

AN ABSTRACT OF THE THESIS OF

Sarah Theresa Ferrer for the degree of Honors Baccalaureate of Science in Biochemistry and Biophysics presented on May 31st, 2013. Title: Mutations of CenH3 Domains and their Effect on Chromocenter Localization and Meiosis in *Neurospora crassa*.

Abstract approved: _____
Michael Freitag

In recent years, non-coding DNA has received much attention within the scientific community. This attention has not only illuminated the mystery behind functionally important regions like the centromere, but has also brought to light additional important questions regarding centromere mechanisms, inheritance, and cellular recognition by the kinetochore, a protein complex that facilitates chromosome segregation. Although these questions have begun to be explored, centromeres still remain elusive, especially in the diverse and important group of filamentous fungi. The key finding that histones, the proteins that organize DNA, vary within centromere regions has led scientists to hypothesize that sequential variations hold the key to the observed altered function and kinetochore recognition. Compared to the normal histone H3, a centromere-specific H3 contains the most obvious sequence variations, especially in the N-terminal tail and the loop 1 region. Studies in humans have demonstrated that a two amino acid residue in the loop 1 region is important for retained centromere localization. In some filamentous fungi, the loop 1 addition is also present, prompting the scrutiny of this region for normal function. The N-terminus has also been shown to contain important element for normal meiotic function. To test these regions, loop 1 and N-terminal chimera were assembled in *Neurospora crassa*, a filamentous fungus. Point mutations were also introduced into the

loop 1 region to check for functional importance. The N-terminal chimera demonstrated centromere localization while the loop 1 chimera was never successfully assembled. The loop 1 point mutants were successfully assembled and are currently in the process of being expressed in *N. crassa*.

Key words: Centromeres, kinetochore, histone, filamentous fungi.

Corresponding e-mail address: sarah.ferrer@cgrb.oregonstate.edu

©Copyright by Sarah Theresa Ferrer
May 31, 2013
All Rights Reserved

Mutations of CenH3 Domains and their Effect on
Chromocenter Localization and Meiosis in *Neurospora crassa*

by

Sarah Theresa Ferrer

A PROJECT

Submitted to

Oregon State University

University Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Biochemistry and Biophysics (Honors Scholar)

Presented May 31, 2013
Commencement June 2013

ACKNOWLEDGEMENT

I would like to thank my mentor, Michael Freitag, for his willingness to help with all aspects of this project and for giving me the opportunity and freedom to conduct my own research.

I would also like to thank Pallavi Phatale for her immense assistance in the lab, shared knowledge, and experience. This project could not have happened without her efforts.

Finally, I would like to thank my mom, Debi Ferrer, and my friend, Tylor Slay, for their proof-reading, speech critiquing, and encouragement.

TABLE OF CONTENTS

| | <u>Page</u> |
|--|-------------|
| INTRODUCTION | 1 |
| Cellular Organization of Genetic Material | 1 |
| Centromeres and Kinetochores | 2 |
| Histone Protein Functionality and Conserved Sequences | 3 |
| The Centromere Specific Histone 3 (CenH3) in Filamentous Fungi | 5 |
| Hypotheses | 8 |
| Proposed Experimentation | 8 |
| Collection and Interpretation of Data | 10 |
| Significance of the Work | 11 |
| METHODS | 13 |
| Generation of Loop 1 Point Mutations | 13 |
| Generation of Chimeric Loop 1 Constructs | 16 |
| Assembly of N-Terminal Chimeric Constructs | 19 |
| RESULTS | 21 |
| DISCUSSION | 23 |
| N-terminal Chimera | 23 |
| Site-directed Mutagenesis | 24 |
| Loop 1 Chimera | 24 |
| CONCLUSION | 26 |
| BIBLIOGRAPHY | 27 |
| APPENDICIES | 28 |
| Appendix A Figures | 28 |
| Appendix B Oligonucleotide Table | 36 |

LIST OF APPENDIX FIGURES

| | <u>Page</u> |
|--|-------------|
| Appendix A | 28 |
| Figure 1: Phylogenic tree of filamentous fungus CenH3..... | 29 |
| Figure 2: CENP-A/H3 loop 1 region overlay | 30 |
| Figure 3: CenH3 sequence loop 1 variations between key species..... | 31 |
| Figure 4: CenH3 protein structure and sequence alignments..... | 32 |
| Figure 5: Site-directed mutants construct plasmid map and sequence..... | 33 |
| Figure 6: N-terminal chimera heterokaryote fluorescence and controls..... | 34 |
| Appendix B | 36 |
| Table 1: List of Oligonucleotides..... | 37 |

DEDICATION

This undergraduate thesis is dedicated to the single most influential person in my life: My mother. Congratulations on your retirement: May your next walk of life be full of joy, fond memories, and fun. I love you.

Mutations of CenH3 Domains and their Effect on Chromocenter Localization and Meiosis in *Neurospora crassa*

INTRODUCTION

Cellular Organization of Genetic Material

DNA is known as the “blueprint for life” because it contains the requisite information to assemble the proteins that make up living organisms. Considering that all this information is conveyed using different combinations of only four bases, the sheer amount of DNA required to support complex organisms is evident. In eukaryotes, DNA is divided into different chromosomes and is wound into chromatin; a highly condensed structure composed of DNA and structural proteins, known as histones. The most basic unit of chromatin is a “nucleosome”, DNA wrapped around a core of histone proteins (Nelson, D. L. and M. M. Cox 2008). These nucleosomes make up all chromosomes but differ regionally in their composition.

The primary function of DNA is generally thought to be coding for RNA that is translated into proteins. While this is arguably the most important function, protein coding genes only account for about two percent of the entire human genome. The remaining DNA is composed of regulatory regions, pseudogenes, functionally important RNAs, regions with undetermined functions, and repetitive regions, including transposable elements. One of these repetitive regions is the centromere, a highly condensed chromatin region, rich in adenine (A) and thymine (T). This is where sister chromatids remain attached during meiosis I. Between species, the DNA sequence for the

centromere region can be highly variable, making it difficult to accurately determine the outer limits of the region (Alberts, B., D. Bray, et al. 2009). Current limitations in sequencing technology and assembly of repetitive regions have also contributed to the difficulty of studying centromeric regions. Although significant advancements in centromere knowledge have been made over the past decade (Cleveland, Mao et al. 2003; Smith, Galazka et al. 2012), centromere regions and the mechanisms they participate in remain unclear.

Centromeres and Kinetochores

Centromeres play a crucial role in cell survival and replication in eukaryotes yet they are not well-understood in the scientific community due to their highly variable, repetitive, and rapidly evolving sequences. However, cellular mechanisms can recognize centromere regions with great accuracy. The kinetochore, a protein complex that facilitates chromosome segregation in cellular division, is localized, assembled, and maintained exclusively on centromeric regions of chromosomes (Cleveland, Mao et al. 2003). The kinetochore is responsible for microtubule recognition and attachment. These microtubules separate replicated DNA in cellular division as well as homologous chromosomes in meiosis. The high variability of centromeric DNA sequences makes it unlikely that the kinetochore recognizes centromeres directly. To determine how this recognition takes place, proteins were studied for centromere localization. Along with over one hundred different centromere associating proteins, researchers found that the histones that help with DNA organization are slightly different in the centromere compared to the rest of the chromosomes; specifically, the H3 subunit of the histone. Not

only has this histone variant been found exclusively in centromeric regions, but it has also been shown to be crucial for recognition by and assembly of subsequent kinetochore proteins (Cleveland, Mao et al. 2003; Dalal, Furuyama et al. 2007; Henikoff and Furuyama 2010; Smith, Phatale et al. 2011).

Histone Protein Functionality and Conserved Sequences

In most eukaryotes, histone proteins are essential for storage of and access to DNA. The core histones (two each of H2A, H2B, H3 and H4) assemble into an octameric complex called the nucleosome, which keeps DNA stable and condensed but can be easily modified by cellular mechanisms; making DNA accessible for transcription and other cellular functions. There are currently two competing structure models for centromere specific nucleosomes that include a specialized version of H3, a histone subunit called CENPA or CenH3. While most nucleosomes are known to form octamers, centromeric nucleosomes were proposed to be tetramers (Furuyama and Henikoff 2006; Dalal, Wang et al. 2007; Furuyama and Henikoff 2009). The octamer model was recently proposed for human centromeres from in vitro studies (Tachiwana, Kagawa et al. 2011). The tetramer model is supported by *Drosophila melanogaster* studies and suggests that each histone complex contains one copy each of Cid (CenH3), H4, H2A and H2B (Dalal, Furuyama et al. 2007; Dalal, Wang et al. 2007; Dimitriadis, Weber et al. 2010). Because both these models have supporting evidence, the idea that centromeric nucleosomes change composition at different times during the cell cycle has become a popular hypothesis (Bui, Dimitriadis et al. 2012; Shivaraju, Unruh et al. 2012). The function of centromere histone proteins is conserved in many of the eukaryotic models that have been studied so

far. In fact, histone proteins are among the most conserved proteins in all of biology, lending credence to their biological importance and making the study of functional variations straightforward. Although histone function is highly conserved, variations have been found in nucleosome structure as well as in modes for centromere inheritance. These variations, as well as general centromere behavior that they influence, are still poorly understood, especially in the diverse group of filamentous fungi (Figure 1).

Although histones are functionally highly conserved, sequence differences have been found between species as well as within histone variants in the same species. Between canonical H3 and CenH3, these differences have been hypothesized to be responsible for the functional differences observed. Although there is sequence and functional variation between these two subunits, the general 3-dimensional protein structure is the same. The histone-folding domain (HFD) of H3 subunits contain four alpha helices separated with looping regions. Flanking the HFD are unstructured N and C-termini. Studies in *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals have suggested that the N-terminal region and the loop 1 region within the HFD possess sufficient variation to account for the observed functional differences (Talbert, Bryson et al. 2004). The recently published crystal structure of CENP-A, the mammalian homologue of CenH3, revealed a two amino acid addition in the loop 1 region of the protein (Tachiwana, Kagawa et al. 2011). These two residues, arginine 80 and glycine 81, form a solvent-accessible loop (Figure 2). Since these residues result in an obvious structural variation, it has been hypothesized that they are important for CENP-A function. To test this hypothesis, point mutations and deletions of these two residues in CENP-A were tested for centromere localization in human cells. The resulting loop 1 mutants showed initial

centromere localization but severely reduced retention over time suggesting that the loop 1 region may be important for centromere maintenance or for subsequent centromere protein recognition (Tachiwana, Kagawa et al. 2011).

The Centromere Specific Histone 3 (CenH3) in Filamentous Fungi

Sequence alignments from publicly available CenH3 genomes show variable N-terminal regions and highly conserved HFDs. Complementation assays have found that replacing the endogenous *Neurospora crassa* NcCenH3 gene with the CenH3 of *Podospora anserina*, *Fusarium graminearum*, and *Aspergillus nidulans* disrupts some mitotic and meiotic ability but retains centromere localization (Phatale *et al.*, in preparation).

Additionally, the highly conserved C-terminal region of NcCenH3 is required for meiotic function.

In the case of three filamentous fungi, loop 1 appears strikingly similar to that of human CENP-A: *N. crassa*, *P. anserina*, and *F. graminearum* all show similarities, as they also have two extra amino acid residues when compared to normal H3 (in this case, residues 79 and 80, an arginine and a proline, respectively). Although these residues are not identical to those in the CENP-A loop 1, the conserved function of the histones makes it highly likely that these residues make up a similar solvent accessible loop. This is further supported by the insertion's position within the protein, the likelihood of proline to induce turns in helices, and the hydrophilic nature of both proline and arginine.

Although there are alternate, divergent sequences that may be important for meiotic function, the loop 1 region of CenH3 was, initially, the most attractive region for further

study. The hypothesized importance of this region is supported by the fact that (1) it is conserved through many species; (2) it is altered in the canonical histone H3; (3) it is accessible to kinetochore elements for recognition.

One particular fungus, *A. nidulans*, has a more divergent CenH3 structure. Specifically, the loop 1 region is more divergent and contains additional amino acid residues not found in similar species. *A. nidulans* does not contain the arginine addition found in humans or in other filamentous fungi but instead contains a lysine and proline addition in the loop 1 region (Figure 3). Interestingly, when *N. crassa* CenH3 is replaced with *A. nidulans* CenH3, the resulting strains are viable but result in barren crosses, suggesting that *A. nidulans* CenH3 lacks one or more specific structures that are functionally important for meiosis in *N. crassa*. Finding this sequence variation could help elucidate the first step of the kinetochore assembly pathway or important features for kinetochore maintenance in meiosis.

Although the loop 1 region may well be important for meiosis, the absence of the hypothesized ‘key’ arginine residue in *A. nidulans* suggests otherwise. *A. nidulans* CenH3 is obviously functional as wild-type strains from this species are both viable and fertile. The fact that this residue is absent in *A. nidulans* suggests that it is not exclusively important for CenH3 function or that changes elsewhere in the protein may compensate for the differences in the loop 1 region. It also suggests that other residues in the loop 1 region and/or other whole regions might also be important for proper function of CenH3 in all species.

One of the additional CenH3 sequence variations that may be important for proper CenH3 function is the N-terminal tail. Between species, the N-terminus varies significantly in length and also in amino acid sequence. For example, *N. crassa* has a relatively long N-terminus while that of *P. anserina* is much shorter. The N-terminus of *A. nidulans* shows similar length to that of *N. crassa* but diverges in a few amino acid residues. Recent studies have shown that *N. crassa* requires certain N-terminal elements for meiosis, as strains with chimeric assemblies of *P. anserina* N-termini and *N. crassa* C-termini of the CenH3 protein are barren (Figure 4). Similar findings in *Arabidopsis thaliana* further support the hypothesis of the importance of the N-terminal tail of CenH3 for meiotic function (Ravi, Kwong et al. 2010; Ravi, Shibata et al. 2011). Although the importance of the N-terminus has been demonstrated, it remains unclear whether the length of the tail or amino acid sequence is responsible for this importance.

Although the N-terminal tail of CenH3 has been shown to be important for protein function in many species, the higher degree of sequence divergence across species suggests that this region is not as functionally important as the CenH3 HFD. The N-terminal tail may assist in retaining structural integrity of CenH3 without necessarily interacting with the chromocenter directly. Some species have demonstrated normal CenH3 function with altered N-termini. In fact, human CENP-A does not even seem to require its short N-terminal tail for normal chromocenter localization (Sullivan, Hechenberger et al. 1994; Shelby, Vafa et al. 1997).

Hypotheses

Based on previous knowledge gained from the study of several different species, the following hypotheses were tested:

1. The loop 1 region of *N. crassa* CenH3, specifically the arginine and proline residues, are important for meiotic function and centromere chromocenter localization.
2. The approximate tail length, not the specific sequence of the hypervariable N-terminal region of CenH3, is important for meiotic function.

Proposed Experimentation

Initially, three experiments were proposed to test the meiotic importance of various CenH3 regions in *N. crassa*. Each experiment involved constructing a specifically mutated and fluorescently tagged *N. crassa* CenH3 construct and subsequent expression of these mutant CenH3s in an otherwise wild-type *N. crassa* strain. The various mutations were carefully selected and based on previous CenH3 research in other species.

The first of these experiments was a site-directed mutagenesis experiment involving the hypothesized key arginine and proline residues (residues 80 and 81 in humans but 79 and 80 *N. crassa*). In a previous study (Tachiwana *et al.* 2011), various mutations were introduced into the loop 1 region of CENP-A. Although there was initial chromocenter localization, the team observed a gradual loss of kinetochore retention over time. It was concluded that the loop 1 region of CENP-A was important for kinetochore retention as it may allow for binding of kinetochore proteins. As this type of cloning experiment had not yet been attempted in filamentous fungi, a similar experiment was designed using *N.*

crassa as a model organism. By introducing various point mutations in the loop 1 region of CenH3, the importance of the mutated residues could be tested. The desired mutations were chosen carefully in order to retain the three-dimensional structure of the protein while altering the chemistry that might allow for various, downstream kinetochore proteins to recognize and/or attach to the loop.

The first point mutation was a substitution of the residue 80 proline for a glycine. This mutation was chosen to mimic the structure and biochemistry of the human CENP-A protein since *N. crassa* retains the arginine residue. The second point mutation was a mutation to two glycine residues. The non-existent side chains should not cause any steric hindrance and the flexibility should yield a similar loop structure as in the wild type loop 1. The hydrophilic nature of glycine should retain the hydrophilic nature of the loop 1 region, keeping it solvent-accessible. The third and final point mutation was a mutation to two alanine residues. The small R-group on alanine should provide minimal steric hindrance but the nonpolar nature of alanine might make the loop less accessible to recognition by kinetochore factors.

The second proposed experiment was a complete loop 1 swap between *A. nidulans* and *N. crassa*. The loop 1 region of *A. nidulans* was replaced with the loop 1 region of *N. crassa*. Because a full swap of CenH3 genes rendered *N. crassa* barren, a partial swap may be sufficient to 'fix' the CenH3 function if the loop 1 region is important for meiosis. If this loop 1 swap was sufficient to 'fix' meiotic function, the addition of a single arginine in the loop 1 region of *A. nidulans* was proposed to test if this single addition was sufficient to fix CenH3 function in *N. crassa*.

The third and final proposed CenH3 experiment was a chimeric construct of the *A. nidulans* N-terminal tail and either the *N. crassa* or *P. anserina* HFD. As previously mentioned, this variable region of CenH3 has been shown to be important for meiosis. However, it is unclear whether the length, the amino acid sequence, or both factor into meiotic ability. By testing fertility of these chimeric constructs when expressed in *N. crassa*, it can be determined if N-terminal length, amino acid sequence, or both factors are important for meiotic ability of NcCenH3.

Collection and Interpretation of Data

From these experiments, two types of data can be collected. First, since the entirety of the mutant constructs are fluorescently tagged, microscopy can be utilized to view mutant CenH3 strains for centromere localization. The mutant CenH3 proteins can also be expressed in strains that already contain tagged kinetochore proteins. This allows visualization of subsequent kinetochore protein assembly on the chromocenter. Because meiosis is sensitive to changes in the CenH3 structure, it will be imperative to also test the meiotic function of constructs. This can be done by crossing mutant CenH3 strains with wild-type strains and checking for viable crosses.

All three of these experiments will be interpreted in the same way: If histone localization on the centromere is normal initially and over time, then the changes made to the CenH3 did not alter attachment and retention of centromere histones at the chromocenter. If subsequent kinetochore proteins localize to chromocenters and are retained over time, then the changes made to the CenH3 did not alter recognition and attachment of kinetochore proteins.

These strains will also be tested for meiotic ability. After the mutant constructs are transformed into *N. crassa*, they can be directly screened for centromere localization. The progeny is, however, heterokaryotic and contains both transformed and untransformed nuclei. To determine if meiosis is possible for transformed progeny, a further cross must be made. By observing homokaryotic progeny for fluorescence, we can determine if the tagged construct was able to undergo meiosis. If tagged, homokaryotic crosses are possible, then the CenH3 mutations introduced were sufficient to fix meiotic function and the region or amino acid residues mutated is the important sequence for CenH3 function. Alternately, if transformants are barren, it will suggest that there are another novel sequences in CenH3 that are responsible for the proper function of the protein. It does not necessarily mean that the region tested is not important for meiosis, but rather that there is an additional sequence that is also important for proper meiotic function.

Significance of the Work

Filamentous fungi are a large and diverse group of organisms that are economically and medically important. Despite this importance, the centromere regions of these organisms have not been thoroughly investigated. The highly conserved nature of histone proteins makes their study in fungi more applicable to a wide range of species. It is not implausible to hypothesize that, like the amino acid sequence, the mechanism for kinetochore recognition and assembly is also similar across species. The importance of the kinetochore complex for the proper attachment of spindles and, thus, the proper separation of sister chromatids has been repeatedly demonstrated. Improper segregation of the sister chromatids can result in severe chromosome abnormalities and, in many

cases, cell death. Proper spindle-kinetochore attachment is highly regulated by way of a cellular division checkpoint. Cells may stop dividing altogether if the spindle attachments are irregular.

Kinetochores are important in the division of both autosomes and germ cells. Although the importance of kinetochores for viability has been repeatedly demonstrated, the mechanism by which the kinetochore assembles and functions is still relatively unclear. The purpose of these experiments is to gain a better overall understanding of which sequence variations can be attributed to the functional differences between NcCenH3 and *N. crassa* canonical histone H3. With recent data demonstrating that functional CenH3 proteins are vital for kinetochore retention, it is highly likely that CenH3 is the anchor upon which kinetochores assemble. Determining which sequence variation is responsible for this function will contribute to the increasing database of centromere research and may help to elucidate the mechanisms responsible for centromere recognition and kinetochore assembly.

METHODS

Generation of Loop 1 Point Mutations

Previously constructed plasmids containing wild-type *N. crassa* CenH3 (NcCenH3) were used for the original site-directed mutagenesis experiment. These plasmids had been tested for accuracy with restriction enzyme digests and Sanger sequencing. The previously described loop 1 point mutations, RG, GG, and AA, were introduced in 20 µl PCR reactions with 10 µl of 2x Phusion enzyme (Thermo Fisher), 10-50 ng of plasmid DNA, and 2 µl primers with specifically designed, overlapping primers containing desired mutation sites. The primers used to introduce each mutation are OMF2858 and OMF2859 (P92G), OMF2860 and OMF2861 (R91G and P92G), and OMF2862 and OMF2863 (R91A and P92A) (Table 1).

Site-directed mutagenesis PCR results in a mixed final product. The template DNA varies from the desired DNA product by adenine methylation (if the original plasmid was amplified in a bacterial strain in which the *dam* gene was present). This allows separation of template and newly synthesized DNA by digestion of DNA with the restriction endonuclease *DpnI*, which fragments methylated template DNA while leaving the unmethylated daughter strands intact. DNA from the *DpnI* digest was directly transformed into *E. coli* NEB 5-alpha cells following standard cloning procedures (Sambrook, Fritsch et al. 1989).

After growing the transformants on selective media, potential clones for each mutated plasmid were selected for further confirmation with Sanger sequencing. All sequenced

clones retained the introduced point mutations. Unfortunately, an unexpected mutation was also inserted into GG and AA mutants further downstream from the desired mutations. It is unclear where this mutation came from but, because it was present in the exact same position in two different clones, it is probable that the mutation was present in some of the template plasmid used. To fix these undesired mutations, the experiment could have been started over using a new NcCenH3 template for PCR. Instead, *Bgl*II and *Pst*I restriction sites were utilized to excise the fragment containing the desired mutations (in clones GG and AA) while excluding the downstream insertion. The RG mutant (with the correct sequence) was now used as the plasmid backbone and digested with *Bgl*II and *Pst*I. All fragments were purified by gel electrophoresis and Qiagen gel extraction kits, and fragments with the GG and AA mutations were ligated with T4 DNA ligase into the RG vector to obtain the two additional desired point mutations with no unwanted additional mutations. These plasmids were then transformed into *E. coli* as previously described.

NcCenH3 untranslated regions (UTRs) are considered important for proper regulation of histone function. Since these flanks were not present in the original template CenH3, they were added separately along with the N-terminal GFP-tagged flank that was necessary to visualize histone localization with fluorescent microscopy. When searching the lab database for available plasmids with useable flanking regions, a couple of considerations had to be made: First, it would be beneficial to have the desired flanks in a plasmid that can be directly expressed in *N. crassa*. It was also requisite that the flank-containing plasmid has single-cutting restriction enzyme sites that were compatible to those in the mutant CenH3 gene plasmids. Finally, having a plasmid that already has a fluorescent tag

attached to the N-terminal flank would reduce cloning steps. The tag must be on the N-terminus as C-terminal fluorescent tagging has been shown to inhibit function of CenH3.

After searching through plasmid maps for a plasmid to fit the previously mentioned parameters, it was determined that one would have to be assembled as the current lab stocks did not include a perfectly compatible plasmid. Some plasmids contained the correct components but were in the wrong backbone while others lacked complementing restriction sites that would be necessary for subsequent cloning steps. Assembling a new plasmid from two pre-existing lab plasmids was the next, most logical step. The first plasmid (pPP74.55) contained the GFP tag, flanking regions, and pBS vector backbone but had two *NotI* restriction enzyme cut sites, one between the CenH3 gene and the C-terminus, and one further upstream in the CenH3 gene itself. The *NotI* site is one of the few useful restriction sites in both the point mutant plasmids and this vector plasmid but it is rendered useless, as the flanks cannot be added with any directionality. To counter this problem, another plasmid containing the 3' UTR and NcCenH3 was utilized (pPP65.15). This plasmid lacked the upstream *NotI* site but possessed the common surrounding cut sites; making it possible to eliminate the 3' *NotI* site while maintaining the same sequence and other, previously mentioned plasmid parameters. Plasmids pPP74.55 and pPP65.15 were digested with *BstBI* and *AscI*, purified by gel electrophoresis and Qiagen gel extraction kit, and ligated together using pPP74.55 as the vector and pPP65.15 as the insert. Ligation reactions were transformed into *E. coli* NEB 5-alpha cells as previously described (Sambrook, Fritsch et al. 1989). Clones were checked using restriction enzyme digests and the resulting positive clones were named pSTF10.

To add the 3' flank, each point mutation plasmid (as the vector) and pSTF10 (as the insert) were digested with *Stu*I and *Spe*I. Plasmid pSTF10 was also digested with *Hinc*III as digestion with only *Stu*I and *Spe*I would have yielded two fragments of equal size, the desired insert and another contaminating fragment from the plasmid backbone. The fragments were ligated together and transformed in *E. coli*. Resulting clones were tested for successful retention of the ligation product with restriction enzyme digests.

The final step of the cloning was attachment of the 5' UTR and GFP tag. Cloning steps were carried out as previously described using the enzymes *Sna*BI and *Not*I to generate fragments. The point mutation plasmids with the 3' UTRs were used as the insert and pSTF10 was used as the vector. The ligations were then transformed in *E. coli*, potential clones were isolated, plasmid DNA extracted, and the mutated regions sequenced using primer OMF1756 to ensure that all clone sequences were complete. The positive plasmids were named pSTF11(RG), pSTF12(GG), and pSTF13(AA) (Figure 5).

Generation of Chimeric Loop 1 Constructs

The initial step in assembling a chimeric gene is introduction of restriction sites. These sites must be inserted into the same location, flanking the splicing region, in both genes used for the construct. Primers were specifically designed to allow AnCenH3 to be digested into three parts using the restriction sites. These primers were OMF2873/OMF2852 (N-terminal end), OMF2851/OMF2854 (loop 1 region), and OMF2853/OMF2855 (C-terminal end). PCR products were isolated by gel extraction kit, ligated together using the Topo blunt ligation protocol, and transformed as previously described.

After selecting clones with kanamycin resistance and isolating plasmid DNA, *EcoRI* was used to identify correct clones. *EcoRI* was used because there are two cut-sites in the pCR4 backbone and one cut site in the AnCenH3 gene. Positive clones will show three bands while negative clones will show two different sized bands. Unfortunately, after the digests were run on a gel, there was no sign of DNA present. It was unclear whether the ligation, transformation, or plasmid isolation failed, so a colony PCR was done on potential clone *E. coli* cultures. By using T3 and T7, primers that would amplify inserted fragments; it was possible to see that some of the clones retained the desired insert. Due to the abnormal liquid culture growth we observed while isolating the plasmids, we hypothesized that our kanamycin plates may have not been strong enough to select against untransformed plasmids.

Since the pCR4 vector for Topo cloning contains genes for both ampicillin and kanamycin resistance, plasmid isolation was attempted again with the positive clones from the colony PCR using ampicillin selective media. Some of the clones did not grow in ampicillin liquid LB media and most of the clones that did grow did not show definitive digestion. There was also a significant amount of genomic DNA contamination. With these setbacks, only the N-terminal segment was successfully cloned into the Topo vector.

In the interest of time, the original scheme was altered. The two remaining PCR products were isolated and digested with their introduced restriction enzyme sites. They were then ligated into the successful N-terminal Topo clone. Unfortunately, these ligations were unsuccessful and the PCR products were not added into the vector. Additionally, there was a mounting concern that a problem existed with either the kanamycin selection plates

used or even with the competent *E. coli* used for transformation. This concern was due to the high level of false-positives and also to the abnormal *E. coli* growth in selective liquid culture. Thus, a new cloning scheme was developed and the experiment was re-started. This cloning scheme was different from the first attempt as the PCR fragments were used to first assemble the AnCenH3 gene with restriction enzyme sites before inserting it into a vector.

Using the previously described PCR amplification, we amplified, isolated, and digested the fragments in the following manner: The N-terminal fragment was digested with *HindIII*, the loop 1 fragment with *HindIII* and *PstI*, and the C-terminal fragment with *PstI*. These three fragments were then ligated together; all at the same time. Because the restriction sites are specific and produce sticky-ends, the fragments can only assemble in the desired orientation. The resulting ligation was digested with *BamHI* and *ApaI* and ligated into a pBS vector before being transformed into *E. coli* as previously described.

After screening potential clones, it was determined that there were still no clones present. In an attempt to solve this problem, a third cloning scheme was developed based upon the hypothesis that the PCR amplification did not yield sufficient DNA over-hang for the restriction enzymes to attach and cut DNA. The new scheme differed from the other schemes as it generated a large, tandem-repeating chain of each fragment using T4 polynucleotide kinase and T4 DNA ligase. It was hypothesized that, by constructing a repeating chain, the restriction enzymes would have sufficient overhang to properly cut DNA. After the initial chain assembly, the three fragment chains were digested with their respective enzymes and ligated together. This total fragment (the AnCenH3 gene) was then amplified using OMF 2873 and OMF 2855. The resulting, 660 bp fragment was

purified with gel electrophoresis, phosphorylated with T4 polynucleotide kinase, and ligated as previously described. This created a repeating chain of the AnCenH3 gene with internal restriction sites. Because the restriction sites necessary for ligation into a vector were located on each end of the gene, it was predicted that these sites would also lack sufficient over-hang to be successfully digested. The repeating AnCenH3 chain was then digested with *ApaI* and *BamHI*, ligated into the pBS vector, and transformed as previously described. The transformation was unsuccessful as no colonies were present on the ampicillin screening plate.

After the third unsuccessful attempt at a loop 1 chimeric assembly, this approach was abandoned. Although it was the intention to eventually revisit the loop 1 assembly experiment, time constraints kept this from happening. Suggestions for future AnCenH3 loop 1 chimeric assemblies will be made in the discussion section.

Assembly of N-Terminal Chimeric Constructs

To determine whether the N-terminus length, amino acid sequence, or both factors affect meiotic ability of CenH3, chimeric constructs were built using a strategy similar to the one proposed in the loop 1 chimeric assembly experiment. Restriction enzyme sites were introduced to assemble two segments: The N-terminal tail of *A. nidulans* CenH3 and the histone folding domain (HFD) of CenH3 (*N. crassa* or *P. anserina*). The following primers were used to amplify the various regions with PCR: (1) the N-terminus from AnCenH3 (OMF2873, OMF3019), (2) the HFD of PaCenH3 (OMF1849, OMF1957) and (3) the HFD of NcCenH3 (OMF1956, OMF2937). These fragments were then ligated into a Topo cloning vector (pCR4) using blunt ligation. However, the nature of blunt

ligation yields a mixture of clones as the insert can be incorporated into the vector in either orientation. To ensure that only clones with the correct insert orientation were used in subsequent cloning steps, restriction enzyme digests with *Pst*I and *Bam*HI were used to distinguish the correct from incorrect inserts. The resulting fragments were slightly different lengths depending on the direction of the insert, allowing visual discrimination of the two conformations. After selecting positive clones, the AnCenH3 N-terminus was combined with the HFD of PaCenH3 or NcCenH3 (NAnCPaCenH3 or NAnCNcCenH3, respectively).

The NAnCPaCenH3 chimera was fluorescently tagged with an N-terminal mCherry and flanked with 5' and 3' UTRs using overlapping PCR and the primers OMF180/OMF182 and OMF180/OMF188, respectively. It was targeted to the endogenous locus of *N. crassa* CenH3 by transformation of strain N3011 (*mat a his-3 mus-51::bar⁺*), following a published electroporation protocol (Colot, Park et al. 2006), to obtain transformant NMF591, in which the normal CenH3 locus was replaced by the chimera.

The GFP-NAnCNcCenH3 chimera was cloned in a *his-3* targeting plasmid (pBM60) to yield plasmid pPP83.6, which contains the P_{CenH3}-GFP-NAnCNcCenH3-3'UTR fragment. This was targeted to the *his-3* locus by transformation into a heterokaryotic Δ NcCenH3 strain (NMF247) to generate NMF590. The NMF589 and NMF590 were screened for localization of fluorescence. Transformants were crossed to NMF162 to obtain homokaryotic progeny.

RESULTS

Overall, experimentation with CenH3 sequences would perhaps have been more effective had efforts been focused on one project at a time. Proposing the completion of all three experiments might have been too optimistic considering the inherent set-backs of molecular biology and time constraints. Although significant advancements were made towards the final goal of expressing all altered CenH3 constructs in *N. crassa* for the purpose of observing chromocenter localization and meiotic ability, this series of experiments has not yet produced data on all of the proposed experiments.

There was, however, a different type of data produced than originally stated. Because many of the experiments required novel cloning techniques, whether or not a certain method worked is also valuable information for future research. Hopefully, this information will be taken into consideration when future cloning endeavors are pursued; limiting similar setbacks that were present in this project. Instead of including this type of results in this section, they were kept in context in the methods section for clarity.

The site-directed mutagenesis constructs have been successfully assembled and are ready for integration and expression in *N. crassa*. After transformation and growth, the fluorescence data will be collected by epifluorescence microscopy, and crosses will be initiated to determine if these point mutations allow normal progression through the sexual cycle. These results will become part of a publication on systematic mutagenesis of the *N. crassa* CenH3 gene (S. Friedman, S. Ferrer, P. Phatale, M. Freitag, unpublished data).

The loop 1 chimeric construct was never successfully assembled. Multiple setbacks and protocol alterations took time without yielding useful results. Although this experiment did not proceed as planned, useful information regarding various cloning methods and their efficacy was obtained (see Discussion).

The N-terminal chimeric constructs were successfully expressed in *N. crassa* and the heterokaryon were observed for fluorescence. In all cases, single chromocenters were observed, suggesting that the chimeric proteins were targeted to the centromeric regions (Figure 6). The heterokaryotic transformants are in the process of being used in crosses to generate homokaryotic progeny. These crosses were fertile and individual spores were isolated. The next step is to cross compatible progeny from these crosses to each other to assay if all N-terminal chimeric constructs are able to undergo normal meiosis. The expectation is that either the length or the amino acid content of the *A. nidulans* N-terminus will interfere with normal meiosis.

DISCUSSION

The following section will be divided into two segments: Discussion of the N-terminal chimeric construct results and discussion of the protocols followed for the site-directed mutagenesis and chimeric loop 1 assembly experiment. Since neither experiment has yet produced noteworthy results, detailed proposal of the subsequent steps to achieve results, ideas to improve future cloning endeavors, areas for protocol improvements, and expected results will be discussed.

N-terminal Chimera

As both chimeras with the N-terminus of *A. nidulans* CenH3 but HFD of *N. crassa* or *P. anserina* CenH3 were targeted to chromocenters, they seem to at least be properly localized to centromeres. To test their function during meiosis homozygous crosses need to be carried out. Homokaryotic progeny have been isolated and these will be screened for the correct mating type to carry out these crosses. The expectation is that either the long N-terminus or the specific amino acid composition of the *A. nidulans* CenH3 domain of the chimeras will interfere with meiosis. Similar results were obtained in previous experiments where tagging of the C-terminus of numerous CenH3 constructs with GFP, FLAG or ten random amino acids resulted in barren crosses (i.e. no spores were produced in normal fruiting bodies; Phatale *et al.*, in preparation). A second set of experiments showed that the short *P. anserina* CenH3 N-terminus interfered with normal function when mCherry was added to this end of the protein, while the longer *N. crassa* N-terminus tolerated the addition of the rather bulky mCherry tag. It was unclear from

these experiments if it was indeed the length or the specific amino acid composition that was responsible for CenH3 meiotic function.

Site-directed Mutagenesis

Nearing completion, the site-directed mutagenesis experiment has been relatively straight-forward. Setbacks were minimal and usually had simple fixes. Initially, there were some issues with some of the more delicate protocols like plasmid isolation and *E. coli* transformation. These matters were most likely lack of experience with sensitive *E. coli* protocols that resulted in decreased efficiency.

Based upon the previously discussed CENP-A loop 1 mutation results it is expected that the *N. crassa* loop 1 mutations behave similarly. Reduced CenH3 retention at centromeres over time is likely, especially with the GG and AA mutants as they lack the biochemical properties of the native loop 1 residues. It will be interesting to see if the RG mutant demonstrates higher retention as it contains the key arginine found in human CENP-A.

Loop 1 Chimera

This experiment was, by far, the most time consuming of the three. Multiple setbacks with cloning, digestions, and transformations took much time and effort. Fortunately, useful information regarding the methods utilized was obtained and documented, facilitating successful future chimeric assembly endeavors.

The next time this project is attempted, it would be beneficial to design different primers. The primers were designed to insert various restriction enzyme cut-sites that would allow for assembly of the chimera. Unfortunately, they were not designed with enough DNA over-hang to allow for restriction enzyme attachment and function. Instead of designing and ordering new primers, compensation for this set-back was attempted with protocol variations. Theoretically, these altered protocols should work but, in practice, their efficiency is very low. As a result, the loop 1 chimeric construct was never successfully created.

An additional method for assembling the loop 1 chimeric construct was discussed but never attempted due to insufficient time. This method, known as Gibson Assembly, utilizes T5 exonuclease, ligase and DNA polymerase to assemble overlapping PCR products. By digesting the 5' end of many sequential blunt-end PCR products, specific overhangs can be made and used in place of primers. Gibson assembly has been shown to effectively assemble several hundred kilobases of sequence with a single-step isothermal reaction (Gibson, D., L. Young, et al. 2009).

From previous CenH3 experimentation, it is expected that the loop 1 substitution will be sufficient for histone localization to the centromere. Whether or not it will be sufficient to fix meiotic function is unclear as the N-terminus of CenH3 has been shown to also be very important for proper function. When the results from the N-terminal chimeric constructs are analyzed, it will allow for a better prediction for this experiment.

CONCLUSION

Although this series of experiments has not yet provided complete results, some important observations have been made. The N-terminal chimeric construct appears to localize normally and the site-directed mutagenesis constructs are ready to be expressed in *N. crassa* and tested for chromocenter localization and meiotic ability. Two new methods for assembling the loop 1 chimeric construct have been outlined for future endeavors.

BIBLIOGRAPHY

- Alberts, B., D. Bray, et al. (2009). *Essential Cell Biology* (3rd edition). Garland Science.
- Bui, M., E. K. Dimitriadis, et al. (2012). "Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo." *Cell* **150**(2): 317-326.
- Cleveland, D. W., Y. Mao, et al. (2003). "Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling." *Cell* **112**(4): 407-421.
- Colot, H. V., G. Park, et al. (2006). "A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors." *Proc Natl Acad Sci U S A*.
- Dalal, Y., T. Furuyama, et al. (2007). "Structure, dynamics, and evolution of centromeric nucleosomes." *Proc Natl Acad Sci U S A* **104**(41): 15974-15981.
- Dalal, Y., H. Wang, et al. (2007). "Tetrameric structure of centromeric nucleosomes in interphase *Drosophila* cells." *PLoS Biol* **5**(8): e218.
- Dimitriadis, E. K., C. Weber, et al. (2010). "Tetrameric organization of vertebrate centromeric nucleosomes." *Proc Natl Acad Sci U S A* **107**(47): 20317-20322.
- Furuyama, T. and S. Henikoff (2006). "Biotin-tag affinity purification of a centromeric nucleosome assembly complex." *Cell Cycle* **5**(12): 1269-1274.
- Furuyama, T. and S. Henikoff (2009). "Centromeric nucleosomes induce positive DNA supercoils." *Cell* **138**(1): 104-113.
- Gibson, D., L. Young, et al. (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." *Nature Methods* **6**, 343-345.
- Henikoff, S. and T. Furuyama (2010). "Epigenetic inheritance of centromeres." *Cold Spring Harb Symp Quant Biol* **75**: 51-60.
- Nelson, D. L. and M. M. Cox (2008). *Lehninger Principles of Biochemistry* (5th edition). Freeman, W. H. & Company.
- Ravi, M., P. N. Kwong, et al. (2010). "The rapidly evolving centromere-specific histone has stringent functional requirements in *Arabidopsis thaliana*." *Genetics* **186**(2): 461-471.
- Ravi, M., F. Shibata, et al. (2011). "Meiosis-specific loading of the centromere-specific histone CENH3 in *Arabidopsis thaliana*." *PLoS Genet* **7**(6): e1002121.
- Sambrook, J., E. F. Fritsch, et al. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.
- Shelby, R. D., O. Vafa, et al. (1997). "Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites." *J Cell Biol* **136**(3): 501-513.
- Shivaraju, M., J. R. Unruh, et al. (2012). "Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast." *Cell* **150**(2): 304-316.
- Smith, K. M., J. M. Galazka, et al. (2012). "Centromeres of filamentous fungi." *Chromosome Res* **20**(5): 635-656.
- Smith, K. M., P. A. Phatale, et al. (2011). "Heterochromatin is required for normal distribution of *Neurospora crassa* CenH3." *Mol Cell Biol* **31**(12): 2528-2542.
- Sullivan, K. F., M. Hechenberger, et al. (1994). "Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere." *J Cell Biol* **127**(3): 581-592.
- Tachiwana, H., W. Kagawa, et al. (2011). "Crystal structure of the human centromeric nucleosome containing CENP-A." *Nature* **476**(7359): 232-235.
- Talbert, P. B., T. D. Bryson, et al. (2004). "Adaptive evolution of centromere proteins in plants and animals." *J Biol* **3**(4): 18.

APPENDIX A

Figures

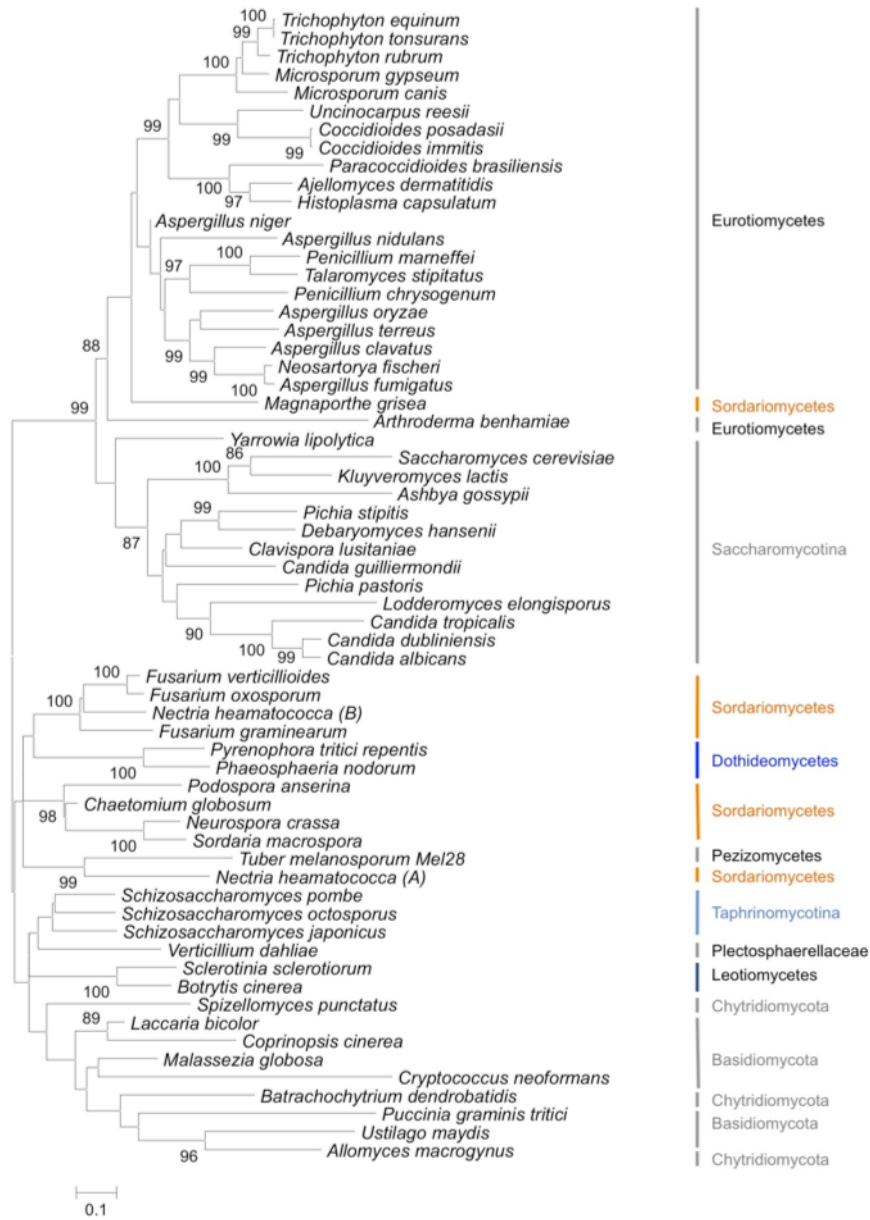


Figure 1 | Phylogenetic tree showing relationship of CenH3 proteins based on sequence variation between diverse groups of filamentous fungi. Of particular interest is the large variation between *A. nidulans* and the other filamentous fungi discussed here, *N. crassa*, *P. anserina* and *F. graminearum*, all belonging to the Sordariomycetes.

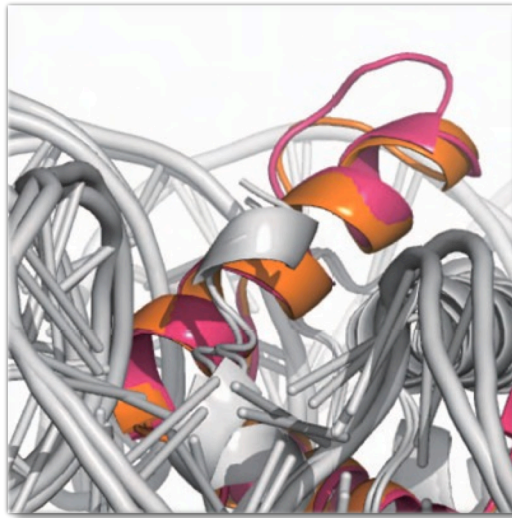


Figure 2 | Visual comparison of the CENP-A solvent accessible loop 1 addition (magenta) in comparison to the canonical H3 loop 1 region (orange). From (Tachiwana, Kagawa et al. 2011). The mammalian RG and fungal RP or LP residues are exposed in the loop and the R, where present, may be modified by methylation.

| | |
|-------------------------|---|
| <i>N. crassa</i> : | REIAMQF <u>RP</u> MD--EEMRWQSQAAILALQEA |
| <i>F. graminearum</i> : | REIALSM <u>RP</u> AD--AGLRWQSQAIMALQEA |
| <i>P. anserina</i> : | REIAITC <u>RP</u> MN--DELRWQSQAIIQALQEA |
| <i>A. nidulans</i> : | REVALDI <u>LP</u> ADVGSELRWQSHAIQALQEA |
| <i>H. sapiens</i> : | REICVKFT <u>RG</u> VD---FNWQAQALLALQEA |

Figure 3 | CenH3 sequence loop 1 variations between key species observed in this study. Residues in red are conserved in most filamentous fungi and underlined residues are hypothesized to be part of a similar solvent accessible loop as the underlined residues in the *H. sapiens* sequence. The amino acid residues are aligned by sequence and not numbered due to large variations in the previous sequence, the N-terminal tail in particular.

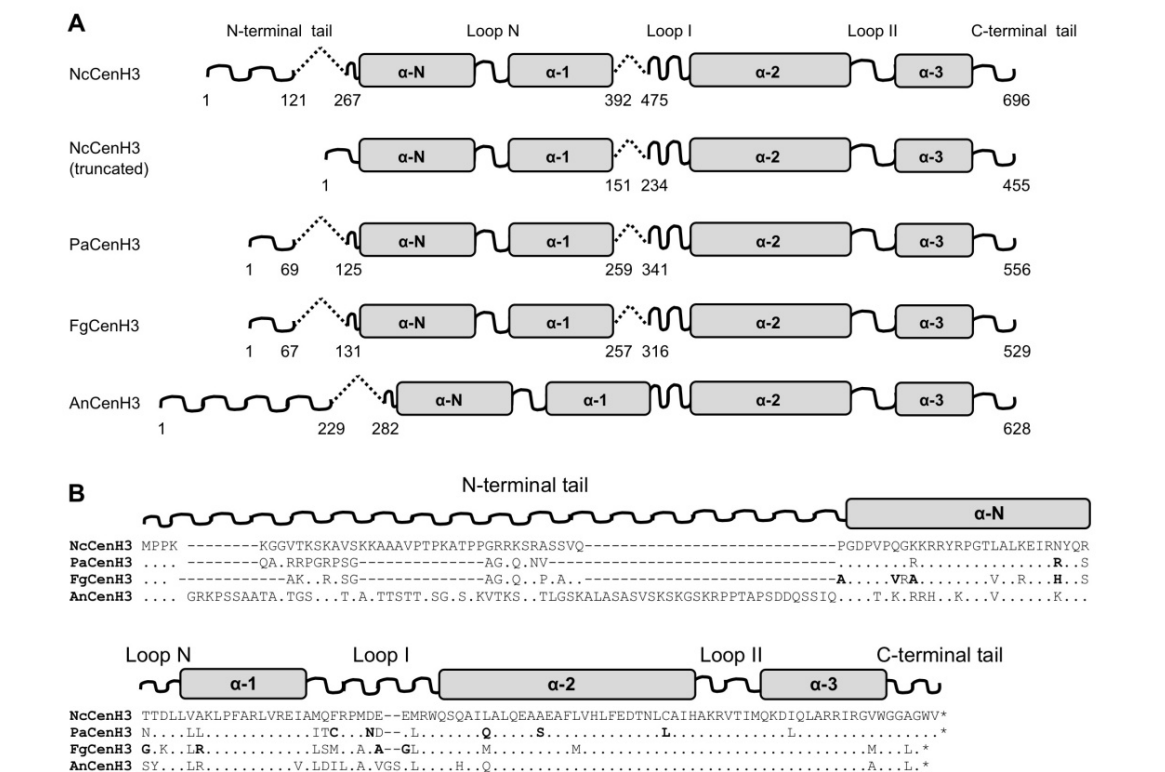


Figure 4 | A) CenH3 protein structure from four filamentous fungi utilized in this study. B) Sequence alignments of CenH3 regions from filamentous fungi used for functional studies. Regions of interest include the hypervariable N-terminal segment as well as the variations in the loop 1 region of the protein (Phatale *et al.*, in preparation). Periods indicate the same residue as seen in the NcCenH3 sequence while a letter indicates an alternate residue in that position. A dash is a placeholder, indicating that there is no amino acid residue present in this position.

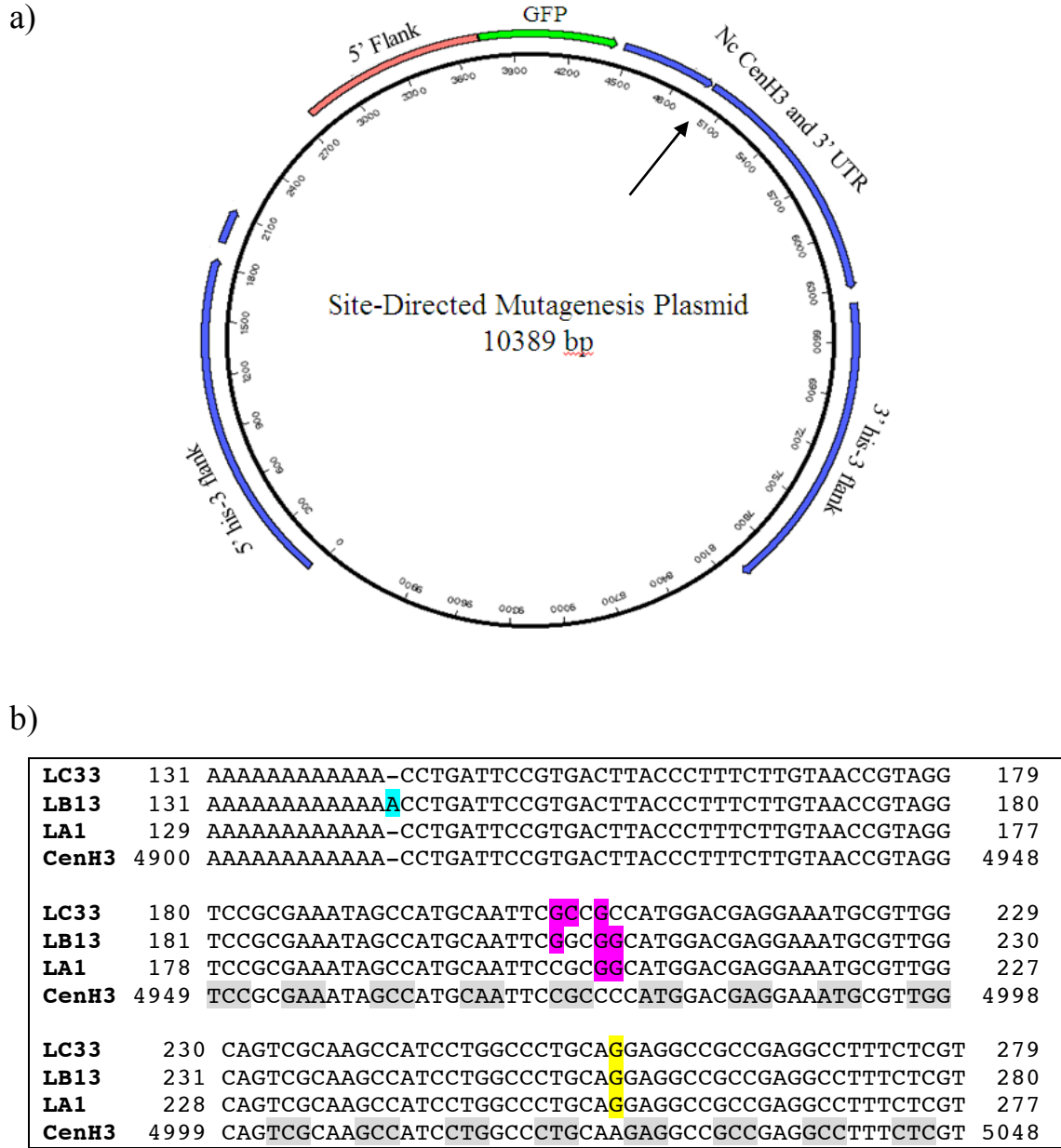
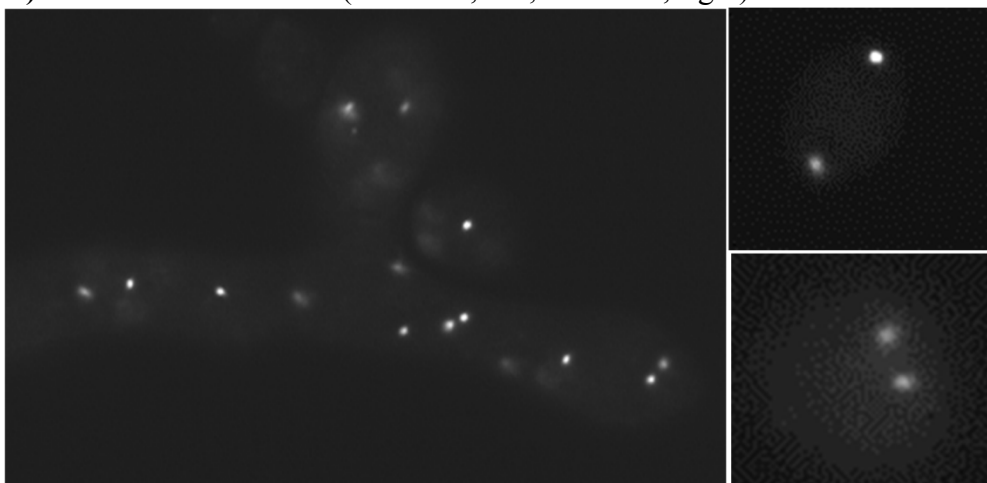


Figure 5 | a) Plasmid map of the three site-directed mutagenesis constructs. The arrow indicates the location of the loop 1 point mutations. The base pairs altered by the site-directed mutagenesis are in positions 4972-4977. b) The point mutant sequences as determined with Sanger sequencing. Cyan highlighted base is an unintended mutation that was acceptable because it was located in an intron (also, polynucleotide stretches are difficult to read so mutation could simply be a sequencing error), fuchsia bases are introduced mutations in the loop 1 region, and the yellow base is an unintended but silent mutation. Reading frame is indicated by the grayed codons present in the CenH3 sequence.

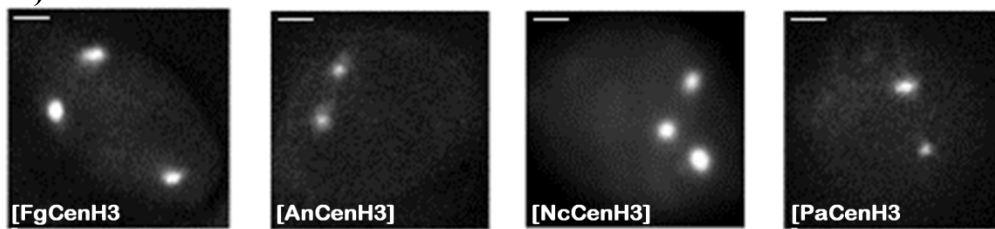
A) GFP-NAnCNcCenH3 (NMF589, left; NMF590, right)



B) mCherryNAnPaCenH3 (NMF591)



C)



D)

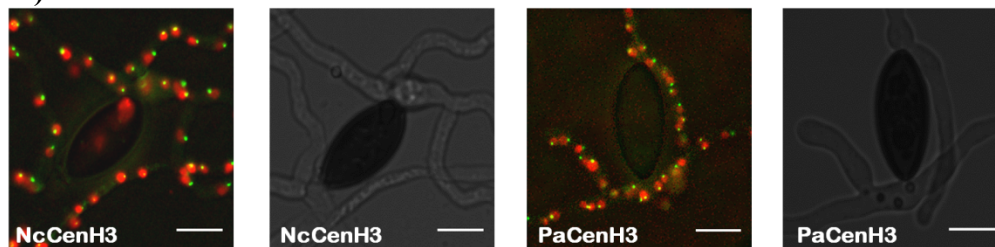


Figure 6 | Chimeric CenH3 is found in chromocenters. a) Localization of GFP-NAnCnNcCenH3 (NMF589, left; NMF590, right) in chromocenters. Images on right are similar magnification as those shown in panel C. b) Localization of mCherryNAnPaCenH3 (NMF591) in chromocenters. Localization was less sharp than for GFP strains but it is still unclear if this is related to the construct or the behavior of the mCherry tag. c) Heterokaryotic transformants express CenH3-GFP that is targeted to a single discrete focus in each nucleus. Representative images of CenH3-GFP expression in heterokaryotic transformants with FgCenH3-GFP, AnCenH3-GFP, NcCenH3-GFP and PaCenH3-GFP (Phatale *et al.*, in preparation). d) CenH3-GFP in homokaryotic progeny with NcCenH3-GFP and PaCenH3-GFP is targeted to centromeric chromocenters. We obtained progeny from crosses of a strain expressing RFP-tagged nuclear linker histone H1 to [NcCenH3-GFP] or [PaCenH3-GFP]. Ascospores were germinated and fluorescence of H1-RFP and CenH3-GFP observed. Crosses with [FgCenH3-GFP] and [AnCenH3-GFP] were barren (Phatale *et al.*, in preparation).

APPENDIX B

Table

Table 1: List of oligonucleotides

| Oligo # | Name | Sequence |
|---------|-----------------|--|
| OMF180 | NcCenH3GlyF | GATGAATGACTAGATGCCGCGGTG |
| OMF182 | NcCenH3GlyR | CCTCCGCCTCCGCCTCCGCCGCTCCGCCTACCCACCCAGCACCTCCCC |
| OMF188 | NcCenH3loxR3 | GCCCCACGCTAAAGCTGTT |
| OMF1756 | BamF_C_NcCenH3 | CTCGGATCCCGTCCCCCAGGGCAAGAAGAGGCGT |
| OMF1849 | CBamF_PaCenH3 | CAGGGGATCCAGTCCCCCAAGGCCGCAA |
| OMF1956 | Nchph5R2 | GATAAGCTTGATATCGAATTCTTACTTGTTTCATACCCACCCAGCACCTCCCCA |
| OMF1957 | Pahph5R2 | GATAAGCTTGATATCGAATTCTTACTTGTTTCACACCCACCCCGCCCCGCCCA |
| OMF2851 | HindAnCenF | CCGAAAGCTTCCATTTGCACGTCTCGT |
| OMF2852 | HindAnCenR | AATGGAAGCTTTCGGAGAAGCAGATCG |
| OMF2853 | PstAnCenF | GGCGCTGCAGGAAGCCGCTGAAGCCTT |
| OMF2854 | PstAnCenR | GCTTCCTGCAGCGCCTGGATTGCATGC |
| OMF2855 | BamTGA_AnCenR | ATGGATCCTCAGCCAAGACCACCCCAAGCT |
| OMF2858 | NcCenRG_mutF | GCCATGCAATTCCGCGGCATGGACGAG |
| OMF2859 | NcCenRG_mutR | CTCGTCCATGCCGCGGAATTGCATGGC |
| OMF2860 | NcCenGG_mutF | GCCATGCAATTCCGCGGCATGGACGAG |
| OMF2861 | NcCenGG_mutR | CTCGTCCATGCCGCGGAATTGCATGGC |
| OMF2862 | NcCenAA_mutF | GCCATGCAATTCCGCGGCATGGACGAG |
| OMF2863 | NcCenAA_mutR | CTCGTCCATGGCGGCGAATTGCATGGC |
| OMF2873 | ApaNde_AnCenF | TAGGGCCCAACATATGCCCCAAAAGGACG |
| OMF2937 | BamF2_C_NcCenH3 | AGCCGGGGATCCCGTCCCCCAGGGCAAGAA |
| OMF3019 | AnCen_NBamR | GGGGTTGGATCCCCGGCTGTGTAGTGT |