulatory targets for the yeast actin-regulating kinase Prk1p. *Mol Biol Cell* 14: 4871-4884.
- Huelsenbeck J, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.
- Huh W, Falvo J, Gerke L, Carroll A, Howson R, Weissman J, O'Shea E. 2003. Global analysis of protein localization in budding yeast. *Nature* 425: 686-691.
- Huse M, Kuriyan J. 2002. The conformational plasticity of protein kinases. *Cell* 109: 275-282.
- Iakoucheva L, Radivojac P, Brown C, O'Connor T, Sikes J, Obradovic Z, Dunker A. 2004. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 32: 1037-1049.

- Ikenoue T, Hikiba Y, Kanai F, Aragaki J, Tanaka Y, Imamura J, Imamura T, Ohta M, Ijichi H, Tateishi K, Kawakami T, Matsumura M, Kawabe T, Omata M. 2004. Different effects of point mutations within the B-raf glycine-rich loop in colorectal tumors on mitogen-activated protein/extracellular signal-regulated kinase kinase/extracellular signal-regulated kinase and nuclear factor kappaB pathway and cellular transformation. *Cancer Res* 64: 3428-3435.
- Ingrell C, Miller M, Jensen O, Blom N. 2007. NetPhosYeast: Prediction of protein phosphorylation sites in yeast. *Bioinformatics* 23: 895-897.
- Inomata K, Nishikawa M, Yoshida K. 1994. The yeast *Saccharomyces kluyveri* as a recipient eukaryote in transkingdom conjugation: behavior of transmitted plasmids in transconjugants. *J Bacteriol* 176: 4770-4773.
- James T, Kauff F, Schoch C, Matheny P, Hofstetter V, Cox C, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch H, Rauhut A, Reeb V, Arnold A, Amtoft A, Stajich J, Hosaka K, Sung G, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis J, Slot J, Wang Z, Wilson A, Schüssler A, Longcore J, O'Donnell K, Mozley Standridge S, Porter D, Letcher P, Powell M, Taylor J, White M, Griffith G, Davies D, Humber R, Morton J, Sugiyama J, Rossman A, Rogers J, Pfister D, Hewitt D, Hansen K, Hambleton S, Shoemaker R, Kohlmeyer J, Volkmann Kohlmeyer B, Spotts R, Serdani M, Crous P, Hughes K, Matsuura K, Langer E, Langer G, Untereiner W, Lücking R, Büdel B, Geiser D, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hibbett D, Lutzoni F, McLaughlin D, Spatafora J, Vilgalys R. 2006. Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature* 443: 818-822.
- Janecke A, Thompson D, Utermann G, Becker C, Hübner C, Schmid E, McHenry C, Nair A, Rüschemdorf F, Heckenlively J, Wissinger B, Nürnberg P, Gal A. 2004. Mutations in RDH12 encoding a photoreceptor cell retinol dehydrogenase cause childhood-onset severe retinal dystrophy. *Nat Genet* 36: 850-854.
- Johnson D, Akamine P, Radzio-Andzelm E, Madhusudan M, Taylor S. 2001. Dynamics of cAMP-dependent protein kinase. *Chem Rev* 101: 2243-2270.
- Johnson L, Noble M, Owen D. 1996. Active and inactive protein kinases: Structural basis for regulation. *Cell* 85: 149-158.
- Jones D, Taylor W, Thornton J. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8: 275-282.
- Jones E. 1977. Proteinase mutants of *saccharomyces cerevisiae*. *Genetics* 85: 23-33.

- Kalyuzhnaya M, Bowerman S, Lara J, Lidstrom M, Chistoserdova L. 2006. *Methylotenera mobilis* gen. nov., sp. nov., an obligately methylamine-utilizing bacterium within the family methylophilaceae. *Int J Syst Evol Microbiol* 56: 2819-2823.
- Kannan N, Neuwald A, Taylor S. 2008. Analogous regulatory sites within the alphaC-beta4 loop regions of ZAP-70 tyrosine kinase and AGC kinases. *Biochim Biophys Acta* 1784: 27-32.
- Karathia H, Vilaprinyo E, Sorribas A, Alves R. 2011. *Saccharomyces cerevisiae* as a model organism: A comparative study. *PLoS ONE* 6: e16015-e16015.
- Katoh K, Kuma K, Miyata T, Toh H. 2005. Improvement in the accuracy of multiple sequence alignment program MAFFT. *Genome Inform* 16: 22-33.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res* 30: 3059-3066.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647-1649.
- Kornev A, Haste N, Taylor S, Ten Eyck L. 2006. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc Natl Acad Sci U S A* 103: 17783-17788.
- Kornev A, Taylor S. 2010. Defining the conserved internal architecture of a protein kinase. *Biochim Biophys Acta* 1804: 440-444.
- Kornev A, Taylor S, Ten Eyck L. 2008. A helix scaffold for the assembly of active protein kinases. *Proc Natl Acad Sci U S A* 105: 14377-14382.
- Kosakovsky Pond S, Frost S. 2005. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22: 1208-1222.
- Kosakovsky Pond S, Murrell B, Fourment M, Frost S, Delpont W, Scheffler K. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol* 28: 3033-3043.
- Kosakovsky Pond S, Posada D, Gravenor M, Woelk C, Frost S. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 23: 1891-1901.

- Krogan N, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis A, Punna T, Peregrín-Alvarez J, Shales M, Zhang X, Davey M, Robinson M, Paccanaro A, Bray J, Sheung A, Beattie B, Richards D, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete M, Vlasblom J, Wu S, Orsi C, Collins S, Chandran S, Haw R, Rilstone J, Gandi K, Thompson N, Musso G, St Onge P, Ghanny S, Lam M, Butland G, Altaf-Ul A, Kanaya S, Shilatifard A, O'Shea E, Weissman J, Ingles C, Hughes T, Parkinson J, Gerstein M, Wodak S, Emili A, Greenblatt J. 2006. Global landscape of protein complexes in the yeast *saccharomyces cerevisiae*. *Nature* 440: 637-643.
- Kucharczyk R, Rytka J. 2001. *Saccharomyces cerevisiae*--a model organism for the studies on vacuolar transport. *Acta Biochim Pol* 48: 1025-1042.
- Kurioka K, Nakagawa K, Denda K, Miyazawa K, Kitamura N. 1998. Molecular cloning and characterization of a novel protein serine/threonine kinase highly expressed in mouse embryo. *Biochim Biophys Acta* 1443: 275-284.
- LaGrassa T, Ungermann C. 2005. The vacuolar kinase Yck3 maintains organelle fragmentation by regulating the HOPS tethering complex. *J Cell Biol* 168: 401-414.
- Landry C, Townsend J, Hartl D, Cavalieri D. 2006. Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. *Mol Ecol* 15: 575-591.
- Lapidus A, Clum A, Labutti K, Kaluzhnaya M, Lim S, Beck D, Glavina Del Rio T, Nolan M, Mavromatis K, Huntemann M, Lucas S, Lidstrom M, Ivanova N, Chistoserdova L. 2011. Genomes of three methylophilic fungi from a single niche reveal the genetic and metabolic divergence of the methylophilaceae. *J Bacteriol* 193:3757-3764.
- Lee H, Chen R, Lee Y, Yoo S, Lee C. 2009. Essential roles of CKIdelta and CKIepsilon in the mammalian circadian clock. *Proc Natl Acad Sci U S A* 106: 21359-21364.
- Lee S, Belyaeva O, Kedishvili N. 2010. Disease-associated variants of microsomal retinol dehydrogenase 12 (RDH12) are degraded at mutant-specific rates. *FEBS Lett* 584: 507-510.
- Lew D. 2003. The morphogenesis checkpoint: How yeast cells watch their figures. *Curr Opin Cell Biol* 15: 648-653.
- Li S, Kane P. 2009. The yeast lysosome-like vacuole: Endpoint and crossroads. *Biochim Biophys Acta* 1793:650-663.
- Ligos J, Gerwin N, Fernandez P, Gutierrez-Ramos J, Bernad A. 1998. Cloning, expression analysis, and functional characterization of PKL12, a member of a new subfamily of ser/thr kinases. *Biochem Biophys Res Commun* 249: 380-384.

- Linder M, Deschenes R. 2007. Palmitoylation: Policing protein stability and traffic. *Nat Rev Mol Cell Biol* 8: 74-84.
- Lutzoni F, Kauff F, Cox C, McLaughlin D, Celio G, Dentinger B, Padamsee M, Hibbett D, James T, Baloch E, Grube M, Reeb V, Hofstetter V, Schoch C, Arnold A, Miadlikowska J, Spatafora J, Johnson D, Hambleton S, Crockett M, Shoemaker R, Sung G, Lücking R, Lumbsch T, O'Donnell K, Binder M, Diederich P, Ertz D, Gueidan C, Hansen K, Harris R, Hosaka K, Lim Y, Matheny B, Nishida H, Pfister D, Rogers J, Rossman A, Schmitt I, Sipman H, Stone J, Sugiyama J, Yahr R, Vilgalys R. 2004. Assembling the fungal tree of life: Progress, classification, and evolution of subcellular traits. *Am J Bot* 91: 1446-1480.
- Luzio J, Pryor P, Bright N. 2007. Lysosomes: Fusion and function. *Nat Rev Mol Cell Biol* 8: 622-632.
- Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D, Freeling M. 2008. Finding and comparing syntenic regions among arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol* 148: 1772-1781.
- Mager W, Winderickx J. 2005. Yeast as a model for medical and medicinal research. *Trends Pharmacol Sci* 26: 265-273.
- Manandhar S, Gharakhanian E. 2013. ENV7 and YCK3, which encode vacuolar membrane protein kinases, genetically interact to impact cell fitness and vacuole morphology. *FEMS Yeast Res.* doi: 10.1111/1567-1364.12128.
- Manandhar S, Ricarte F, Cocca S, Gharakhanian E. 2013. *Saccharomyces cerevisiae* Env7 is a novel serine/threonine kinase 16-related protein kinase and negatively regulates organelle fusion at the lysosomal vacuole. *Mol Cell Biol* 33: 526-542.
- Manning G, Plowman G, Hunter T, Sudarsanam S. 2002. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 27: 514-520.
- Marchette L, Thompson D, Kravtsova M, Ngansop T, Mandal M, Kasus-Jacobi A. 2010. Retinol dehydrogenase 12 detoxifies 4-hydroxynonenal in photoreceptor cells. *Free Radic Biol Med* 48: 16-25.
- Masterson L, Mascioni A, Traaseth N, Taylor S, Veglia G. 2008. Allosteric cooperativity in protein kinase A. *Proc Natl Acad Sci U S A* 105: 506-511.
- Matsuyama A, Arai R, Yashiroda Y, Shirai A, Kamata A, Sekido S, Kobayashi Y, Hashimoto A, Hamamoto M, Hiraoka Y, Horinouchi S, Yoshida M. 2006. ORFeome cloning and global analysis of protein localization in the fission yeast *Schizosaccharomyces pombe*. *Nat Biotechnol* 24: 841-847.

- Meikle P, Hopwood J, Clague A, Carey W. 1999. Prevalence of lysosomal storage disorders. *JAMA: the Journal of the American Medical Association* 281: 249-254.
- Michaillat L, Mayer A. 2013. Identification of genes affecting vacuole membrane fragmentation in *saccharomyces cerevisiae*. *PLoS ONE* 8: e54160-e54160.
- Michelitsch M, Weissman J. 2000. A census of glutamine/asparagine-rich regions: Implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci U S A* 97: 11910-11915.
- Moore M, Adams J, Taylor S. 2003. Structural basis for peptide binding in protein kinase A. role of glutamic acid 203 and tyrosine 204 in the peptide-positioning loop. *J Biol Chem* 278: 10613-10618.
- Munoz R, Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer K, Glöckner F, Rosselló-Móra R. 2011. Release LTPs104 of the All-Species Living Tree. *Syst Appl Microbiol* 34: 169-170.
- Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol* 30: 1196-1205.
- Myers L, Gustafsson C, Bushnell D, Lui M, Erdjument-Bromage H, Tempst P, Kornberg R. 1998. The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev* 12: 45-54.
- Nakagomi S, Barsoum M, Bossy-Wetzel E, Sütterlin C, Malhotra V, Lipton S. 2008. A Golgi fragmentation pathway in neurodegeneration. *Neurobiol Dis* 29: 221-231.
- Nixon R, Yang D, Lee J. 2008. Neurodegenerative lysosomal disorders: A continuum from development to late age. *Autophagy* 4: 590-599.
- Nolen B, Taylor S, Ghosh G. 2004. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell* 15: 661-675.
- Odawara M, Kadowaki T, Yamamoto R, Shibasaki Y, Tobe K, Accili D, Bevins C, Mikami Y, Matsuura N, Akanuma Y. 1989. Human diabetes associated with a mutation in the tyrosine kinase domain of the insulin receptor. *Science* 245: 66-68.
- Odorizzi G, Cowles C, Emr S. 1998. The AP-3 complex: A coat of many colours. *Trends Cell Biol* 8: 282-288.
- Ohta S, Takeuchi M, Deguchi M, Tsuji T, Gahara Y, Nagata K. 2000. A novel transcriptional factor with ser/thr kinase activity involved in the transforming growth factor (TGF)-beta signalling pathway. *Biochem J* 350 Pt 2: 395-404.

- Oliveira L. 2013. The cloning and characterization of the novel genes *ENV10* and *ENV11* in *S. cerevisiae*.
- Pearce D, Ferea T, Nosel S, Das B, Sherman F. 1999. Action of *BTNI*, the yeast orthologue of the gene mutated in Batten disease. *Nat Genet* 22: 55-58.
- Peleg M, Corradini M. 2011. Microbial growth curves: What the models tell us and what they cannot. *Crit Rev Food Sci Nutr* 51: 917-945.
- Perrault I, Hanein S, Gerber S, Barbet F, Ducroq D, Dollfus H, Hamel C, Dufier J, Munnich A, Kaplan J, Rozet J. 2004. Retinal dehydrogenase 12 (RDH12) mutations in leber congenital amaurosis. *Am J Hum Genet* 75: 639-646.
- Pflug A, de-Oliveira T, Bossemeyer D, Engh R. 2011. Mutants of protein kinase A that mimic the ATP-binding site of aurora kinase. *Biochem J* 440: 85-93.
- Pinto R, Caseiro C, Lemos M, Lopes L, Fontes A, Ribeiro H, Pinto E, Silva E, Rocha S, Marcão A, Ribeiro I, Lacerda L, Ribeiro G, Amaral O, Sá Miranda M. 2004. Prevalence of lysosomal storage diseases in Portugal. *Eur J Hum Genet* 12:87-92.
- Platt F, Boland B, van der Spoel A. 2012. The cell biology of disease: Lysosomal storage disorders: The cellular impact of lysosomal dysfunction. *J Cell Biol* 199: 723-734.
- Plummer L, Hildebrandt E, Porter S, Rogers V, McCracken J, Schmidt W. 2006. Mutational analysis of the ras converting enzyme reveals a requirement for glutamate and histidine residues. *J Biol Chem* 281: 4596-4605.
- Pond S, Frost S, Muse S. 2005. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* 21: 676-679.
- Poorthuis B, Wevers R, Kleijer W, Groener J, de Jong J, van Weely S, Niezen-Koning K, van Diggelen O. 1999. The frequency of lysosomal storage diseases in the Netherlands. *Hum Genet* 105: 151-156.
- Poupetová H, Ledvinová J, Berná L, Dvůřáková L, Kozich V, Elleder M. 2010. The birth prevalence of lysosomal storage disorders in the Czech Republic: Comparison with data in different populations. *J Inherit Metab Dis* 33: 387-396.
- Preti M, Ribeyre C, Pascali C, Bosio M, Cortelazzi B, Rougemont J, Guarnera E, Naef F, Shore D, Dieci G. 2010. The telomere-binding protein Tbf1 demarcates snoRNA gene promoters in *Saccharomyces cerevisiae*. *Mol Cell* 38: 614-620.
- Prowse C, Deal M, Lew J. 2001. The complete pathway for catalytic activation of the mitogen-activated protein kinase, ERK2. *J Biol Chem* 276: 40817-40823.

- Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, Guo H, Jona G, Breitzkreutz A, Sopko R, McCartney R, Schmidt M, Rachidi N, Lee S, Mah A, Meng L, Stark MJR, Stern D, De Virgilio C, Tyers M, Andrews B, Gerstein M, Schweitzer B, Predki P, Snyder M. 2005. Global analysis of protein phosphorylation in yeast. *Nature* 438: 679-674.
- Rajagopalan H, Bardelli A, Lengauer C, Kinzler K, Vogelstein B, Velculescu V. 2002. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature* 418: 934-934.
- Ralph MR, Menaker M. 1988. A mutation of the circadian system in golden hamsters. *Science* 241: 1225-1227.
- Raymond C, Howald-Stevenson I, Vater C, Stevens T. 1992. Morphological classification of the yeast vacuolar protein sorting mutants: Evidence for a prevacuolar compartment in class E vps mutants. *Mol Biol Cell* 3: 1389-1402.
- Rhind N, Chen Z, Yassour M, Thompson D, Haas BJ, Habib N, Wapinski I, Roy S, Lin M, Heiman D, Young S, Furuya K, Guo Y, Pidoux A, Chen H, Robbertse B, Goldberg J, Aoki K, Bayne E, Berlin A, Desjardins C, Dobbs E, Dukaj L, Fan L, FitzGerald M, French C, Gujja S, Hansen K, Keifenheim D, Levin J, Mosher R, Müller C, Pfiffner J, Priest M, Russ C, Smialowska A, Swoboda P, Sykes S, Vaughn M, Vengrova S, Yoder R, Zeng Q, Allshire R, Baulcombe D, Birren B, Brown W, Ekwall K, Kellis M, Leatherwood J, Levin H, Margalit H, Martienssen R, Nieduszynski C, Spatafora J, Friedman N, Dalgaard J, Baumann P, Niki H, Regev A, Nusbaum C. 2011. Comparative functional genomics of the fission yeasts. *Science* 332: 930-936.
- Ricarte F, Menjivar R, Chhun S, Soreta T, Oliveira L, Hsueh T, Serranilla M, Gharakhanian E. 2011. A genome-wide immunodetection screen in *S. cerevisiae* uncovers novel genes involved in lysosomal vacuole function and morphology. *PLoS ONE* 6: e23696-e23696.
- Richards T, Leonard G, Soanes D, Talbot N. 2011. Gene transfer into the fungi. *Fungal Biol Rev* 25: 98-110.
- Robbins E, Gonatas N. 1964. The ultrastructure of a mammalian cell during the mitotic cycle. *J Cell Biol* 21: 429-463.
- Robinson J, Klionsky D, Banta L, Emr S. 1988. Protein sorting in *Saccharomyces cerevisiae*: Isolation of mutants defective in the delivery and processing of multiple vacuolar hydrolases. *Mol Cell Biol* 8: 4936-4948.
- Rocks O, Peyker A, Kahms M, Verveer P, Koerner C, Lumbierres M, Kuhlmann J, Waldmann H, Wittinghofer A, Bastiaens P. 2005. An acylation cycle regulates localization and activity of palmitoylated ras isoforms. *Science* 307: 1746-1752.

- Roth A, Wan J, Bailey A, Sun B, Kuchar J, Green W, Phinney B, Yates J, Davis N. 2006. Global analysis of protein palmitoylation in yeast. *Cell* 125: 1003-1013.
- Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc* 5: 725-738.
- Rubenstein E, Schmidt M. 2007. Mechanisms regulating the protein kinases of *Saccharomyces cerevisiae*. *Eukaryot cell* 6: 571-583.
- Santos-Rosa H, Leung J, Grimsey N, Peak-Chew S, Siniossoglou S. The yeast lipin Smp2 couples phospholipid biosynthesis to nuclear membrane growth. *EMBO J* 24: 1931-1941.
- Sasser T, Qiu Q, Karunakaran S, Padolina M, Reyes A, Flood B, Smith S, Gonzales C, Fratti R. 2012. Yeast lipin 1 orthologue pah1p regulates vacuole homeostasis and membrane fusion. *J Biol Chem* 287: 2221-2236.
- Sawasaki Y, Inomata K, Yoshida K. 1996. Trans-kingdom conjugation between *Agrobacterium tumefaciens* and *Saccharomyces cerevisiae*, a bacterium and a yeast. *Plant Cell Physiol* 37: 103-106.
- Scannell D, Butler G, Wolfe K. 2007. Yeast genome evolution--the origin of the species. *Yeast* 24: 929-942.
- Scheeff E, Bourne P. 2005. Structural evolution of the protein kinase-like superfamily. *PLoS Comput Biol* 1: e49-e49.
- Schuster-Böckler B, Schultz J, Rahmann S. 2004. HMM logos for visualization of protein families. *BMC Bioinformatics* 5: 7.
- Schwartz S, Kent W, Smit A, Zhang Z, Baertsch R, Hardison R, Haussler D, Miller W. 2003. Human-mouse alignments with BLASTZ. *Genome Res* 13: 103-107.
- Scott S, Klionsky D. 1998. Delivery of proteins and organelles to the vacuole from the cytoplasm. *Curr Opin Cell Biol* 10: 523-529.
- Sharifpoor S, van Dyk D, Costanzo M, Baryshnikova A, Friesen H, Douglas AC, Youn J, VanderSluis B, Myers C, Papp B, Boone C, Andrews B. 2012. Functional wiring of the yeast kinome revealed by global analysis of genetic network motifs. *Genome Res* 4: 791-801.
- Shorter J, Warren G. 2002. Golgi architecture and inheritance. *Annu Rev Cell Dev Biol* 18: 379-420.

- Sigrist C, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P. 2002. PROSITE: A documented database using patterns and profiles as motif descriptors. *Brief Bioinform* 3: 265-274.
- Sigrist C, de Castro E, Cerutti L, Cuče B, Hulo N, Bridge A, Bougueleret L, Xenarios I. 2013. New and continuing developments at PROSITE. *Nucleic Acids Res* 41: D344-D347.
- Sim K, Creamer T. 2002. Abundance and distributions of eukaryote protein simple sequences. *Mol Cell Proteomics* 1: 983-995.
- Sipiczki M. 2000. Where does fission yeast sit on the tree of life?. *Genome Biol* 1(2): REVIEWS1011.
- Smotrys J, Linder M. 2004. Palmitoylation of intracellular signaling proteins: Regulation and function. *Annu Rev Biochem* 73: 559-587.
- Smythe E, Ayscough K. 2003. The Ark1/Prk1 family of protein kinases. regulators of endocytosis and the actin skeleton. *EMBO Rep* 4: 246-251.
- Stairs D, Gardner H, Ha S, Copeland N, Gilbert D, Jenkins N, Chodosh L. 1998. Cloning and characterization of Krct, a member of a novel subfamily of serine/threonine kinases. *Hum Mol Genet* 7: 2157-2166.
- Stapp J, Huang K, Lemmon S. 1997. The yeast adaptor protein complex, AP-3, is essential for the efficient delivery of alkaline phosphatase by the alternate pathway to the vacuole. *J Cell Biol* 139: 1761-1774.
- Storch S, Pohl S, Bräulke T. 2004. A dileucine motif and a cluster of acidic amino acids in the second cytoplasmic domain of the batten disease-related CLN3 protein are required for efficient lysosomal targeting. *J Biol Chem* 279: 53625-53634.
- Sütterlin C, Hsu P, Mallabiabarrena A, Malhotra V. 2002. Fragmentation and dispersal of the pericentriolar Golgi complex is required for entry into mitosis in mammalian cells. *Cell* 109: 359-369.
- Takahashi M, Frost C, Oyadomari K, Pinho M, Sao D, Chima-Okereke O, Gharakhanian E. 2008. A novel immunodetection screen for vacuolar defects identifies a unique allele of VPS35 in *S. cerevisiae*. *Mol Cell Biochem* 311: 121-136.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.
- Taylor J, Berbee M. 2006. Dating divergences in the fungal tree of life: Review and new analyses. *Mycologia* 98: 838-849.

- Taylor S, Zhang P, Steichen J, Keshwani M, Kornev A. 2013. PKA: Lessons learned after twenty years. *Biochim Biophys Acta* 1834: 1271-1278.
- Taylor S, Knighton D, Zheng J, Sowadski J, Gibbs C, Zoller M. 1993. A template for the protein kinase family. *Trends Biochem Sci* 18: 84-89.
- Taylor S, Radzio-Andzelm E, Hunter T. 1995. How do protein kinases discriminate between serine/threonine and tyrosine? Structural insights from the insulin receptor protein-tyrosine kinase. *FASEB J* 9: 1255-1266.
- Taylor S, Yang J, Wu J, Haste N, Radzio-Andzelm E, Anand G. 2004. PKA: A portrait of protein kinase dynamics. *Biochim Biophys Acta* 1697: 259-269.
- Taylor S, Keshwani M, Steichen J, Kornev A. 2012. Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos Trans R Soc Lond B Biol Sci* 367: 2517-2528.
- Taylor S, Kornev A. 2011. Protein kinases: Evolution of dynamic regulatory proteins. *Trends Biochem Sci* 36: 65-77.
- Taylor S, Shaw A, Hu J, Meharena H, Kornev A. 2013. Pseudokinases from a structural perspective. *Biochem Soc Trans* 4: 981-986.
- Thompson D, Janecke A, Lange J, Feathers K, Hübner C, McHenry C, Stockton D, Rammesmayer G, Lupski J, Antinolo G, Ayuso C, Baiget M, Gouras P, Heckenlively J, den Hollander A, Jacobson S, Lewis R, Sieving P, Wissinger B, Yzer S, Zrenner E, Utermann G, Gal A. 2005. Retinal degeneration associated with RDH12 mutations results from decreased 11-cis retinal synthesis due to disruption of the visual cycle. *Hum Mol Genet* 14: 3865-3875.
- Thompson J, Higgins D, Gibson T. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.
- Titz B, Thomas S, Rajagopala S, Chiba T, Ito T, Uetz P. 2006. Transcriptional activators in yeast. *Nucleic Acids Res* 34: 955-967.
- Torkamani A, Kannan N, Taylor S, Schork N. 2008. Congenital disease SNPs target lineage specific structural elements in protein kinases. *Proc Natl Acad Sci U S A* 105: 9011-9016.
- Ułamek-Kozioł M, Furmaga-Jabłońska W, Januszewski S, Brzozowska J, Sciślewska M, Jabłoński M, Pluta R. 2013. Neuronal autophagy: Self-eating or self-cannibalism in alzheimer's disease. *Neurochem Res* 38: 1769-1773.

- Valiev M, Yang J, Adams J, Taylor S, Weare J. 2007. Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations. *J Phys Chem B* 111: 13455-13464.
- Vergarajauregui S, Puertollano R. 2006. Two di-leucine motifs regulate trafficking of mucolin-1 to lysosomes. *Traffic* 7: 337-353.
- Vinayagam A, Stelzl U, Foulle R, Plassmann S, Zenkner M, Timm J, Assmus H, Andrade-Navarro M, Wanker E. 2011. A directed protein interaction network for investigating intracellular signal transduction. *Sci Signal* 4: rs8.
- Vowels J, Payne G. 1998. A dileucine-like sorting signal directs transport into an AP-3-dependent, clathrin-independent pathway to the yeast vacuole. *EMBO J* 17: 2482-2493.
- Wainright P, Hinkle G, Sogin M, Stickel S. 1993. Monophyletic origins of the metazoa: An evolutionary link with fungi. *Science* 260: 340-342.
- Wang X, Hoekstra M, DeMaggio A, Dhillon N, Vancura A, Kuret J, Johnston G, Singer R. 1996. Prenylated isoforms of yeast casein kinase I, including the novel Yck3p, suppress the *gcs1* blockage of cell proliferation from stationary phase. *Mol Cell Biol* 16: 5375-5385.
- Warren G. 1993. Membrane partitioning during cell division. *Annu Rev Biochem* 62: 323-348.
- Watanabe M, Watanabe D, Nogami S, Morishita S, Ohya Y. 2009. Comprehensive and quantitative analysis of yeast deletion mutants defective in apical and isotropic bud growth. *Curr Genet* 55: 365-380.
- Watari H, Blanchette-Mackie E, Dwyer N, Glick J, Patel S, Neufeld E, Brady R, Pentchev P, Strauss J. 1999. Niemann-Pick C1 protein: obligatory roles for N-terminal domains and lysosomal targeting in cholesterol mobilization. *Proc Natl Acad Sci U S A* 96:805-810.
- Weiss E, Bishop A, Shokat K, Drubin D. 2000. Chemical genetic analysis of the budding-yeast p21-activated kinase Cla4p. *Nat Cell Biol* 10: 677-685.
- Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18: 691-699.
- Wiemken A, Schellenberg M, Urech K. 1979. Vacuoles: The sole compartments of digestive enzymes in yeast (*Saccharomyces cerevisiae*). *Arch Microbiol* 123: 23-35.

- Wilson W, Wang Z, Roach P. 2002. Systematic identification of the genes affecting glycogen storage in the yeast *Saccharomyces cerevisiae*: Implication of the vacuole as a determinant of glycogen level. *Mol Cell Proteomics* 1: 232-242.
- Wohlbach D, Kuo A, Sato T, Potts K, Salamov A, Labutti K, Sun H, Clum A, Pangilinan J, Lindquist E, Lucas S, Lapidus A, Jin M, Gunawan C, Balan V, Dale B, Jeffries T, Zinkel R, Barry K, Grigoriev I, Gasch AP. 2011. Comparative genomics of xylose-fermenting fungi for enhanced biofuel production. *Proc Natl Acad Sci U S A* 108: 13212-13217.
- Wong E, Cuervo A. 2010. Autophagy gone awry in neurodegenerative diseases. *Nat Neurosci* 13: 805-811.
- Yan B, Sun Y. 1997. Glycine residues provide flexibility for enzyme active sites. *J Biol Chem* 272: 3190-3194.
- Yang J, Garrod S, Deal M, Anand G, Woods V, Taylor S. 2005. Allosteric network of cAMP-dependent protein kinase revealed by mutation of Tyr204 in the P+1 loop. *J Mol Biol* 346: 191-201.
- Yang J, Wu J, Steichen J, Kornev A, Deal M, Li S, Sankaran B, Woods V, Taylor S. 2012. A conserved glu-arg salt bridge connects coevolved motifs that define the eukaryotic protein kinase fold. *J Mol Biol* 415: 666-679.
- Yu Y, Wootton J, Altschul S. 2003. The compositional adjustment of amino acid substitution matrices. *Proc Natl Acad Sci U S A* 100: 15688-15693.
- Yu Y, Altschul S. 2005. The construction of amino acid substitution matrices for the comparison of proteins with non-standard compositions. *Bioinformatics* 21: 902-911.
- Yuen S, Davies H, Chan T, Ho J, Bignell G, Cox C, Stephens P, Edkins S, Tsui W, Chan A, Futreal P, Stratton M, Wooster R, Leung S. 2002. Similarity of the phenotypic patterns associated with BRAF and KRAS mutations in colorectal neoplasia. *Cancer Res* 62: 6451-6455.
- Zeng G, Cai M. 1999. Regulation of the actin cytoskeleton organization in yeast by a novel serine/threonine kinase Prk1p. *J Cell Biol* 144: 71-82.
- Zeng G, Yu X, Cai M. 2001. Regulation of yeast actin cytoskeleton-regulatory complex Pan1p/Sla1p/End3p by serine/threonine kinase Prk1p. *Mol Biol Cell* 12: 3759-3772.
- Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9: 40.

Zhang Y, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins* 57: 702-710.

Zhao Y, Hu J, Miao G, Qu L, Wang Z, Li G, Lv P, Ma D, Chen Y. 2013. Transmembrane protein 208: A novel ER-localized protein that regulates autophagy and ER stress. *PLoS ONE* 8: e64228-e64228.

Zhao N, Pang B, Shyu C, Korkin D. 2011. Charged residues at protein interaction interfaces: Unexpected conservation and orchestrated divergence. *Protein Sci* 20: 1275-1284.

Zuccotto F, Ardini E, Casale E, Angiolini M. 2010. Through the "gatekeeper door": Exploiting the active kinase conformation. *J Med Chem* 53: 2681-2694.