Loss of inner kinetochore genes is associated with the transition to an unconventional point centromere in budding yeast

Nagarjun Vijay Corresp. 1

1 Computational Evolutionary Genomics Lab, Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal, Madhya Pradesh, India

Corresponding Author: Nagarjun Vijay
Email address: nagarjun@iiserb.ac.in

Background: The genomic sequences of centromeres as well as the set of proteins that recognize and interact with centromeres are known to quickly diverge between lineages potentially contributing to post-zygotic reproductive isolation. However, the actual sequence of events and processes involved in the divergence of the kinetochore machinery is not known. The patterns of gene loss that occur during evolution concomitant with phenotypic changes have been used to understand the timing and order of molecular changes.

Methods: We screened the high quality genomes of twenty budding yeast species for the presence of well-studied kinetochore genes. Based on the conserved gene order and complete genome assemblies we identified gene loss events. Subsequently, we searched the intronic regions to identify any un-annotated genes or gene remnants to obtain additional evidence of gene loss.

Results: Our analysis identified the loss of four genes of the inner kinetochore constitutive centromere-associated network (CCAN) in both the Naumovozyma species for which genome assemblies are available. Surprisingly, this collective loss of four genes of the CCAN coincides with the origin of unconventional centromeres in N.*castellii* and N.*dairenensis*. Our study suggests a tentative link between the origin of unconventional point centromeres and turnover of kinetochore genes in budding yeast.

Loss of inner kinetochore genes is associated with the transition to an unconventional point centromere in budding yeast

Nagarjun Vijay[1]
[1]Computational Evolutionary Genomics Lab, Department of Biological Sciences, IISER Bhopal, Bhauri, Madhya Pradesh, India
*Correspondence: nagarjun@iiserb.ac.in

Keywords: point centromere, unconventional centromeres, kinetochore, interaction network, speciation

Abstract

Background: The genomic sequences of centromeres as well as the set of proteins that recognize and interact with centromeres are known to quickly diverge between lineages potentially contributing to post-zygotic reproductive isolation. However, the actual sequence of events and processes involved in the divergence of the kinetochore machinery is not known. The patterns of gene loss that occur during evolution concomitant with phenotypic changes have been used to understand the timing and order of molecular changes.

Methods: We screened the high quality genomes of twenty budding yeast species for the presence of well-studied kinetochore genes. Based on the conserved gene order and complete genome assemblies we identified gene loss events. Subsequently, we searched the intronic regions to identify any un-annotated genes or gene remnants to obtain additional evidence of gene loss.

Results: Our analysis identified the loss of four genes of the inner kinetochore constitutive centromere-associated network (CCAN) in both the Naumovozyma species for which genome assemblies are available. Surprisingly, this collective loss of four genes of the CCAN coincides with the origin of unconventional centromeres in N.*castellii* and N.*dairenensis*. Our study suggests a tentative link between the origin of unconventional point centromeres and turnover of kinetochore genes in budding yeast.

Introduction

Increasing availability of genomic datasets across diverse species has allowed the use of comparative genomic approaches to study kinetochore evolution. Such comparative approaches

have led to interesting evolutionary insights from species with unique kinetochores (Drinnenberg and Akiyoshi 2017). The kinetochore associated protein interaction network has undergone frequent changes in gene content and also exhibits strong signatures of adaptive evolution (Malik and Henikoff 2001; Talbert et al. 2004; Schueler et al. 2010). The frequent turnover in the gene content of the kinetochore machinery while essentially performing the same function has even been equated to a "Ship of Theseus" (Drinnenberg et al. 2016). Kinetochore genes that form protein complexes also show patterns of co-evolution with interacting components through correlated gene loss events (Tromer 2017). Large-scale systematic screening for kinetochore genes in the genomes of numerous eukaryotes has shown that majority of the components of the CCAN (constitutive centromere-associated network) are lost in many lineages (van Hooff et al. 2017). Another unexpected observation is that the CENH3 (CSE4 homolog) gene which performs a central role in kinetochore function is lost in many insect species. This potentially recurrent loss of the CENH3 gene in insect species coincides with their transition to holocentricity (Drinnenberg et al. 2014). Interestingly, it has been shown that C.*elegans* holocentromeres are actually point centromeres that are dispersed at transcription factor hotspots (Steiner and Henikoff 2014). Hence, it has been speculated that changes in the gene content of kinetochores might be functionally related to transitions in the centromere type (Drinnenberg and Akiyoshi 2017).

Centromere sequences in the vast majority of species are repeat rich regions that are thought to be defined epigenetically and/or through recognition of dyad rich regions with non-B-form DNA structures (Kasinathan and Henikoff 2018). Despite advances in genome sequencing and assembly methods, the high repeat content of these centromeres makes it harder to assemble and study them. Hence, as a proxy to the study of centromere regions comparison of tandem repeats across large phylogenetic distances have been performed to understand the evolution of centromere sequences (Melters et al. 2013).

One of the most interesting transitions during centromere evolution is the origin of genetically defined point centromeres in budding yeasts (Malik and Henikoff 2009). These ~150bp long centromeres found in Saccharomycetaceae have been easier to study due to the lack of repetitive regions, availability of complete high quality genome assemblies for multiple closely related

67 species, gene knockout libraries and ease of experimental manipulation. Hence, budding yeast
68 species are a popular system to study the evolution of centromeres and the set of proteins that
69 interact with them (Roy and Sanyal 2011).
70
71 In addition to changing their sequences, centromeres are also known to change their genomic
72 positions without any change in the karyotype (Montefalcone et al. 1999). However, centromere
73 evolution can also be accompanied by changes in the karyotype (O'Neill et al. 2004). Within
74 mammals, centromeres are known to have undergone multiple re-positioning events (Rocchi et
75 al. 2012). Centromere repositioning events seem to be fairly common and could have a role in
76 driving speciation or at the very least have a non-negligible role in affecting the local genomic
77 selection landscape. Recently, it has been shown that in Naumovozyma an unconventional
78 centromere has come into existence at a location that is largely distinct from that expected based
79 on synteny with other Saccharomycetaceae species (Kobayashi et al. 2015). The Naumovozyma
80 genus has been suggested as model for comparative genomics and study of adaptive evolution
81 due to the various phenotypic differences compared to other yeast species (Karademir Andersson
82 and Cohn 2017).
83
84 In this study, we first screened the genome assemblies of eight Pre-WGD and twelve Post-WGD
85 yeast species for the presence/absence of homologs of 67 kinetochore genes. We find evidence
86 for the concurrent loss of multiple genes from the CTF19 protein complex in Naumovozyma
87 species and corresponding sequence divergence of the N-terminal region of the CSE4 gene that
88 interacts with the CTF19 complex. In contrast to this we see high levels of sequence
89 conservation of the C-terminal region of the CSE4 gene that mediates an interaction between the
90 kinetochore and centromere. Our analysis finds an association between gene loss events and the
91 origin of novel centromeres in Naumovozyma species.
92
93 **Materials and Methods**
94
95 **Gene presence/absence screening**
96
97 We compiled a list of 67 kinetochore associated genes in the yeast S. *cerevisiae* by downloading
98 genes annotated with the GO term GO: 0000776 in Ensembl 91. Orthologs of these genes in all
99 twenty yeast species were identified from the Yeast Gene Order Browser (YGOB). We screened

the genomes of all twenty yeast species for the presence of all 67 genes. Kinetochore genes are known to evolve at a very fast rate, making it hard to identify orthologs of these genes even in closely related species (van Hooff et al. 2017). Despite our precautions, it is possible that these genes have evolved at a very fast rate making it unfeasible to establish homology of these genes. Fortunately, the YGOB provides not only the order of the genes but also the intergenic sequences between genes. Based on flanking genes with conserved synteny in other species we identified the intergenic regions that correspond to the location of the missing genes. We checked these intergenic sequences for the presence of open reading frames (ORFs). The ORFs that were found in the intergenic regions did not show any homology to the genes that we have inferred to be lost. Based on the evidence from identification of syntenic regions using YGOB and additional screening of intergenic regions we are fairly confident of these gene loss events.

Multiple sequence alignments of CSE4 gene

The complete open reading frame of the CSE4 gene was used for multiple sequence alignment. To ensure that the results we see is not the result of alignment artifacts we performed the multiple sequence alignment at the nucleotide sequence level and amino acid residue level using three different programs (Clustal omega, M-coffee and Guidance with PRANK as the aligner). The presence of additional sequences in the genus Naumovozyma was seen in all of the multiple sequence alignments that we have analysed. It is possible that longer ORF's have been incorrectly annotated as the CDS for both Naumovozyma species. However, the second methionine codon in the CDS occurs at the 226th and 40th residue from the currently annotated start codon in N. *castellii* & N. *dairenensis* respectively. If the second methionine is the correct start codon, the N. *castellii* protein would be just 39 residues. This suggests that the correct ORF is annotated in Naumovozyma species. To further rule out the possibility of erroneous annotation of shorter ORF's in non-naumovozyma species we took the flanking introns and searched them for ORF's. We found that the longest ORF that could be identified in the sequence spanning across both flanking introns was the currently annotated ORF itself. This further supports the validity of the annotation and multiple sequence alignments generated by us.

Results

Patterns of kinetochore gene loss

Our study system consists of eight Pre-WGD and twelve Post-WGD yeast species at varying evolutionary distances (see **Fig. 1**). Based on our screening of twenty yeast genomes we identified seven genes that are lost in Pre-WGD species and eight genes that have lost both copies in Post-WGD species (see **Table S1**). Intriguingly, five of the eight genes (see **Table S1**) that have lost both their copies in the Post-WGD species are also the ones that are lost in the Pre-WGD species. This hints at the dispensability of these genes over the course of evolution. The gene presence/absence matrix for six selected genes that have lost both gene copies in Post-WGD species is presented in **Fig. 2**. First, we focus on the set of four genes (NKP1, NKP2, IML3, and CHL4) that are lost in Naumovozyma. As the name suggests, the Non-essential Kinetochore Protein genes NKP1 & NKP2 produce proteins that localise to kinetochores and produce viable single gene mutants (Cheeseman et al. 2002). However, both single gene mutants are known to show elevated rates of chromosome loss (Fernius and Marston 2009). Despite being non-essential genes, at least one copy of both of these genes is found in 15 of the 20 yeast species screened in this study. A broader phylogenetic search for homologs of these genes has shown repeated loss of these two genes in various taxa (Tromer 2017).

The CHL4 (CHromosome Loss 4) single gene knockouts are also viable but are known to show increased levels of chromosome loss, miss-segregation and abnormal kinetochores (Roy et al. 1997). However, based on the chromatin state CHL4 mutant cells are known to show two distinct (high and low) levels of mitotic mobility (Roy and Sanyal 2011). This suggests that CHL4 mutants can be compensated through changes in the chromatin state. IML3 (Increased Minichromosome Loss) protein forms a stable heterodimer with CHL4 protein (Hinshaw and Harrison 2013). Although IML3 and CHL4 are known interactors, the phylogenetic distribution of IML3 seems to be more restricted than that of CHL4. The association of cohesin with the pericentromeric regions is ensured by the action of IML3 and CHL4. The lack of these two genes leads to reduced cohesin binding at the pericentromere that results in miss-segregation of chromosomes (Fernius and Marston 2009). While the gene loss events in Naumovozyma are very interesting, they are not the only kinetochore genes that we identify as missing both copies among the twenty species screened. Two other genes (CNN1 & WIP1) of the CCAN are lost in another yeast species T. *blattae* (see **Fig. 2**).

163

164 **N-terminal divergence of CSE4 gene in Naumovozyma**

165

166 Detailed study of the interactions between kinetochore proteins in S. *cerevisiae* has led to a better
167 understanding of their individual roles (Measday et al. 2005; Baetz et al. 2006). However, the
168 evolution of the kinetochore network across Eukaryotes has been shown to be a complex process
169 that requires further investigation (van Hooff et al. 2017). In the current study, we have focussed
170 on well-studied genetic interactions (see **Fig. 3A**) of the selected genes that have lost both copies
171 in Post-WGD yeast species. Chromatin Immuno-Precipitation (ChIP) of CSE4 protein followed
172 by sequencing of DNA fragments was used by Kobayashi et al. (2015) to identify all the CSE4
173 binding sites in the N. *castellii* genome. Based on this experimental data for CSE4 along with
174 similar ChIP-seq data for the NDC10, NDC80 and CEP3 proteins the locations of the new
175 centromeres in N. *castellii* have been validated (Kobayashi et al. 2015). The multiple sequence
176 alignment of amino-acid sequences across the study species (see **Fig. 3B**) shows that the CSE4
177 genes of N. *castellii* and N. *dairenensis* have a stretch of approximately 10 amino acid residues
178 at the very beginning of the CSE4 gene that are absent in other species. It is known that the C-
179 terminal of the CSE4 protein binds to centromeric DNA sequence and the N-terminal interacts
180 with the CTF19 complex (Chen et al. 2000). The N-terminal region has overall reduced sequence
181 conservation across all species compared to the C-terminal region. The N-terminal of the CSE4
182 protein has been shown to play a role in the ubiquitin-mediated proteolysis of the CSE4 protein
183 (Au et al. 2013). More recently, sumoylation and ubiquitination of the N-terminal have been
184 shown to be required to prevent mislocalization of CSE4 to non-centromeric chromatin (Ohkuni
185 et al. 2018). The lysine 65 residue (K65) in the CSE4 gene has been identified as the residue
186 important for proper localisation. Changes in the N-terminal region of CSE4 gene could have led
187 to changes in the post-translational modifications resulting in changes in the localisation patterns
188 and subsequent movement of centromeres in Naumovozyma.

189

190 In Post-WGD species, most (58 out of 67) of the kinetochore genes have lost at least one of the
191 copies that resulted during the WGD event. Only nine genes have retained both copies in at least
192 one Post-WGD species. This set of nine genes could have been retained to maintain a certain
193 dosage of expression or might have diverged in function through neo or sub functionalisation

(Zhang 2003; Qian and Zhang 2014). The YBP1gene is the only kinetochore gene that is present in all Pre-WGD species that has also retained both copies in all Post-WGD species. HRR25 is found in two copies in K. *africana*, K. *naganishii*, N. *castellii*, N. *dairenensis*, T. *blattae*, T. *phaffii,* and V. *polyspora*. The CIN8 gene has retained two copies each in both T. *phaffii* and V. *polyspora*. MAD1 gene is present in two copies only in V. *polyspora* and RTS1 occurs twice only in T. *blattae*. BUB1 is present in two copies in all Post-WGD species except T. *blattae*, T. *phaffii* and V. *polyspora*. The gene CRM1 has the most intricate pattern of gene loss with single copies found in K. *naganishii*, S. *uvarum*, S. *cerevisiae*, S. *kudriavzevii*, S. *mikatae* and T. *blattae*.

The MPS1 gene has retained two copies in N. *castellii* but has lost one of the copies in all other Post-WGD species including N. *dairenensis*. Interestingly, the copy that has been lost in N. *dairenensis* is different from the copy that has been lost in the other Post-WGD species (see **Table S1**). This suggests a more recent potentially independent gene loss in N. *dairenensis*. MIF2 has retained two copies only in N. *castellii*. These gene copy differences within Naumovozyma might reflect the ongoing process of kinetochore evolution in this genus. It is of special interest to note that MPS1 gene is required for correct localisation of kinetochore components (Vigneron 2004). Experimental evaluation of the roles of these genes in terms of gene dosage as well as compensation ability with respect to chromosome segregation phenotypes would be able to provide a better idea about the molecular details.

A recent study has established that centromere regions are defined by the presence of Non-B-form DNA structures resulting from the presence of dyad symmetry in the nucleotide sequence (Kasinathan and Henikoff 2018). This same study found that while the new Naumovozyma centromeres are enriched for dyad symmetries and non-B-form DNA, the dyad symmetry was less and the SIST DNA melting and cruciform extrusion scores were lower in Naumovozyma (sensu lato) compared to sensu strictu species (see Figure 6 of Kasinathan and Henikoff 2018). However, they utilised the contrast of sensu stricto vs sensu lato with very few species to make this comparison. Using a larger sample size (centromeres from various budding yeast species) we show that the GC content shows a significant negative correlation (Pearson's correlation coefficient r of -0.33, p-value: 1.157e-6) with dyad density (see **Fig. 4**). This correlation is

reflective of the genome-wide pattern seen in S. *cerevisiae* (Lisnić et al. 2005). By establishing the presence of this correlation even for the nucleotide sequences of centromeres we are able to understand some of the variance seen in dyad density across budding yeast species.

The unconventional centromeres in the yeast species N. *castellii* and N. *dairenensis* are in most cases located at a different genomic locus compared to the point centromeres found in other budding yeast species such as S. *cerevisiae* (Kobayashi et al. 2015). This movement of centromeres has partly been attributed to chromosomal re-arrangements. We show that the GC content of the introns corresponding to the older (S. *cerevisiae* like) centromeres is very high and dyad density is low (see **Fig. 4**; red coloured circles and squares). On the other hand the new unconventional centromeres that have been identified in Naumovozyma (see **Fig. 4**; blue coloured circles and squares) are having higher dyad densities and lower GC content. We note that the pattern is the same for both Naumovozyma species (see **Fig. 4**; filled squares for N. *castellii* and filled circles for N. *dairenensis*). This change in dyad density and GC content probably reflects a divergence that occurred after the change in the location of the centromere.

Discussion

The role of gene-flow during the process of speciation has been the focus of considerable debate especially with respect to natural hybrid zones (Barton and Hewitt 1985; Slatkin 1987). The recent availability of large-scale genomic datasets has accentuated the role of gene-flow resulting in the development of models of speciation with gene flow (Feder et al. 2012). Prevalence of such unexpected gene-flow across relatively large evolutionary distances has important implications for how reproductive isolation builds up between species (Harrison and Larson 2014; Payseur and Rieseberg 2016). This has motivated re-evaluation of the speciation phases at which karyotype and centromere evolution are important (Brown and O'Neill 2010; Faria and Navarro 2010). Another aspect that has received considerable attention is the nucleotide sequence divergence of centromere proximal regions (Stump et al. 2005; Ellegren et al. 2012). However, there is no consensus on the role, if any, that restricted recombination at centromeres has on maintaining species boundaries (Noor and Bennett 2009; Cruickshank and Hahn 2014).

Nucleotide sequence divergence at the centromeres themselves as well as in the genes that recognise and bind to centromeres have been proposed as potential mechanisms for the build-up

257 of post-zygotic reproductive isolation (Borodin et al. 2001). The paradoxical behaviour of
258 centromeres to evolve rapidly while still being inherited stably makes them good candidates for
259 loci that contribute to the process of speciation (Henikoff et al. 2001). However, the observation
260 that Drosophila species produce fertile offspring despite the extensive divergence of the
261 sequence of centromere binding proteins seems to contradict this idea (Sainz et al. 2003; but see
262 Thomae et al. 2013). Computational identification and functional characterisation of the effects
263 of sequence divergence, loss and duplication of the kinetochore genes in diverse species will
264 help clarify the role of the centromere in facilitating reproductive isolation.
265
266 In this study, we provide evidence that suggests that both copies of four genes of the CCAN are
267 lost in Naumovozyma budding yeast species that have transitioned to unconventional
268 centromeres (Kobayashi et al. 2015). The loss of these CCAN genes that are from the CTF19
269 complex is potentially mediated by changes in the N-terminal region of the CSE4 gene (see **Fig.**
270 **3B**). Although we have no experimental data to suggest that the changes in the N-terminal region
271 of the CSE4 gene are a consequence of changes in the sequence of the centromere being bound
272 by the C-terminal region, it is one potential scenario that could explain the observed association
273 between gene loss and transition to unconventional centromeres. In contrast to the well-studied
274 case of loss of CENH3 (CSE4 homolog) gene in multiple insect lineages (Drinnenberg et al.
275 2014) and the multiple cases of exon gain and loss (Fan et al. 2013), we show that transition to a
276 new centromere sequence might be sufficient for the loss of CCAN genes. Understanding the
277 sequence of events involved in the loss of the CCAN genes and origin of the unconventional
278 centromere in budding yeast species would help understand the loss of CCAN genes in various
279 eukaryotic lineages (van Hooff et al. 2017). The CSE4 gene has an additional sequence in
280 Naumovozyma compared to other species suggesting that the unconventional centromeres might
281 potentially be remnants of the ancestral state of point centromeres. We discuss few other
282 potential scenarios for the sequence of events that could have led to the association observed.

283 The dispensability of the genes that are lost in Naumovozyma is supported by the viability of S.
284 *cerevisiae* knockouts and independent loss in the two Pre-WGD Eremothecium species. It is
285 possible that initially these genes were sequentially lost and the origin of novel centromeres was
286 an adaptive response to compensate for the lost genes. Yet, the loss of three out of these four
287 genes in Eremothecium species (E. *gossypii* and E. *cymbalariae*) does not seem to be associated

with any noticeable changes in centromere type. The evolution of novel centromeres could just be one possible solution to the loss of these genes from the CCAN and Eremothecium species might have come up with a different solution. Further phenotypic characterisation of Eremothecium species might shed light on this.

It has been shown that loss of RNAi leads to a shortening of the centromeres and is an important determinant of centromere evolution in fungi (Yadav et al. 2018). The N. *castellii* species in addition to having unconventional point centromeres have also been shown to have a functional RNA interference pathway (Drinnenberg et al. 2009). Despite having a role in fission yeast heterochromatin specification at the centromeres, siRNA's have not been found to have any centromere specific role in N. *castellii* (Kobayashi et al. 2015). Nonetheless, it would be interesting to know what role the presence of RNAi machinery in Naumovozyma might have had in the loss of CCAN genes. The Drinnenberg et al. study (2009) reports the gene expression fold change of ago1 knockout and dcr1 knockout in N. *castellii* compared to the wild type. However, none of the genes showing a fold-change greater than two are part of the kinetochore gene set.

The function of the genes lost in Naumovozyma may be performed by genes that are found in Naumovozyma but are absent in S. *cerevisiae*. Naumovozyma species have 46 genes that are absent in the other 18 species analysed in this study. To evaluate whether any of these genus specific genes could have taken over the role performed by the four missing genes, we identified homologs of each of these genes by performing blastp search against the non-redundant protein database with an e-value cut-off of 10^{-3}. None of the identified homologs had a characterised role related to the kinetochore machinery.

The current study does not perform any experimental characterisation of phenotypes across the twenty yeast species that would be caused by knocking out the genes that we identify as lost. However, kinetochore genes have been studied extensively in S. *cerevisiae*. Single and double gene knockouts have been phenotypically characterised in great detail (Measday et al. 2005). This functional data from S. *cerevisiae* can be extrapolated to other species by supplementing it with computational predictions. Such extrapolation would, of course, have to be experimentally verified by subsequent studies in the focal species. Hence, our study is merely focussed on

319 demonstrating strong patterns of associations that suggests kinetochore network rewiring in
320 Naumovozyma.
321
322 Stable separation of chromosomes into daughter cells requires proper functioning of the
323 kinetochore machinery. Reduced efficiency in such segregation of chromosomes would result in
324 reduced fertility. The high levels of gene turnover seen in kinetochore genes suggest frequent
325 changes in the interactome of kinetochore genes. Such frequent changes in gene content could
326 lead to differences in the interactome of closely related species. These differences in the
327 interactome can have a prominent role in mediating speciation through reproductive isolation.
328 Reconstruction of the sequence of events leading to turnover of kinetochore machinery and
329 centromere sequence at the molecular level in budding yeast species might serve as good test
330 case for understanding its role in speciation.
331
332 **Conclusions**
333 Rapid changes in the genomic sequence of centromeres and associated kinetochore proteins
334 between closely related species are thought to have an important role in speciation. Yet, it is not
335 conclusively known whether the centromere sequence divergence occur first or kinetochore
336 proteins evolve to use a different centromere sequence. We look at patterns of kinetochore gene
337 loss in twenty species of yeast to identify major lineage specific events. Interestingly, the loss of
338 four genes of the CCAN coincides with the origin of unconventional centromeres in N.*castellii*
339 and N.*dairenensis*. We speculate that loss of these genes might have occurred subsequent to the
340 divergence of the centromere sequences as their role might have been taken over by other genes.
341 While our results cannot establish the sequence of events, the identified lineage specific loss of
342 kinetochores genes that are known to functionally interact serves as a molecular footprint of
343 genetic changes that contribute to reproductive isolation between species.
344
345 **References**
346
347 Au, W. C., A. R. Dawson, D. W. Rawson, S. B. Taylor, R. E. Baker, and M. A. Basrai. 2013. A
348 novel role of the N terminus of budding yeast histone H3 variant Cse4 in ubiquitin-mediated
349 proteolysis.
350 Baetz, K., V. Measday, and B. Andrews. 2006. Revealing hidden relationships among yeast
351 genes involved in chromosome segregation using systematic synthetic lethal and synthetic

dosage lethal screens.
Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.
Borodin, P. M., S. Henikoff, K. Ahmad, and H. S. Malik. 2001. Speciation and centromere evolution [3] (multiple letters).
Brown, J. D., and R. J. O'Neill. 2010. Chromosomes, Conflict, and Epigenetics: Chromosomal Speciation Revisited. Annu. Rev. Genomics Hum. Genet. 11:291–316.
Cheeseman, I. M., S. Anderson, M. Jwa, E. M. Green, J. seog Kang, J. R. Yates, C. S. M. Chan, D. G. Drubin, and G. Barnes. 2002. Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p.
Chen, Y., R. E. Baker, K. C. Keith, K. Harris, S. Stoler, and M. Fitzgerald-Hayes. 2000. The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain. Mol. Cell. Biol. 20:7037–7048.
Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133–3157.
Delport, W., A. F. Y. Poon, S. D. W. Frost, and S. L. Kosakovsky Pond. 2010. Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457.
Drinnenberg, I. A., and B. Akiyoshi. 2017. Evolutionary Lessons from Species with Unique Kinetochores. Pp. 111–138 *in* Centromeres and Kinetochores.
Drinnenberg, I. A., D. deYoung, S. Henikoff, and H. S. ingh Malik. 2014. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. Elife 3.
Drinnenberg, I. A., S. Henikoff, and H. S. Malik. 2016. Evolutionary Turnover of Kinetochore Proteins: A Ship of Theseus?
Drinnenberg, I. A., D. E. Weinberg, K. T. Xie, J. P. Mower, K. H. Wolfe, G. R. Fink, and D. P. Bartel. 2009. RNAi in budding yeast. Science (80-.). 326:544–550.
Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, A. Künstner, H. Mäkinen, K. Nadachowska-Brzyska, A. Qvarnström, S. Uebbing, and J. B. W. Wolf. 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature 491:756–60. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
Fan, X. Y., L. Yu, H. L. Xu, and Y. Li. 2013. Multiple intron gain and loss events occurred during the evolution of Cenp-A gene. Chinese Sci. Bull. 58:2174–2178.
Faria, R., and A. Navarro. 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends Ecol. Evol. 25:660–9.
Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–50. Elsevier.
Fernius, J., and A. L. Marston. 2009. Establishment of cohesion at the pericentromere by the Ctf19 kinetochore subcomplex and the replication fork-associated factor, Csm3. PLoS Genet. 5.
Harrison, R. G., and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. Pp. 795–809 *in* Journal of Heredity.
Henikoff, S., K. Ahmad, and H. S. Malik. 2001. The centromere paradox: Stable inheritance with rapidly evolving DNA.
Hinshaw, S. M., and S. C. Harrison. 2013. An Iml3-Chl4 Heterodimer Links the Core Centromere to Factors Required for Accurate Chromosome Segregation. Cell Rep. 5:29–36.
Hinshaw, S. M., V. Makrantoni, S. C. Harrison, and A. L. Marston. 2017. The Kinetochore

Receptor for the Cohesin Loading Complex. Cell 171:72–84.e13.
Karademir Andersson, A., and M. Cohn. 2017. Naumovozyma castellii: an alternative model for budding yeast molecular biology. Yeast 34:95–109.
Kasinathan, S., and S. Henikoff. 2018. Non-B-Form DNA Is Enriched at Centromeres. Mol. Biol. Evol.
Kobayashi, N., Y. Suzuki, L. W. Schoenfeld, C. A. Müller, C. Nieduszynski, K. H. Wolfe, and T. U. Tanaka. 2015. Discovery of an Unconventional Centromere in Budding Yeast Redefines Evolution of Point Centromeres. Curr. Biol. 25:2026–2033.
Kosakovsky Pond, S. L., and S. D. W. Frost. 2005a. Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531–2533.
Kosakovsky Pond, S. L., and S. D. W. Frost. 2005b. Not so different after all: A comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22:1208–1222.
Lisnić, B., I. K. Svetec, H. Šarić, I. Nikolić, and Z. Zgaga. 2005. Palindrome content of the yeast Saccharomyces cerevisiae genome. Curr. Genet. 47:289–297.
Malik, H. S., and S. Henikoff. 2001. Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics 157:1293–1298.
Malik, H. S., and S. Henikoff. 2009. Major Evolutionary Transitions in Centromere Complexity.
Malvezzi, F., G. Litos, A. Schleiffer, A. Heuck, K. Mechtler, T. Clausen, and S. Westermann. 2013. A structural basis for kinetochore recruitment of the Ndc80 complex via two distinct centromere receptors. EMBO J. 32:409–423.
Measday, V., K. Baetz, J. Guzzo, K. Yuen, T. Kwok, B. Sheikh, H. Ding, R. Ueta, T. Hoac, B. Cheng, I. Pot, A. Tong, Y. Yamaguchi-Iwai, C. Boone, P. Hieter, and B. Andrews. 2005. Systematic yeast synthetic lethal and synthetic dosage lethal screens identify genes required for chromosome segregation. Proc. Natl. Acad. Sci. 102:13956–13961.
Melters, D. P., K. R. Bradnam, H. A. Young, N. Telis, M. R. May, J. Ruby, R. Sebra, P. Peluso, J. Eid, D. Rank, J. Garcia, J. L. DeRisi, T. Smith, C. Tobias, J. Ross-Ibarra, I. Korf, and S. W. Chan. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol. 14:R10.
Montefalcone, G., S. Tempesta, M. Rocchi, and N. Archidiacono. 1999. Centromere repositioning. Genome Res. 9:1184–1188.
Noor, M. A. F., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity (Edinb). 103:439–44.
O'Neill, R. J., M. D. B. Eldridge, and C. J. Metcalfe. 2004. Centromere dynamics and chromosome evolution in marsupials. Pp. 375–381 *in* Journal of Heredity.
Ohkuni, K., R. Levy-Myers, J. Warren, W.-C. Au, Y. Takahashi, R. E. Baker, and M. A. Basrai. 2018. N-terminal Sumoylation of Centromeric Histone H3 Variant Cse4 Regulates Its Proteolysis To Prevent Mislocalization to Non-centromeric Chromatin. G3 (Bethesda). g3.300419.2017.
Payseur, B. A., and L. H. Rieseberg. 2016. A genomic perspective on hybridization and speciation.
Qian, W., and J. Zhang. 2014. Genomic evidence for adaptation by gene duplication. Genome Res. 24:1356–1362.
Rocchi, M., N. Archidiacono, W. Schempp, O. Capozzi, and R. Stanyon. 2012. Centromere repositioning in mammals. Heredity (Edinb). 108:59–67.
Roy, B., and K. Sanyal. 2011. Diversity in requirement of genetic and epigenetic factors for

centromere function in fungi. Eukaryot. Cell 10:1384–1395.
Roy, N., A. Poddar, A. Lohia, and P. Sinha. 1997. The mcm17 mutation of yeast shows a size dependent segregational defect of a mini-chromosome. Curr. Genet. 32:182–189.
Sainz, A., J. A. Wilder, M. Wolf, and H. Hollocher. 2003. Drosophila melanogaster and D. simulans rescue strains produce fit offspring, despite divergent centromere-specific histone alleles. Heredity (Edinb). 91:28–35.
Schueler, M. G., W. Swanson, P. J. Thomas, and E. D. Green. 2010. Adaptive evolution of foundation kinetochore proteins in primates. Mol. Biol. Evol. 27:1585–1597.
Slatkin, M. 1987. Gene Flow and the Geographic Structure of Natural Populations. Science 236:787–792.
Steiner, F. A., and S. Henikoff. 2014. Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. Elife 2014.
Stump, A. D., M. C. Fitzpatrick, N. F. Lobo, S. Traoré, N. Sagnon, C. Costantini, F. H. Collins, and N. J. Besansky. 2005. Centromere-proximal differentiation and speciation in Anopheles gambiae. Proc. Natl. Acad. Sci. U. S. A. 102:15930–5.
Talbert, P. B., T. D. Bryson, and S. Henikoff. 2004. Adaptive evolution of centromere proteins in plants and animals. J. Biol. 3:18.
Thomae, A. W., G. O. M. Schade, J. Padeken, M. Borath, I. Vetter, E. Kremmer, P. Heun, and A. Imhof. 2013. A Pair of Centromeric Proteins Mediates Reproductive Isolation in Drosophila Species. Dev. Cell 27:412–424.
Tromer, E. C. 2017. Evolution of the Kinetochore Network in Eukaryotes. Utrecht University.
van Hooff, J. J., E. Tromer, L. M. van Wijk, B. Snel, and G. J. Kops. 2017. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. EMBO Rep. e201744102.
Vigneron, S. 2004. Kinetochore Localization of Spindle Checkpoint Proteins: Who Controls Whom? Mol. Biol. Cell 15:4584–4596.
von Mering, C., M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, and B. Snel. 2003. STRING: A database of predicted functional associations between proteins.
Yadav, V., S. Sun, R. B. Billmyre, B. C. Thimmappa, T. Shea, R. Lintner, G. Bakkeren, C. A. Cuomo, J. Heitman, and K. Sanyal. 2018. RNAi is a critical determinant of centromere evolution in closely related fungi. Proc. Natl. Acad. Sci. 201713725.
Zhang, J. 2003. Evolution by gene duplication : an update. Evolution (N. Y). 18:292–298.

Figure legends

Figure 1: Phylogenetic relationship between budding yeast species. The whole genome duplication event separating the Pre-WGD and Post-WGD species is denoted by a green dot. Point centromeres are common to all twenty budding yeast species and their origin is depicted by a dark blue dot. Origin of unconventional point centromere in Naumovozyma is depicted by a light blue dot.

Figure 2: Patterns of kinetochore gene loss events across the twenty species. Gene IDs of NKP1, NKP2, IML3, CHL4, CNN1 and WIP1 from YGOB are provided for each species.

The genes missing in both Post-WGD copies are coloured red. Missing in only one Post-WGD copy is coloured light blue. Missing in the Pre-WGD copy is coloured orange. Present in Pre-WGD is coloured light green.

Figure 3: Known genetic and physical interactions between lost kinetochore genes. (A) Network representing the known/predicted biological interactions between various kinetochore genes in S. *cerevisiae* as annotated in the String database (von Mering et al. 2003).**(B)** Multiple sequence alignment (Clustal omega) of the amino acid residues in the CSE4 gene from twenty yeast species. Naumovozyma species show additional stretches of amino acids that are missing in the other yeast species.

Figure 4: Negative correlation between GC content and dyad density at the point centromeres across budding yeast species. The filled squares represent N. *castellii* and filled circles show N. *dairenensis*. Red colour is used for the intronic regions corresponding to the older (S. cerevisiae like) centromeres. Blue colour is used for the new unconventional centromeres that have been identified in Naumovozyma. Centromeres of the remaining budding species are coloured black except for four of the centromeres that are coloured brown (S. *uvarum* centromeres that have gaps in the genome assembly).

Figure 1

Phylogenetic relationship between budding yeast species

The whole genome duplication event separating the Pre-WGD and Post-WGD species is denoted by a green dot. Point centromeres are common to all twenty budding yeast species and their origin is depicted by a dark blue dot. Origin of unconventional point centromere in *Naumovozyma* is depicted by a light blue dot.

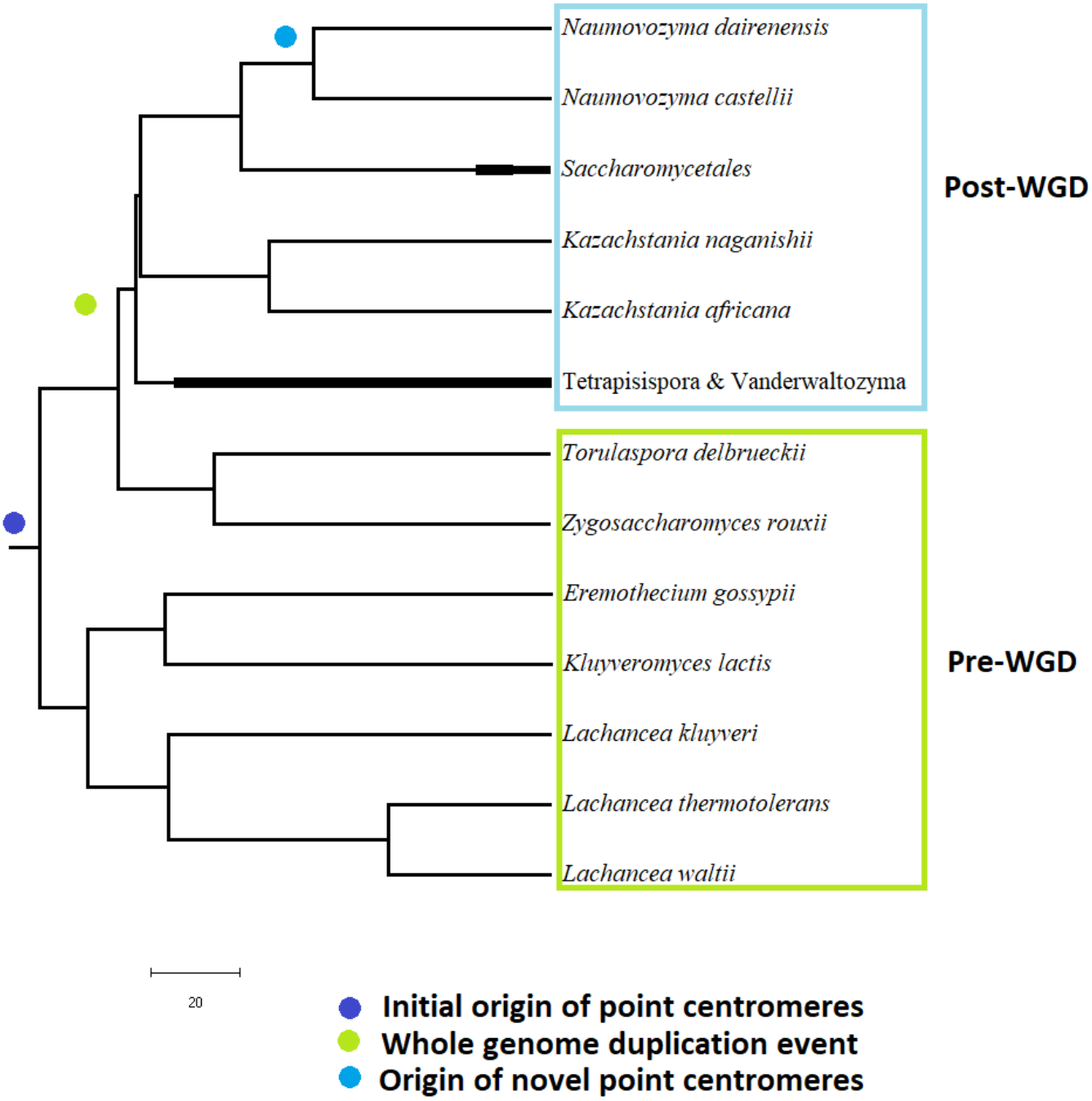
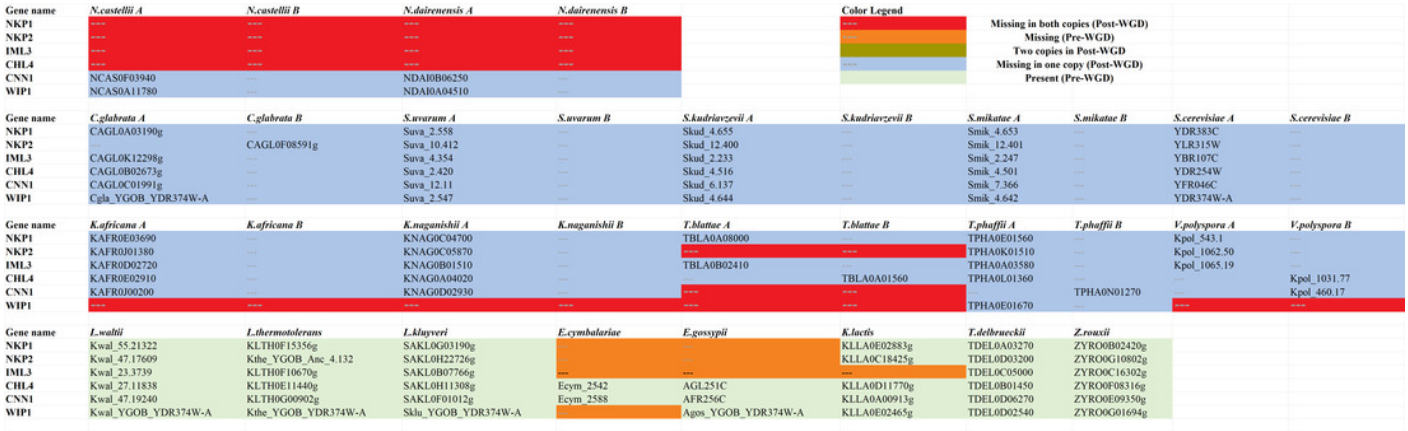

Naumovozyma dairenensis
Naumovozyma castellii
Saccharomycetales
Kazachstania naganishii
Kazachstania africana
Tetrapisispora & Vanderwaltozyma
Post-WGD
Torulaspora delbrueckii
Zygosaccharomyces rouxii
Eremothecium gossypii
Kluyveromyces lactis
Lachancea kluyveri
Lachancea thermotolerans
Lachancea waltii
Pre-WGD
20
Initial origin of point centromeres
Whole genome duplication event
Origin of novel point centromeres

Figure 2

Patterns of kinetochore gene loss events across the twenty species

Gene IDs of NKP1, NKP2, IML3, CHL4, CNN1 and WIP1 from YGOB are provided for each species. The genes missing in both Post-WGD copies are coloured red. Missing in only one Post-WGD copy is coloured light blue. Missing in the Pre-WGD copy is coloured orange. Present in Pre-WGD is coloured light green.



Gene name	*N.castellii A*	*N.castellii B*	*N.dairenensis A*	*N.dairenensis B*
NKP1	---	---	---	---
NKP2	---	---	---	---
IML3	---	---	---	---
CHL4	---	---	---	---
CNN1	NCAS0F03940	---	NDAI0B06250	---
WIP1	NCAS0A11780	---	NDAI0A04510	---

Color Legend: Missing in both copies (Post-WGD); Missing (Pre-WGD); Two copies in Post-WGD; Missing in one copy (Post-WGD); Present (Pre-WGD)

Gene name	*C.glabrata A*	*C.glabrata B*	*S.uvarum A*	*S.uvarum B*	*S.kudriavzevii A*	*S.kudriavzevii B*	*S.mikatae A*	*S.mikatae B*	*S.cerevisiae A*	*S.cerevisiae B*
NKP1	CAGL0A03190g	---	Suva_2.558	---	Skud_4.655	---	Smik_4.653	---	YDR383C	---
NKP2	---	CAGL0F08591g	Suva_10.412	---	Skud_12.400	---	Smik_12.401	---	YLR315W	---
IML3	CAGL0K12298g	---	Suva_4.354	---	Skud_2.233	---	Smik_2.247	---	YBR107C	---
CHL4	CAGL0B02673g	---	Suva_2.420	---	Skud_4.516	---	Smik_4.501	---	YDR254W	---
CNN1	CAGL0C01991g	---	Suva_12.11	---	Skud_6.137	---	Smik_7.366	---	YFR046C	---
WIP1	Cgla_YGOB_YDR374W-A	---	Suva_2.547	---	Skud_4.644	---	Smik_4.642	---	YDR374W-A	---

Gene name	*K.africana A*	*K.africana B*	*K.naganishii A*	*K.naganishii B*	*T.blattae A*	*T.blattae B*	*T.phaffii A*	*T.phaffii B*	*V.polyspora A*	*V.polyspora B*
NKP1	KAFR0E03690	---	KNAG0C04700	---	TBLA0A08000	---	TPHA0E01560	---	Kpol_543.1	---
NKP2	KAFR0J01380	---	KNAG0C05870	---	---	---	TPHA0K01510	---	Kpol_1062.50	---
IML3	KAFR0D02720	---	KNAG0B01510	---	TBLA0B02410	---	TPHA0A03580	---	Kpol_1065.19	---
CHL4	KAFR0E02910	---	KNAG0A04020	---	---	TBLA0A01560	TPHA0L01360	---	---	Kpol_1031.77
CNN1	KAFR0J00200	---	KNAG0D02930	---	---	---	---	TPHA0N01270	---	Kpol_460.17
WIP1	---	---	---	---	---	---	TPHA0E01670	---	---	---

Gene name	*L.waltii*	*L.thermotolerans*	*L.kluyveri*	*E.cymbalariae*	*E.gossypii*	*K.lactis*	*T.delbrueckii*	*Z.rouxii*
NKP1	Kwal_55.21322	KLTH0F15356g	SAKL0G03190g	---	---	KLLA0E02883g	TDEL0A03270	ZYRO0B02420g
NKP2	Kwal_47.17609	Kthe_YGOB_Anc_4.132	SAKL0H22726g	---	---	KLLA0C18425g	TDEL0D03200	ZYRO0G10802g
IML3	Kwal_23.3739	KLTH0F10670g	SAKL0B07766g	---	---	---	TDEL0C05000	ZYRO0C16302g
CHL4	Kwal_27.11838	KLTH0E11440g	SAKL0H11308g	Ecym_2542	AGL251C	KLLA0D11770g	TDEL0B01450	ZYRO0F08316g
CNN1	Kwal_47.19240	KLTH0G00902g	SAKL0F01012g	Ecym_2588	AFR256C	KLLA0A00913g	TDEL0D06270	ZYRO0E09350g
WIP1	Kwal_YGOB_YDR374W-A	Kthe_YGOB_YDR374W-A	Sklu_YGOB_YDR374W-A	---	Agos_YGOB_YDR374W-A	KLLA0E02465g	TDEL0D02540	ZYRO0G01694g

Figure 3

Known genetic and physical interactions between lost kinetochore genes

(A) Network representing the known/predicted biological interactions between various kinetochore genes in S. *cerevisiae* as annotated in the String database (von Mering et al. 2003) .**(B)** Multiple sequence alignment (Clustal omega) of the amino acid residues in the CSE4 gene from twenty yeast species. *Naumovozyma* species show additional stretches of amino acids that are missing in the other yeast species.

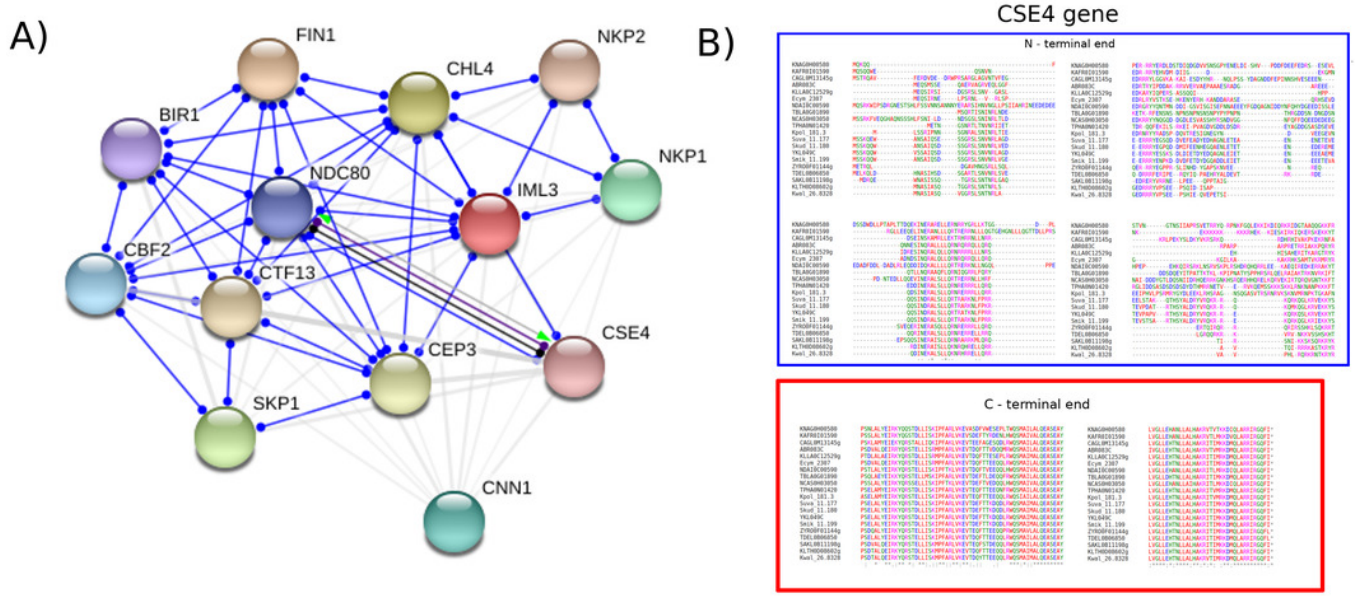


Figure 4

Negative correlation between GC content and dyad density at the point centromeres across budding yeast species

The filled squares represent N. *castellii* and filled circles show N. *dairenensis*. Red colour is used for the intronic regions corresponding to the older (S. cerevisiae like) centromeres. Blue colour is used for the new unconventional centromeres that have been identified in Naumovozyma. Centromeres of the remaining budding species are coloured black except for four of the centromeres that are coloured brown (S. *uvarum* centromeres that have gaps in the genome assembly).