Title (88 chars < 120): Punctuated evolution and transitional hybrid network in an ancestral cell cycle of fungi

Authors: Edgar M. Medina[1,2], Jonathan J. Turner[3], Jan M. Skotheim[3], and Nicolas E. Buchler[1,2]

[1] Department of Biology, Duke University, Durham, NC, 27708, USA
[2] Center for Genomic and Computational Biology, Duke University, Durham, NC, 27710, USA
[3] Department of Biology, Stanford University, Stanford, CA, 94305, USA

Contact: nicolas.buchler@duke.edu

Abstract: (130 words < 150)

Although cell cycle control is an ancient, conserved, and essential process, some core animal and fungal cell cycle regulators share no more sequence identity than non-homologous proteins. Here, we show that evolution along the fungal lineage was punctuated by the early acquisition and entrainment of the SBF transcription factor. Cell cycle evolution in the fungal ancestor then proceeded through a hybrid network containing both SBF and its ancestral animal counterpart E2F, which is still maintained in many basal fungi. We hypothesize that a viral SBF may have initially hijacked cell cycle control by activating transcription via the *cis*-regulatory elements targeted by the ancestral cell cycle regulator E2F, much like extant viral oncogenes. Consistent with this hypothesis, we show that SBF can regulate promoters with E2F binding sites in budding yeast.

Impact statement (15-30 words): Cell cycle network evolution in a fungal ancestor was punctuated by arrival of a viral DNA-binding protein that was permanently incorporated into the G1/S regulatory network controlling cell cycle entry.

Keywords: cell cycle / evolution / fungi / chytrid / virus

INTRODUCTION

The networks regulating cell division in yeasts and animals are highly similar in both physiological function and network structure (Figure 1) (Cross et al., 2011; Doonan and Kitsios, 2009). For example, the cell cycle controls proliferation in response to a variety of internal and external signals during the G1 phase, between cell division and DNA replication. These input signals, including cell growth, are integrated into a gradual increase in cyclin dependent kinase (Cdk) activity, which triggers a feedback loop at the basis of the all-or-none irreversible decision to proliferate (Bertoli et al., 2013).

Many of the molecular mechanisms underlying G1 regulation are highly conserved. In animal cells, Cyclin D, in complex with either Cdk4 or Cdk6, initiates cell cycle entry by phosphorylating the retinoblastoma protein, pRb. This begins the inactivation of pRb and the concomitant activation of the E2F transcription factors that induce transcription of downstream cyclins E and A, which complete the inhibition of pRb thereby forming a positive feedback loop (Bertoli et al., 2013). Similarly, in budding yeast, the G1 cyclin Cln3-Cdk1 complex initiates the transition by phosphorylating and partially inactivating Whi5, an inhibitor of the SBF transcription factor (Costanzo et al., 2004; de Bruin et al., 2004; Nasmyth and Dirick, 1991; Ogas et al., 1991; Sidorova and Breeden, 1993). This allows for SBF-dependent transcription of the downstream G1 cyclins *CLN1* and *CLN2*, which also inactivate Whi5 to complete a positive feedback loop (Skotheim et al., 2008). Thus, both the biochemical function of G1 regulators and their specific targets are highly conserved (Figure 1).

Many of the individual proteins performing identical roles are unlikely to be true orthologs, *i.e.*, it cannot be inferred from sequence identity that the proteins evolved from a common ancestral gene. In yeast, a single cyclin-dependent kinase, Cdk1, binds distinct cyclin partners to perform all the functions of three non-orthologous animal Cdks (Cdk2, 4 and 6) during cell cycle entry (Liu and Kipreos, 2000). Furthermore, no member of the transcription factor complex SBF-Whi5 exhibits amino acid sequence identity or structural similarity to any member of the E2F-pRb complex (Cross et al., 2011; Hasan et al., 2013; Taylor et al., 1997). Finally, Cdk inhibitors such as Sic1 and p27 play analogous roles in yeast and mammals despite a

total lack of sequence identity (Cross et al., 2011). Taken together, these examples imply significant evolution of cell cycle regulatory proteins in fungi and/or animals while the network topology remains largely intact. While identification of network topology is restricted to a few model organisms and is not as broad as sequence analysis, the similar network topology in budding yeast and animals suggests that this feature is more conserved than the constituent regulatory proteins (Cross et al., 2011; Doonan and Kitsios, 2009).

The shared presence of E2F-pRb within plants (Archaeaplastida) and animal (Metazoa) lineages indicates that this regulatory complex, rather than the fungal SBF-Whi5 complex, was present in the last eukaryotic common ancestor (Cao et al., 2010; Doonan and Kitsios, 2009; Fang et al., 2006; Harashima et al., 2013). The divergence of G1 regulator sequences is surprising because fungi and animals are more closely related to one another than either is to plants. This fungal-metazoan difference raises the question as to where the fungal components came from. Fungal components could either be rapidly evolved ancestral regulators or have a distinct evolutionary history, which would suggest convergent evolution of regulatory networks.

Here, we examine conserved and divergent features of eukaryotic cell cycle regulation. In contrast to previous work that considered a protein family of cell cycle regulators in isolation (Cao et al., 2010; 2014; Eme et al., 2011; Gunbin et al., 2011; Ma et al., 2013; Wang et al., 2004), we studied the evolutionary history of an entire regulatory network. We examined a greater number of genomes covering most of eukaryotic diversity, including Excavata, Haptophyta, Cryptophyta, SAR (Stramenopiles, Alveolata & Rhizaria), Archaeplastida (plants), Amoebozoa, Apusozoa and the Opisthokonta (animals and fungi). This survey allowed us to estimate the cell cycle repertoire of the last eukaryotic common ancestor (LECA), a prerequisite to clarifying the evolutionary transitions of the cell cycle components of both animals and fungi.

Our results indicate that LECA likely had complex cell cycle regulation involving at least one Cdk, multiple cyclin families, activating and inhibitory E2F transcription factors, and pRb-family pocket proteins. The LECA repertoire helps establish that the emergence of SBF-Whi5 is abrupt and distinguishes fungi from all other eukaryotes.

103 We also show that basal fungi can have both ancestral E2F-pRb and fungal SBF-Whi5
104 components. Thus, fungal evolution appears to have proceeded through a hybrid
105 network before abruptly losing the ancestral components in the lineage leading to
106 Dikarya. This supports the hypothesis that network structure, but not the individual
107 components, has been conserved through the transition to fungi and argues against
108 the case of convergent evolution.
109
110 Finally, our data confirm that SBF shows homology to KilA-N, a poorly characterized
111 domain present in prokaryotic and eukaryotic DNA viruses. Thus, SBF is not derived
112 from E2F (its functional analog) and likely emerged through horizontal gene transfer
113 in the fungal ancestor. That SBF can still regulate promoters with E2F binding sites
114 in budding yeast suggests that a viral SBF may have initially hijacked cell cycle
115 control, activating transcription via the *cis*-regulatory elements targeted by the
116 ancestral cell cycle regulator E2F, much like extant viral oncogenes.
117
118 **RESULTS**
119
120 **Reconstruction of complex cell cycle control in the ancestral eukaryote**
121 Recent work shows that the last eukaryotic common ancestor (LECA) already had a
122 complex repertoire of proteins (Dacks and Field, 2007; Eichinger et al., 2005;
123 Merchant et al., 2007). Indeed, all sequenced eukaryotic lineages have lost entire
124 gene families that were present in LECA (Fritz-Laylin et al., 2010). In contrast to the
125 growing consensus that LECA had an extensive repertoire of proteins, the prevailing
126 view of the LECA cell cycle is that it was based on a simple oscillator constructed
127 with relatively few components (Coudreuse and Nurse, 2010; Nasmyth, 1995).
128 According to the 'simple' LECA cell cycle model, an ancestral oscillation in Cyclin B-
129 Cdk1 activity drove periodic DNA replication and DNA segregation, while other
130 aspects of cell cycle regulation, such as G1 control, may have subsequently evolved
131 in specific lineages. The model was motivated by the fact that Cdk activity of a single
132 Cyclin B is sufficient to drive embryonic cell cycles in frogs (Murray and Kirschner,
133 1989) and fission yeast (Stern and Nurse, 1996), and that many yeast G1 regulators
134 have no eukaryotic orthologs (Figure 1).
135
136 To determine the complexity of LECA cell cycle regulation, we examined hundreds of
137 diverse eukaryotic genomes (Fungi, Amoebozoa, Excavata, SAR, Haptophyta,

138 Cryptophyta). We first built sensitive profile Hidden Markov Models (Eddy, 2011) for
139 each of the gene families of cell cycle regulators from model organisms *Arabidopsis*
140 *thaliana, Homo sapiens, Schizosaccharomyces pombe,* and *Saccharomyces*
141 *cerevisiae.* These HHMs were then used to query the sequenced eukaryotic genomes
142 for homologs of both fungal and animal cell cycle regulators (Figures 2-3; see
143 Methods). Phylogenetic analyses were performed on the detected homologs for
144 accurate sub-family assignment of the regulators and infer their evolutionary history
145 (see Methods). If LECA regulation were simple, we would expect little conservation
146 beyond the Cyclin B-Cdk1 mitotic regulatory module. However, if LECA regulation
147 were more complex, we would expect to see broad conservation of a wider variety of
148 regulators.
149
150 We did not find either of the fungal regulators, SBF and Whi5, in any Archaeplastida,
151 Amoebozoa, SAR, Haptophyta, Cryptophyta, Excavata or Metazoa. However, we did
152 find the cyclin sub-families (A, B, D, and E) known to regulate the cell cycle in
153 metazoans across all major branches of eukaryotes (for cyclin phylogeny see Figure
154 3-figure supplement 1). We also found examples of all three sub-families of E2F
155 transcription factors (E2F1-6, DP, E2F7/8) and the pRb family of pocket proteins (for
156 E2F/DP and pRb phylogeny see Figure 3-figure supplement 2 and Figure 3-figure
157 supplement 3). Nearly all species contain the APC specificity subunits Cdc20 and
158 Cdh1/Fzr1, which regulate exit from mitosis and maintain low Cdk activity in G1 (for
159 Cdc20-family APC phylogeny see Figure 3-figure supplement 4). Taken together,
160 these data indicate that LECA cell cycle regulation was based on multiple cyclin
161 families, as well as regulation by the APC complex and members of the pRb and E2F
162 families.
163
164 Members of the Cdk1-3 family (i.e. CdkA in plants) are also broadly conserved across
165 eukaryotes, suggesting they were the primary LECA cell cycle Cdks (for CDK
166 phylogeny see Figure 3-figure supplement 5). Other cell cycle Cdk families in animal
167 (Cdk4/6) and plant (CdkB) are thought to be specific to those lineages. However, we
168 found CdkB in Stramenopiles, which may have arrived via horizontal transfer during
169 an ancient secondary endosymbiosis with algae as suggested by (Cavalier-Smith,
170 1999). We excluded from our analysis other families of cyclin-Cdks including Cdk7-9,
171 which regulate transcription and RNA processing, and Cdk5 (yeast Pho85), which
172 regulate cell polarity, nutrient regulation, and contribute to cell cycle regulation in

yeast (Cao et al., 2014; Guo and Stiller, 2004; Ma et al., 2013; Moffat and Andrews, 2004). While interesting and important, an extensive examination of these cyclin-Cdk families is beyond the scope of this work.

A hybrid E2F-pRb-SBF-Whi5 network on the path of fungal evolution

To identify the possible origins of SBF and Whi5, we searched for sequence homologs across eukaryotic genomes (see Methods). We were unable to find any eukaryotic homologs of SBF or Whi5 outside of fungi, with one exception that we discuss in the next section. The emergence of the new fungal components SBF, which includes the large APSES family, and Whi5 is abrupt and occurs near the split of basal fungi from metazoans (Figures 4-5; for SBF only, SBF+APSES, and Whi5 phylogeny see Figure 5-figure supplement 1, Figure 5-figure supplement 2, and Figure 5-figure supplement 3). The precise location remains unclear because we have only 1 Nuclearid genome and Microsporidia are fast-evolving fungal parasites with reduced genomes (Cuomo et al., 2012). Interestingly, the new regulators (SBF and Whi5) and ancestral regulators (E2F and Rb) co-exist broadly across basal fungi and the lineages formerly known as "zygomycetes". At one extreme, Chytridiomycota (*e.g.*, *Spizellomyces punctatus*) can have both fungal and animal cell cycle regulators, which likely represents the ancestral hybrid network. However, this hybrid network was evolutionarily unstable, as different constellations of components are present in the extant zygomycetes and basal fungi. For example, the zygomycetes have lost pRb while retaining E2F, which was then abruptly lost in the transition to Dikarya. The instability of the hybrid network suggests some functional redundancy between the new SBF-Whi5 and ancient E2F-Rb pathway.

The SBF and E2F family of transcription factors are unlikely to be orthologs.

The DNA-binding domains of SBF/MBF (Taylor et al., 2000; Xu et al., 1997) and E2F/DP (Zheng et al., 1999) are structurally classified as members of the winged-helix-turn-helix (wHTH) family, which is found in both prokaryotes and eukaryotes (Aravind and Koonin, 1999; Aravind et al., 2005; Gajiwala and Burley, 2000). Although the DNA-binding domains of E2F/DP and SBF/MBF are both classified as wHTH proteins, they show important differences in sequence, overall structure, and

207 mode of protein-DNA complex formation that lead us to conclude that it is highly
208 unlikely that they are orthologs.
209
210 The strongest arguments against SBF-E2F orthology are based on structural biology.
211 Many wHTH transcription factors, including the E2F/DP family, have a 'recognition
212 helix' that interacts with the major or minor grooves of the DNA. The E2F/DP family
213 has an RRXYD DNA-recognition motif in its helix that is invariant within the E2F/DP
214 family and is responsible for interacting with the conserved, core GCGC motif (Zheng
215 et al., 1999) (see Figure 6A: red structure). The RRXYD recognition motif is strikingly
216 conserved in E2F/DP across all eukaryotes, including the E2F/DP proteins uncovered
217 in basal fungi (Figure 6B, left). In contrast, the first solved SBF/MBF structure, Mbp1
218 from *S. cerevisiae* in the absence of DNA, suggested Mbp1 recognizes its MCB (Mlu I
219 cell cycle box, ACGCGT) binding site via a recognition helix (Taylor et al., 1997; Xu
220 et al., 1997). A recent structure of PCG2, an SBF/MBF homolog in the rice blast
221 fungus *Magnaporthe oryzae*, in complex with its MCB binding site does not support
222 this proposed mode of DNA binding (Liu et al., 2015). In striking contrast to many
223 wHTH structures, in which the recognition helix is the mediator of DNA binding
224 specificity, the wing of PCG2 binds to the minor groove to recognize the MCB binding
225 site. The two glutamines in the wing (Q82, Q89) are the key elements that recognize
226 the core MCB binding motif CGCG (Figure 6A, blue structure). Family-specific
227 conservation in the DNA-binding domain is observed for all members of the SBF and
228 APSES family, including basal fungal sequences (Figure 6B, right). In summary, the
229 incongruences in sequence, structure, and mode of DNA-interaction between E2F/DP
230 and SBF/MBF families strongly suggest that SBF is not derived from E2F.

231 **Viral origin and evolution of the fungal SBF and APSES family**

232 Since SBF is unlikely to be orthologous to the E2F family of transcription factors, we
233 therefore sought an alternative evolutionary origin. SBF is a member of a larger
234 family of transcription factors in yeast that includes Xbp1, Bqt4, and the APSES
235 family (Acm1, Phd1, Sok2, Egf1, StuA); see Figure 5. Previous work has shown that
236 the DNA-binding domain of the APSES and SBF proteins is homologous to a viral
237 KilA-N domain (Iyer et al., 2002). KilA-N is a member of a core set of "viral hallmark
238 genes" found across diverse DNA viruses that infect eubacteria, archaea, and
239 eukaryotes (Koonin et al., 2006). Outside the fungal SBF/APSES sub-family, little is
240 known about the KilA-N domain structure, its DNA-binding recognition sequence, and
241 function (Brick et al., 1998). The wide distribution of DNA viruses and KilA-N across

242 the three domains of life suggests that the fungal ancestor likely acquired SBF via
243 horizontal gene transfer.
244
245 We were unable to find ancestral SBF-like proteins in other eukaryotes—with the
246 notable exception of *Trichomonas vaginalis*, a parasitic excavate that has over 100
247 homologs. These potential SBF homologs were compared to the Uniprot protein
248 database, and their best sequence matches were to Mimivirus, a large double-
249 stranded DNA virus of the Nucleo-Cytoplasmic Large DNA Viruses (Yutin et al.,
250 2009). Upon closer inspection, both fungal SBF sequences and *T. vaginalis* Mimivirus
251 sequences share the viral KilA-N domain. To gain insight into the possible
252 evolutionary origins of the SBF subfamily, we aligned all KilA-N sequences from the
253 Uniprot and PFAM database (Finn et al., 2014) to all our fungal SBF sequences and
254 built a phylogenetic tree (Figure 7). There are three major phylogenetic groups of
255 KilA-N domains: those found in eukaryotic viruses, prokaryotic viruses, and the
256 fungal SBF family. Our current phylogeny is unable to distinguish whether the SBF
257 family arrived in a fungal ancestor through a eukaryotic virus or a phage-infected
258 bacterium. Structural and functional characterization of existing viral KilA-N domains
259 could help distinguish between these two hypotheses. However, our phylogeny does
260 show that *T. vaginalis* obtained its KilA-N proteins independently from the fungal
261 ancestor because its KilA-N domains are clearly placed within the eukaryotic DNA
262 viruses (*e.g.*, Mimivirus) rather than with the fungal subfamily (Figure 7).

263 **SBF ancestor could regulate E2F-target genes**

264
265 Of all the members of the SBF/APSES family, the most likely candidate to be a
266 "founding" TF is SBF, as it is the only member present in all fungi. In budding yeast
267 and other fungi, SBF functions in G1/S cell cycle regulation and binds a consensus
268 site CGCGAA (Gordân et al., 2011), which overlaps with the consensus site
269 GCGSSAAA for the E2F family (Rabinovich et al., 2008). The APSES regulators, Xbp1,
270 and MBF in budding yeast bind TGCA, TCGA, ACGCGT motifs, respectively. A viral
271 origin of the SBF/APSES family—with the founding member involved in cell cycle
272 control—suggests the hypothesis that perhaps the founder TF functioned like a DNA
273 tumor virus protein and hijacked cell cycle control to promote proliferation.
274
275 For the viral TF (SBF) to hijack cell cycle control in the fungal ancestor, it must have
276 been able to both bind E2F regulatory regions and then activate the expression of

genes under E2F in a cell cycle-regulated fashion. The overlap between the conserved E2F and SBF consensus sites suggests that ancestral SBF could bind E2F regulatory regions. However, a single base pair substitution in the SBF motif can reduce gene expression by up to ~95% (Andrews and Moore, 1992) and flanking regions outside the core are often important for binding affinity and gene expression (Nutiu et al., 2011). To test whether yeast SBF can bind a canonical E2F binding site, we inserted consensus E2F binding sites in the budding yeast genome; see Figure 8. The hijacking hypothesis would be supported *in vivo* if E2F binding sites could generate SBF-dependent cell cycle regulated gene expression. We used the well-studied *CLN2* promoter, which has three binding sites for SBF (SCB, Swi6-dependent cell cycle box) in a nucleosome-depleted region. Removal of these SCB sites is known to eliminate cell cycle-dependent gene expression (Bai et al., 2010). We changed all three SCBs to the E2F binding site variant GCGCGAAA known to regulate the histone gene cluster in mammals (Rabinovich et al., 2008). We observed significant oscillations in GFP expression, which were coordinated with the cell cycle. Importantly, the amplitude of these oscillations was dependent on the budding yeast SBF, but not MBF, and disappeared when the 3 binding sites were removed. Taken together, this set of *in vivo* experiments lends support to the hijacking hypothesis, where an ancestral SBF took control of several E2F-regulated genes.

DISCUSSION

Cell division is an essential process that has been occurring in an uninterrupted chain for billions of years. Thus, one expects strong conservation in the regulatory network controlling the eukaryotic cell division cycle. Consistent with this idea, cell cycle network structure is highly similar in budding yeast and animal cells. However, many components performing similar functions, such as the central SBF and E2F transcription factors, lack sequence identity suggesting a significant degree of evolution or independent origin. To identify axes of conservation and evolution in eukaryotic cell cycle regulation, we examined a large number of genome sequences in Archaeplastida, Amoebozoa, SAR, Haptophyta, Cryptophyta, Excavata, Metazoa and Fungi. Across eukaryotes, we found a large number of proteins homologous to metazoan rather than fungal G1/S regulators. Our analysis indicates that the last eukaryotic common ancestor likely had complex cell cycle regulation based on Cdk1, Cyclins D, E, A and B, E2F, pRb and APC family proteins.

312

313 In contrast, SBF was not present in the last common eukaryotic ancestor, and
314 abruptly emerged, with its regulator Whi5, in fungi likely due to the co-option of a
315 viral protein at the base of the fungal lineage. The origin of Whi5 is unclear because
316 we found no homologs outside of fungi. Whi5 is a mostly unstructured protein,
317 which, like pRb, recruits transcriptional inhibitor proteins to specific sites on DNA via
318 transcription factor binding (Huang et al., 2009; Wang et al., 2009). The relatively
319 simple structure of Whi5 suggests that it may have been subsequently co-opted as a
320 phosphopeptide to entrain SBF activity to cell cycle regulated changes in Cdk activity
321 (Figure 9).
322

323 The replacement of E2F-Rb with SBF-Whi5 at the core of the cell cycle along the
324 fungal lineage raises the question as to how such a drastic change to fundamental
325 regulatory network could evolve? One answer can be found in the evolution of
326 transcription factors. When the function of an essential transcription factor does
327 change, it often leaves behind a core part of its regulon for another factor (Brown et
328 al., 2009; Gasch et al., 2004; Lavoie et al., 2009). This process of handing off
329 transcription factor function has been observed to proceed through an intermediate
330 state, present in some extant genomes, in which both factors perform the function
331 (Tanay et al., 2005). The logic of proceeding through an intermediate state has been
332 well-documented for the regulation of genes expressed only in yeast of mating type
333 **a** (asgs) (Tsong et al., 2006). In the ancestral yeast and many extant species, asgs
334 expression is activated by a protein only present in **a** cells, while in other yeasts,
335 expression is repressed by a protein only present in α cells and **a**/α diploids. The
336 replacement of the ancestral positive regulation by negative regulation occurred via
337 yeast that contained both systems illustrating how an essential function can evolve
338 through a hybrid state (Baker et al., 2012).
339

340 Clearly, something similar happened during cell cycle evolution. It appears that the
341 replacement of the E2F-pRb transcription regulatory complex with the SBF-Whi5
342 complex proceeded via a hybrid intermediate that preserved its function. In the
343 hybrid intermediate, E2F-Rb and SBF-Whi5 may have evolved to be parallel
344 pathways whose functions overlapped to such an extent that the previously essential
345 E2F-Rb pathway could be lost in the transition to Dikarya. Interestingly, many basal
346 fungi have preserved rather than lost this hybrid intermediate, which suggests that

347 each pathway may have specialized functions. Chytrids exhibit both animal (e.g.
348 centrioles, flagella, amoeboid movement) and fungal features (e.g. cell walls, hyphal
349 morphology) whose synthesis needs to be coordinated with cell division. The
350 preservation of the hybrid network in chytrids could then be explained if animal and
351 fungal features are regulated by the E2F-Rb and SBF-Whi5 pathways respectively.
352
353 The origin of the hybrid network at the base of Fungi is abrupt and may have been
354 initiated by the arrival of SBF via virus. Many tumor viruses activate cell
355 proliferation. For example, the DNA tumor viruses Adenovirus and SV40 highjack cell
356 proliferation in part by activating the expression of E2F-dependent genes by binding
357 pRb to disrupt inhibition of E2F (DeCaprio, 2009). While the specific mechanisms
358 may differ, when SBF entered the fungal ancestor cell it might have activated the
359 transcription of E2F target genes. Rather than inhibiting the inhibitor of E2F, SBF
360 may have directly competed for E2F binding sites with transcriptionally inactive Rb-
361 E2F complexes (Figure 9). Consistent with this model, we have shown here that SBF
362 can directly regulate gene expression in budding yeast via a consensus E2F binding
363 site. Thus, the cooption of a viral protein generated a hybrid network to ultimately
364 facilitate dramatic evolution of the core cell cycle network in fungi.
365

366 **MATERIALS AND METHODS**

367 **Identification of potential protein family homologs.**
368 We used Profile-Hidden Markov Models (profile-HMMs) to detect homologs for each of
369 the families studied, using the HMMER 3 package (Eddy, 2011). Profile-HMMs are
370 sensitive tools for remote homology detection. Starting with a set of diverse yet
371 reliable protein homologs is fundamental for detecting remote protein homology and
372 avoiding "model poisoning" (Johnson et al., 2010). To this end, we used reliable
373 training-set homologs from the cell cycle model organisms *Arabidopsis thaliana,*
374 *Homo sapiens, Schizosaccharomyces pombe,* and *Saccharomyces cerevisiae*, to build
375 the profile-HMMs used to detect homologs. Our profile-HMM search used a stringent
376 e-value threshold of 1e-10 to detect putative homologs in the "best" filtered protein
377 sets (where available) of our 100+ eukaryotic genomes (see Supplementary File 1A
378 for genome details). All putative homologs recovered through a profile-HMM search
379 were further validated (or rejected) using an iterative search algorithm (Jackhmmer)

380 against the annotated SwissProt database using the HMMER web server (Finn et al.,
381 2011).
382
383 Our profile-HMM for E2F/DP family only detects E2F or DP, where as our profile-HMM
384 for SBF/MBF family only detects SBF/MBF (or APSES). The same protein was never
385 identified by both profile-HMMs because the sequence profiles and the structure are
386 different. In the case of basal fungi, which have both E2F/DP and SBF/MBF, all
387 proteins classified as an E2F/DP had clear homology to E2F or DP (see alignment in
388 new Figure 6B) and all proteins that we classified as SBF/MBF had clear homology to
389 SBF/MBF (see alignment in new Figure 6B). Thus, we have not misclassified E2Fs or
390 SBF/MBFs in basal fungi and we have not reached the limit of homology detection.

391 **Phylogenetic-based classification of protein homologs in sub-families.**

392 A phylogenetic analysis and classification was built in four stages. In the first stage,
393 we used MAFFT-L-INS-i (-maxiterate 1000) to align the sequences of eukaryotic
394 protein family members (Katoh and Standley, 2013). We then used probabilistic
395 alignment masking using ZORRO (Wu et al., 2012) to create different datasets with
396 varying score thresholds. Next, we used ProtTest 3 to determine the empirical
397 amino-acid evolutionary model that best fit each of our protein datasets using
398 several criteria: Akaike Information Criterion, corrected Akaike Information Criterion,
399 Bayesian Information Criterion and Decision Theory (Darriba et al., 2011). Last, for
400 each dataset and its best-fitting model, we ran different phylogenetic programs that
401 use maximum-likelihood methods with different algorithmic approximations (RAxML
402 and PhyML) and Bayesian inference methods (PhyloBayes-MPI) to reconstruct the
403 phylogenetic relationships between proteins.
404
405
406 For RAxML analyses, the best likelihood tree was obtained from five independent
407 maximum likelihood runs started from randomized parsimony trees using the
408 empirical evolutionary model provided by ProtTest. We assessed branch support via
409 rapid bootstrapping (RBS) with 100 pseudo-replicates. PhyML 3.0 phylogenetic trees
410 were obtained from five independent randomized starting neighbor-joining trees
411 (RAND) using the best topology from both NNI and SPR moves. Non-parametric
412 Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-aLRTs) and
413 parametric *à la Bayes* aLRTs (aBayes) were calculated to determine branch support
414 from two independent PhyML 3.0 runs. For Bayesian inference we used PhyloBayes

415 (rather than the more frequently used MrBayes) because it allows for site-specific
416 amino-acid substitution frequencies, which better models the level of heterogeneity
417 seen in real protein data (Lartillot and Philippe, 2004; Lartillot et al., 2009). We
418 performed Phylobayes analyses by running three independent chains under CAT and
419 the exchange rate provided by ProtTest 3 (e.g. CAT-LG), four discrete gamma
420 categories, and with sampling every 10 cycles. Proper mixing was initially confirmed
421 with Tracer v1.6 (Rambaut et al., 2014). The first 1000 samples were discarded as
422 burn-in, and convergence was assessed using bipartition frequencies and summary
423 statistics provided by bpcomp and tracecomp from Phylobayes. These were visually
424 inspected with an R version of AWTY (https://github.com/danlwarren/RWTY)
425 (Nylander et al., 2008). The best phylogenies are shown in Figure 3-figure
426 supplement 1-5 and Figure 5-figure supplement 1-3, and were used to tentatively
427 classify sequences into sub-families and create Figures 2-5.
428
429 We note that the confidence of each node in the phylogenetic trees was assessed
430 using multiple, but complementary support metrics: (1) posterior probability for the
431 Bayesian inference, (2) rapid bootstrap support (Stamatakis, 2006; Stamatakis et
432 al., 2008) for RAxML, and (3) non-parametric Shimodaira-Hasegawa-like
433 approximate likelihood ratio tests (SH-aLRTs) and parametric *à la Bayes* aLRTs
434 (aBayes) for PhyML. These different support metrics complement each other in their
435 advantages and drawbacks. SH-aLRT is conservative enough to avoid high false
436 positive rates but performs better compared to bootstrapping (Guindon et al., 2010;
437 Simmons and Norton, 2014). aBayes is powerful compared to non-parametric tests,
438 but has a tendency to increase false-positive rates under serious model violations,
439 something that can be balanced with SH-aLRTs (Anisimova and Gascuel, 2006;
440 Anisimova et al., 2011).
441

442 **Strain construction:**

443

444 Our *CLN2pr-GFP-CLN2PEST* constructs were all derived from pLB02-0mer (described
445 in (Bai et al., 2010) and obtained from Lucy Bai). To create pLB02-CLN2, a synthetic
446 DNA fragment (IDT, Coralville, IA) encompassing a region of the *CLN2* promoter
447 from 1,130 bp to 481 bp upstream of the *CLN2* ORF was digested with BamHI and
448 SphI and ligated into pLB02-0mer digested with the same enzymes. To create
449 pLB02-E2F, which contains E2F binding sites, the same procedure was applied to a

450 version of the promoter fragment in which the SCBs at 606bp, 581bp, and 538bp
451 upstream of the ORF were replaced with the E2F binding site consensus sequence
452 GCGCGAAA (Thalmeier et al., 1989). All these plasmids were linearized at the BbsI
453 restriction site in the *CLN2* promoter and transformed. Both *swi4Δ* and *mbp1Δ*
454 strains containing pLB02-0mer, pLB02-Cln2, pLB02-E2F fluorescent expression
455 reporters were produced via mating lab stocks using standard methods. JE103 was a
456 kind gift from Dr. Jennifer Ewald. Plasmids and strains are listed in Supplementary
457 File 1B and 1C, respectively.

458 **Imaging and analysis:**

459

460 Imaging proceeded essentially as described in Bean *et al*., 2006. Briefly, early log-
461 phase cells were pre-grown in SCD and gently sonicated and spotted onto a SCD
462 agarose pad (at 1.5%), which was inverted onto a coverslip. This was incubated on a
463 heated stage on a Zeiss Observer Z.1 while automated imaging occurred (3 minute
464 intervals, 100-300 ms fluorescence exposures). Single-cell time-lapse fluorescence
465 intensity measurements were obtained using software described in (Doncic and
466 Skotheim, 2013; Doncic et al., 2011), and oscillation amplitudes were obtained
467 manually from the resulting traces. The single-cell fluorescence intensity traces used
468 mean cellular intensity with the median intensity of the entire field of view
469 subtracted, to control for any fluctuations in fluorescent background. The resulting
470 measurements were analyzed in R.

471

472 **ACKNOWLEDGEMENTS**

473 We thank F. Cross and A. Robinson-Mosher for extensive discussions, and R. Gordan,
474 J. Heitman, D. Lew, J. Nevins, S. Rubin for critical comments on the manuscript. We
475 thank J. Stajich for providing access to the Bioinformatics cluster at the Institute for
476 Integrative Genome Biology at UC Riverside supported by the College of Natural and
477 Agricultural Sciences, the National Science Foundation, and Alfred P Sloan
478 Foundation.

479 **References:**

480

481 Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown,
482 M.W., Burki, F., Dunthorn, M., Hampl, V., et al. (2012). The revised classification of
483 eukaryotes. Journal of Eukaryotic Microbiology *59*, 429–493.

484 Andrews, B.J., and Moore, L. (1992). Mutational analysis of a DNA sequence involved
485 in linking gene expression to the cell cycle. Biochem. Cell Biol. *70*, 1073–1080.

486 Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for
487 branches: A fast, accurate, and powerful alternative. Systematic Biology *55*, 539–
488 552.

489 Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., and Gascuel, O. (2011).
490 Survey of branch support methods demonstrates accuracy, power, and robustness of
491 fast likelihood-based approximation schemes. Systematic Biology *60*, 685–699.

492 Aravind, L., and Koonin, E.V. (1999). DNA-binding proteins and evolution of
493 transcription regulation in the archaea. Nucleic Acids Res *27*, 4658–4670.

494 Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., and Iyer, L.M. (2005). The
495 many faces of the helix-turn-helix domain: Transcription regulation and beyond⋆.
496 FEMS Microbiol Rev *29*, 231–262.

497 Bai, L., Charvin, G., Siggia, E.D., and Cross, F.R. (2010). Nucleosome-Depleted
498 Regions in Cell-Cycle-Regulated Promoters Ensure Reliable Gene Expression in Every
499 Cell Cycle. Dev Cell *18*, 544–555.

500 Baker, C.R., Booth, L.N., Sorrells, T.R., and Johnson, A.D. (2012). Protein
501 modularity, cooperative binding, and hybrid regulatory states underlie transcriptional
502 network diversification. Cell *151*, 80–95.

503 Bertoli, C., Skotheim, J.M., and de Bruin, R.A.M. (2013). Control of cell cycle
504 transcription during G1 and S phases. Nat Rev Mol Cell Biol *14*, 518–528.

505 Brick, D.J., Burke, R.D., Schiff, L., and Upton, C. (1998). Shope fibroma virus RING
506 finger protein N1R binds DNA and inhibits apoptosis. Virology *249*, 42–51.

507 Brown, V., Sabina, J., and Johnston, M. (2009). Specialized Sugar Sensing in Diverse
508 Fungi. Curr Biol.

509 Cao, L., Chen, F., Yang, X., Xu, W., Xie, J., and Yu, L. (2014). Phylogenetic analysis
510 of CDK and cyclin proteins in premetazoan lineages. BMC Evol Biol *14*, 10.

511 Cao, L., Peng, B., Yao, L., Zhang, X., Sun, K., Yang, X., and Yu, L. (2010). The
512 ancient function of RB-E2F pathway: insights from its evolutionary history. Biol
513 Direct *5*, 55–.

514 Cavalier-Smith, T. (1999). Principles of Protein and Lipid Targeting in Secondary
515 Symbiogenesis: Euglenoid, Dinoflagellate, and Sporozoan Plastid Origins and the
516 Eukaryote Family Tree. Journal of Eukaryotic Microbiology *46*, 347–366.

517 Costanzo, M., Nishikawa, J.L., Tang, X., Millman, J.S., Schub, O., Breitkreuz, K.,
518 Dewar, D., Rupes, I., Andrews, B., and Tyers, M. (2004). CDK activity antagonizes
519 Whi5, an inhibitor of G1/S transcription in yeast. Cell *117*, 899–913.

520 Coudreuse, D., and Nurse, P. (2010). Driving the cell cycle with a minimal CDK
521 control network. Nature *468*, 1074–1079.

522 Cross, F.R., Buchler, N.E., and Skotheim, J.M. (2011). Evolution of networks and
523 sequences in eukaryotic cell cycle control. Philosophical Transactions of the Royal
524 Society B: Biological Sciences *366*, 3532–3544.

525 Cuomo, C.A., Desjardins, C.A., Bakowski, M.A., Goldberg, J., Ma, A.T., Becnel, J.J.,
526 Didier, E.S., Fan, L., Heiman, D.I., Levin, J.Z., et al. (2012). Microsporidian genome
527 analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res
528 *22*, 2478–2488.

529 Dacks, J.B., and Field, M.C. (2007). Evolution of the eukaryotic membrane-trafficking
530 system: origin, tempo and mode. J Cell Sci *120*, 2977–2985.

531 Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast
532 selection of best-fit models of protein evolution. Bioinformatics *27*, 1164–1165.

533 de Bruin, R.A.M., McDonald, W.H., Kalashnikova, T.I., Yates, J., and Wittenberg, C.
534 (2004). Cln3 activates G1-specific transcription via phosphorylation of the SBF bound
535 repressor Whi5. Cell *117*, 887–898.

536 DeCaprio, J.A. (2009). How the Rb tumor suppressor structure and function was
537 revealed by the study of Adenovirus and SV40. Virology *384*, 274–284.

538 Doncic, A., and Skotheim, J.M. (2013). Feedforward Regulation Ensures Stability and
539 Rapid Reversibility of a Cellular State. Mol Cell.

540 Doncic, A., Falleur-Fettig, M., and Skotheim, J.M. (2011). Distinct interactions select
541 and maintain a specific cell fate. Mol Cell *43*, 528–539.

542 Doonan, J.H., and Kitsios, G. (2009). Functional evolution of cyclin-dependent
543 kinases. Mol Biotechnol *42*, 14–29.

544 Eddy, S.R. (2011). Accelerated profile HMM searches. PLoS Comput Biol *7*,
545 e1002195.

546 Eichinger, L., Pachebat, J.A., Glöckner, G., Rajandream, M.-A., Sucgang, R.,
547 Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q., et al. (2005). The genome
548 of the social amoeba Dictyostelium discoideum. Nature *435*, 43–57.

549 Eme, L., Trilles, A., Moreira, D., and Brochier-Armanet, C. (2011). The phylogenomic
550 analysis of the anaphase promoting complex and its targets points to complex and
551 modern-like control of the cell cycle in the last common ancestor of eukaryotes. BMC
552 Evol Biol *11*, 265.

553 Fang, S.-C., de los Reyes, C., and Umen, J.G. (2006). Cell size checkpoint control by
554 the retinoblastoma tumor suppressor pathway. PLoS Genet *2*, e167.

555 Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger,
556 A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families
557 database. Nucleic Acids Res *42*, D222–D230.

558 Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive
559 sequence similarity searching. Nucleic Acids Res *39*, W29–W37.

560 Fritz-Laylin, L.K., Prochnik, S.E., Ginger, M.L., Dacks, J.B., Carpenter, M.L., Field,
561 M.C., Kuo, A., Paredez, A., Chapman, J., Pham, J., et al. (2010). The genome of
562 Naegleria gruberi illuminates early eukaryotic versatility. Cell *140*, 631–642.

563 Gajiwala, K.S., and Burley, S.K. (2000). Winged helix proteins. Curr Opin Struct Biol
564 *10*, 110–116.

565 Gasch, A.P., Moses, A.M., Chiang, D.Y., Fraser, H.B., Berardini, M., and Eisen, M.B.
566 (2004). Conservation and evolution of cis-regulatory systems in ascomycete fungi.
567 PLoS Biol *2*, e398.

568 Gordân, R., Murphy, K.F., McCord, R.P., Zhu, C., Vedenko, A., and Bulyk, M.L.
569 (2011). Curated collection of yeast transcription factor DNA binding specificity data
570 reveals novel structural and gene regulatory insights. Genome Biol *12*, R125.

571 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O.
572 (2010). New algorithms and methods to estimate maximum-likelihood phylogenies:
573 assessing the performance of PhyML 3.0. Systematic Biology *59*, 307–321.

574 Gunbin, K.V., Suslov, V.V., Turnaev, I.I., Afonnikov, D.A., and Kolchanov, N.A.
575 (2011). Molecular evolution of cyclin proteins in animals and fungi. BMC Evol Biol *11*,
576 224.

577 Guo, Z., and Stiller, J.W. (2004). Comparative genomics of cyclin-dependent kinases
578 suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs. BMC
579 Genomics *5*, 69.

580 Hallmann, A. (2009). Retinoblastoma-related proteins in lower eukaryotes. Commun
581 Integr Biol *2*, 538–544.

582 Harashima, H., Dissmeyer, N., and Schnittger, A. (2013). Cell cycle control across
583 the eukaryotic kingdom. Trends Cell Biol *23*, 345–356.

584 Hasan, M.M., Brocca, S., Sacco, E., Spinelli, M., Papaleo, E., Lambrughi, M.,
585 Alberghina, L., and Vanoni, M. (2013). A comparative study of Whi5 and
586 retinoblastoma proteins: from sequence and structure analysis to intracellular
587 networks. Front Physiol *4*, 315.

588 Huang, D., Kaluarachchi, S., Van Dyk, D., Friesen, H., Sopko, R., Ye, W., Bastajian,
589 N., Moffat, J., Sassi, H., Costanzo, M., et al. (2009). Dual regulation by pairs of
590 cyclin-dependent protein kinases and histone deacetylases controls G1 transcription
591 in budding yeast. PLoS Biol *7*, e1000188.

592 Iyer, L.M., Koonin, E.V., and Aravind, L. (2002). Extensive domain shuffling in
593 transcription regulators of DNA viruses and implications for the origin of fungal
594 APSES transcription factors. Genome Biol *3*, RESEARCH0012.

595 Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed
596 heuristic and iterative HMM search procedure. BMC Bioinformatics *11*, 431.

597 Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software
598 Version 7: Improvements in Performance and Usability. Mol Biol Evol *30*, 772–780.

599 Koonin, E.V., Senkevich, T.G., and Dolja, V.V. (2006). The ancient Virus World and
600 evolution of cells. Biol Direct *1*, 29.

601 Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site
602 heterogeneities in the amino-acid replacement process. Mol Biol Evol *21*, 1095–1109.

603 Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian
604 software package for phylogenetic reconstruction and molecular dating.
605 Bioinformatics *25*, 2286–2288.

606 Lavoie, H., Hogues, H., and Whiteway, M. (2009). Rearrangements of the
607 transcriptional regulatory networks of metabolic pathways in fungi. Curr Opin
608 Microbiol *12*, 655–663.

609 Liu, J., and Kipreos, E.T. (2000). Evolution of cyclin-dependent kinases (CDKs) and
610 CDK-activating kinases (CAKs): differential conservation of CAKs in yeast and
611 metazoa. Mol Biol Evol *17*, 1061–1074.

612 Liu, J., Huang, J., Zhao, Y., Liu, H., Wang, D., Yang, J., Zhao, W., Taylor, I.A., and
613 Peng, Y.-L. (2015). Structural basis of DNA recognition by PCG2 reveals a novel DNA
614 binding mode for winged helix-turn-helix domains. Nucleic Acids Res *43*, 1231–1240.

615 Ma, Z., Wu, Y., Jin, J., Yan, J., Kuang, S., Zhou, M., Zhang, Y., and Guo, A.-Y.
616 (2013). Phylogenetic analysis reveals the evolution and diversification of cyclins in
617 eukaryotes. Molecular Phylogenetics and Evolution *66*, 1002–1010.

618 Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman,
619 G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Maréchal-Drouard, L., et al. (2007).
620 The Chlamydomonas genome reveals the evolution of key animal and plant
621 functions. Science *318*, 245–250.

622 Moffat, J., and Andrews, B. (2004). Late-G1 cyclin-CDK activity is essential for
623 control of cell morphogenesis in budding yeast. Nat Cell Biol *6*, 59–66.

624 Murray, A.W., and Kirschner, M.W. (1989). Cyclin synthesis drives the early
625 embryonic cell cycle. Nature *339*, 275–280.

626 Nasmyth, K. (1995). Evolution of the cell cycle. Philos Trans R Soc Lond, B, Biol Sci
627 *349*, 271–281.

628 Nasmyth, K., and Dirick, L. (1991). The role of SWI4 and SWI6 in the activity of G1
629 cyclins in yeast. Cell *66*, 995–1013.

630 Nutiu, R., Friedman, R.C., Luo, S., Khrebtukova, I., Silva, D., Li, R., Zhang, L.,
631 Schroth, G.P., and Burge, C.B. (2011). Direct measurement of DNA affinity
632 landscapes on a high-throughput sequencing instrument. Nat Biotechnol *29*, 659–
633 664.

634 Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., and Swofford, D.L. (2008). AWTY
635 (are we there yet?): a system for graphical exploration of MCMC convergence in
636 Bayesian phylogenetics. Bioinformatics *24*, 581–583.

637 Ogas, J., Andrews, B.J., and Herskowitz, I. (1991). Transcriptional activation of

638 CLN1, CLN2, and a putative new G1 cyclin (HCS26) by SWI4, a positive regulator of
639 G1-specific transcription. Cell *66*, 1015–1026.

640 Rabinovich, A., Jin, V.X., Rabinovich, R., Xu, X., and Farnham, P.J. (2008). E2F in
641 vivo binding specificity: Comparison of consensus versus nonconsensus binding sites.
642 Genome Res *18*, 1763–1777.

643 Rambaut, A., Suchard, M.A., Xie, D., and Drummond, A.J. (2014). Tracer v1.6,
644 Available from http://beast.bio.edu.ac.uk/Tracer.

645 Sidorova, J., and Breeden, L. (1993). Analysis of the SWI4/SWI6 protein complex,
646 which directs G1/S-specific transcription in Saccharomyces cerevisiae. Mol Cell Biol
647 *13*, 1069–1077.

648 Simmons, M.P., and Norton, A.P. (2014). Divergent maximum-likelihood-branch-
649 support values for polytomies. Molecular Phylogenetics and Evolution *73*, 87–96.

650 Skotheim, J.M., Di Talia, S., Siggia, E.D., and Cross, F.R. (2008). Positive feedback
651 of G1 cyclins ensures coherent cell cycle entry. Nature *454*, 291–296.

652 Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic
653 analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.

654 Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm
655 for the RAxML Web servers. Systematic Biology *57*, 758–771.

656 Stern, B., and Nurse, P. (1996). A quantitative model for the cdc2 control of S phase
657 and mitosis in fission yeast. Trends Genet *12*, 345–350.

658 Tanay, A., Regev, A., and Shamir, R. (2005). Conservation and evolvability in
659 regulatory networks: The evolution of ribosomal regulation in yeast. Proc Natl Acad
660 Sci USA.

661 Taylor, I.A., Treiber, M.K., Olivi, L., and Smerdon, S.J. (1997). The X-ray structure of
662 the DNA-binding domain from the Saccharomyces cerevisiae cell-cycle transcription
663 factor Mbp1 at 2.1 A resolution. J Mol Biol *272*, 1–8.

664 Thalmeier, K., Synovzik, H., and Mertz, R. (1989). Nuclear factor E2F mediates basic
665 transcription and trans-activation by E1a of the human MYC promoter. Genes & ….

666 Travesa, A., Kalashnikova, T.I., de Bruin, R.A.M., Cass, S.R., Chahwan, C., Lee, D.E.,
667 Lowndes, N.F., and Wittenberg, C. (2013). Repression of G1/S Transcription Is
668 Mediated via Interaction of the GTB Motifs of Nrm1 and Whi5 with Swi6. Mol Cell Biol
669 *33*, 1476–1486.

670 Tsong, A.E., Tuch, B.B., Li, H., and Johnson, A.D. (2006). Evolution of alternative
671 transcriptional circuits with identical logic. Nature *443*, 415–420.

672 van den Heuvel, S., and Dyson, N.J. (2008). Conserved functions of the pRB and E2F
673 families. Nat Rev Mol Cell Biol *9*, 713–724.

674 Wang, G., Kong, H., Sun, Y., Zhang, X., Zhang, W., Altman, N., DePamphilis, C.W.,
675 and Ma, H. (2004). Genome-wide analysis of the cyclin family in Arabidopsis and

676 comparative phylogenetic analysis of plant cyclin-like proteins. Plant Physiol *135*,
677 1084–1099.

678 Wang, H., Carey, L.B., Cai, Y., Wijnen, H., and Futcher, B. (2009). Recruitment of
679 Cln3 cyclin to promoters controls cell cycle entry via histone deacetylase and other
680 targets. PLoS Biol *7*, e1000189.

681 Wu, M., Chatterji, S., and Eisen, J.A. (2012). Accounting for alignment uncertainty in
682 phylogenomics. PLoS ONE *7*, e30288.

683 Xu, R.M., Koch, C., Liu, Y., Horton, J.R., Knapp, D., Nasmyth, K., and Cheng, X.
684 (1997). Crystal structure of the DNA-binding domain of Mbp1, a transcription factor
685 important in cell-cycle control of DNA synthesis. Structure *5*, 349–358.

686 Yutin, N., Wolf, Y.I., Raoult, D., and Koonin, E.V. (2009). Eukaryotic large nucleo-
687 cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral
688 genome evolution. Virol. J. *6*, 223.

689 Zheng, N., Fraenkel, E., Pabo, C.O., and Pavletich, N.P. (1999). Structural basis of
690 DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. Genes
691 Dev *13*, 666–674.

692

693
694

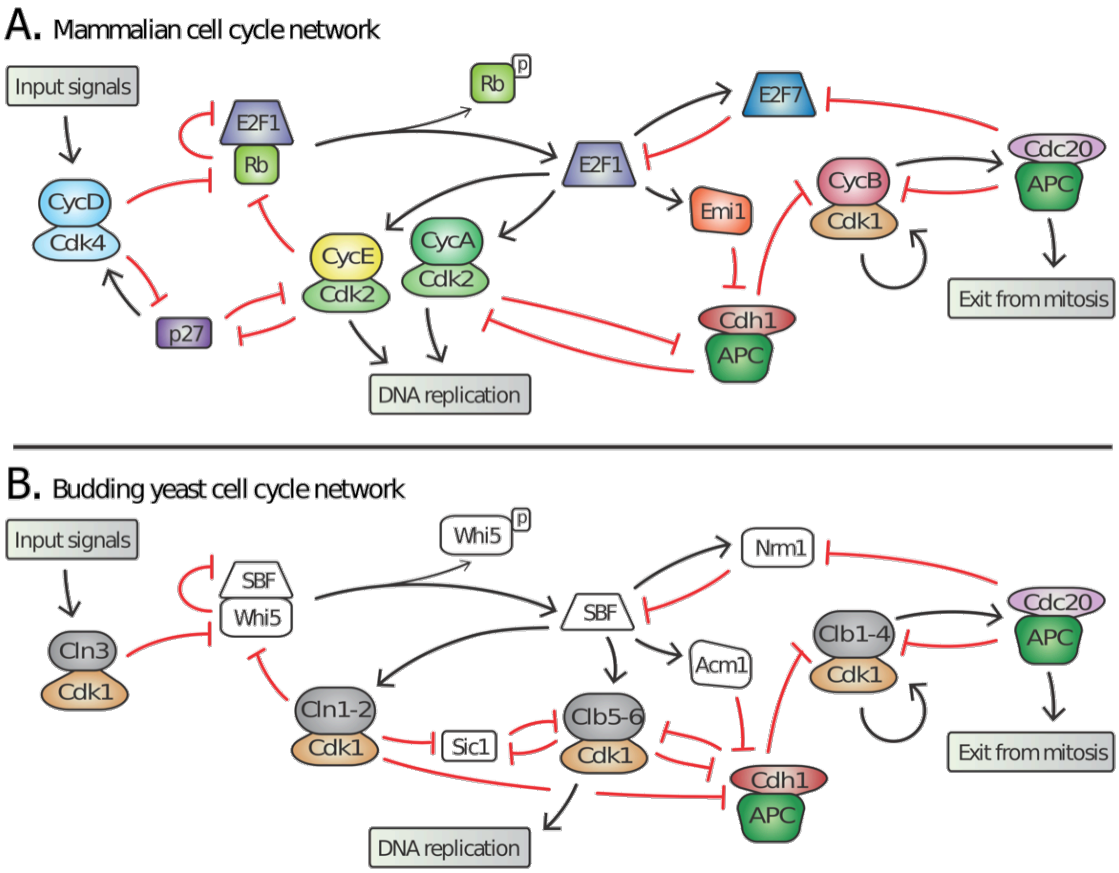


Figure 1. Topology of G1/S regulatory network in mammals and budding yeast is conserved, yet many regulators exhibit no detectable sequence homology. Schematic diagram illustrating the extensive similarities between animal (A) and budding yeast (B) G1/S cell cycle control networks. Similar coloring denotes members of a similar family or sub-family. Fungal components colored white denote proteins with no identifiable animal orthologs.

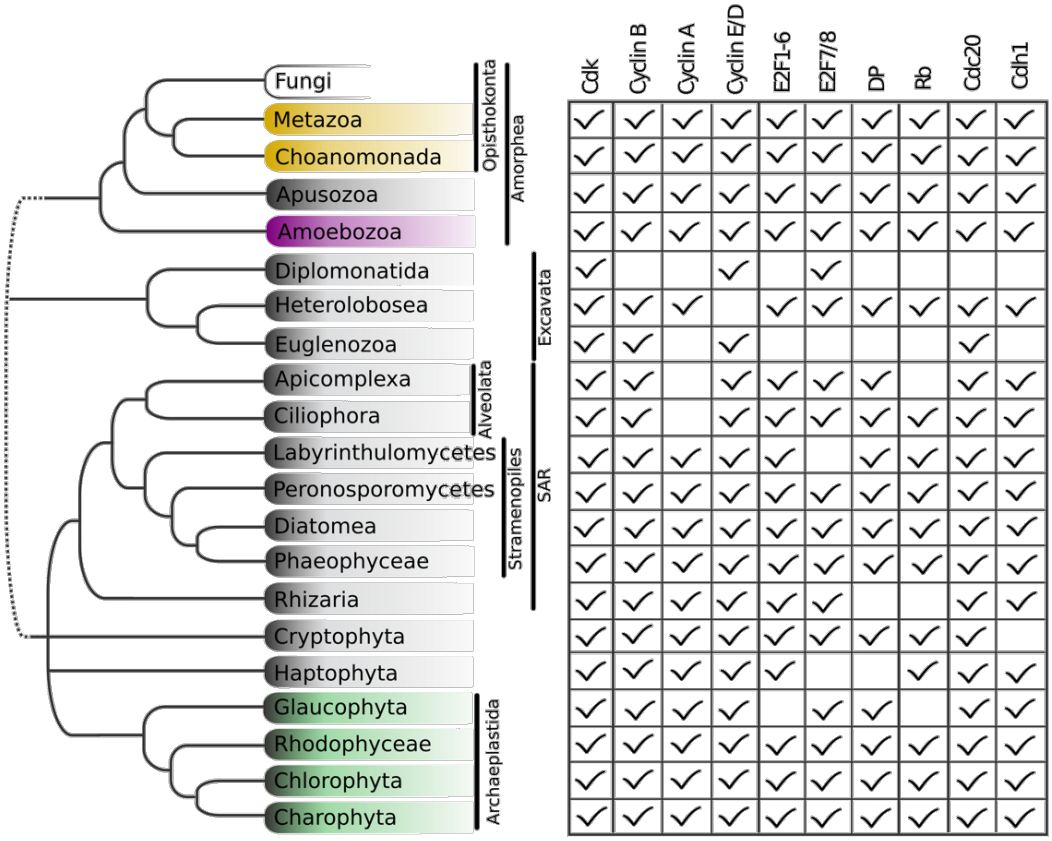


Figure 2. Animal and plant G1/S regulatory network was present in the last eukaryotic common ancestor. Distribution of cell cycle regulators across the eukaryotic species tree (Adl et al., 2012). Animals (Metazoa) and yeasts (Fungi) are sister groups (Opisthokonta), and are distantly related to plants (Charophyta), which are members of the Archaeplastida. Check marks indicate the presence of at least one member of a protein family in at least one sequenced species from the corresponding group. See Figure 3 for a complete list of components in all species analyzed.

			(CycB3)		(CycO/F/J/G/I)						
	(H. sapiens)	**Cdk1-4,6**	**CycB**	**CycA**	**CycE/D**	**E2F1-6**	**E2F7-8**	**DP**	**Rb**	**Cdc20**	**Fzr1**
Metazoa	*H. sapiens (human)*	5	3	2	13	6	2	3	3	2	1
	G. gallus (chicken)	4	2	2	13	6	2	2	3	2	1
	D. rerio (zebrafish)	5	3	2	12	5	2	3	3	1	2
	B. floridae (lancelet)	5	1	1	6	2	2	1	2	1	2
	C. intestinalis (sea squirt)	5	1	2	5		1	1	1	1	1
	S. purpuratus (urchin)	3	2	1	6	2		1	3	2	1
	L. gigantea (sea snail)	3	2	1	7	2	1	1	2	2	1
	D. melanogaster (fly)	3	2	1	3	2		1	2	1	2
	C. elegans (nematode)	3	4	1	2	2	1	1	1	1	1
	N. vectensis (anemone)	3	3	1	5	5	3	1	1	2	1
	A. queenslandica (sponge)	2	1	1	5	2	1	1	3	2	1
	T. adhaerens (placozoa)	3	2	1	3	2	1	1	2	1	1
Choanomonada	*Sphaeroforma arctica*		1		3	2		2	1	1	1
	Capsaspora owczarzaki		1	1	4	1	1	1	1	1	1
	Monosiga brevicollis	1	1	1	2	1	1	1	1	1	1
	Salpingoeca rosetta	2	1	1	2	2	1	1	2	1	1
Apusozoa	*Thecamonas trahens*	1	1	1	3	1	1	1	1	1	1
Amoebozoa	*Dictyostelium discoideum*	1	1	1	1	1		1	1	1	1
	Dictyostelium purpureum	1	1	1	1	1		1	1	1	1
	Dictyostelium fasciculatum	1	1	1	1	1	1	1	1	1	1
	Polysphondylium pallidum	1	1	1		1	2		1	1	1
	Entamoeba histolytica	6	2		3						2
	Entamoeba nuttalli	6	1		1						2
Diplomonatida	*Giardia intestinalis*	2			2		1				
Heterolobosea	*Naegleria gruberi*	3	3	3		1	1	1	1	3	1
Euglenozoa	*Trypanosoma brucei*	3	1		2					3	
	Leishmania major	3	2		1					2	
	Leishmania donovani	2	2		1					2	
Apicomplexa	*Symbiodinium minutum*	7	2							1	1
	Plasmodium falciparum	2									1
	Cryptosporidium muris	3			2	1	1	1			1
	Toxoplasma gondii	2									1
Ciliophora	*Tetrahymena thermophila*	7	5		11	2	2	3	2	1	8
Labyrinthulomycetes	*Aurantiochytrium limacinum*	3	1	1	3			1	1	1	1
	Schizochytrium aggregatum	3	1	1	2					1	1
	Aplanochytrium kerguelense	3	1	1	2	1		1	1	2	1
Peronosporomycetes	*Phytophthora infestans*	5	1	1	3	1	1	1	1	1	1
	Pythium ultimatum	1	1	1	4	1	1	1	1	1	1
	Hyaloperonospora parasitica	3		2	4			2	1	1	
	Saprolegnia parasitica	3	1	2	4	1	1	2	1	4	1
Diatomea	*Fragilariopsis cylindrus*	2	4	1	10	1	1	1		2	1
	Phaeodactylum tricornutum	2	1	1	10	2	1	1	1	2	1
	Thalassiosira pseudonana	2	3	1	26	1	1	1	1	2	2
Eustigmatales	*Nannochloropsis gaditana*	3	2	1	2			1	1	1	
Phaeophyceae	*Ectocarpus siliculosus*	3	2	1	2	1	1	1	1	2	1
Pelagophyceae	*Aureococcus anophaefferens*	2	2		2	1		1		2	1
Rhizaria	*Bigelowiella natans*	1	2	1	3	1	2			2	1
Cryptophyta	*Guillardia theta*	4	3	1	3	1	1	2	1	3	
Haptophyta	*Emiliania huxleyi*	2	2	3	1	2			1	6	2
Glaucophyta	*Cyanophora paradoxa*	1	1	1	1		1	1		1	2
Rhodophyceae	*Porphyridium cruentum*	3	1	1	1	1	1	1		1	1
	Cyanidioschyzon merolae	2		1	2	1	1	2	1	1	1
Chlorophyta	*Ostreococcus tauri*	2	1	1	2	1	1	1	1	1	1
	Ostreococcus lucimarinus	2	1	1	2	1	1	1	1	1	1
	Micromonas pusilla	2	1	1	2	1	1	1	1	1	1
	Coccomyxa subellipsoidea	4	1	1	1	1	1	1	1	1	1
	Volvox carteri	3	1	1	4	1		1	1	1	1
	Chlamydomonas reinhardtii	3	1	1	4	1		1	1	1	1
Charophyta	*Physcomitrella patens*	9	2	8	2	3	2	2	3	6	4
	Selaginella moellendorffii	3	1	3	3	2	1	1	2	3	2
	Brachypodium distachyon	5	5	6	14	3	2	2	2	5	2
	Oryza sativa	5	6	6	12	4	2	2	2	3	2
	Arabidopsis thaliana	5	11	10	11	3	3	2	1	6	3
	(A. thaliana)	**CdkA/B**	**CycB**	**CycA**	**CycD**	**E2FA-C**	**DEL**	**DP**	**RBR**	**Cdc20**	**Fzr1**
					(CycSDS)						

Figure 3. Comparative genomic data of G1/S regulators across eukaryotes. We developed profile-HMMs to detect cell division cycle regulators in eukaryotic genomes. For each cell cycle regulatory family (*e.g.*, cyclins), we used molecular phylogeny to classify eukaryotic sequences into sub-families (*e.g.*, Cyclins B, Cyclin A, Cyclins E/D). See Methods for details and Figure 3-supplement 1 (Cyclin), Figure 3-supplement 2 (E2F/DP), Figure 3-supplement 3 (pRb), Figure 3-supplement 4 (Cdc20-family), and Figure 3-supplement 5 (CDK) for final phylogenies. Each entry lists the number of sub-family members (column) for each eukaryotic genome (row). Grey rows list the sub-family gene names in *H. sapiens* and *A. thaliana*. Additional cyclin sub-family members are listed in parentheses.

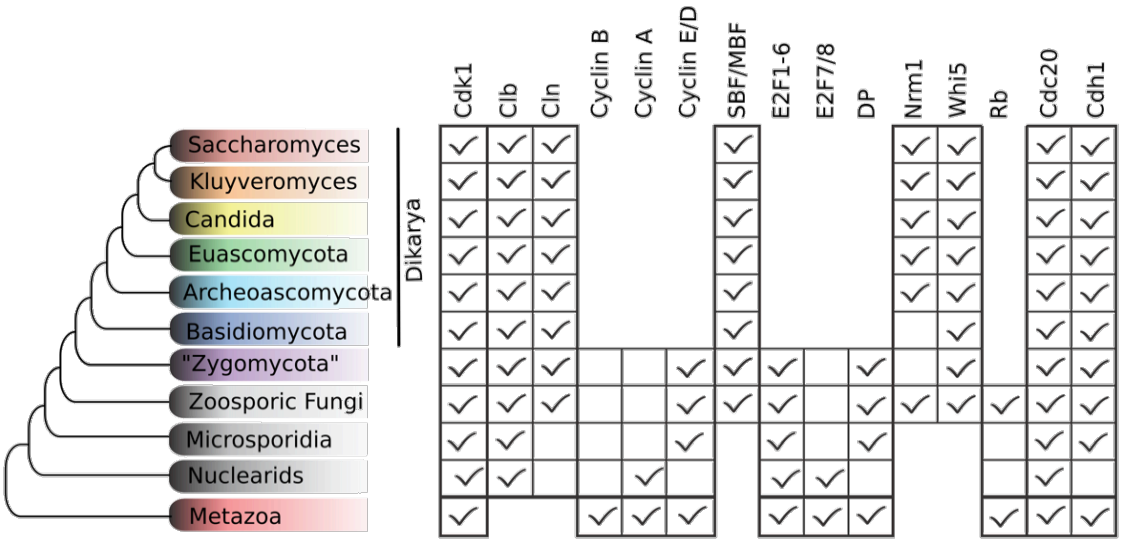


Figure 4. Fungal ancestor evolved novel G1/S regulators, which eventually replaced ancestral cyclins, transcription factors, and inhibitors in Dikarya. Basal fungi and "Zygomycota" contain hybrid networks comprised of both ancestral and fungal specific cell cycle regulators. Check marks indicate the presence of at least one member of a protein family in at least one sequenced species from the group. See Figure 5 for a complete list of components in all fungal species analyzed.

(S. cerevisiae)	CLB	CLN				SBF/MBF	APSES	Xbp1					Nrm1	Whi5	
Saccharomyces cerevisiae	6	3				3	2	1					1	2	
Saccharomyces mikatae	6	3				3	2	1					1	2	
Saccharomyces bayanus	6	3				3	2	1					1	2	
Candida glabrata	6	3				3	2	1					1	1	
Zygosaccharomyces rouxii	3	3				3	1	1					1	1	
Kluyveromyces waltii	3	4				3	1	1					1	1	
Kluyveromyces thermotolerans	3	3				3	1	1					1	1	
Saccharomyces kluyveri	3	3				3	1	1					1	1	
Ashbya gossypii	3	2				3	1	1					1	1	
Kluyveromyces lactis	3	3				3	1	1					1	1	
Wickerhamomyces anomalus	3	3				3	2	1					1	1	
Candida parapsilosis	2	3				3	2	1					1		
Candida albicans	2	3				3	2	1					1		
Candida tropicalis	2	3				3	1	1					1		
Candida guilliermondii	2	3				3	2	1					1	1	
Debaryomyces hansenii	2	3				3	2	1					1		
Candida lusitaniae	2	4				3	2	1					1		
Yarrowia lipolytica	2	2				2	2	1					1	1	
Neurospora crassa	2	1				2	1	1	1				1	2	
Podospora anserina	2	1				2	1	1	1				1	2	
Magnaporthe grisea	2	1				2	1	1	1				1	1	
Fusarium graminearum	2	1				2	1	1	1				1	2	
Cladonia grayi	2	1				2	1	1	1				1	2	
Aspergillus nidulans	2	1				2	1	1	1					1	
Cocciodioides immitis	2	1				2	1	1	1					2	
Saitoella complicata	2	1				2	1	1	1				1	2	
Pneumocystis jirovecii	2	1				2	1		1						
Taphrina deformans	2	1				2	1	1	1					1	
Schizosaccharomyces pombe	4	1				3			1				1	1	
Schizosaccharomcyes octosporus	4					3			1				1	1	
Schizosaccharomyces japonicus	4	1				3			1				1	1	
(S. pombe)	Cdc13	Puc1				Cdc10/Res	APSES		Bqt4					Whi5	
Coprinopsis cinerea	3	1				2		1	1					1	
Laccaria bicolor	3	1				2		1	1					2	
Schizophyllum commune	2	1				2		1	1						
Phanerochaete chrysosporium	2	1				2			1					2	
Cryptococcus neoformans	2	1				2			1					1	
Ustilago maydis	2	1				2	1	1	1					2	
Puccinia graminis	2	1				2	1	1	1					1	
Mortierella alpina	3	1			1	4	1	1	2	1		2		3	
Mortierella verticillata	3	1			1	3	1	1	2	1		1		1	
Rhizophagus irregularis	1					2	1	1		1		1			
Umbelopsis ramanniana	1				1	3	3	1		1				2	
Lichtheimia hyalospora	1				1	4	9	1		2		3		5	
Mucor circineolloides	1					4	6	2		1		2		5	
Phycomyces blakesleeanus	1					4	6	1		2		2		4	
Rhizopus oryzae	1					6	10	1		2		1		4	
Coemansia reversa	4				1	2	1								
Conidiobolus coronatus	1				2	1	2	1							
Allomyces macrogynus	1				4	6	3								
Catenaria anguillulae	1				1	2	1								
Gonapodya prolifera	3	1			8	2									1
Piromyces sp E2	3	1				1	2	1							
Batrachochytrium dendrobatidis	2	1			2	3							1	1	1
Spizellomyces punctatus	1	1			3	3				1		1		2	1
Rozella allomycis	1				2	1				1		1			1
Encephalitozoon cuniculi	1				1					1		1			
Nosema ceranae	1				1					1		1			
Edhazardia aedis	1				1					1		1			
Vittaforma corneae	1				1					1		1			
Nematocida parisii	1				1					1		1			
Fonticula alba	2				2						1	1			
Thecamonas trahens			1	1	3					1	1	1			1
Capsaspora owczarzaki			1	1	4					1	1	1			1
H. sapiens (human)			3	2	13					6	2	3			3
(H. sapiens)			CycB	CycA	CycE/D					E2F1-6	E2F7-8	DP			pRB

Figure 5. Comparative genomic data of G1/S regulators across fungi. We developed profile-HMMs to detect fungal-specific cell division cycle regulators in eukaryotic genomes. For each cell cycle regulatory family (*e.g.*, SBF/APSES), we used molecular phylogeny to classify eukaryotic sequences into sub-families (*e.g.*, SBF/MBF, APSES, Xbp1, Bqt4). Corresponding table shows the number of regulators of each class for all species analyzed. See Methods for details and Figure 5-supplement 1 (SBF only), Figure 5-supplement 2 (SBF+APSES), and Figure 5-supplement 3 (Whi5) for final phylogenies. Grey rows list the sub-family gene names in *S. cerevisiae*, *S. pombe*, and *H. sapiens*

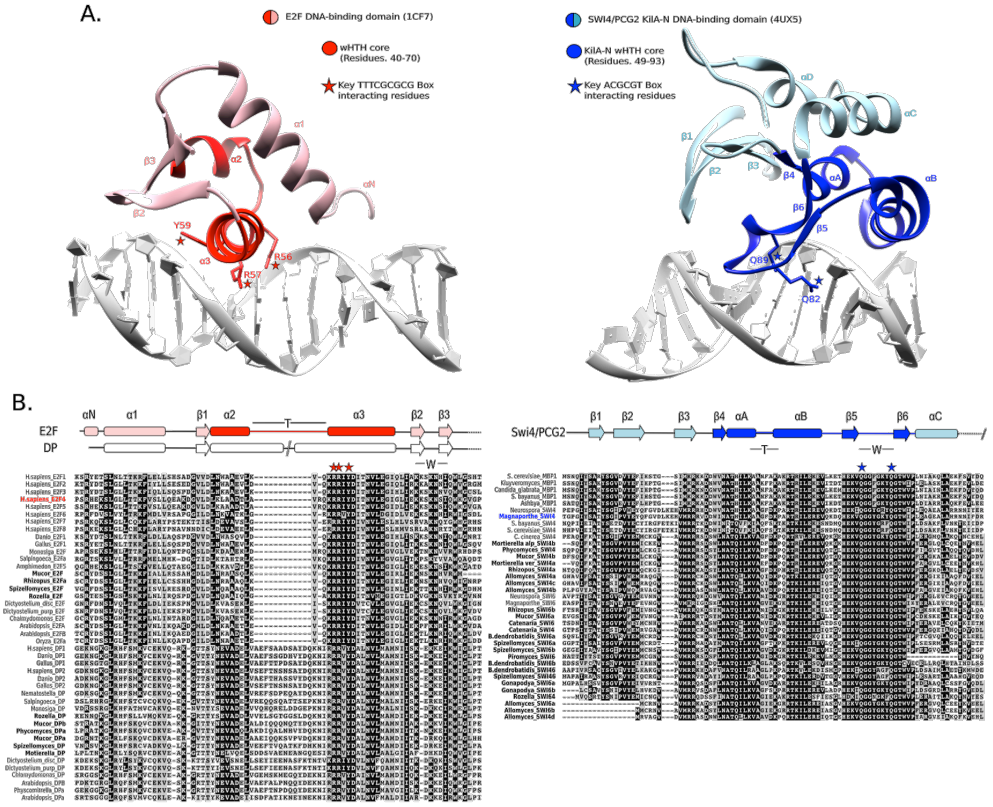
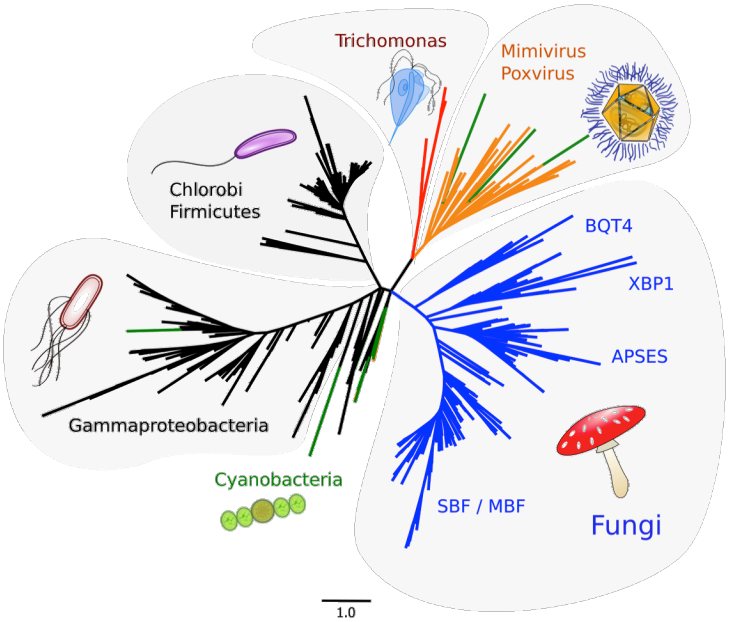


Figure 6. E2F and SBF show incongruences in sequence, structure, and mode of DNA binding. (A) Although both proteins share a winged helix-turn-helix (wHTH) domain, the E2F/DP and SBF/MBF superfamilies do not exhibit significant sequence identity or structural similarity to suggest a common recent evolutionary origin according to CATH or SCOP databases. Furthermore, each wHTH has a different mechanism of interaction with DNA: the arginine and tyrosine side-chains of recognition helix-3 of E2F (E2F4 from *Homo sapiens* (Zheng et al., 1999)) interact with specific CG nucleotides, where as the glutamine side-chains of the "wing" of SBF/MBF (PCG2 from *Magnaporthe oryzae* (Liu et al., 2014)) interact with specific CG nucleotides. (B) Sequence alignment of the DNA binding domain of representative eukaryotic E2F/DP (left) and fungal SBF/MBF (right). The corresponding secondary structure is above the sequence alignment. Evolutionary conserved residues of sequence aligned DNA binding domains are highlighted in black. Bold sequence names correspond to E2F/DP and SBF/MBF sequences from basal fungi. Colored sequence names correspond to sequences of the structures shown in panel A. PDB IDs for the structures used are shown in parentheses. W = wing; T= turn.



765
766 **Figure 7. Viral origin of yeast cell cycle transcription factor SBF.** Maximum
767 Likelihood unrooted phylogenetic tree depicting relationships of fungal SBF-family
768 proteins (blue), KilA-N domains in prokaryotic (black, green), and eukaryotic DNA
769 viruses (orange, red). The particular origin of the HGT event that originated the
770 SBF/APSES family is unclear. We found no significant rapid bootstrapping support
771 (RBS ≥80%) for internal branches. Inference was performed under GTR+G model of
772 evolution. Scale bar in substitutions per site; (See Methods).
773

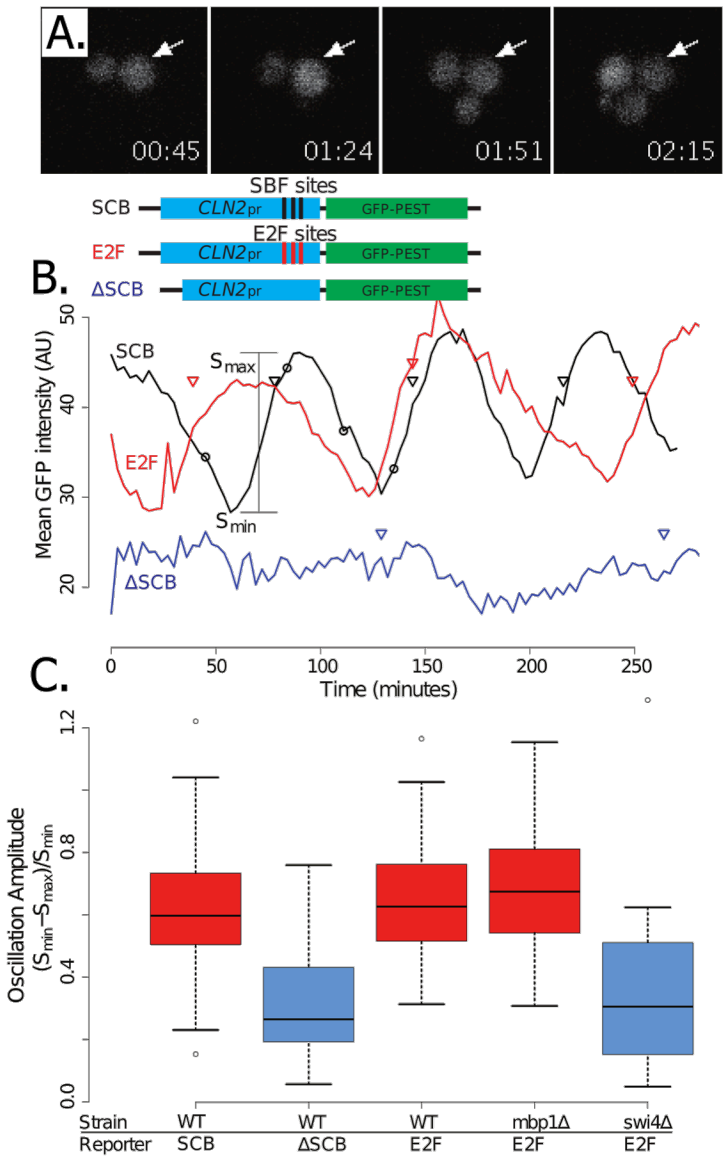


Figure 8. Yeast cell cycle transcription factor SBF can regulate cell cycle-dependent transcription via E2F binding sites (A) Fluorescence images of cells expressing a destabilized GFP from the SBF-regulated *CLN2* promoter. (B) Oscillation of a transcriptional reporter in budding yeast. Characteristic time series of GFP expression from a *CLN2* promoter (SCB), a *CLN2* promoter where the SBF binding sites were deleted (ΔSCB), or a *CLN2* promoter where the SBF binding sites were replaced with consensus E2F binding sites (E2F). Oscillation amplitudes were quantified by scaling the mean fluorescence intensity difference from peak to trough divided by the trough intensity $(S_{max} - S_{min})/S_{min}$. Circles denote time points corresponding to (b). Triangles denote budding events. (C) Distribution of oscillation amplitudes for different genotypes and GFP reporters. *swi4Δ* and *mbp1Δ* strains have deletions of the SBF and MBF DNA-binding domain subunits respectively. t-test comparisons within and across red and blue categories yield p-values > 0.3 or < 0.01 respectively. Boxes contain 25th, median and 75th percentiles, while whiskers extend to 1.5 times this interquartile range.

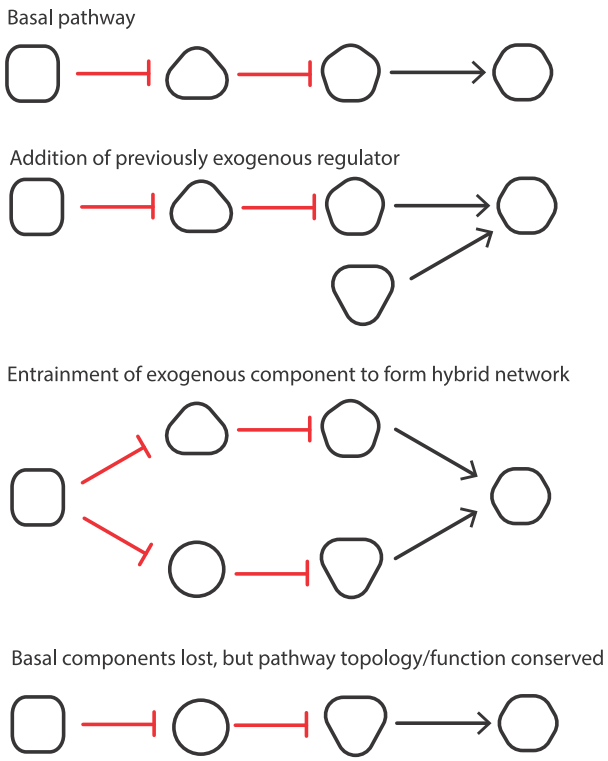


Figure 9. Punctuated evolution of a conserved regulatory network. Evolution can replace components in an essential pathway by proceeding through a hybrid intermediate. Once established, the hybrid network can evolve dramatically and lose previously essential regulators, while sometimes retaining the original network topology.

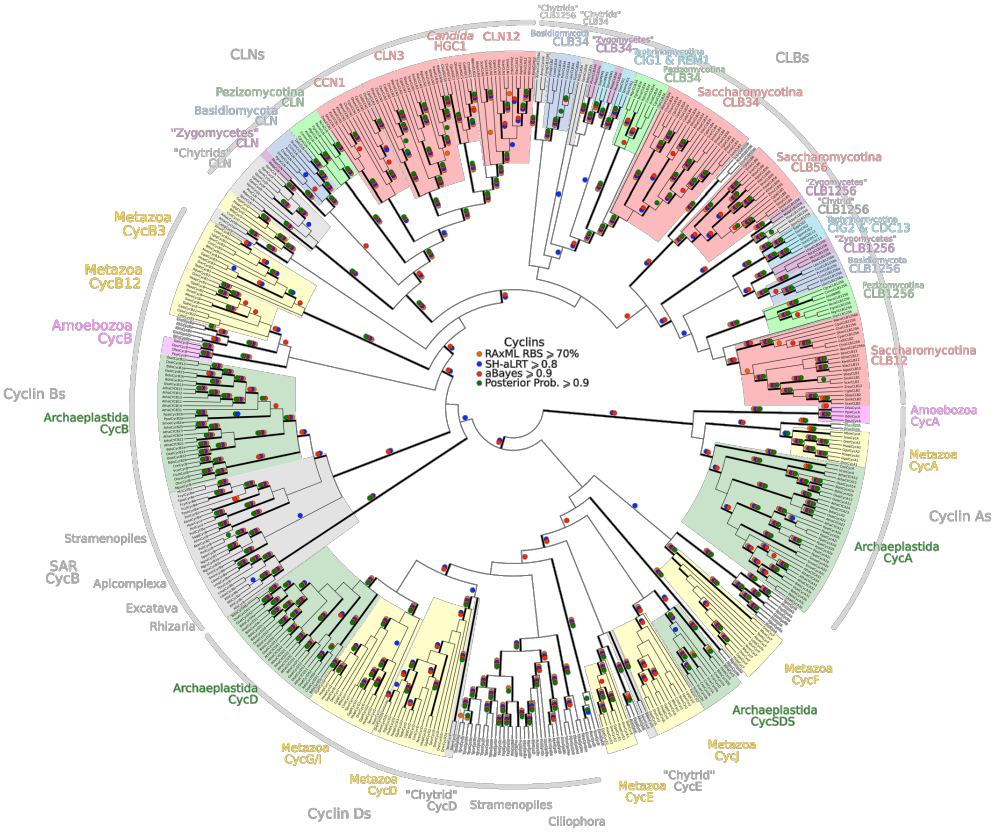


Figure 3 - supplement 1: Reduced phylogeny of eukaryotic cell cycle cyclins. The cell division cycle (CDC) cyclin family consists of several sub-families with a well-characterized cyclin box: (1) CycA (*H. sapiens, A. thaliana*), (2) CycB and CLB (*H. sapiens, A. thaliana, S. cerevisiae, S. pombe*), (3) CycD (*H. sapiens, A. thaliana*), (4) CycE (*H. sapiens*), and (5) CLN (*S. cerevisiae, S. pombe*). We combined CycA, CycB, CycD, CycE, CLB, and CLN sequences from *H. sapiens* (10 cyclins)*, A. thaliana* (13 cyclins)*, S. cerevisiae* (9 cyclins)*,* and *S. pombe* (5 cyclins) to create a eukaryotic CDC cyclin profile-HMM (pCYC.hmm) following the procedure outlined in the Methods. Our HMM profile was sensitive enough to discover known, but uncharacterized cyclin sub-families (CycO, CycF, CycG, CycI, CycJ, CycSDS) as *bona fide* CDC cyclins. A domain threshold of E-20 was used to identify potential CDC cyclin homologs. We first made a phylogeny of all cyclins to classify them. This dataset was then manually pruned to remove long-branches and problematic lineages. Our reduced CDC cyclin dataset has a total of 499 sequences. Columns with the top 10% Zorro score (496 positions) were used in our alignment. Confidence at nodes was assessed with multiple support metrics using different phylogenetic programs under LG model of evolution (aBayes and SH-aLRT metrics with PhyML, RBS with RAxML, Bayesian Posterior Probability with Phylobayes (23,163 sampled trees, meandiff= 0.01, maxdiff= 0.5)). Colored dots in branches indicate corresponding branch supports. Thick branches indicate significant support by at least two metrics, one parametric and one non-parametric; branch support thresholds are shown in the center of the figure; see Methods.

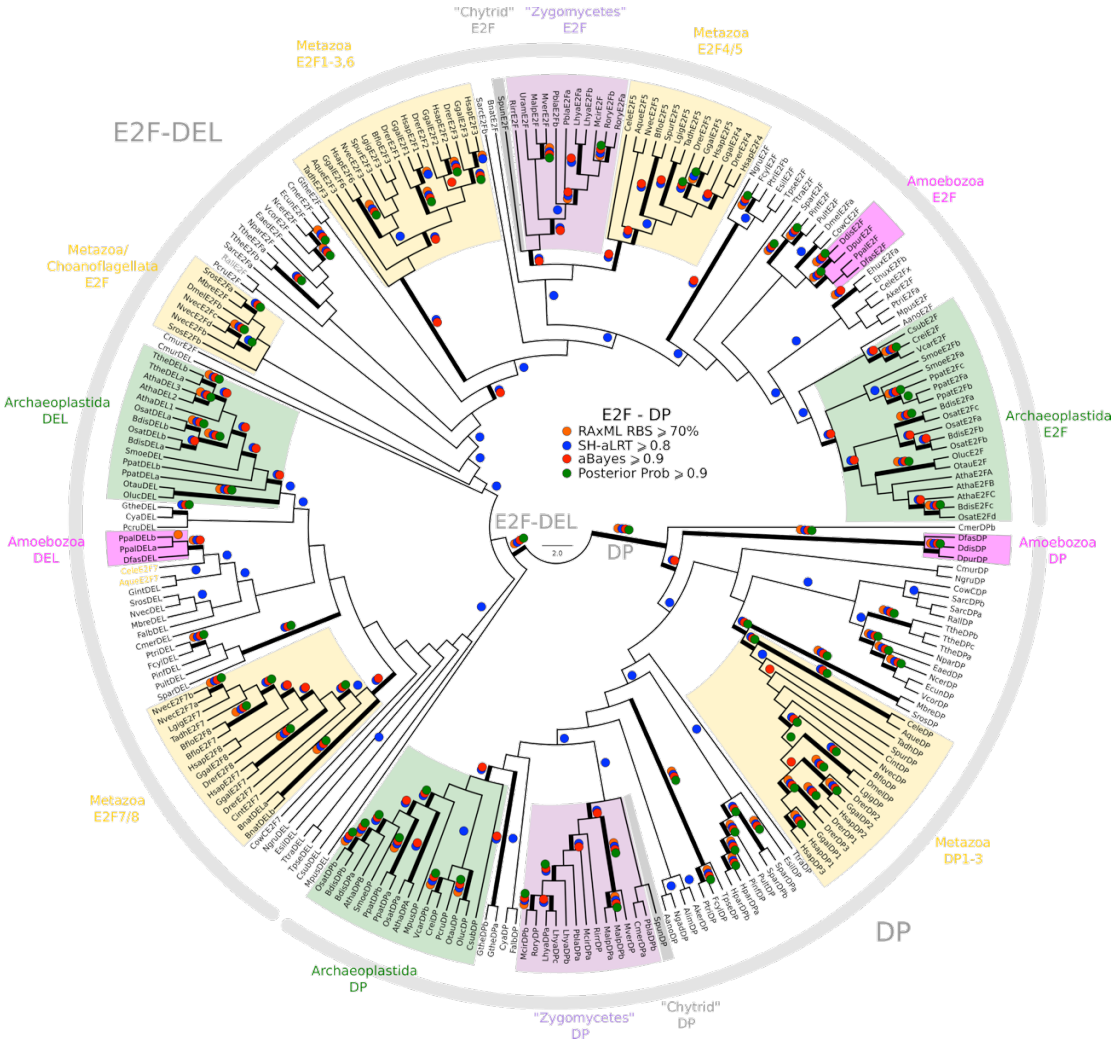


Figure 3 - supplement 2: Phylogeny of eukaryotic E2F/DP transcription factors. E2F-DP is a winged helix-turn-helix DNA-binding domain that is conserved across eukaryotes (van den Heuvel and Dyson, 2008). There are three sub-families within the E2F-DP family: (1) the E2F subfamily, (2) the E2F7-8/DEL subfamily, and (3) the DP subfamily. The E2F family consists of E2F1-6 (*H. sapiens*) and E2FA-C (*A. thaliana*). The E2F7-8/DEL family consists of E2F7-8 (*H. sapiens*) and DEL1-3 (*A. thaliana*). The DP family consists of DP1-3 (*H. sapiens*) and DPA-B (*A. thaliana*). The members of E2F form heterodimers with DP, whereas the DEL family has two DNA-binding domains and does not require DP to bind DNA. We used the E2F_TDP.hmm profile from PFAM to uncover members of the E2F/DP family across eukaryotes. A domain threshold of E-10 was used to identify potential E2F/DP homologs. Our E2F/DP dataset (pE2FDP.fasta) has 248 sequences. Columns with the top 8% Zorro score (284 positions) were used in our alignment. Confidence at nodes was assessed with multiple support metrics using different phylogenetic programs under LG model of evolution (aBayes and SH-aLRT metrics with PhyML, RBS with RAxML, Bayesian Posterior Probability with Phylobayes (53,009 sampled trees, meandiff=0.0064, maxdiff=0.18)). Colored dots in branches indicate corresponding branch supports. Thick branches indicate significant support by at least two metrics, one parametric and one non-parametric; branch support thresholds are shown in the center of the figure; see Methods.

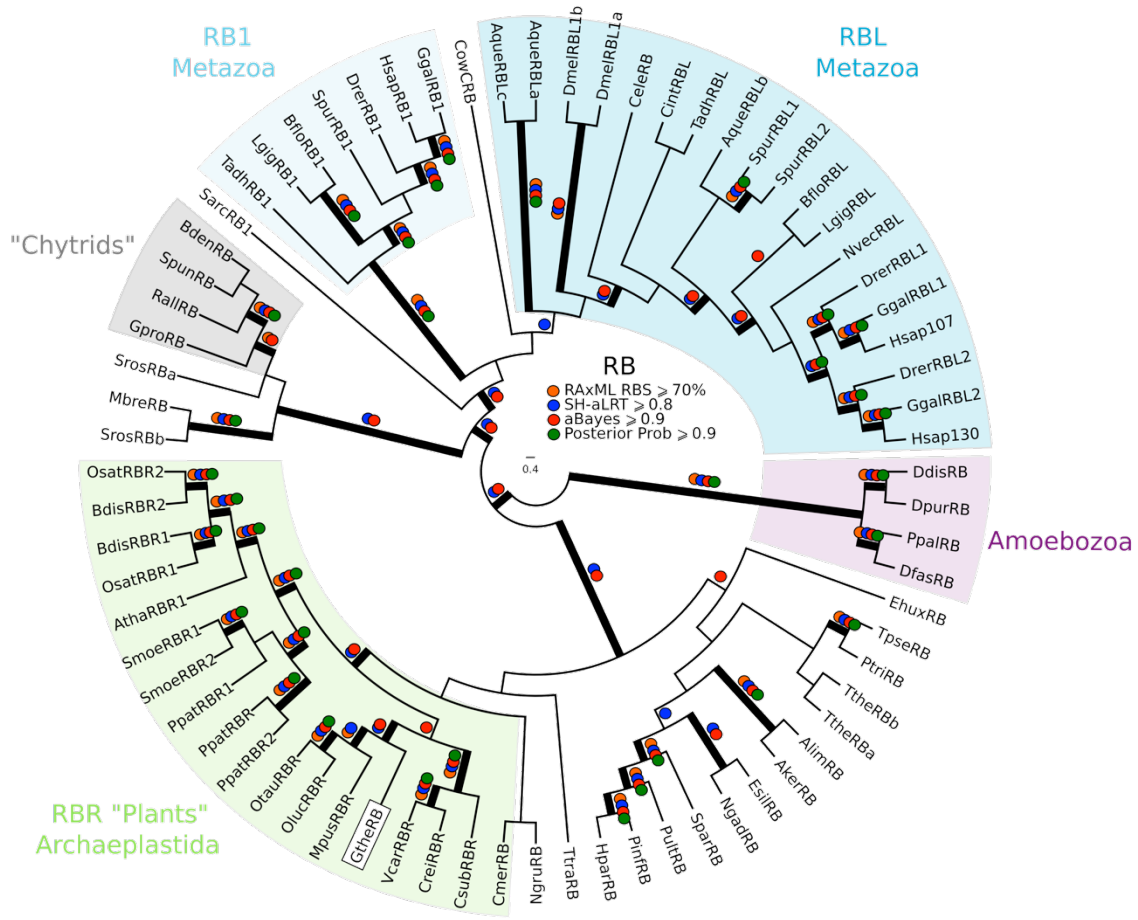


Figure 3 - supplement 3: Phylogeny of eukaryotic Rb inhibitors. *H. sapiens* has Rb1, RBL1 (p107), and RBL2 (p130), and *A. thaliana* has RBR1. The model fungi *S. cerevisiae* and *S. pombe* do not have any obvious retinoblastoma pocket proteins. We needed more eukaryotic sequences to create a robust HMM profile (pRb.hmm) for the pRb family. Based on the pRb sequences collected in (Hallmann, 2009), we built a profile-HMM using putative pRb homologs from *H. sapiens, G. gallus, C. intestinalis, D. melanogaster, C. elegans, N. vectensis, T. adhaerens* (metazoa)*; B. dendrobatidis* (fungi); *D. discoideum, D. purpureum, T. pseudonana, P. tricornutum, N. gruberi, E. huxleyi* (protists)*; C. merolae, O. tauri, O. lucimarinus, M. pusilla, V. carteri, C. reinhartdii, P. patens, S. moellendorfii, A. thaliana* (plants). A domain threshold of E-20 was used to identify pRB homologs. Our pRB dataset (pRb.fasta) has 72 sequences. Columns with the top 15% Zorro score (566 positions) were used in our alignment. Confidence at nodes was assessed with multiple support metrics using different phylogenetic programs under LG model of evolution (aBayes and SH-aLRT metrics with PhyML, RBS with RAxML, Bayesian Posterior Probability with Phylobayes (23,219 sampled trees, meandiff=0.0035, maxdiff=0.067)). Colored dots in branches indicate corresponding branch supports. Thick branches indicate significant support by at least two metrics, one parametric and one non-parametric; branch support thresholds are shown in the center of the figure; see Methods.

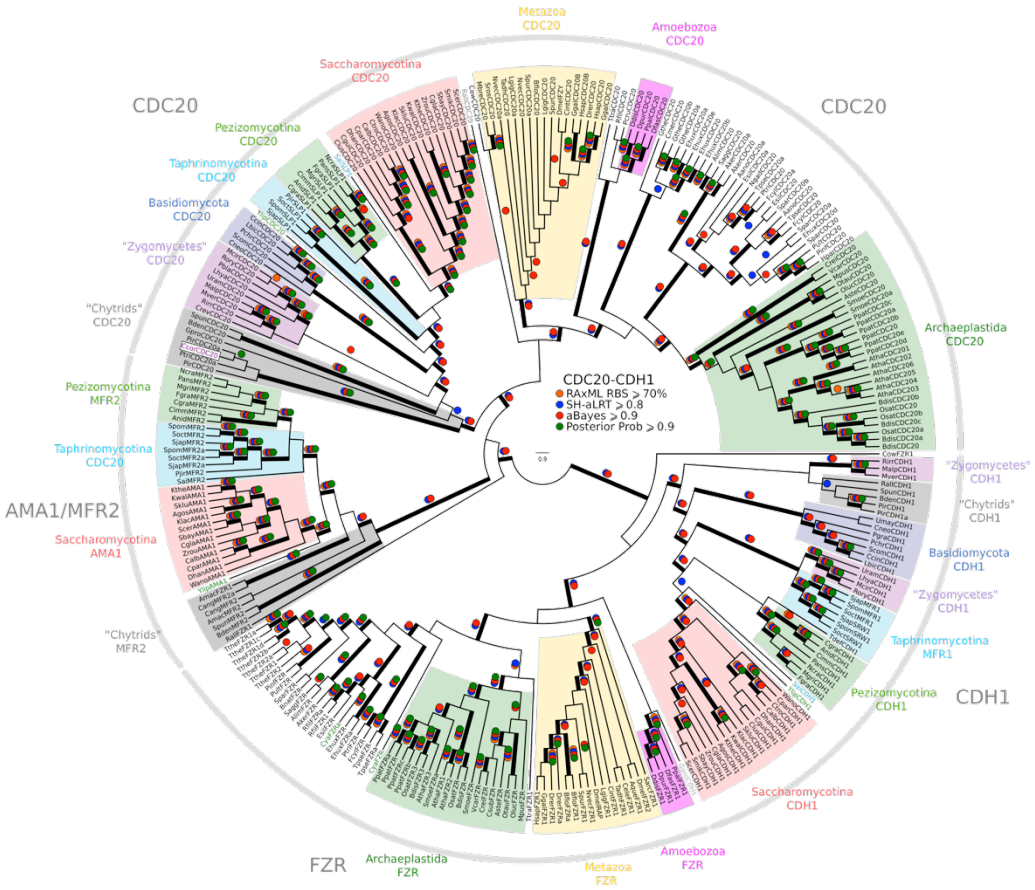


Figure 3 - supplement 4: Reduced phylogeny of eukaryotic Cdc20-family APC regulators. We combined CDC20 and CDH1/FZR1 sequences from *H. sapiens* (3 members), *A. thaliana* (9 members), and *S. cerevisiae* (3 members) to create a eukaryotic CDC20-family APC regulator profile-HMM (pCDC20.hmm) following the procedure outlined in the Methods. A domain threshold of E-50 was used to identify CDC20 homologs. Our pCDC20 dataset (pCDC20.fasta) has 350 sequences. This dataset was manually pruned to remove long-branches and problematic lineages. Our reduced CDC20 dataset has a total of 289 sequences. Columns with the top 20% Zorro score (608 positions) were used in our alignment. Confidence at nodes was assessed with multiple support metrics using different phylogenetic programs under LG model of evolution (aBayes and SH-aLRT metrics with PhyML, RBS with RAxML, Bayesian Posterior Probability with Phylobayes (13,638 sampled trees, meandiff=0.015, maxdiff=0.37)). Colored dots in branches indicate corresponding branch supports. Thick branches indicate significant support by at least two metrics, one parametric and one non-parametric; branch support thresholds are shown in the center of the figure; see Methods.

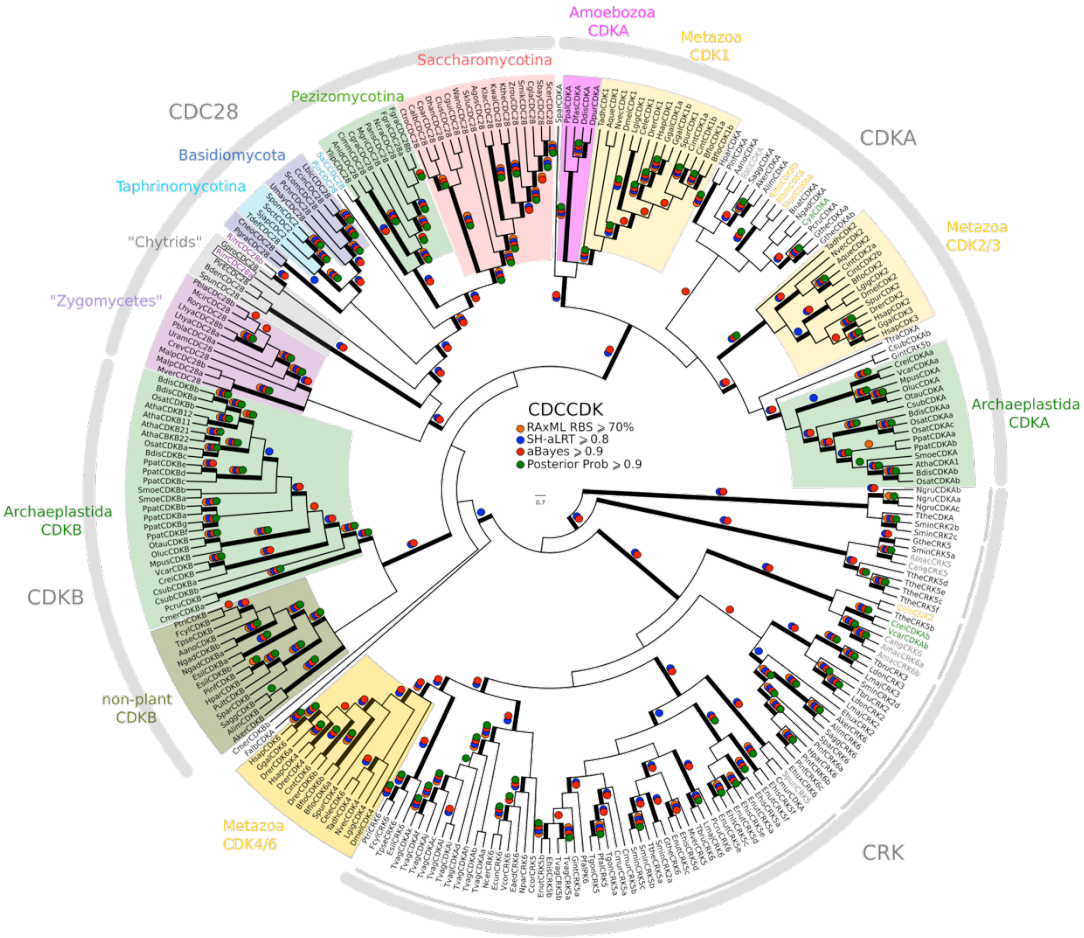


Figure 3 - supplement 5: Phylogeny of eukaryotic cyclin-dependent kinases. To create a profile-HMM (pCDCCDK.hmm) for eukaryotic cell cycle CDK, we combined Cdk1-3, Cdk4, Cdk6 sequences from *H. sapiens,* CdkA and CdkB from *A. thaliana,* Cdc28 from *S. cerevisiae,* and Cdc2 from *S. pombe*. A domain threshold of E-20 was used to identify CDK homologs. Our cell cycle CDK dataset (pCDCCDK.fasta) has 272 sequences. Columns with the top 15% Zorro score (473 positions) were used in our alignment. Confidence at nodes was assessed with multiple support metrics using different phylogenetic programs under LG model of evolution (aBayes and SH-aLRT metrics with PhyML, RBS with RAxML, Bayesian Posterior Probability with Phylobayes (28,193 sampled trees, meandiff=0.015, maxdiff=0.53)). Colored dots in branches indicate corresponding branch supports. Thick branches indicate significant support by at least two metrics, one parametric and one non-parametric; branch support thresholds are shown in the center of the figure; see Methods.

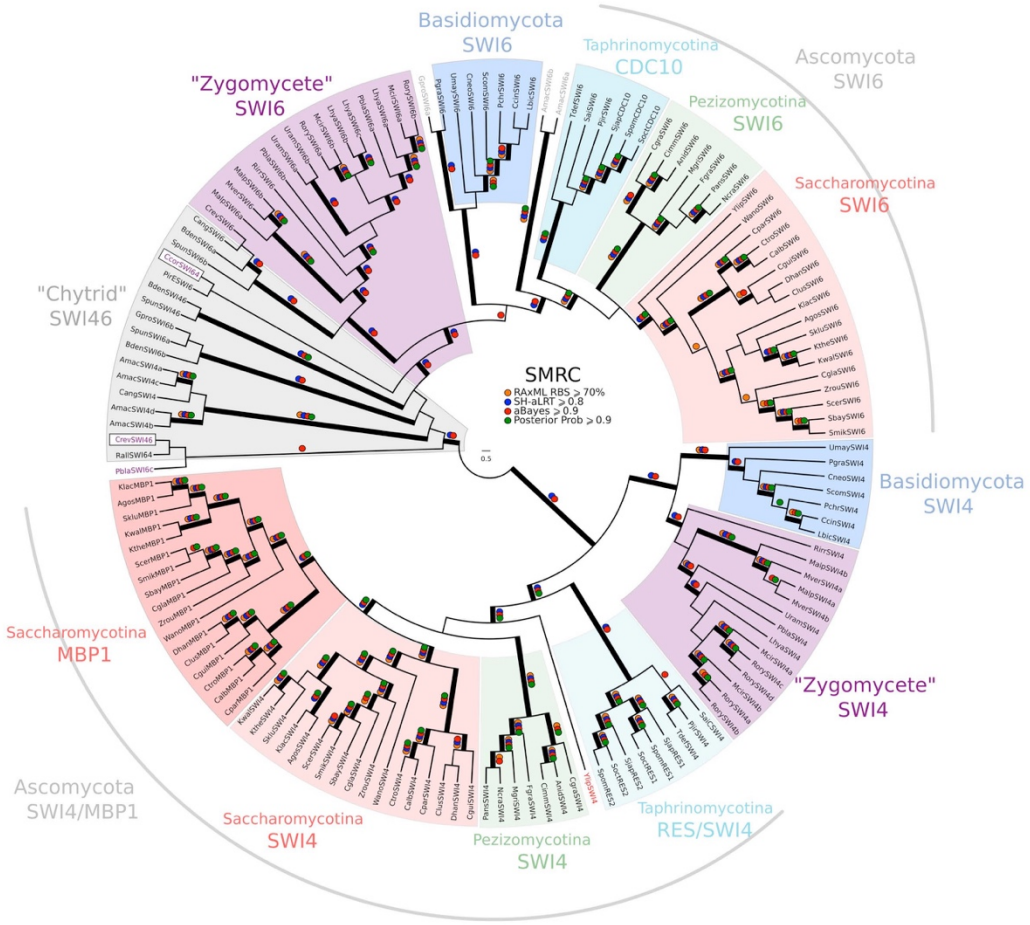


Figure 5 - supplement 1: Phylogeny of fungal SBF transcription factors. SBF and MBF are transcription factors that regulate G1/S transcription in budding and fission yeast. To detect SMRC (Swi4/6 Mbp1 Res1/2 Cdc10) across fungi, we built a sensitive profile-HMM (pSMRC.hmm) by combining well-characterized SMRC sequences from *S. cerevisiae, C. albicans, N. crassa, A. nidulans,* and *S. pombe*. A domain threshold of E-10 was used to identify SMRC homologs. Our SMRC dataset (pSMRC.fasta) has 147 sequences. Columns with the top 20% Zorro score (709 positions) were used in our alignment. Confidence at nodes was assessed with multiple support metrics using different phylogenetic programs under LG model of evolution (aBayes and SH-aLRT metrics with PhyML, RBS with RAxML, Bayesian Posterior Probability with Phylobayes (19,457 sampled trees, meandiff=0.0056, maxdiff=0.145)). Colored dots in branches indicate corresponding branch supports. Thick branches indicate significant support by at least two metrics, one parametric and one non-parametric; branch support thresholds are shown in the center of the figure; see Methods.

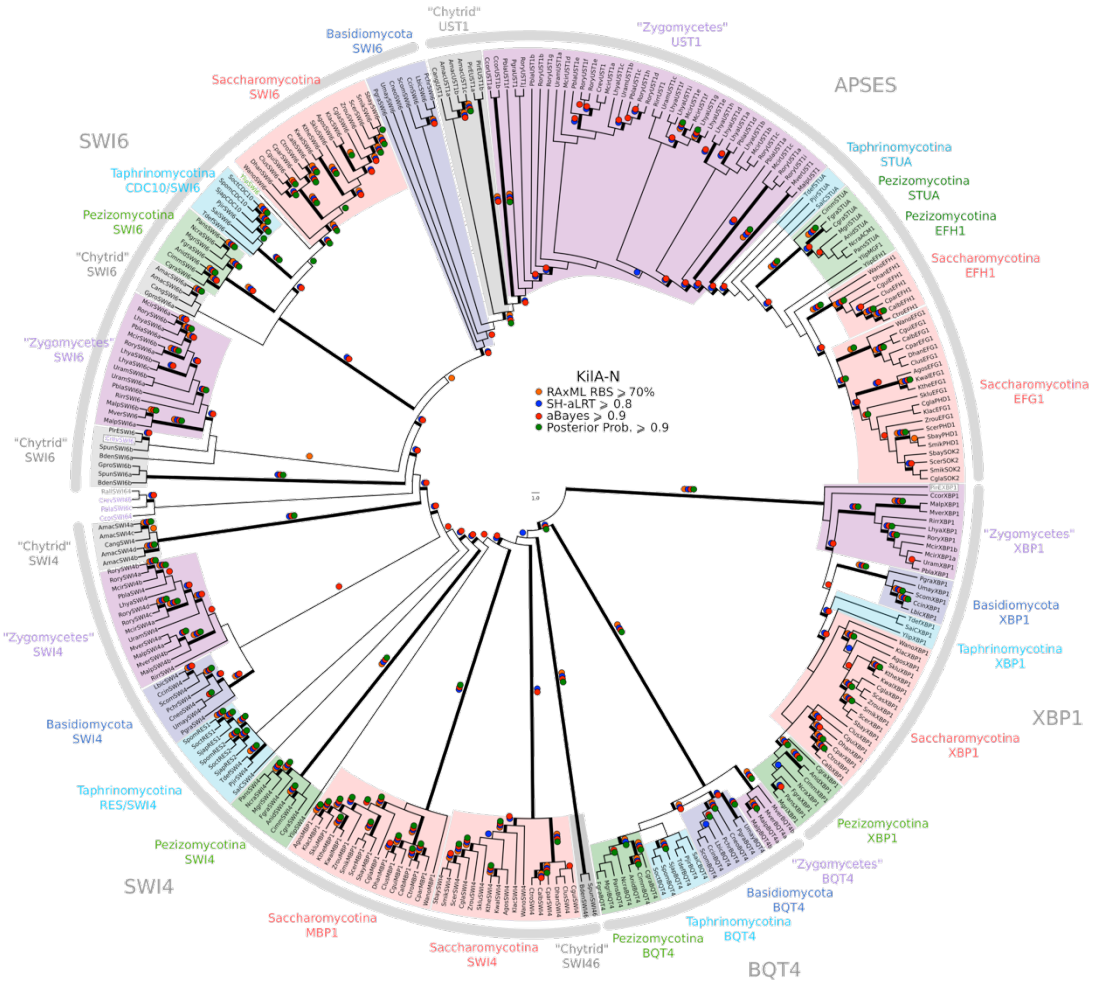


Figure 5 - supplement 2: Phylogeny of fungal SBF and APSES transcription factors. SBF and APSES transcription factors (Asm1, Phd1, Sok2, Efg1, StuA) share a common DNA-binding domain (KilA-N), which is derived from DNA viruses. During our search for SBF and APSES homologs, we consistently detected two additional fungal sub-families with homology to KilA-N: XBP1 (family name taken from *S. cerevisiae*) and BQT4 (family name taken from *S. pombe*). To detect APSES, XBP1, and BQT4 homologs, we built profile-HMMs (APSES.hmm, XBP1.hmm, and BQT4.hmm) by combining APSES, XBP1, and BQT4 homologs from *S. cerevisiae, C. albicans, N. crassa, A. nidulans,* and *S. pombe*. A domain threshold of E-10 was used to identify APSES, XBP1, and BQT4 homologs. Our final dataset (pKILA.fasta) contains all fungal KILA sub-families (SBF, APSES, XBP1, BQT4) and has a total of 301 sequences. Columns with the top 10% Zorro score (447 positions) were used in our alignment. Confidence at nodes was assessed with multiple support metrics using different phylogenetic programs under LG model of evolution (aBayes and SH-aLRT metrics with PhyML, RBS with RAxML, Bayesian Posterior Probability with Phylobayes (15,251 sampled trees, meandiff=0.012, maxdiff=0.25)). Colored dots in branches indicate corresponding branch supports. Thick branches indicate significant support by at least two metrics, one parametric and one non-parametric; branch support thresholds are shown in the center of the figure; see Methods.

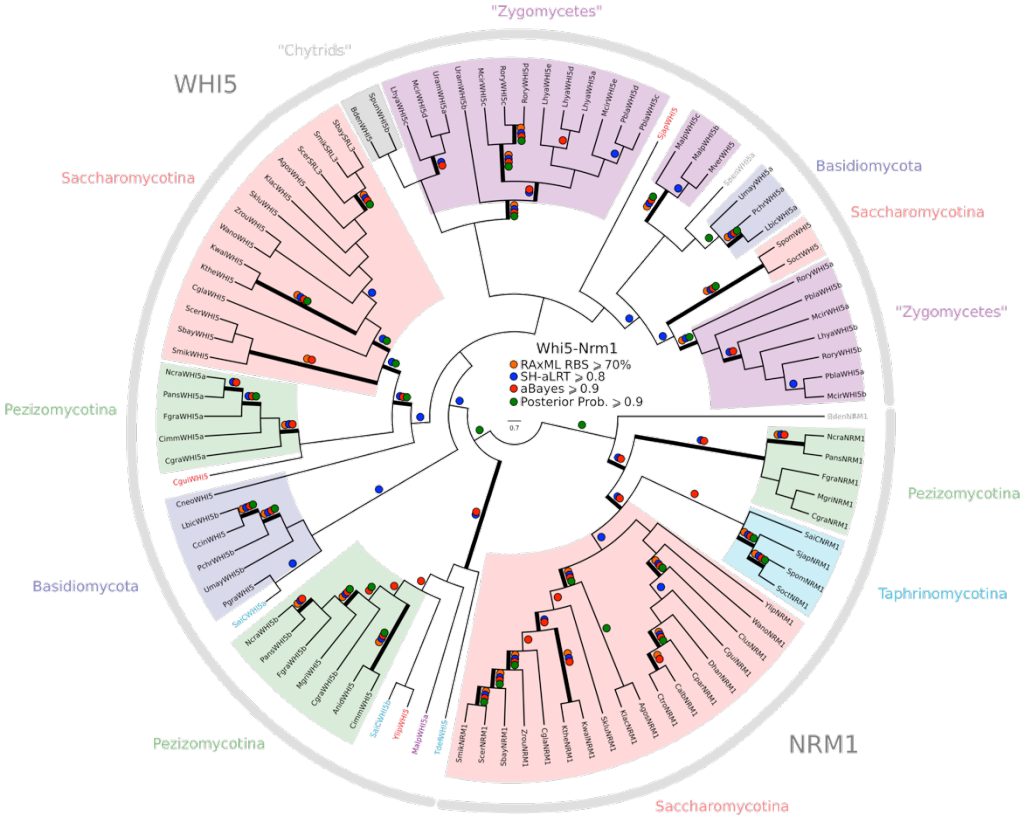


Figure 5 - supplement 3: Phylogeny of fungal Whi5 inhibitors. WHI5 and NRM1 are a yeast-specific protein family that has been identified and functionally characterized across *S. cerevisiae*, *C. albicans*, and *S. pombe.* Both WHI5 and NRM1 are fast evolving proteins. There is a small conserved region of 25 amino-acids (known as the GTB domain) that is responsible for interacting with Swi6/Cdc10 (Travesa et al., 2013). Unfortunately, the Whi5.hmm profile from PFAM is unable to detect an SRL3 paralogue in *S. cerevisiae* or the NRM1 orthologues in *A. gossypii* or *C. albicans.* We built a more sensitive profile-HMM (pWHI5.hmm) by combining WHI5/NRM1 sequences across ascomycetes (including SRL3 from Saccharomyces genomes and NRM1 from Candida genomes). A domain threshold of E-05 was used to identify WHI5 homologs. Our WHI5 dataset (pWHI5.fasta) has 98 sequences. Columns with the top 15% Zorro score (260 positions) were used in our alignment. Confidence at nodes was assessed with multiple support metrics using different phylogenetic programs under the LG model of evolution (aBayes and SH-aLRT metrics with PhyML, RBS with RAxML, Bayesian Posterior Probability with Phylobayes (77,696 sampled trees, meandiff=0.0068, maxdiff=0.11)). Colored dots in branches indicate corresponding branch supports. Thick branches indicate significant support by at least two metrics, one parametric and one non-parametric; branch support thresholds are shown in the center of the figure; see Methods.

967 **Supplementary File 1A: List of eukaryotic genomes.** We downloaded and
968 analyzed the following annotated genomes using the "best" filtered protein sets when
969 available. We gratefully acknowledge the Broad Institute, the DOE Joint Genome
970 Institute, Génolevures, PlantGDB, SaccharomycesGD, AshbyaGD, DictyBase, JCV
971 Institute, Sanger Institute, TetrahymenaGD, PythiumGD, AmoebaDB,
972 NannochloroposisGD, OrcAE, TriTryDB, GiardiaDB, TrichDB, CyanophoraDB, and
973 CyanidioschizonDB for making their annotated genomes publicly available. We
974 especially thank D. Armaleo, I. Grigoriev, T. Jeffries, J. Spatafora, S. Baker, J.
975 Collier, and T. Mock for allowing us to use their unpublished data.
976
977 **Supplementary File 1B: Plasmids.**

978 **Supplementary File 1C: Strains.** All yeast strains were derived from W303 and
979 constructed using standard methods.