INNER KINETOCHORE COMPOSITIONS ACROSS DIVERSE CENTROMERE TYPES IN BUDDING YEASTS

A Thesis

Presented to

The Faculty of the Department of Molecular Biology

Colorado College

By

Mai Tien Nguyen

May 2024

Prof. Sara Hanson
Primary Thesis Advisor

____

Prof. Olivia Hatton
Secondary Thesis Advisor

____

Abstract

Accurate chromosome segregation is essential for eukaryotic cell division, ensuring the faithful transmission of genetic information and preventing aneuploidy. This process relies on the precise attachment of microtubules to kinetochores, multiprotein structures assembled on specific chromosomal regions known as centromeres. While functionally conserved, the composition of the kinetochore complex, especially the inner kinetochore, varies across eukaryotes. The inner kinetochore proteins interact with centromeric DNA, which is structurally diverse and rapidly evolving. Together, despite plasticity in organization and composition, the centromere-kinetochore complex works towards the shared goal of accurate chromosome segregation. Centromere diversity can be observed in budding yeasts, where centromeres vary from short, sequence-specific point centromeres to long, epigenetically maintained regional centromeres. However, our understanding of the inner kinetochore compositions remains limited in budding yeasts. Given their substantial centromere and genetic diversity, budding yeasts are an ideal group for studying the co-evolution of centromere structures and inner kinetochore compositions at the species level. To inventory inner kinetochore compositions in budding yeasts with varying centromere types, we developed "mign", a tool written in Python, to automate the homolog identification of 20 inner kinetochore proteins across 338 species. The resulting inventory reveals that proteins binding to point centromeres in a sequence-specific manner are found in species with regional centromeres, even though regional centromeres lack sequence conservation and inherit centromere function independently of DNA sequence. Additionally, the inner kinetochore inventory in the Saccharomycodaceae family is similar to species with point

centromeres, making them promising subjects for point centromere research. Discovering point centromeres in this family could imply an earlier origin of point centromeres than currently postulated. Exploring the co-evolutionary dynamics of centromeres and inner kinetochore compositions adds valuable insights into how distinct modes of genome evolution have shaped the diversity of kinetochores. Furthermore, understanding centromere and kinetochore plasticity is important for understanding their conserved role in chromosome segregation. Our data provide insights and guidance for future wet lab projects aimed at validating and characterizing inner kinetochore inventory in budding yeasts and conducting functional analyses of kinetochore proteins.

Introduction

Cell division is fundamental for the proper functioning and growth of organisms through the inheritance of genetic information. Chromosome segregation ensures accurate genetic transmission during cell division in eukaryotes (Cleveland et al., 2003). This highly conserved process becomes particularly important as dysregulation in chromosome segregation is implicated in tumorigenesis, genomic instability, and developmental disorders (Maiato & Silva, 2023). Accurate chromosome segregation is facilitated by centromeric DNA and its associated kinetochore protein complex (Figure 1). The centromere, a specialized chromosomal region, acts as the primary constriction where spindle microtubules attach to separate sister chromatids during mitotic and meiotic divisions (Cleveland et al., 2003). The microtubules and centromeric DNA are connected by the kinetochore, a multiprotein structure that assembles on centromeric DNA (Cleveland et al., 2003). The centromere-kinetochore complex guarantees the faithful segregation of sister chromatids by ensuring the precise attachment of spindle microtubules.

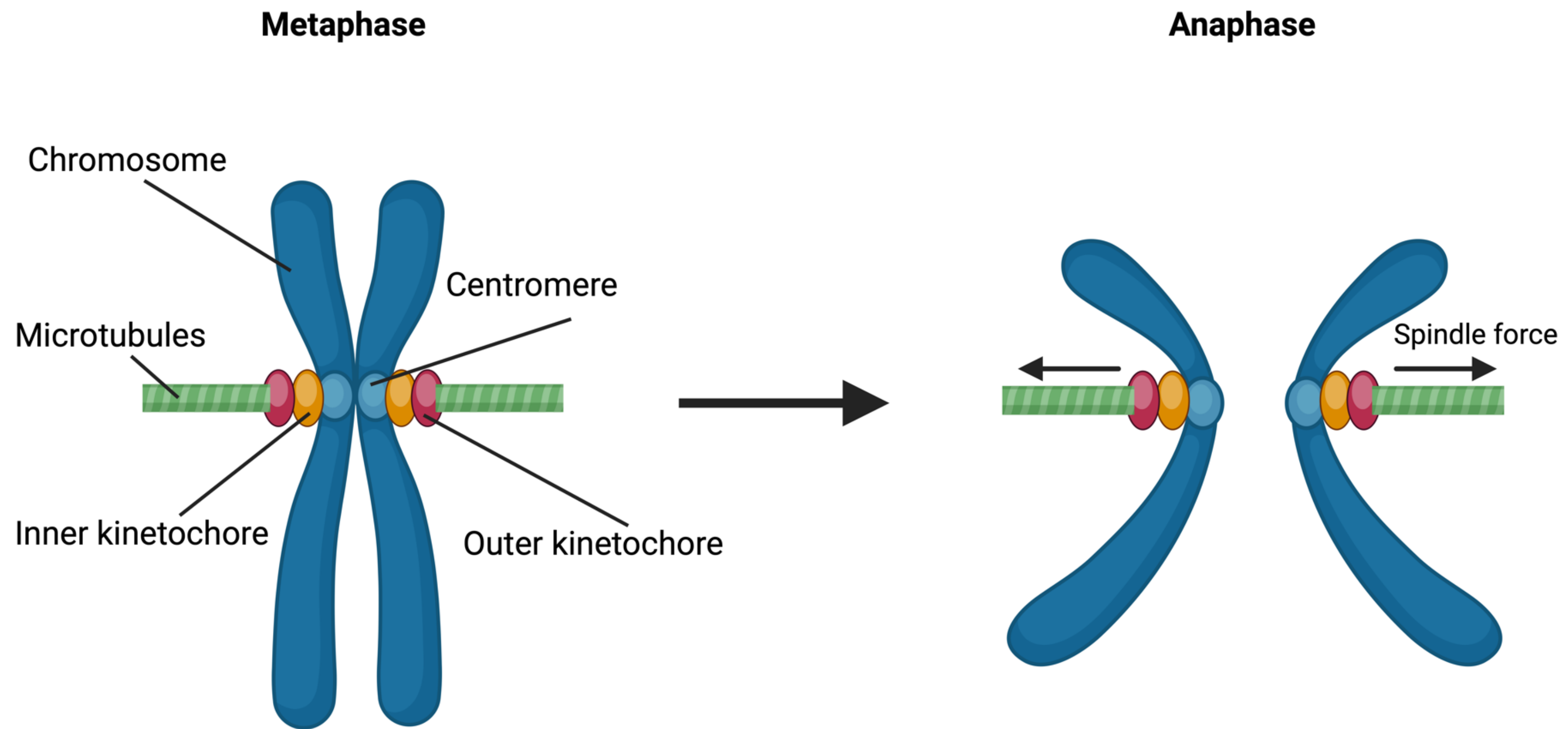

Figure 1. Centromeres and kinetochores facilitate chromosome segregation.
In metaphase, the kinetochore complex, a multiprotein structure (shown in yellow and red), attaches to centromeres, a region of chromosomal DNA (shown in light blue), on sister chromatids. Microtubules (shown in green) from opposite poles of the spindle attach to the kinetochores of sister chromatids. During anaphase, the microtubules exert pulling forces on the kinetochores, leading to the separation of sister chromatids and their migration toward opposite poles of the cell. Created with BioRender.com.

Diversity in organization of the centromeres

Despite functional conservation, centromeres are not only some of the most rapidly evolving DNA sequences in the genome, but also structurally diverse (Guin et al., 2020; Sridhar & Fukagawa, 2022). Within eukaryotes, centromeres can be broadly categorized into two types: point centromeres and regional centromeres. Point centromeres, like those in the budding yeast *Saccharomyces cerevisiae*, are characterized by their short length that can be wrapped around a single histone octamer, capturing one spindle microtubule (Figure 2). This centromere type features conserved DNA elements: CDEI, CDEII, and CDEIII (Drinnenberg et al., 2016; Guin et al., 2020). Both CDEI and CDEIII contain highly conserved nucleotide sequences, where a single point mutation is sufficient to disrupt centromere function, such as a single-base pair mutation in the CCG motif of CDEIII (Espelin et al., 2003; Guin et al., 2020). While CDEI is not essential for centromere function, both the conserved length of the highly AT-rich CDEII and CDEIII play important roles (Espelin et al., 2003; Guin et al., 2020). On the other hand, regional centromeres are more complex and extend beyond a single nucleosome, capturing multiple microtubules and lacking conservation of the underlying DNA sequences (Drinnenberg et al., 2016; Guin et al., 2020). The absence of sequence conservation suggests DNA sequence–independent inheritance of centromere function, with regional centromeres being epigenetically defined (Drinnenberg et al., 2016; Guin et al., 2020). Diverse regional centromere types exist, from the short regionals in the human pathogenic budding yeast *Candida albicans* and the long regionals in the fission yeast *Schizosaccharomyces pombe* to the several megabase satellite arrays in humans (Figure 2). While centromere sequences vary

among eukaryotes, a common feature is the presence of the histone H3 variant CENP-A (blue circles, Figure 2), which plays a role in epigenetically determining centromeres and is essential for kinetochore specification and assembly (Drinnenberg et al., 2016).

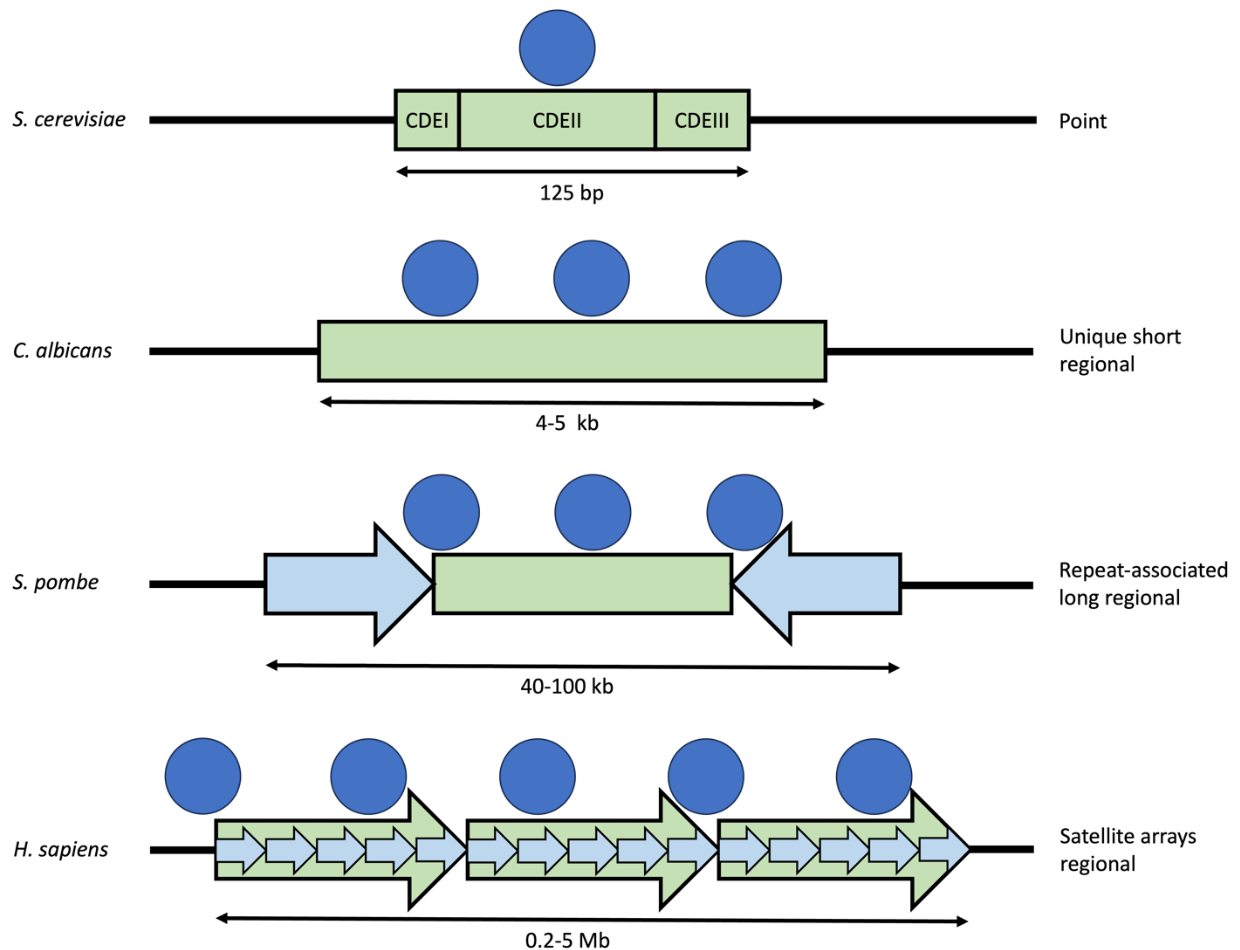

Figure 2. Centromere organization among different eukaryotes.
Centromeres vary in size and sequence across eukaryotes and can be broadly categorized as either point or regional. Point centromeres, such as those in *S. cerevisiae*, are characterized by short and defined sequences known as centromere DNA elements (CDEs). Regional centromeres are determined epigenetically and span more than several kilobases. Repetitive sequences are depicted by light blue arrows. Dark blue circles represent histone H3 variant CENP-A. Diagrams are not drawn to scale.

Diversity in composition of the kinetochore complex

To achieve functional goals, the kinetochore, comprising over 100 proteins with around 30 core structural components, exhibits a highly conserved basic organization in eukaryotes (Sridhar & Fukagawa, 2022). Early studies of kinetochore composition performed in yeast and vertebrate cell lines revealed a high degree of similarity between yeast and vertebrate kinetochores, especially the inner complex, despite millions of years of divergence between the two lineages (Drinnenberg et al., 2016). However, evolutionary analysis also indicates significant divergence in kinetochores across lineages since the last eukaryotic common ancestor (LECA), with variations in protein loss, duplication, invention, and sequence diversification (van Hooff et al., 2017). The structural components of the kinetochore can broadly be divided into two parts: the inner kinetochore, which interacts with centromeric chromatin and is assembled throughout the cell cycle, and the outer kinetochore, which is responsible for microtubule binding and is assembled only in mitosis. The outer kinetochore includes the KMN (Knl1, Mis12, and Ndc80 complexes) network and the Ska/Dam1 complex (Figure 3). While outer kinetochore components are well-conserved across eukaryotic evolution, the inner kinetochore, serving as the recruitment platform for the kinetochore complex, displays greater diversity (Sridhar & Fukagawa, 2022). The inner kinetochore consists of the CENP-A and the constitutive centromere associated network (CCAN). In the budding yeast system, the CBF3 complex is also included (Figure 3). The plasticity of the inner kinetochore composition is multifaceted, ranging from the absence of CENP-M in fungi and the presence of novel components, such as subcomplexes in the CBF3 complex, in species with point centromeres (Drinnenberg et al., 2016) (Figure 3).

Despite compositional diversity and rapid protein evolution, the kinetochore continues to assemble on centromeric DNA and maintains its conserved function in mediating chromosome segregation across eukaryotes (Drinnenberg et al., 2016).

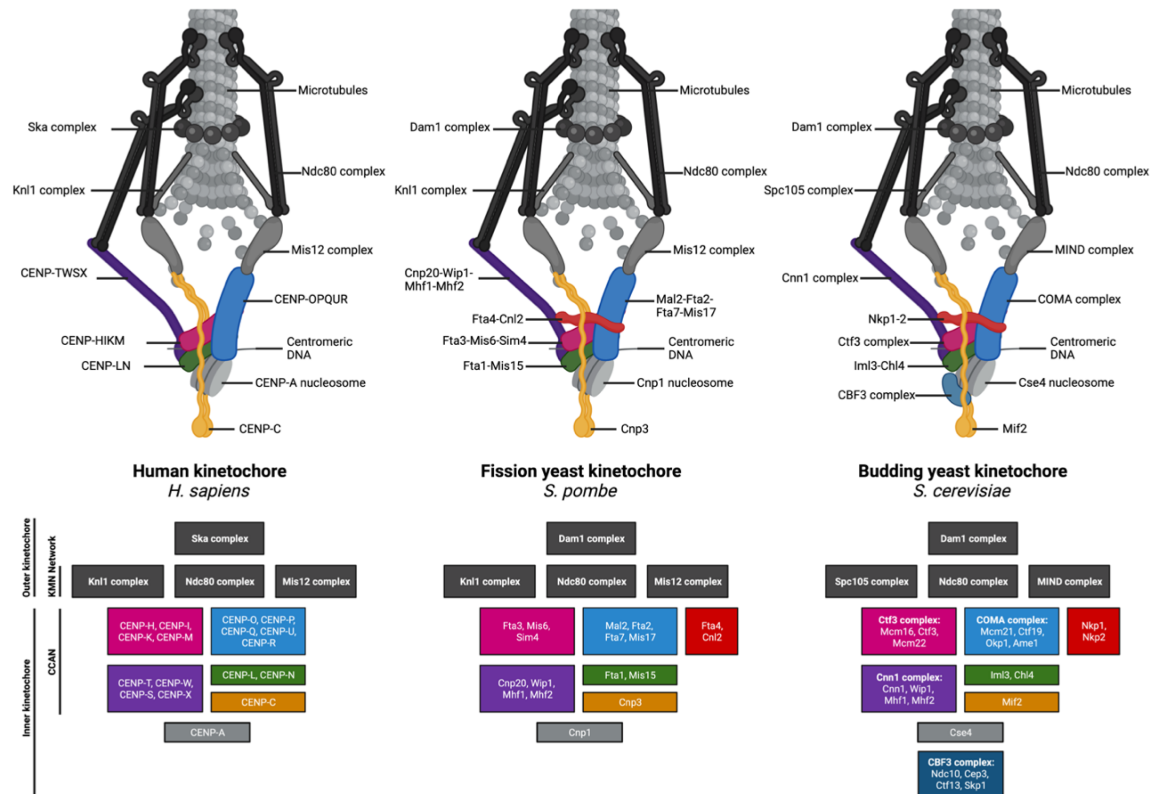

Figure 3. Kinetochore composition among different eukaryotes.

The kinetochore consists of both inner and outer components. The inner kinetochore includes the constitutive centromere-associated network (CCAN), responsible for recruiting the outer kinetochore KMN network. While the kinetochore's fundamental function is conserved across eukaryotes to ensure accurate chromosome segregation, the composition varies across different systems, particularly in the inner kinetochore elements. Homologous complexes between human, fission yeast, and budding yeast share the same color codes. Kinetochore homologs are mentioned in the corresponding positions. Created with BioRender.com.

Kinetochore recruitment in different centromere types

For the point centromeres of *S. cerevisiae*, a defined centromeric DNA sequence is both necessary and sufficient for centromere function and kinetochore assembly (McAinsh & Marston, 2022). The point centromere's CDEIII is the binding site for the CBF3 complex, which is required for the initiation of kinetochore assembly (Espelin et al., 2003). All four of the CBF3 protein subunits are necessary for DNA binding and for cell viability (Espelin et al., 2003). Three CBF3 proteins, Ndc10, Cep3, and Ctf13 are in direct contact with DNA, whereas Skp1 mediates the phosphorylation-dependent activation of Ctf13 (Espelin et al., 2003). Additionally, the CBF3 complex is important for localization of the centromere-specific histone Cse4 (CENP-A) localization. With CBF3 bound to CDEIII, the histone chaperone Scm3 binds to Ndc10 and mediates the deposition of Cse4 at the centromere (McAinsh & Marston, 2022).

On the other hand, regional centromeres lack defined centromeric DNA sequences and consist of repetitive DNA elements, such as tandem repeats and retrotransposons (Guin et al., 2020). These centromeres are epigenetically defined by the presence of the centromere-specific histone CENP-A (Cse4) (McAinsh & Marston, 2022). At regional centromeres, existing CENP-A promotes the assembly of new CENP-A through an epigenetic loop (McAinsh & Marston, 2022). The structural component CENP-C directly binds to the CENP-A nucleosome, recruiting the Mis18 complex, which in turn binds the HJURP (Scm3) chaperone to promote CENP-A deposition (McAinsh & Marston, 2022). Despite the intrinsic differences between the epigenetically maintained regional centromeres and sequence-specified point centromeres, both are universally defined by the presence of a conserved histone H3 variant Cse4/CENP-A. Kinetochore

position is typically determined by Cse4/CENP-A-containing nucleosomes, upon which CCAN components assemble. CCAN recruits outer kinetochore components including Ndc80 complexes that bind microtubules (Sridhar & Fukagawa, 2022).

Inner kinetochore compositions in budding yeasts

The assembly of the inner kinetochore onto centromeres highlights the significance of understanding how the centromere and inner kinetochore coordinate to ensure accurate chromosome segregation. The budding yeast (Saccharomycotina) subphylum, a genetically diverse group comparable to plant and animal lineages (Shen et al., 2018) (Figure 4), is a valuable platform to study the interactions between centromeres and the inner kinetochore. This subphylum includes not only the widely studied baker's yeast *S. cerevisiae* but also the common human commensal and opportunistic pathogen *Candida albicans*, along with over 1,000 other known species, with ongoing discoveries (Shen et al., 2018). Within budding yeasts, there has been an evolutionary transition in centromeres from a regional type to point centromeres, with the postulated single origin of the point centromere at the divergence of *S. cerevisiae*'s clade, Saccharomycetaceae (Guin et al., 2020) (Figure 4). Beyond this clade, earlier diverged budding yeasts are anticipated to possess regional centromeres, as exemplified by the confirmed short regional centromeres identified in *C. albicans* in the CUG-SER1 clade (Guin et al., 2020). While *S. cerevisiae*'s kinetochore composition is widely employed as a model for centromere and inner kinetochore studies, there exists a large group of yeasts whose diversity of inner kinetochore compositions remains insufficiently explored.

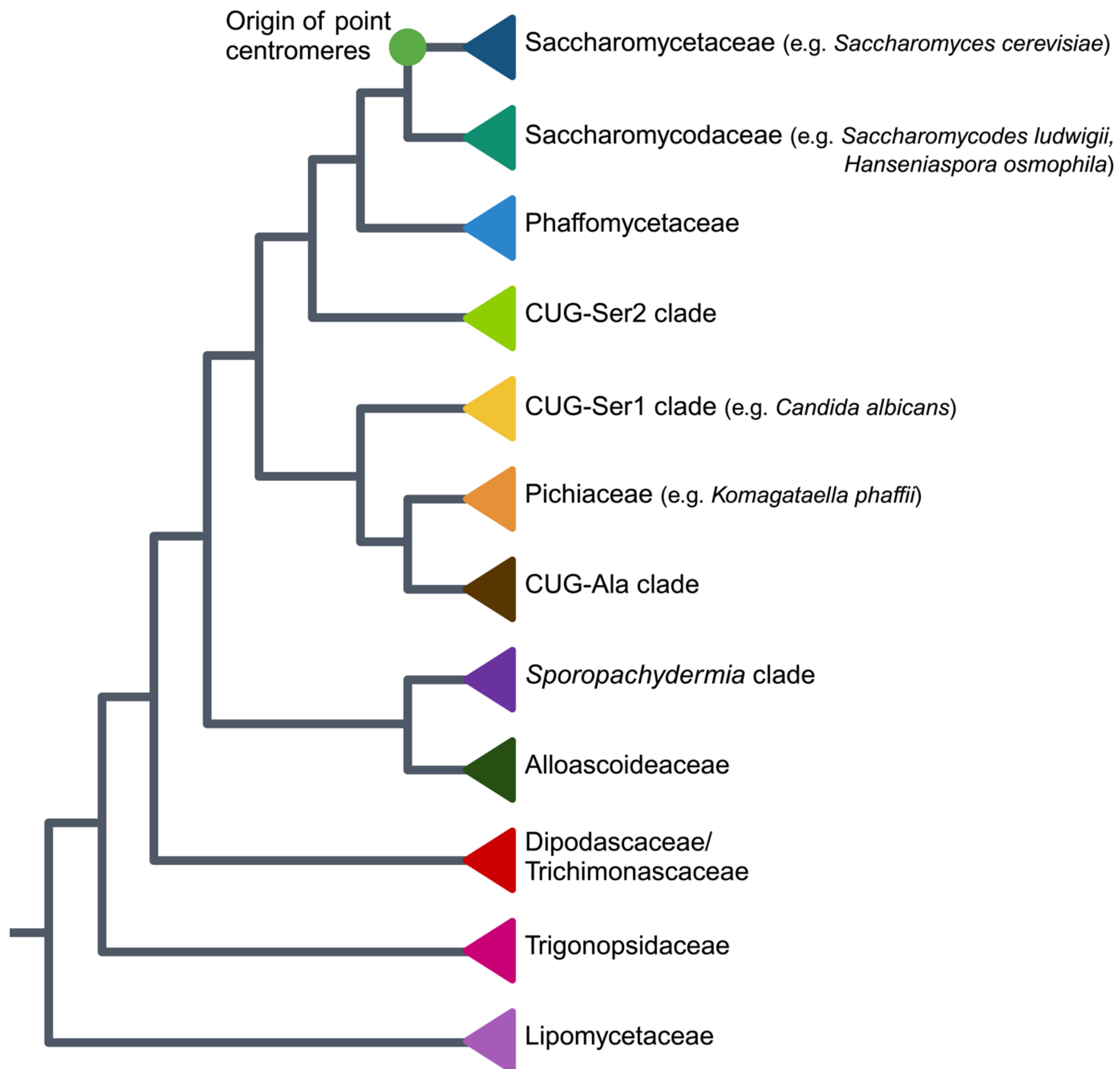

Figure 4. Budding yeast phylogenetic tree.

Phylogenetic tree showing the divergence of 12 budding yeast families/clades. Common model budding yeasts *Saccharomyces cerevisiae*, *Candida albicans*, and *Komagataella phaffii* are shown next to their respective family. Saccharomycodaceae species *Saccharomycodes ludwigii* and *Hanseniapsora osmophila* are shown. The green circle shows the currently postulated origin of point centromeres at the divergence of the Saccharomycetaceae family. Phylogenetic tree adapted from Shen et al.

Our study focuses on this diverse group of budding yeasts. Many yeast kinetochore proteins have orthologs in animal cells, indicating conservation of essential kinetochore structure elements throughout evolution (Espelin et al., 2003). Given diverse centromeres found in budding yeasts and the assembly of the inner kinetochore on these varied centromeres, budding yeasts offer an opportunity to investigate potential relationships between centromere type and inner kinetochore composition at the species level. Here, we compiled the inventory of 20 inner kinetochores in 338 diverse budding yeast genomes. Using the inventory, we explored the co-evolutionary dynamics between the inner kinetochores and the centromeres, particularly whether the transition from regional to point centromeres correlates with distinct inner kinetochore compositions. Our findings revealed no clear patterns between centromere types and inner kinetochore compositions in budding yeasts. However, we observed an interesting pattern in the point-centromere-associated complex CBF3 within the Saccharomycodaceae family. Upon further analyses, we hypothesize that there are five species in the Saccharomycodaceae family possessing point centromeres, suggesting a potential earlier origin of point centromeres in budding yeasts. Additionally, we unexpectedly detected the presence of the CBF3 proteins in species without point centromeres, challenging traditional associations with sequence-defined point centromeres.

Methods

mign

To streamline the data mining process and facilitate the analysis of biological sequences, mign was developed and written in Python (version 3.10.11). The workflow of mign is illustrated in Figure 5. An amino acid or nucleotide query sequence and taxonomic identifications (taxIDs) are provided by the user. Using the provided taxIDs, NCBI Datasets and Dataformat (version 15.10.0) (Sayers et al., 2022) are used to gather genome and proteome data for all species associated with the provided taxIDs in the NCBI database, creating a species search set. Only species with reference genomes are included, while hybrid species are excluded from the species search set. Two separate databases are constructed: one containing the genome of all included species, and another containing the protein sequences of these species. It should be noted that some species do not have annotated protein sequences in the NCBI database, and consequently, they are not included the protein sequence database.

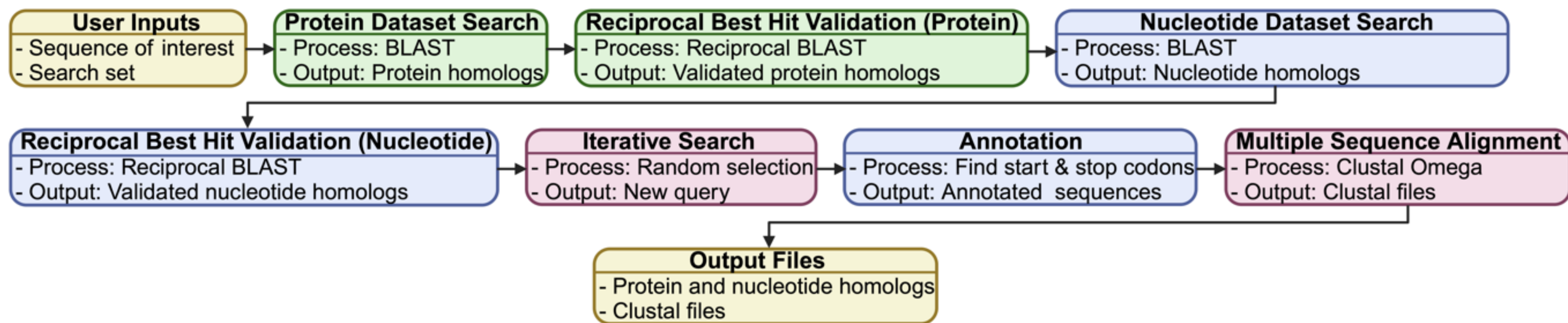

Figure 5. Workflow of mign, a tool to automate the identification of homologs.

Diagram of the step-by-step workflow of the tool mign, designed for streamlining the data mining process and facilitating the analysis of biological sequences. Each step is color-coded: yellow for inputs and outputs, green for protein sequence processing, blue for nucleotide sequence processing, and red for protein and nucleotide sequence processing.

To identify homologous sequences related to the query, BLAST+ (version 2.14.0) (Altschul et al., 1990) was used to perform BLAST searches to carry out an iterative search process. Starting with provided query sequence, it is used to search for homologous proteins in the protein database using BLASTP. To validate the results, reciprocal best hit searches are conducted against the protein dataset of the query sequence's species, considering only sequences where the query is the top hit in the reciprocal BLASTP results. Subsequently, the genome database is updated to exclude species with validated homologous proteins. For species lacking annotated genomes or those without hits from the protein search, TBLASTN is performed using the query against the updated genome database. Validation is again executed through reciprocal best hit searches against the protein dataset of the query sequence's species, considering only sequences where the query is the top hit in the reciprocal BLASTX results. To further the comprehensiveness of the search, iterative searches are performed. This involved selecting a random species from the protein hit results, with the condition that the chosen species must belong to a different family rank than all previous species that were used as query. ETE Toolkits (ete3 version 3.1.1) (Huerta-Cepas et al., 2016) is used to sort species in the protein hit list into their respective family rank. The homologous protein from the selected species then served as a new query, initiating further searches for additional homologous protein and nucleotide sequences. This iterative process continues until no new sequences are discovered, there are no new species available from other family ranks to select from, or there are no species in the protein search results to select from.

To extend the length of nucleotide sequence hits obtained during homolog detection, an annotation process is implemented to identify the stop and start codons within the nucleotide homologs. For sequences presenting multiple non-overlapping hits on the same scaffold, manual annotation is required, with an output file reporting the sequences. Automated annotation is conducted for the remaining nucleotide sequence hits. To identify start and stop codons, their proximity to both sequence termini and scaffold boundaries are considered. The stop codon is the nearest stop in the 3' direction of the sequence. For start codon, the search starts with finding the nearest stop codon in the 5' direction of the sequence, progressing towards the 3' direction until encountering the first start codon. In cases where a stop or a start codon is not found, the scaffold's extremities are designated as the sequence's start and end points. Subsequently, annotated nucleotide sequences are translated into their corresponding amino acid counterparts for further analysis.

To explore the evolutionary relationships of the found homolog sequences, Clustal Omega (version 1.2.3) (Sievers et al., 2011) is used to perform multiple sequence alignment. The protein hits and the translated nucleotide hits are included in this analysis.

The source code is available at Github at https://github.com/maitiennguyen/mign.

Detecting homologs of kinetochore proteins in budding yeasts

To investigate the distribution of kinetochore proteins among budding yeasts, *S. cerevisiae*'s homologs were used as the initial query. For any given protein, if homologs were not easily detected, the code was ran again using *Schizosaccharomyces pombe*'s

homolog as the initial query. The accession numbers corresponding to the specific protein homolog sequences used here can be found in Table 1. The results from both runs were combined into a single set of homologs. 560 budding yeasts were included in the search set, but only 338 species are reported here due to the consideration of known phylogenetic relationships. Interactive Tree Of Life (iTOL version 6) (Letunic & Bork, 2021) was used to create figures depicting the phylogenetic tree of budding yeasts.

Table 1: Initial Query Protein Accession Numbers

	S. cerevisiae	*Sch. pombe*
Ndc10	NP_011656.3	-
Cep3	NP_013891.1	-
Ctf13	NP_013812.1	-
Skp1	AJV18975.1	-
Mif2	NP_012834.1	-
Iml3/Fta1	NP_009665.3	NP_594755.1
Chl4	NP_010540.3	-
Cnn1/Cnp20	NP_116704.1	NP_595115.1
Wip1/New1	NP_001032576.1	NP_001343107.1
Mhf1	NP_076910.1	-
Mhf2	NP_878060.1	-
Mcm16	NP_015371.1	-
Ctf3	NP_013485.1	-
Mcm22	NP_012669.1	-
Mcm21	NP_010604.4	-
Ctf19	NP_015307.1	-
Okp1/Fta7	NP_011695.1	NP_587733.1
Ame1/Mis17	NP_009770.3	NP_001018835.1
Nkp1/Fta4	NP_010671.4	NP_587962.1
Nkp2/Cnl2	NP_013419.1	NP_593395.1

Comparing DNA binding motifs in CBF3 proteins

To determine the consensus DNA binding domain features of CBF3 proteins in species with point centromeres, MEME Suite (version 5.5.3) (Bailey et al., 2009) was used to perform motif analyses focusing on the DNA binding domain regions of homologs within Saccharomycetaceae. Notably, Skp1 was excluded from this analysis because it does not directly interact with centromere DNA. For each protein within the CBF3 complex, *S. cerevisiae*'s sequence was used as a reference in InterPro (version 94.0) (Paysan-Lafosse et al., 2023) to identify their respective DNA binding domains. Subsequently, the alignment of all homologs in Saccharomycetaceae was used to guide the trimming of the sequences to focus on the DNA binding region. These trimmed sequences were compiled into a single file, and MEME analysis was performed to identify motifs of the CBF3 proteins in species with point centromeres. The number of motifs was increased in each MEME run until most, if not all, key residues are included.

To compare the DNA binding domains of CBF3 protein sequences in budding yeasts with point centromeres to those with other centromere types, the homologs of each protein were scanned for the identified motifs within Saccharomycetaceae. MAST analyses using MEME Suite were performed to scan for the identified motifs in homologs of other budding yeasts. The data collection included the presence of these motifs and the $-\log_{10}$ of their corresponding p-values.

Mapping point centromeres in *H. osmophila*

To determine the locations of centromeres in *H. osmophila*, gene synteny and DNA motif analyses were performed. This involved referencing the centromere locations

in *Saccharomycodes ludwigii* (Papaioannou et al., 2021) and the ancestral centromere locations within Yeast Gene Order Browser (YGOB version 7) (Byrne & Wolfe, 2005) to identify the genes flanking point centromeres. For each set of flanking genes, BLASTP was performed against *H. osmophila* protein dataset to pinpoint the flanking genes necessary for locating candidate centromere sequences. In instances where *Sd. ludwigii* and Ancestral centromere locations are not syntenic, all combinations of flanking genes were examined and the shortest intergenic sequence as the candidate was selected.

To look for the consensus features of point centromeres *in H. osmophila*, MEME and MAST analyses were carried out using MEME Suite. Initially, candidate centromere sequences that do not contain genes were compiled and a MEME analysis was performed. Upon identifying a motif similar to those found in Saccharomycetaceae centromeres, it was used as a guide for MAST motif scanning and trimming other candidate sequences that contained genes. These sequences were trimmed such that they would not contain any genes. All candidate sequences were then combined with 178 point centromere sequences in Saccharomycetaceae. Another round of MEME analysis was performed, followed by a final trimming of the centromere sequences, aligning them with the start positions of the CDEI and CDEIII motifs.

Results

No overall pattern between inner kinetochore compositions and centromere types

To explore the coevolutionary relationship between inner kinetochore composition and the transition of centromere types in budding yeasts, we conducted an inventory of inner kinetochore composition in 338 species. Our inventory revealed that while not all individual subunits of every inner kinetochore complex are consistently present, at least some members of each complex are found across all major budding yeast groups, regardless of point or regional centromeres (Figure 6). Given that all subunits of Mif2, Iml3-Chl4, Cnn1, Ctf3, COMA, and Nkp1-2 complexes have homologs in *Sch. pombe* (Schleiffer et al., 2012), their presence across the subphylum is expected. Even though centromere type transitioned from a long, epigenetically-defined to a short, sequence-defined structure, the presence or absence of individual inner kinetochore proteins is sporadic (Figure 6). Overall, we observed no clear pattern when it came to the presence of inner kinetochore complexes as centromere type transitions from regional to point centromeres.

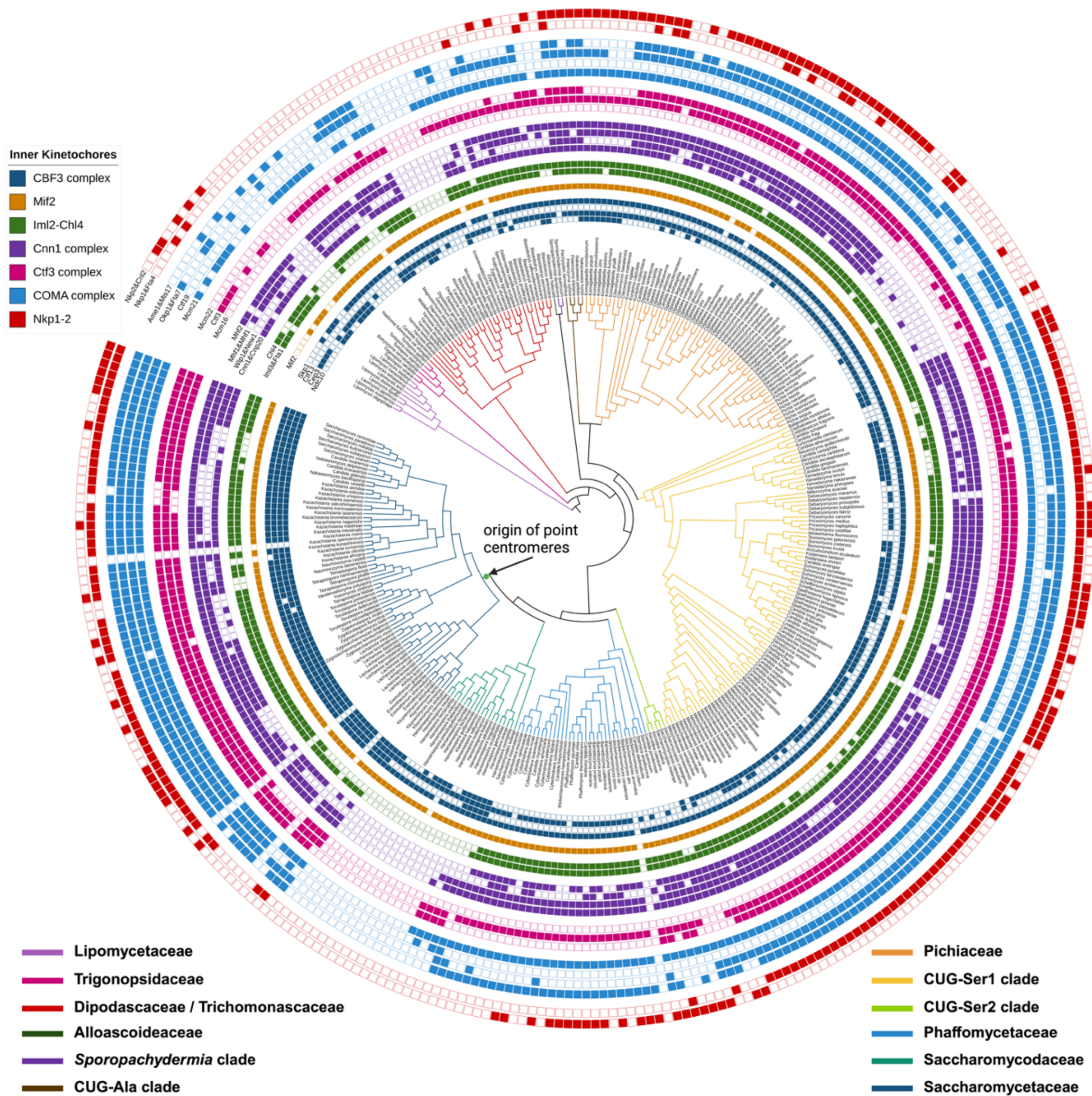

Figure 6. Phylogenetic distribution of 20 inner kinetochore proteins across 338 budding yeasts.

Presence and absence of 20 inner kinetochore proteins in 338 budding yeasts. Homologs were identified using mign. S. cerevisiae homolog for each protein served as the initial query. Proteins that additionally use Sch. pombe as the initial query are denoted by the inclusion of Sch. pombe's homolog name. The colored branches of the phylogenetic tree, adapted from Shen et al., denotes clades. Filled squares indicate presence, while non-filled squares indicate absence. Proteins within the same complex are grouped together and share square coloring. The point centromere is positioned at the divergence of Saccharomycetaceae. Created with iTOL (Letunic & Bork, 2021).

Interestingly, even though we used both *S. cerevisiae*'s and *Sch. pombe*'s homologs as the initial query sequence in out homology searches, the Ctf3 subunit Mcm16 can only be identified in species with point centromeres, the Saccharomycetaceae family. Similarly, the COMA subunit Ctf19 can only be identified in the Saccharomycetaceae family, in addition to three species in the Saccharomycodaceae family with unknown centromere type. Given that Mcm16 and Ctf9 have homologs in *Sch. pombe* (Schleiffer et al., 2012), their absence in the majority of budding yeasts is unexpected.

CBF3 DNA binding subunits found in species without point centromeres

Subunits of a kinetochore complex with highly similar phylogenetic profiles tend to co-occur across genomes, as they are expected to have evolved as a functional unit (van Hooff et al., 2017). The CBF3 complex is the most extensively studied kinetochore complex in *S. cerevisiae* and contains four proteins subunits: Ndc10, Cep3, Ctf13, and Skp1, all of which are necessary for DNA binding to point centromere and for cell viability (Espelin et al., 2003). Ndc10, Cep3, and Ctf13 bind to point centromeric DNA in a sequence-specific manner, as judged by DNA cross-linking in vitro (Espelin et al., 2003), suggesting potential co-evolution with point centromeres. Additionally, these three proteins have highly similar phylogenetic profiles, which mean they are likely to have co-evolved as a functional unit (van Hooff et al., 2017). Our inventory found that the full CBF3 complex, all four subunits, are typically found in Saccharomycetaceae species (Figure 6). This finding is consistent with the current understanding that the CBF3 complex is associated with point centromeres. However, we are unable to identify

the complete complex in two Saccharomycetaceae species: *Tetrapisispora iriomotensis* and *Yueomyces sinesis*. Examining individual CBF3 subunits revealed that Ndc10, Cep3, and Skp1 are present across all major budding yeast groups, while Ctf13 is typically found in Saccharomycetaceae. Skp1 is evolutionarily highly conserved and involved in essential cell cycle functions beyond the kinetochore (Connelly & Hieter, 1996). Therefore, its presence across the budding yeast subphylum is consistent with these functions (Figure 6). The presence of Ndc10 and Cep3, which bind to point centromeres in a sequence-specific manner, in species like *C. albicans* with regional centromeres is surprising (Figure 6). On the other hand, Ctf13 presence is isolated to Saccharomycetaceae, in addition to five species in Saccharomycodaceae with unknown centromere type, suggesting an expected correlation with point centromere (Figure 6).

Saccharomycodaceae inner kinetochore resembles point centromere species

The Saccharomycodaceae family stands as the closest evolutionary relative to the Saccharomycetaceae family (Figure 4, 6). While it is bioinformatically proposed that *Sd. ludwigii* in Saccharomycodaceae have point centromeres on all of its seven chromosomes (Papaioannou et al., 2021), the centromere type of other Saccharomycodaceae species remains unknown. As previously noted, all four subunits of the CBF3 complex are necessary for DNA binding to a point centromere. As a result, all four subunits tend to co-occur in species with point centromeres. This pattern of co-occurrence can also be found in five Saccharomycodaceae species: *Sd. ludwigii*, *H. osmophila*, *H. gamundiae, Hanseniaspora occidentalis*, and *Hanseniaspora vineae* (Figure 6). Therefore, besides *Sd. ludwigii,* which has been predicted to have point

centromeres using bioinformatics *(Papaioannou et al., 2021)*, we hypothesize that the other four species may also have point centromeres.

Interestingly, outside of the five Saccharomycodaceae species with all four CBF3 subunits, we found that complexes Iml3-Chl4, Cnn1, Ctf3, COMA, and Nkp1-2 are absent in other Saccharomycodaceae species (Figure 6). This subset of Saccharomycodaceae species missing the majority of the inner kinetochore proteins, including the CBF3 subunit Ctf13, shares a common ancestor (Figure 6). The shared ancestry suggests that if this subset of species truly do not have the inner kinetochore proteins, it likely occurred as a single loss event. Another potential explanation for the absence of these proteins is a functional transition in the inner kinetochore, where rapid DNA sequence divergence makes homolog detection challenging. This transition may have occurred at the divergence of this subset of species, given the shared ancestry.

CBF3 DNA binding motifs in point centromeres

Given that the CBF3 complex subunits Ndc10 (Cho & Harrison, 2011; Perriches & Singleton, 2012), Ctf13 (Yan et al., 2018), and Cep3 (Strunnikov et al., 1995) directly bind to centromeric DNA in point centromeres, yet are present in species without known point centromeres, we next examined the DNA binding domain motifs of these three proteins within Saccharomycetaceae. By performing MEME analysis (Bailey et al., 2009) on all identified sequences for each subunit within Saccharomycetaceae, we identified the DNA binding domain motifs for Ndc10, Ctf13, and Cep3.

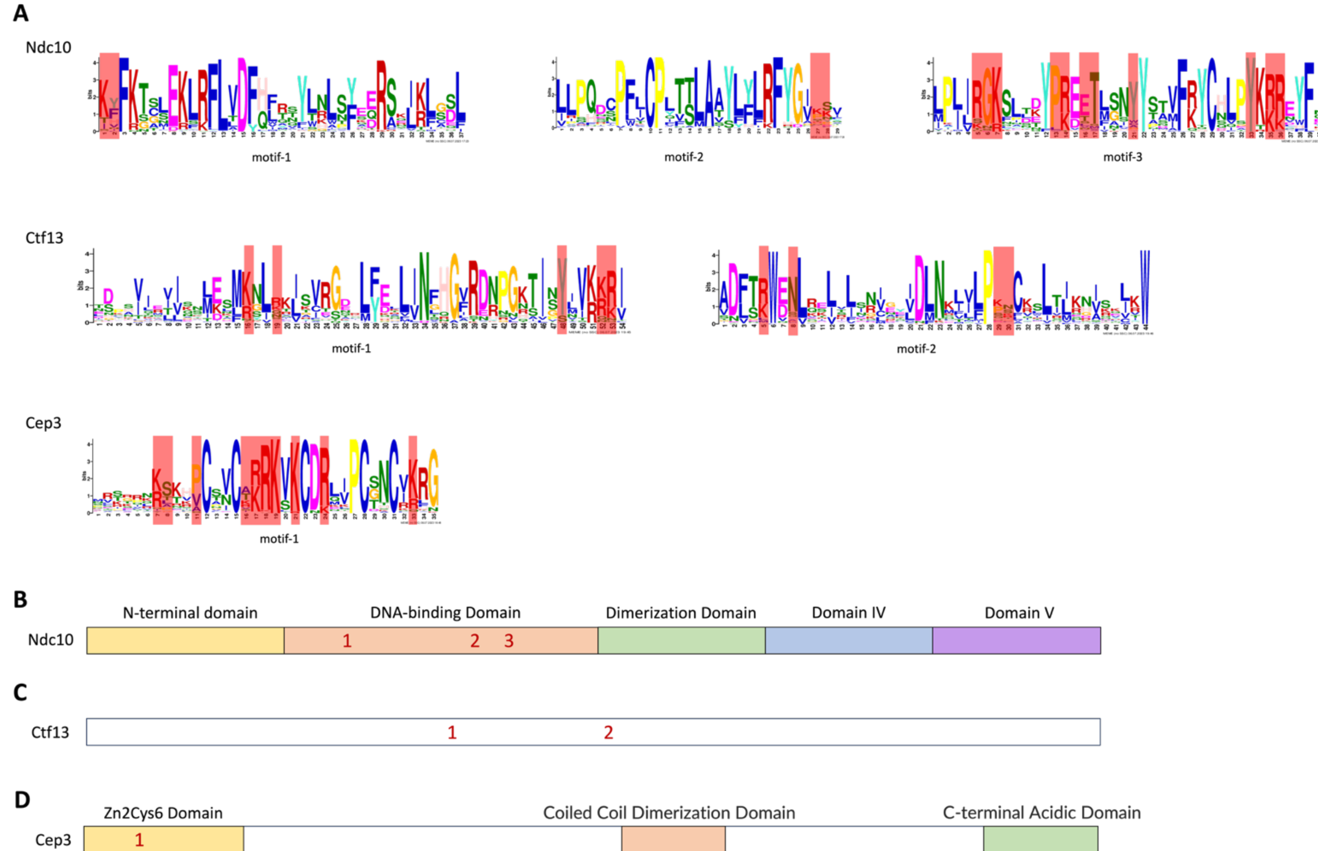

Figure 7. DNA binding motifs of Ndc10, Ctf13, and Cep3 in Saccharomycetaceae.

DNA binding domain motifs of CBF3 within the Saccharomycodaceae for three proteins: (A) Ndc10, Ctf13, and Cep3. Motifs of the proteins are identified using MEME analysis. Key DNA binding residues are highlighted in red. Key residues of the proteins are identified using S. cerevisiae sequences as reference. For Saccharomycetaceae, the number of motifs increases in each MEME run until most, if not all, key residues are included. Protein schematics of (B) Ndc10, (C) Ctf13, and (D) Cep3 are annotated with domain architectures and motif(s) location.

Ndc10 has three motifs representing its DNA binding domain (Figure 7 A, B), Ctf13 has two motifs (Figure 7 A, C), and Cep3 has one motif (Figure 7 A, D). In Ndc10 motifs 1 and 2, there are no distinctive conserved residues within Saccharomycetaceae at known DNA binding sites. On the other hand, Ndc10 motif 3 shows a high degree of conservation, particularly at the residue patterns RGK and YKRR, as well as residues P (proline), R (arginine), T (threonine), and Y (tyrosine). For Ctf13, the residue pattern KKR found in motif 1 is conserved within Saccharomycetaceae, along with residue K (lysine) at the first and residue Y (tyrosine) at the third known DNA binding site. However, no residues stand out as conserved in motif 2. In Cep3, the residue pattern K/RRKVK is conserved at known DNA binding sites within Saccharomycetaceae, along with residue R (arginine) at the ninth and K (lysine) at the tenth known DNA binding site. These conserved DNA binding sites likely play an important role in the DNA binding characteristics of Ndc10, Ctf13, and Cep3.

CBF3 DNA binding motifs in species without point centromeres

Given our identification of the CBF3 complex subunits Ndc10, Ctf13, and Cep3 in species that diverged before the proposed origin of point centromeres, we next investigated if the DNA binding domain motifs associated with point centromeres are present in species with unknown or regional centromere types. Using MAST analysis on the identified sequences of these subunits across the budding yeast subphylum, we search for motifs in the 338 species, along with their associated $-\log_{10}$(p-value) value (Figure 8). We interpret a motif presence as statistically significant if detected, with the highest p-value recorded at 1e-04 for all detections.

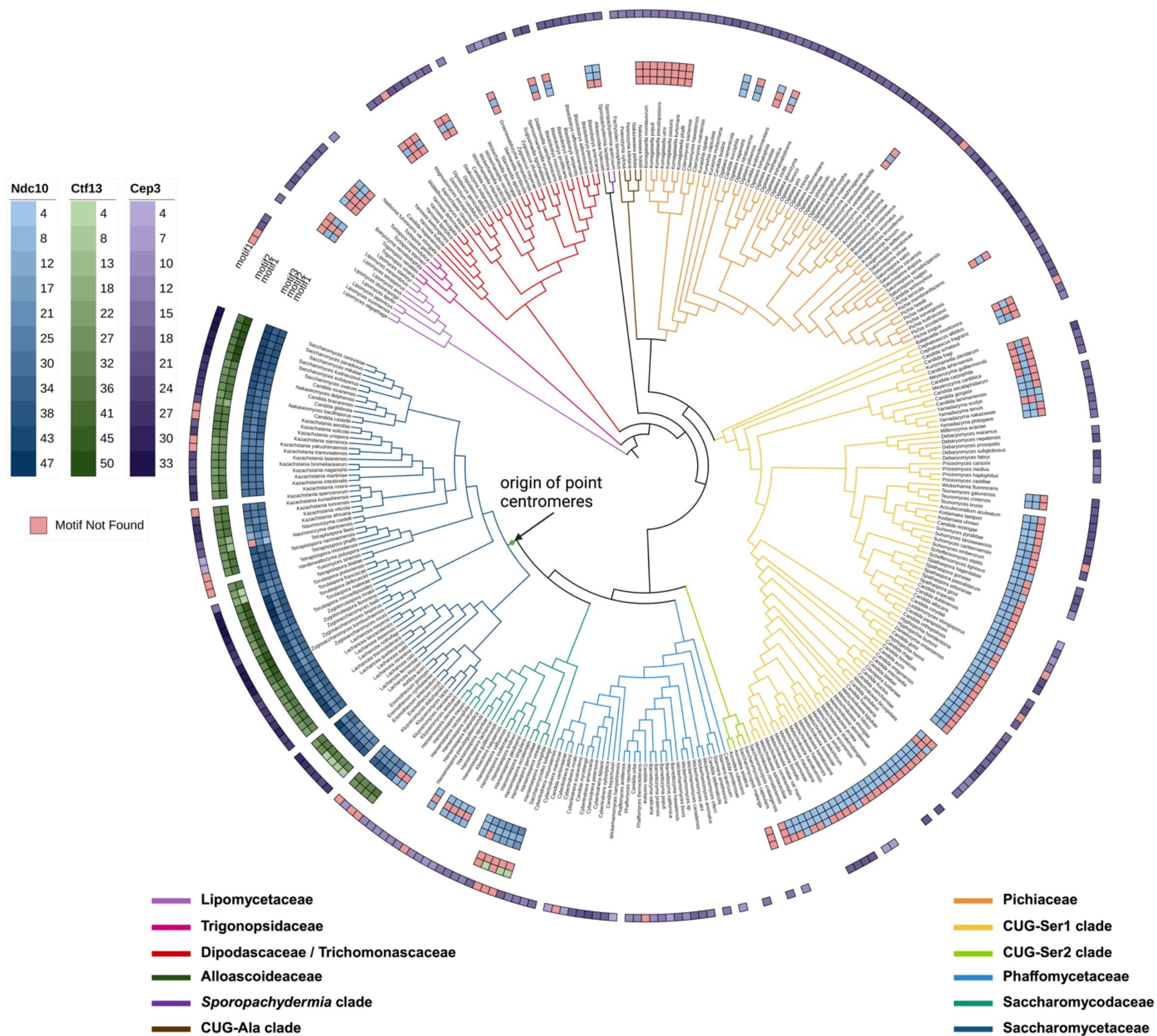

Figure 8. CBF3 DNA binding motifs associated with point centromeres across budding yeasts.

Presences and absences of DNA binding motifs associated with point centromere in 338 budding yeasts for Ndc10, Ctf13, and Cep3. The homologs of each protein in budding yeasts are scanned for the identified DNA binding motifs in the Saccharomycetaceae family using MAST analyses. The track indicates the $-\log_{10}$(p-value) for the found Saccharomycetaceae motifs in each species. Each protein is color-coded: blue for Ndc10, green for Ctf13, and purple for Cep3. Squares, colored according to the protein, indicate motif presence. Red indicates motif absence in the homolog. Absence of squares signifies no homolog found for the species. The colored branches of the phylogenetic tree, adapted from Shen et al., denotes clades. Created with iTOL (Letunic & Bork, 2021).

Cep3 motif 1 and Ndc10 motifs 1 and 2 are detectable across all major budding yeast groups (Figure 8). Cep3 motif 1 is consistently found across the subphylum, whereas Ndc10 motifs 1 and 2 are consistently present in the CUG-Ser1 clade, which has regional centromeres. In the case of Ctf13, its homologs can only be identified in Saccharomycetaceae and Saccharomycodaceae species (Figure 6), with Ctf13 motif 2 being detectable in both groups (Figure 8).

Naumovozyma dairenensis and *Naumovozyma castellii* were previously found to have nonconventional point centromeres with unique CDEs, differing from other species in Saccharomycetaceae (Kobayashi et al., 2015). A previous study has shown that the core DNA binding domain of their Ndc10 has undergone a more rapid change compared with other budding yeasts with standard centromeres in order to adapt to the new type of point centromere (Kobayashi et al., 2015). Our inability to detect Ndc10 motif 3 in *N. dairenensis* is consistent with this (Figure 8). Interestingly, motif 3 is successfully identified in *N. castellii*.

Despite successfully identifying protein homologs in certain species, there are cases where we could not detect DNA binding motifs in these homologs. For instance, Ndc10 motif 3 was not identified outside of Saccharomycetaceae and Saccharomycodaceae (Figure 8). This finding is consistent with the conservation of DNA binding residues in Ndc10 motif 3 in species with point centromeres (Figure 7 A). In the two groups where we identified Ctf13 homologs, Ctf13 motif 1 is exclusive to Saccharomycetaceae (Figure 8). Our inability to detect the DNA binding domain motifs in these instances suggests that these motifs may play a functional role in point centromeres.

CBF3 DNA binding motifs in Saccharomycodaceae

We observed that five species in Saccharomycodaceae have inner kinetochore patterns consistent with point centromeres. First, these species have homologs for all four CBF3 subunits (Figure 6), a characteristic that is typically found in species with point centromeres. Second, we observed the presence of at least one Saccharomycetaceae motif in these five species for the three CBF3 subunits known to directly bind to centromeric DNA in point centromeres: Ndc10, Ctf13, and Cep3 (Figure 8). As a result, we hypothesized that these species may have previously unidentified point centromeres. We therefore examined Saccharomycodaceae de novo motifs within the DNA binding domains of these three proteins and compared them to Saccharomycetaceae motifs. This comparison aims to examine whether there is potential conservation of protein function between the two groups. To achieve this, we conducted MEME analysis on identified sequences for each subunit within the family. Subsequently, TOMTOM analysis was carried out to compare these motifs with those identified in Saccharomycetaceae (Figure 9).

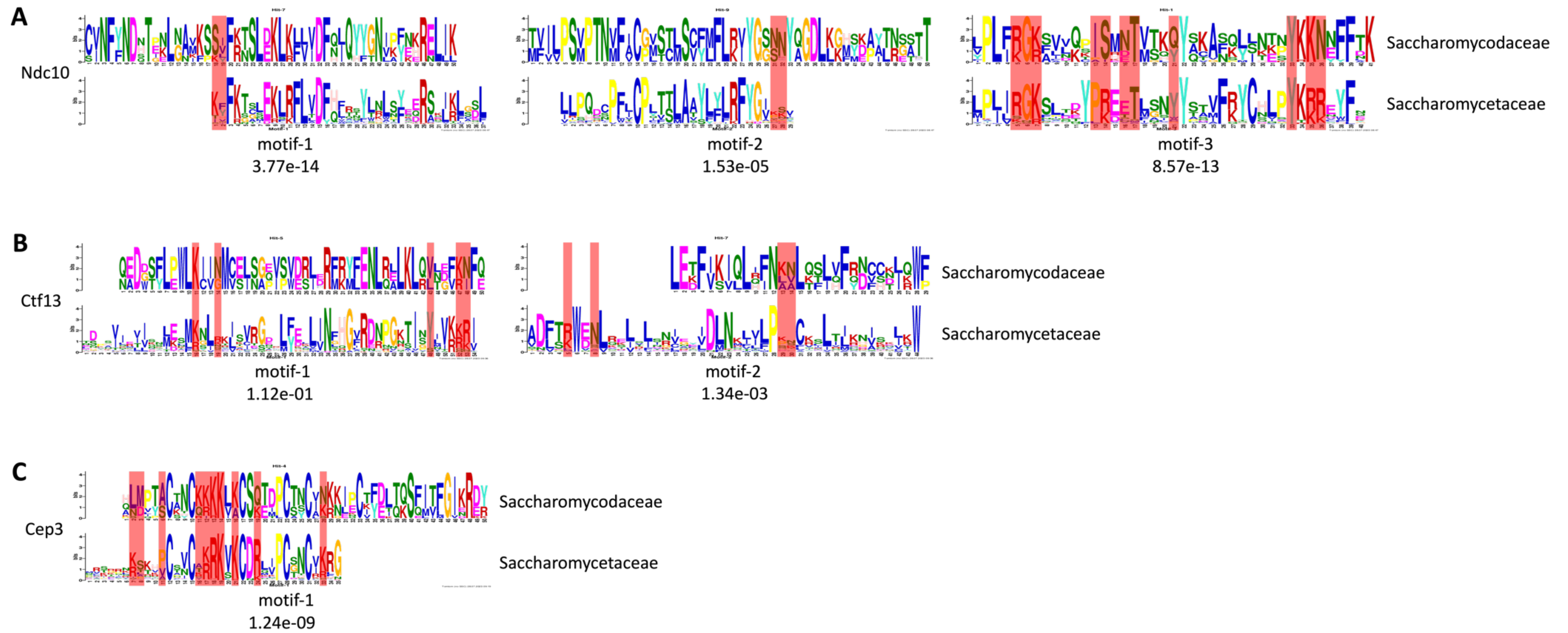

Figure 9. DNA binding motif alignments of Ndc10, Ctf13, and Cep3 between Saccharomycetaceae and Saccharomycodaceae.

Alignments of DNA binding domain motifs of CBF3 between Saccharomycodaceae and Saccharomycetaceae for three proteins: (A) Ndc10, (B) Ctf13, and (C) Cep3. Motifs of the proteins are identified using MEME analysis. For Saccharomycodaceae, 10 motifs were identified for each protein and aligned with Saccharomycetaceae using TOMTOM. Alignment pair with best p-value were considered. Key DNA binding residues are highlighted in red. Key residues of the proteins are identified using S. cerevisiae sequences as reference. The p-value of the alignment is below each motif.

The three DNA binding motifs identified in Ndc10 within Saccharomycodaceae show statistically significant similarity to those identified in Saccharomycetaceae. Their p-values are 3.77e-14, 1.53e-05, 8.57e-13 for motif 1, 2, and 3 respectively (Figure 9 A). For Cep3, the alignment of motif 1 statistically significant (p-value of 1.24e-09) between Saccharomycetaceae and Saccharomycodaceae (Figure 9 C). This shows that Ndc10 and Cep3 DNA binding region are conserved within the two families. In contrast, two of the DNA binding sites within Ctf13 motif 2 in Saccharomycetaceae cannot be aligned with the motif 2 found in Saccharomycodaceae (Figure 9 B). P-value for motifs 1 and 2 alignments (1.12e-01 and 1.34e-03, respectively) are relatively low when compared to the alignments for Ndc10 and Cep3 motifs. As a result, Ctf13's DNA binding domains are not very well conserved between Saccharomycetaceae and Saccharomycodaceae.

In Ndc10 motif 3, DNA binding residue patterns RGK and YKRR found in Saccharomycetaceae are also present in Saccharomycodaceae, although in Saccharomycodaceae, YKRR is replaced by YKKN (Figure 9 A). Additionally, other DNA binding sites that are well conserved in motif 3 within Saccharomycetaceae, as identified previously, are similarly conserved in Saccharomycodaceae. At DNA binding sites in Cep3 motif 1, the residue pattern K/RRKVK observed in Saccharomycetaceae is similarly present in Saccharomycodaceae (Figure 9 C). However, V (valine) is substituted with L (leucine) instead. For Ctf13, in both Saccharomycetaceae and Saccharomycodaceae, residue K (lysine) is conserved at the first DNA binding site in motif 1 (Figure 9 B). The residue pattern KKR within Saccharomycetaceae Ctf13 motif 1 is replaced with FKN in Saccharomycodaceae. At DNA binding positions three and four in Ctf13 motif 2, residues K (lysine) and N (asparagine) are found in both groups.

Overall, these DNA binding sites are worth noting for the functional analysis of Ndc10, Ctf13, and Cep3 in Saccharomycodaceae.

Identifying point centromeres *in H. osmophila*

A previous study has shown bioinformatic evidence of point centromere motifs found in *Sd. ludwigii*, a Saccharomycodaceae species (Papaioannou et al., 2021). Based on our inner kinetochore inventory, we further hypothesize that in addition *Sd. ludwigii*, four more Saccharomycodaceae species may also have point centromeres: *H. osmophila*, *H. gamundiae, Hanseniaspora occidentalis*, and *Hanseniaspora vineae*. Therefore, we were interested in identifying de novo point centromeres in a Saccharomycodaceae species.

		CDEI	CDEII	CDEIII
Candidate Centromere 1	(+) 7403235..7403399	...ATCACGTG...	132 bp (50.1% AT)	...GTTTTAAAAATCCGAAAAGTTAATG...
Candidate Centromere 2	(−) 4274366..4274522	...GTCACATG...	124 bp (67.5% AT)	...GTTTTAAAAGTCCGAAAATATTTTT...
Candidate Centromere 3	(−) 2954926..2955094	...AGCACGTG...	136 bp (56.8% AT)	...GTTTATGAAGCCCGAAATACTTTTT...
Candidate Centromere 4	(+) 6352054..6352126	...GGCATCTG...	40 bp (41.1% AT)	...GATGCAAGAGTTCGAATCTCTTAGC...
Candidate Centromere 5	(−) 1882949..1883123	...GTCACGTG...	142 bp (41.7% AT)	...GCTCTGAAACCCCCAACTGAACTTG...

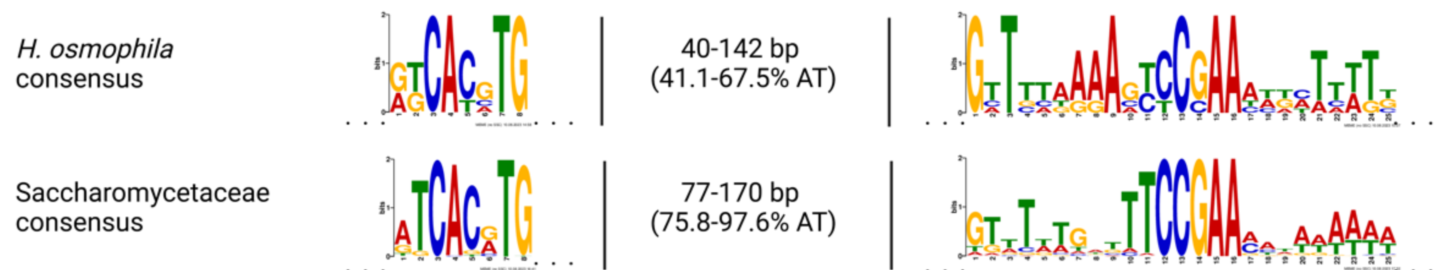

Figure 10. Features of predicted point centromeres in *H. osmophila*.

Sequences of the centromere DNA elements (CDEs) I and III, along with length and AT content of CDEII of each predicted centromere. The orientation of each centromere relative to the genomic scaffold (+/-) is indicated to the left of the sequence. The consensus DNA motifs of CDEI and CDEIII, as well as the ranges of length and AT content for CDEII in all sequences, are shown at the bottom. These are compared to the corresponding CDEI and CDEIII motifs and length and AT content ranges of CDEII in centromere sequences in the Saccharomycetaceae family.

Similar to *Sd. ludwigii, H. osmophila* is one of the five Saccharomycodaceae species that has all four subunits in the CBF3. In particular, its assembled genome comes with comprehensive gene annotations (Seixas et al., 2017). Therefore, with the hypothesis that *H. osmophila* has point centromeres, we performed gene synteny and DNA motif analyses to determine the locations of five possible centromeres in *H. osmophili*a (Esteve-Zarzoso et al., 2001). The hypothesized CDEI and CDEIII consensus in *H. osmophila* show significant similarity to those in Saccharomycetaceae, sharing highly conserved consensus TCACGTG for CDEI and TCCGAA for CDEIII (Figure 10). However, the hypothesized CDEII in *H. osmophila* have lower AT content and shorter in length compared to CDEII in Saccharomycetaceae. The CDEII AT content in Saccharomycetaceae ranges from 75.8% to 97.6%, whereas the CDEII AT content in *H. osmophila* only ranges from 41.1% to 67.5% (Figure 10). The length of CDEII ranges from 77 to 170 bp in Saccharomycetaceae, but only 40 to 142 bp in *H. osmophila*. It is important to note that the directionality of CDEIII is upstream of CDEII in all five of *H. osmophila* centromeres, while CDEI is upstream of CDEII in Saccharomycetaceae. Despite the similarity of the CDEI and CDEIII consensus to those in Saccharomycetaceae, the proposed point centromeres in *H. osmophila* require further wet lab confirmation.

Discussion

Despite the diversity in inner kinetochore composition and its interaction with equally diverse centromeric DNA (Drinnenberg et al., 2016; Guin et al., 2020; Sridhar & Fukagawa, 2022; van Hooff et al., 2017), there remains a gap in our understanding of how inner kinetochore composition changes during the transition from epigenetically defined regional to sequence-defined point centromeres. In this work, we conducted an inventory of inner kinetochore proteins across different centromere types at the species level in budding yeasts to explore the co-evolutionary dynamics between centromere type and inner kinetochore composition. Our findings show that the presence and absence of the proteins are sporadic (Figure 6), with no clear inner kinetochore composition patterns emerging as centromere type transitions from regional to point centromeres. While the overall picture of inner kinetochore composition lacks a clear pattern, our inventory provides a foundation for future research interested in studying these proteins in budding yeasts.

While we carefully inventory the homologous groups of each of the inner kinetochore proteins, we must also acknowledge the potential for false positives and/or false negatives: instances of incorrectly assigned presences or absences. In particular, we are concerned about false absences where a protein homolog in a given species may have diverged too extensively to be recognized during homology searches, resulting in a false absence. This divergence could arise from a transition in protein function in that species, wherein the structure of the protein has changed. A false positive can occur when sequences share similar regions by chance rather than evolutionary relatedness. Our iterative homology search approach increases the risk of

false positives, as one false positive can lead to subsequent incorrect identification. Such false positive or negative could potentially disrupt any underlying patterns in inner kinetochore composition as centromere type transitions from regional to point centromeres. Therefore, further experiments will be needed to validate the proposed inventory of budding yeasts inner kinetochore.

From our inventory, we surprisingly detected the presence of the CBF3 complex outside of the Saccharomycetaceae family. Traditionally, the CBF3 complex is linked with sequence-defined point centromeres, where its subunits Ndc10, Ctf13, and Cep3 directly bind to centromeric DNA in a sequence-specific manner (Espelin et al., 2003; Sridhar & Fukagawa, 2022). However, we found these proteins in species lacking point centromeres (Figure 6). Specifically, Ndc10 and Cep3 are widely distributed and are found in species known to have regional centromeres, such as *Candida albicans* and *Komagataella phaffii* (Figure 6). Since regional centromeres function independently of DNA sequence (Guin et al., 2020), the presence of these DNA binding proteins in species with regional centromeres suggests alternative mechanisms through which the proteins contribute to centromere-kinetochore function in a less sequence-specific manner. Another possibility is that the CBF3 complex has essential non-centromeric roles, leading to its conservation across different centromere types. This latter hypothesis is supported by previous analyses of the CBF3 complex, which identified essential non-centromeric roles for the complex in *S. cerevisiae*: spindle regulation and cytokinesis (Bouck & Bloom, 2005). Specifically, Ndc10 is directly involved in maintaining spindle stability during anaphase and coordinates the completion of cell division after chromosome segregation (Bouck & Bloom, 2005). While CBF3 is known to

localize to the spindle midzone during anaphase (Bouck & Bloom, 2005), its DNA binding functions beyond the centromere remain unknown.

We further investigated whether the DNA binding domain motifs of Ndc10, Ctf13, and Cep3 in species with point centromeres could also be found in species without. Our analysis revealed that the DNA binding domain motifs of these proteins can sporadically be identified in species without point centromeres (Figure 8). Given that we found the centromere-binding motifs in species with regional centromeres, it is possible that the complex gained its sequence specificity role in chromosome segregation as centromere type transitioned from regional to point centromeres. It will be interesting to learn whether Ndc10 and Cep3 also have roles in chromosome segregation, spindle regulation, and cytokinesis in species with regional centromeres.

Currently, a single origin of point centromere is postulated to be at the divergence of the Saccharomycetaceae family (Guin et al., 2020), suggesting that only species within this family have point centromeres. However, recent bioinformatic analyses of *Saccharomycodes ludwigii*, a species within the Saccharomycodaceae family closely related to Saccharomycetaceae, suggests the presence of putative point centromeres (Papaioannou et al., 2021). If experimentally validated, this finding challenges the current understanding, indicating that point centromeres may exist beyond the Saccharomycetaceae family. Our investigation into the CBF3 complex supports this hypothesis. While the CBF3 complex is widespread throughout the budding yeast subphylum, we observed a unique characteristic typically associated with species with point centromeres: the presence of all four CBF3 subunits (Figure 6). This distinct co-occurrence pattern is also observed in five Saccharomycodaceae species: *Sd. ludwigii*,

H. osmophila, *H. gamundiae, H. occidentalis*, and *H. vineae* (Figure 6).). Only *Sd. ludwigii* centromeres has previously been examined. Therefore, the presence of this pattern suggests the potential presence of point centromeres in these five Saccharomycodaceae species. To explore this further, we conducted a de novo analysis of the DNA binding domain motifs in these five Saccharomycodaceae species, comparing them with motifs found in species with known point centromeres (Figure 7 A). We found residues of interest for the centromere-binding functional analysis of Ndc10, Ctf13, and Cep3 in Saccharomycodaceae. Our CBF3 inventory and DNA binding domain motifs analyses further support the presence of point centromeres in *Sd. ludwigii* and suggest that four additional Saccharomycodaceae species may also possess point centromeres.

Among the Saccharomycodaceae species, aside from the five we hypothesize to have point centromeres, the remaining species do not have the complete CBF3 complex (Figure 6). Interestingly, Ctf13 and complexes Iml3-Chl4, Cnn1, Ctf3, COMA, and Nkp1-2 are all absent in this subset of Saccharomycodaceae species (Figure 6). This widespread absence of most inner kinetochore proteins suggests potential false negatives in our homology search methodology, although this is inconsistent given our successful identification of the Mif2 protein and other CBF3 subunits in these species. To ensure we are not overlooking any potential homologs, we manually conducted homology searches using various query sequences from closely related species within both the Saccharomycetaceae and Phaffomycetaceae families. However, this manual search also failed to identify homologs for the subset of Saccharomycodaceae species. Given that they share a common ancestor (Figure 6), we suggest that there may be a

biological explanation for this widespread absence. As the distinct co-occurrence pattern of all four CBF3 subunits not found in this subset of species, it is likely that they either do not have point centromeres altogether or that the sequences of their point centromeres have diverged from those found in the other five species and in Saccharomycetaceae. Another possible explanation is that there was a single event where these inner kinetochore proteins were lost in this subset of Saccharomycodaceae species, and other proteins have assumed their functions within the inner kinetochore.

Given the availability of *H. osmophila*'s assembled genome and its comprehensive gene annotations, we identified potential point centromeres in this species. Our analysis found conserved CDEI and CDEIII consensus between *H. osmophila* and species with known point centromeres (Figure 10), suggesting the presence of point centromeres in *H. osmophila*. However, the AT content of CDEII in *H. osmophila* is lower than what is typically found in species with known point centromeres (Figure 10). Furthermore, the directionality of CDEIII is upstream of CDEII in all five of *H. osmophila*'s centromeres, whereas in Saccharomycetaceae, CDEI is upstream. Due to these differences and relying on bioinformatic data for conserved features, additional experimental validation is needed to confirm that *H. osmophila* has point centromeres. For example, conducting chromatin immunoprecipitation sequencing (ChIP-Seq) could help determine whether Cse4 is enriched at the loci corresponding to the proposed centromeres, which would confirm the identity of these regions as centromeres.

In summary, our study revealed no clear pattern in the inner kinetochore composition as budding yeasts transition from regional to point centromeres. Surprisingly, we detected the inner kinetochore centromere-binding complex CBF3 in

species lacking point centromeres. This suggests a broader functional role beyond chromosome segregation or a less centromeric DNA sequence-specific role, particularly in species with regional centromeres. Furthermore, our CBF3 inventory for Saccharomycodaceae supports the hypothesis that *Sd. ludwigii* have point centromeres, and we propose that four additional species in this family may share this feature. To further explore this hypothesis, we found potential point centromeres in *H. osmophila* through gene synteny and motif analysis. However, as our findings are based on bioinformatic data, additional experiments are needed to validate the functional implications of the inventory and the presence of point centromeres in *H. osmophila*. Our inventory provides a foundation for future research interested in studying the kinetochore proteins in budding yeasts and conducting functional analyses of these proteins.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, *37*(suppl_2), W202–W208. https://doi.org/10.1093/nar/gkp335

Bouck, D. C., & Bloom, K. S. (2005). The kinetochore protein Ndc10p is required for spindle stability and cytokinesis in yeast. *Proceedings of the National Academy of Sciences*, *102*(15), 5408–5413. https://doi.org/10.1073/pnas.0405925102

Byrne, K. P., & Wolfe, K. H. (2005). The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*, *15*(10), 1456–1461. https://doi.org/10.1101/gr.3672305

Cho, U.-S., & Harrison, S. C. (2011). Ndc10 is a platform for inner kinetochore assembly in budding yeast. *Nature Structural & Molecular Biology*, *19*(1), 48–55. https://doi.org/10.1038/nsmb.2178

Cleveland, D. W., Mao, Y., & Sullivan, K. F. (2003). Centromeres and Kinetochores: From Epigenetics to Mitotic Checkpoint Signaling. *Cell*, *112*(4), 407–421. https://doi.org/10.1016/S0092-8674(03)00115-6

Connelly, C., & Hieter, P. (1996). Budding Yeast SKP1 Encodes an Evolutionarily Conserved Kinetochore Protein Required for Cell Cycle Progression. *Cell*, *86*(2), 275–285. https://doi.org/10.1016/S0092-8674(00)80099-9

Drinnenberg, I. A., Henikoff, S., & Malik, H. S. (2016). Evolutionary turnover of kinetochore proteins: A ship of Theseus? *Trends in Cell Biology*, *26*(7), 498–510. https://doi.org/10.1016/j.tcb.2016.01.005

Espelin, C. W., Simons, K. T., Harrison, S. C., & Sorger, P. K. (2003). Binding of the Essential Saccharomyces cerevisiae Kinetochore Protein Ndc10p to CDEII. *Molecular Biology of the Cell*, *14*(11), 4557–4568. https://doi.org/10.1091/mbc.E02-08-0533

Esteve-Zarzoso, B., Peris-Torán, M. J., Ramón, D., & Querol, A. (2001). Molecular characterisation of Hanseniaspora species. *Antonie van Leeuwenhoek*, *80*(1), 85–92. https://doi.org/10.1023/A:1012268931569

Guin, K., Sreekumar, L., & Sanyal, K. (2020). Implications of the Evolutionary Trajectory of Centromeres in the Fungal Kingdom. *Annual Review of Microbiology*, *74*, 835–853. https://doi.org/10.1146/annurev-micro-011720-122512

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, *33*(6), 1635–1638. https://doi.org/10.1093/molbev/msw046

Kobayashi, N., Suzuki, Y., Schoenfeld, L. W., Müller, C. A., Nieduszynski, C., Wolfe, K. H., & Tanaka, T. U. (2015). Discovery of an Unconventional Centromere in Budding Yeast Redefines Evolution of Point Centromeres. *Current Biology*, *25*(15), 2026–2033. https://doi.org/10.1016/j.cub.2015.06.023

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Maiato, H., & Silva, S. (2023). Double-checking chromosome segregation. *The Journal of Cell Biology*, *222*(5), e202301106. https://doi.org/10.1083/jcb.202301106

McAinsh, A. D., & Marston, A. L. (2022). The four causes: The functional architecture of centromeres and kinetochores. *Annual Review of Genetics*, *56*, 279–314. https://doi.org/10.1146/annurev-genet-072820-034559

Papaioannou, I. A., Dutreux, F., Peltier, F. A., Maekawa, H., Delhomme, N., Bardhan, A., Friedrich, A., Schacherer, J., & Knop, M. (2021). Sex without crossing over in the yeast Saccharomycodes ludwigii. *Genome Biology*, *22*(1), 303. https://doi.org/10.1186/s13059-021-02521-w

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, *51*(D1), D418–D427. https://doi.org/10.1093/nar/gkac993

Perriches, T., & Singleton, M. R. (2012). Structure of Yeast Kinetochore Ndc10 DNA-binding Domain Reveals Unexpected Evolutionary Relationship to Tyrosine Recombinases*. *Journal of Biological Chemistry*, *287*(7), 5173–5179. https://doi.org/10.1074/jbc.C111.318501

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for

biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Schleiffer, A., Maier, M., Litos, G., Lampert, F., Hornung, P., Mechtler, K., & Westermann, S. (2012). CENP-T proteins are conserved centromere receptors of the Ndc80 complex. *Nature Cell Biology*, *14*(6), 604–613. https://doi.org/10.1038/ncb2493

Seixas, I., Barbosa, C., Salazar, S. B., Mendes-Faia, A., Wang, Y., Güldener, U., Mendes-Ferreira, A., & Mira, N. P. (2017). Genome Sequence of the Nonconventional Wine Yeast Hanseniaspora guilliermondii UTAD222. *Genome Announcements*, *5*(5), e01515-16. https://doi.org/10.1128/genomeA.01515-16

Shen, X.-X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A. B., Wisecaver, J. H., Wang, M., Doering, D. T., Boudouris, J. T., Schneider, R. M., Langdon, Q. K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R.-I., Čadež, N., Libkind, D., … Rokas, A. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, *175*(6), 1533-1545.e20. https://doi.org/10.1016/j.cell.2018.10.023

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539. https://doi.org/10.1038/msb.2011.75

Sridhar, S., & Fukagawa, T. (2022). Kinetochore Architecture Employs Diverse Linker Strategies Across Evolution. *Frontiers in Cell and Developmental Biology*, *10*. https://www.frontiersin.org/articles/10.3389/fcell.2022.862637

Strunnikov, A. V., Kingsbury, J., & Koshland, D. (1995). CEP3 encodes a centromere protein of Saccharomyces cerevisiae. *The Journal of Cell Biology*, *128*(5), 749–760. https://doi.org/10.1083/jcb.128.5.749

van Hooff, J. J., Tromer, E., van Wijk, L. M., Snel, B., & Kops, G. J. (2017). Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Reports*, *18*(9), 1559–1571. https://doi.org/10.15252/embr.201744102

Yan, K., Zhang, Z., Yang, J., McLaughlin, S. H., & Barford, D. (2018). Architecture of the CBF3–centromere complex of the budding yeast kinetochore. *Nature Structural & Molecular Biology*, *25*(12), 1103–1110. https://doi.org/10.1038/s41594-018-0154-1