Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites

Alan M. Moses*†, Muluye E. Liku‡, Joachim J. Li§, and Richard Durbin*

*Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1HH, United Kingdom; and Departments of ‡Biochemistry and §Microbiology and Immunology, University of California, San Francisco, CA 94143

Edited by Philip P. Green, University of Washington School of Medicine, Seattle, WA, and approved September 25, 2007 (received for review February 6, 2007)

Evolutionary change in gene regulation is a key mechanism underlying the genetic component of organismal diversity. Here, we study evolution of regulation at the posttranslational level by examining the evolution of cyclin-dependent kinase (CDK) consensus phosphorylation sites in the protein subunits of the prereplicative complex (RC). The pre-RC, an assembly of proteins formed during an early stage of DNA replication, is believed to be regulated by CDKs throughout the animals and fungi. Interestingly, although orthologous pre-RC components often contain clusters of CDK consensus sites, the positions and numbers of sites do not seem conserved. By analyzing protein sequences from both distantly and closely related species, we confirm that consensus sites can turn over rapidly even when the local cluster of sites is preserved, consistent with the notion that precise positioning of phosphorylation events is not required for regulation. We also identify evolutionary changes in the clusters of sites and further examine one replication protein, Mcm3, where a cluster of consensus sites near a nucleocytoplasmic transport signal is confined to a specific lineage. We show that the presence or absence of the cluster of sites in different species is associated with differential regulation of the transport signal. These findings suggest that the CDK regulation of MCM nuclear localization was acquired in the lineage leading to *Saccharomyces cerevisiae* after the divergence with *Candida albicans*. Our results begin to explore the dynamics of regulatory evolution at the posttranslational level and show interesting similarities to recent observations of regulatory evolution at the level of transcription.

DNA replication | MCM3 | phosphorylation

The contribution of regulatory evolution to biological diversity is increasingly well appreciated (1–4). The identification of changes in transcriptional regulatory proteins (5, 6) and, more frequently, the *cis*-elements they recognize in noncoding DNA (reviewed in ref. 7), has provided mechanistic insight into the evolution of gene regulation.

Genes are regulated at multiple levels, however. In eukaryotes, posttranslational regulation of protein activity by phosphorylation is of particular importance (8). Although little is known in general about the evolution of this type of regulation, comparative studies of posttranslational modification sites in phosphorylase (9, 10) and fructose 1-6-bisphosphatase (11) revealed that they were not conserved between homologues.

Recent studies have applied computational approaches to databases of protein sequences to perform comparative studies on larger scales. For example, targets of protein kinase A were predicted based on conservation of consensus sites between *Candida albicans* and *Saccharomyces cerevisiae* (12). Another study examined regulation of cell-cycle proteins in four species and proposed coevolution between posttranslational regulation by phosphorylation and transcriptional regulation (13).

Phosphoregulation plays a critical role in cell-cycle control (14–16). For example, it has been found in several species that after the initiation of DNA replication, to ensure that a single round of DNA replication occurs in each eukaryotic cell cycle, a subset of the DNA replication machinery (the pre-RC) is directly inhibited by cyclin-dependent kinase (CDK) (17, 18).

Here, we examine the evolution of regulation of the pre-RC by CDKs. Several features of this system make it attractive for evolutionary analysis. First, the pre-RC proteins are found in single copy in many animals and fungi (17), so it is relatively easy to identify their orthologs in most species. Also, human CDKs have been shown to rescue yeast CDK mutations (19, 20), suggesting little change in the functional capabilities of the kinase. Finally, CDK is a proline-directed serine/threonine kinase (21) with a well defined consensus site S/T-P-X-R/K (where X is any amino acid). Evolutionary loss of the critical S/T or P is likely to preclude phosphorylation by CDK in that species.

In some cases, the specific consensus sites likely to be phosphorylated by CDK *in vivo* have been determined through a combination of experimental methods; we refer to these sites as "characterized." In addition, CDK target proteins often contain multiple CDK consensus sites closely spaced in their primary amino acid sequence; we refer to these as "clusters." Previous studies have noted that, even when clusters of characterized sites are found in orthologous pre-RC components, the individual consensus sites are not always conserved in position or number (22, 23). We refer to this as "turnover" of sites and suggest that it is consistent with regulation through mechanisms that impose loose constraints on spacing and number of phosphorylation sites (ref. 24; see *Discussion*).

Our analysis of evolutionary changes in CDK consensus sites in pre-RC proteins reveals examples of both turnover of characterized sites in preserved clusters and lineage-specific changes in the clusters of sites. We suggest that the CDK regulation of nuclear localization of the pre-RC component Mcm3 (25) was gained on the lineage leading to *S. cerevisiae* after the divergence from *C. albicans*, and we provide experimental support for this model.

Results

Signatures of CDK Regulation in Pre-RC Proteins. To get a broad sense of the conservation of CDK regulation, we obtained

Author contributions: A.M.M., M.E.L., J.J.L., and R.D. designed research; A.M.M. and M.E.L. performed research; J.J.L. and R.D. contributed new reagents/analytic tools; A.M.M., M.E.L., and J.J.L. analyzed data; and A.M.M., J.J.L., and R.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Abbreviations: CDK, cyclin-dependent kinase; dn/ds, nonsynonymous to synonymous substitution rates; RC, replicative complex.

†To whom correspondence should be addressed. E-mail: am8@sanger.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0700997104/DC1.

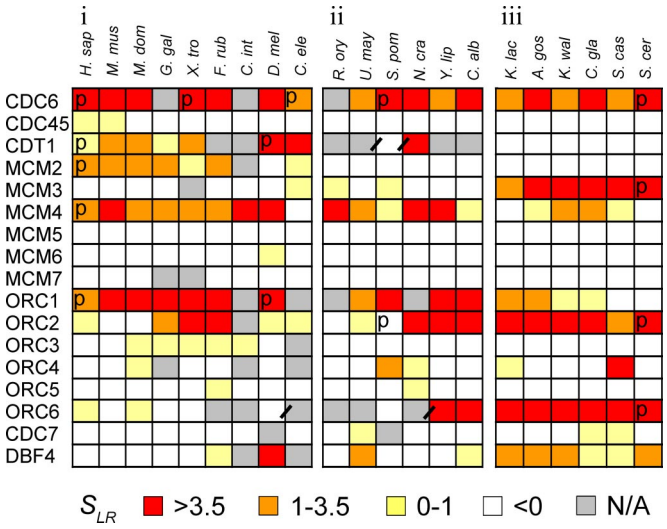

Fig. 1. Enrichment and clustering of CDK consensus motifs in pre-RC proteins from diverse animals and fungi. Each row represents the value of the S_{LR} statistic for the protein indicated on the left in the species indicated above the column. Gray boxes represent cases where a confident ortholog could not be identified, there was no single ortholog in that species, or the ortholog was truncated. Orthologs in *i* were taken from TreeFam, orthologs in *ii* where assigned as described in *Methods*, and orthologs in *iii* were taken from the Yeast Gene Order Browser. ''P''s indicate CDK targets where consensus sites have been characterized. Diagonal bars indicate a species boundary across which reliable sequence alignments were not possible. *H. sap*, *Homo sapiens*; *M. mus*, *Mus musculus*; *M. dom*, *Monodelphis domestica*; *G. gal*, *Gallus gallus*; *X. tro*, *Xenopus tropicalis*; *F. rub*, *Takifugu rubripes*; *C. int*, *Ciona intestinalis*; *D. mel*, *Drosophila melanogaster*; *C. ele*, *Caenorhabditis elegans*; *R. ory*, *Rhizopus orysae*; *U. may*, *Ustilago maydis*; *S. pom*, *Schizosaccharomyces pombe*; *N. cra*, *Neurospora crassa*; *Y. lip*, *Yarrowia lipolytica*; *C. alb*, *C. albicans*; *K. lac*, *K. lactis*; *A. gos*, *A. gossypii*; *K. wal*, *K. waltii*; *C. gla*, *C. glabrata*; *S. cas*, *Saccharomyces castelii*; *S. cer*, *S. cerevisiae*.

sequences and orthologs for pre-RC proteins (see *Methods*) from 21 species with complete genome sequences publicly available, selected so that their phylogenetic positions were informative amongst the animals and fungi. For each protein, we identified experimentally verified CDK targets where consensus sites had been characterized (''P''s in Fig. 1 and refs. 22, 23, and 25–38) and also calculated the S_{LR} statistic (see *Methods*), which measures the overrepresentation and spatial clustering of strong (S/T-P-X-R/K, where X is any amino acid) and weak (S/T-P) CDK consensus matches (Fig. 1), which we have shown to be predictive of CDK regulation (39). Because the pre-RC is expected to be regulated by CDKs in all these species, a simple expectation is that the same proteins would be targets in all species. Indeed, we find proteins that have high values of S_{LR} across many (Orc1, Mcm4) or all (Cdc6) of the species examined, suggesting that regulation has been preserved since a common ancestor. However, other proteins (Orc2) show less consistent patterns, while some (Orc6, Mcm2, Mcm3) show lineage-specific patterns. In these cases, the changes in statistical signal could be due to either bona fide changes in regulation or incorrect classification by our statistical method, but in at least one of these cases, we see a functional difference corresponding to the statistical difference (see below).

Turnover of Functional CDK Consensus Sites. Even when regulation appears conserved, as has been noted in previous studies (22, 23), we found that the numbers and positions of CDK consensus sites were not always conserved. A striking example of this is the linker region of ORC1, which contains a strong cluster of CDK consensus matches in all of the animals and most of the fungi (Figs. 1 and 2*a*). Sites in this region are phosphorylated by CDK in *Drosophila* (23) and are involved in CDK-regulated localization and degradation of ORC1 in mammalian cells (38, 40, 41). Despite the persistence of the cluster over long evolutionary distances, examination of the numbers and positions of individual CDK consensus sites (Fig. 2*a*) reveals rapidly changing organization.

It is possible that this apparent turnover of sites is due simply to difficulties in comparing highly diverged amino acid sequences, or that consensus matches in clusters do not all represent functional sites and are not constrained. To rule these out, we examined the evolution of experimentally characterized consensus sites. We consider a consensus site characterized if there is some *in vivo* (including cell culture) evidence of phosphorylation and/or function in a CDK-regulated process [''P'' in Fig. 1 and supporting information (SI) Table 1]. We examined these sites in alignments of orthologs from closely related species (see *Methods*), where most residues are unchanged and we have high-confidence in multiple alignments (84%, 74%, and 64% identical for yeast, mammals, and *Drosophilae*, respectively).

We found in each clade that characterized consensus sites accumulated on average fewer substitutions than the flanking residues (rates were 20%, 60%, and 27% of flanking regions for yeast, mammals, and *Drosophila*, respectively; SI Fig. 4*a*). Interestingly, despite this evidence for constraint, we also found that of the 55 experimentally characterized CDK consensus sites, 9 had substitutions in the critical S/T or P of the CDK consensus in these closely related species (not conserved, Fig. 2*b*). These include a previously reported nonconserved CDK site in the N terminus of mammalian CDC6 (42). We also noted five sites that changed between strong and weak consensus matches in these alignments (Fig. 2*b*). Thus, microevolutionary changes in functional CDK sites provide a potential mechanism for the changes in number and positions of consensus sites observed over long evolutionary distances.

The linker region of mammalian ORC1 (boxed region in Fig. 2*a* and ref. 40) provides an extreme example of this evolutionary turnover (Fig. 2*c*). Of three strong and one weak characterized consensus sites (ref. 38 and Fig. 2 *a* and *c iii*, *iv*, *vi*, and *viii*), only one is conserved over the mammals (Fig. 2*c iv*), although it is additionally modulated by alternative splicing in mouse (43). Furthermore, one of these sites appeared within the divergence of the primates (Fig. 2*c viii*). In addition to these changes in characterized sites, we noted a region containing human-specific losses of consensus matches and a human polymorphism appears to disrupt an ancestral consensus match (Fig. 2*c x*).

To test for constraint in the linker region of ORC1 more formally, we computed the ratio of nonsynonymous to synonymous substitution rates (dn/ds) (see *Methods*) for this region, and found it to be 0.98. Consistent with the rapid changes in the consensus sites, we could not reject the hypothesis of no constraint (dn/ds = 1, $P > 0.91$, see *Methods*).

Such weak constraint and rapid turnover of consensus sites in the mammals is surprising given that that cluster of sites in this region of ORC1 appears to have been retained since the common ancestor of the animals (Figs. 1 and 2*a*). We therefore sought to detect constraint on the cluster of sites. We reconstructed the sequence of the ancestral ORC1 linker region (see *Methods*) and found it to have more consensus sites and stronger clustering than the extant human sequence (5 strong, 11 weak, $S_{LR} = 9.89$ vs. 3 strong, 9 weak, $S_{LR} = 4.50$). We then simulated the evolution of the ancestral sequence, using a general protein model (see *Methods*). Constraint on the cluster of sites should lead to greater and less variable values of S_{LR}, so we compared a composite statistic (the difference between the mean and standard deviation of the S_{LR} over the seven species) in the real mammalian sequences to the simulations, and found it to be significantly greater (Fig. 2*e*, $P < 0.005$, $n = 5{,}000$). These

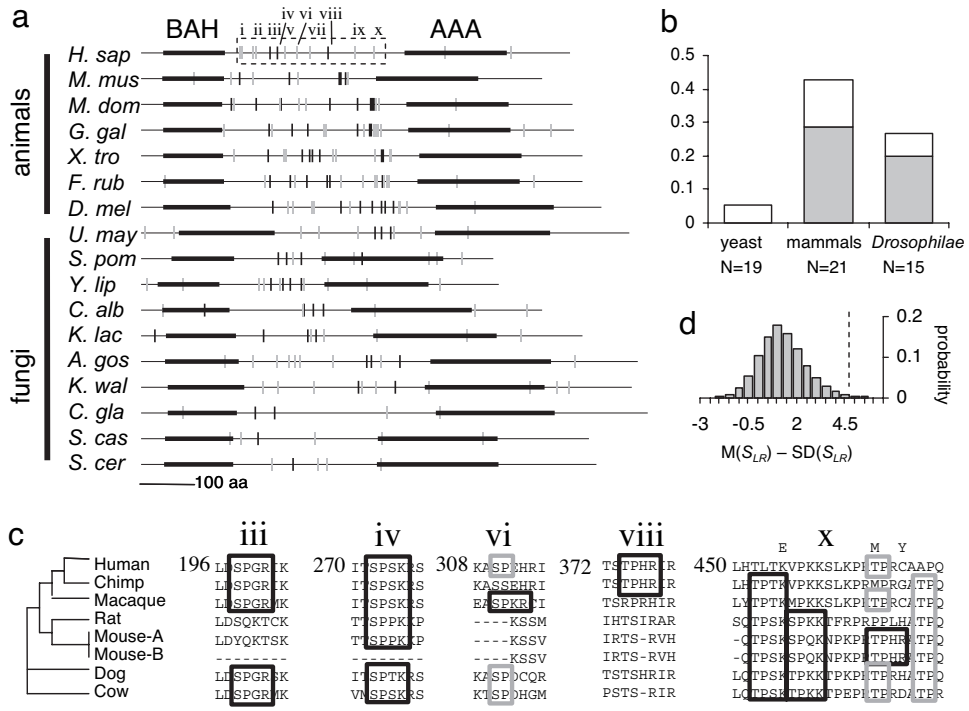

Fig. 2. Turnover of CDK consensus sites. (*a*) Schematic view of ORC1 orthologs. Black and gray ticks represent matches to the strong and weak CDK consensus, respectively. *iii*, *iv*, *vi*, and *viii* indicate the characterized CDK consensus sites. Thickened regions of the sequences represent BAH and AAA pfam domains (68), respectively. Boxed region indicates the human linker region. (*b*) Percentage of characterized CDK consensus sites that are either not conserved (gray) or change between strong and weak consensus (white) in alignments of closely related species. (*c*) Alignments of seven mammals for consensus sites in the linker region of ORC1. *iii*, *iv*, *vi*, and *vii* are experimentally characterized sites (38). Mouse-A and Mouse-B indicate alternative transcripts for the mouse gene. Text above the human sequence in *x* indicates polymorphisms within the human population. Black and gray boxes indicate matches to the strong and weak CDK consensus, respectively; numbers are as in *a*. (*d*) Comparison of the observed value (dotted trace) of a composite statistic to the distribution obtained from simulations indicates constraint at the level of the cluster of sites. See *Results* for details.

simulation results support the model that the cluster of CDK sites in ORC1 evolves under purifying selection, even though little constraint is apparent at the amino acid level.

Lineage Specific Regulatory Evolution. In contrast to cases like ORC1 where cluster of CDK consensus sites is largely conserved, other pre-RC proteins show considerable variation in S_{LR} across species (Fig. 1), despite the importance of proper regulation (17, 18). For example, Mcm3 is a CDK target in *S. cerevisiae* (35), but CDK regulation has not been reported in *S. pombe* or human.

Consistent with the hypothesis of lineage-specific regulation of MCM3, we find a dramatic statistical change in the clustered CDK consensus sites on the lineage leading to *S. cerevisiae* (Fig. 1). This change is due to a cluster of consensus sites in the C-terminal region of *S. cerevisiae* Mcm3 (*Sc*-Mcm3-CTR) that was found to be critical for the CDK-mediated shuttling of the MCM complex in and out of the nucleus in that species (25, 44, 45). Indeed, mutation of the CDK consensus sites in the *Sc*-Mcm3-CTR abolished its ability to confer regulated nuclear localization to a GFP reporter construct (25). Interestingly, in contrast to *S. cerevisiae*, the MCM protein complex is constitutively nuclear in *S. pombe* and human (46). We therefore decided to test whether the changes in CDK consensus sites were associated with lineage-specific changes in regulation.

We first sought to rule out that the changes in CDK consensus sites could be explained by statistical fluctuations. To do so, we obtained Mcm3 orthologs from six additional fungi to improve resolution within the Ascomycetes (see *Methods*). We then used maximum parsimony to reconstruct the ancestral organization of these sequences and infer the gains and losses of CDK consensus sites along each branch (Fig. 3*a*; see *Methods*). For the strong consensus, we inferred 13 gains in the clade containing *S. cerevisiae*, significantly greater than the 5.37 expected if gains were randomly distributed proportional to the evolutionary distance on each branch ($P = 0.0037$, see *Methods*). For the weak consensus, we inferred 13 gains in this clade, which also greater than the 10.36 expected but is not statistically significant ($P = 0.24$). These data show that gains of strong consensus matches are nonrandomly distributed along the tree and suggest that the CDK-regulated shuttling of MCMs in and out of the nucleus in *S. cerevisiae* is due at least in part to changes in CDK consensus sites that occurred after the divergence from *C. albicans* (Fig. 3*a*).

This model predicts that the region homologous to the *Sc*-Mcm3-CTR from species outside this clade would not confer regulated localization to a GFP reporter construct. We therefore inserted the homologous region of *C. albicans* Mcm3 into such a construct (Fig. 3 *b* and *c*, see *Methods*) and tested its localization in *S. cerevisiae* in cells arrested in G_1 (by alpha factor) or G_2 (by nocodazole). Although the *S. cerevisiae* construct showed nuclear localization in the G_1 but not the G_2 arrest (Fig. 3*d*, compare *iv* with *viii*), the *C. albicans* construct was constitutively nuclear (Fig. 3*d*, compare *i* with *v*), confirming a functional difference in this region of the protein between these species. To further resolve the evolutionary events that lead to regulated localization of MCMs in *S. cerevisiae*, we performed similar experiments, using the C-terminal region of Mcm3 from *Candida glabrata* (Fig. 3*c*, *ii* and *vi*) and *Kluveromyces lactis* (Fig. 3*c*, *iii* and *vii*) and found that these showed regulated nuclear localization, consistent with the origin of this regulation in the ancestor of the *S. cerevisiae* clade.

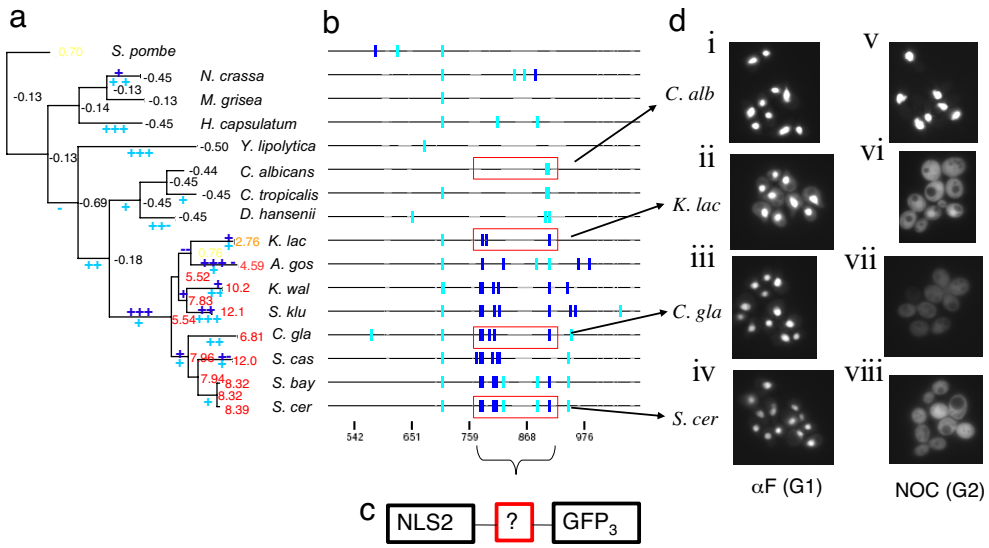

Fig. 3. Lineage-specific evolution in the C terminus of Mcm3. (*a*) Phylogenetic tree relating 16 Ascomycete fungi, with maximum-likelihood branch lengths in amino acid substitutions per site. Blue or cyan plus signs and minus signs above or below the branches represent inferred gain and loss events along that branch for strong or weak CDK consensus matches. Numbers are the values of S_{LR} computed by using the extant sequences at leaf nodes, or using ancestral reconstructions at internal nodes, colored as in Fig. 1. (*b*) Schematic view of a multiple alignment (created by using T-Coffee) of the C-terminal region of Mcm3. Blue and cyan symbols represent matches to the strong and weak CDK consensus, respectively. Gray regions indicate gaps in the aligned sequences, and red boxes indicate regions inserted into the reporter construct. Scale indicates the position in the multiple sequence alignment. (*c*) Schematic of the GFP reporter construct used in *d*. See ref. 25 for details. (*d*) GFP localization assays show nuclear localization of the *C. albicans* reporter construct in both alpha factor (*i*) and nocodazole arrested (*v*) cells. This is in contrast to CDK-regulated localization of the *S. cerevisiae*, *K. lactis* (*ii* and *vi*), and *C. glabrata* (*iii* and *vii*) constructs, which show nuclear localization in alpha factor but not nocodazole arrested cells and therefore suggest evolution of CDK regulation since the divergence of *C. albicans*. *M.*, *Magnaporthe*; *H.*, *Histoplasma*; *C.*, *Candida*; *D.*, *Debaryomyces*; *S. klu*, *S. kluyveri*. Other abbreviations are as in Fig. 1.

Taken together, our experiments and sequence analysis of the Mcm3 C terminus are consistent with the model that the functional CDK consensus sites that regulate nuclear localization arose in the common ancestor of the *S. cerevisiae* clade after divergence from *C. albicans*. It is important to note that the *Sc*-Mcm3-CTR contains other regulatory sequences (ref. 25 and SI Fig. 5*a*), including a Crm1-dependent export signal, which also appeared at that time, and a basic nuclear localization signal, which is shared by all ascomycetes and may be important for the observed nuclear localization of the MCMs in *S. pombe* (46). Consistent with this model, we identify a basic nuclear localization signal, but no leucine rich export signals in the homologous region of the *C. albicans* protein (SI Fig. 5 *b* and *c*). To rule out the possibility that there were cryptic export signals or CDK sites further downstream of the region we defined as homologous to the *Sc*-Mcm3-CTR, we also performed all of the experiments, using the entire C terminus from each species and found similar results (data not shown).

Discussion

Inhibition of the pre-RC by CDKs to prevent rereplication is an ancient feature of the eukaryotic cell cycle (17). Our results suggest that, even though this regulatory logic is preserved, its mechanistic implementation can evolve rapidly.

For example, we found that, on average, 16% (11–21% ± SE) of characterized CDK consensus sites in pre-RC components in budding yeast, human, and *Drosophila* are not conserved in alignments of closely related species. In ORC1, the presence of polymorphisms in the human population suggests that the resculpting of regulatory regions continues.

Traditional models of phosphoregulation invoke allosterically driven conformational changes as a consequence of phosphorylation, which presumably require modification at precise positions in the protein structure. More recent analyses of phosphoregulation suggest alternative regulatory paradigms involving multiple phosphorylation sites that do not need to be conserved (24). Clusters of multiple phosphorylation sites can modulate interactions (25, 47–49) or provide specific dynamic properties (50–52) and these mechanisms may not depend on the specific locations or numbers of sites (24).

Consistent with this model, we found statistical evidence for constraint at the level of the cluster of consensus sites in the linker region of ORC1, despite weak constraint at the amino acid level. In clusters, when new consensus sites appear via point mutations, constraints on the ancestral sites may be relaxed, allowing them to accumulate destructive substitutions. Interestingly, this stabilizing selection model was first proposed for transcriptional enhancer elements in DNA, where, despite little similarity in primary sequence, orthologous enhancers could drive similar expression patterns by preserving clusters of transcription factor binding sites (53, 54).

In addition to turnover of consensus sites in conserved clusters, we found cases of entire clusters that are not conserved over evolution. We observed lineage-specific accumulation of consensus sites in the C terminus of *S. cerevisiae* Mcm3, which we showed was associated with functional differences in localization of a reporter construct (Fig. 3). We also note that the C-terminal cluster of consensus sites in yeast Cdc6 (29) shows a similar pattern, appearing even more recently (SI Fig. 5*d*). Because CDK inhibits the pre-RC through multiple regulatory mechanisms (35), we suggest that new mechanisms may evolve without drastic negative consequences. Thus, a possible explanation for these lineage-specific changes is "regulatory network turnover" (55), in which interactions are gained and lost in the context of a preserved regulatory logic.

Finally, we note that the accretion of regulatory motifs in the Mcm3 C terminus is analogous to the evolutionary gain of transcription factor binding sites in enhancers (56). In extending this model to phosphorylation sites, we suggest that the cooption of a new target into an existing regulatory network by acquisition

of motifs for preexisting, *trans*-acting factors is a general mechanistic basis for evolutionary increases in regulatory specificity and, perhaps, organismal complexity.

Methods

Proteins, Orthologs, and Clustering of CDK Sites. For the animal and yeast genomes used in Fig. 1 *i* and *iii*, protein sequences and ortholog assignments were obtained from the TreeFam database (57) and Yeast Gene Order Browser (58), respectively. To assign orthologs for the species not included in these databases (Fig. 1 *ii*), we obtained amino acid sequences from J. Stajich (University of California, Berkeley, CA; http://fungal.genome.duke.edu). We then aligned the fungal and animal orthologs (from TreeFam or Yeast Gene Order Browser), using T-Coffee software (59), created profile-hidden Markov models, and searched the additional genomes for matches to these profiles, using the HMMer package (http://hmmer.janelia.org, using the -forward option). We took the top hit as the ortholog in each case, except for CDC6, where the top hit was the same as the top hit for ORC1 in some of the fungi, so we took the second hit. Where a protein was present in multiple copies in a species (e.g., CDC7 in *S. pombe*), we excluded that protein for that species from further analyses (gray box in Fig. 1). If the HMMer e-value was >0.001 or the protein was truncated relative to other orthologs, we deemed the ortholog low confidence (gray box in Fig. 1).

For each protein in each species, we computed S_{LR}, a log likelihood ratio statistic, which measures clustering and enrichment of motifs in a sequence. Briefly, this statistic compares the likelihood of the observed motifs and their spacing under a model that includes clusters to that under the genomic background frequency or a model, including clusters of weak sites only (for details, see ref. 39). We computed the background frequencies of these motifs in each of the genomes studied. We reported the analysis shown in Fig. 1 by using other statistical measures and found similar results (SI Fig. 6).

Alignments of Closely Related Species. We obtained ortholog assignments and protein sequences for each of the characterized CDK targets from budding yeast in *S. paradoxis*, *S. mikatae*, and *S. bayanus* from SGD (60), from human in mammals from TreeFam or from *Drosophila* from 12 *Drosophilae* (V. Iyer, D. Pollard, and M. Eisen, personal communication). These were aligned with T-Coffee, and truncated orthologs were removed, except in the case of mammalian CDT1, where only the N-terminal region was available. Alignments of all of the characterized sites are available as SI Dataset 1.

To compute the dn/ds, we obtained coding DNA sequences and inserted the gaps from the protein alignments into these. For the linker region of ORC1 (which we took to be amino acids 196–470 in the human sequence), we used paml (61) to compute maximum-likelihood branch lengths with either an unknown dn/ds or dn/ds fixed at 1, assuming the phylogeny (((human,chimp),macaque),(mouse,rat)),dog,cow). We compared two times the difference in likelihoods to a χ^2 distribution with one degree of freedom. Human SNPs and alternative mouse transcripts for ORC1 were obtained from Ensembl (version 41; ref. 62). We note that dn/ds for the clusters of CDK sites were higher on average than the whole proteins, with ORC1 showing the highest value (data not shown).

Simulations of Orc1 Evolution. To obtain the distribution of the difference of the mean and standard deviation of S_{LR} for the ORC1 linker used the following procedure. We extracted the amino acid alignment and used paml (61), using the mammalian phylogeny described above to obtain the maximum-likelihood estimates for the branch lengths (in amino acid substitutions per site) and to reconstruct the ancestral sequence. We then used the ROSE sequence evolution software (63) to simulate (with default parameters for protein evolution) along the estimated tree starting from the ancestral sequence. Finally, we computed the average and standard deviation of the S_{LR} in the simulated sequences for the extant species.

Reconstruction of Ancestral Mcm3 CDK Matches. Because we wanted to reconstruct the ancestral organization of CDK matches in Mcm3 over longer evolutionary distances where we were no longer confident in the alignment of individual residues, we devised the following parsimony method. First, we obtained protein predictions for six additional Ascomycete genomes (http://fungal.genome.duke.edu/), assigned orthologs as above, made a multiple alignment of the protein sequences, using T-Coffee, and used paml to obtain maximum-likelihood estimates of the branch lengths for the tree topology shown in Fig. 3*A*. We then searched the aligned sequences for matches to the CDK consensus and created an "alignment" of CDK consensus matches by treating any CDK match within five amino acid residues as another in a different species as "aligned." For Mcm3, this yielded 31 aligned "columns," where there was a match to either the strong or weak CDK consensus in at least one species. Based on this, we used the "classical parsimony" algorithm (64) to reconstruct the ancestral states, either "strong match," "weak match," or "background" and infer the number of gains and losses for strong and weak matches along each branch.

Although the current view supports the clade containing *K. lactis*, *Ashbya gossypii*, *Kluveromyces waltii*, and *Saccharomyces kluyveri* as a sister to the clade containing *S. cerevisiae* (65, 66) the placement of the species (Fig. 3*A*) is not yet conclusively established (66). We therefore repeated the analysis using a multifurcation at this node and found similar results regarding the asymmetry, but observed variation in the estimates of CDK consensus gain and loss events on each branch (data not shown).

To calculate the expected number of gains in the *Saccharomyces* clade under the hypothesis of symmetrically distributed changes, we assume the number of background positions is large relative to the number of matches and that gains of matches are rare (no multiple hits). The expected number of gains in a subclade c is then Poisson with mean $= n_g \times t_c/t$, where t is the sum of the branch lengths (tree length), t_c is the sum of the branches in the clade c, and n_g is the number of gains inferred along the whole tree. To calculate the ancestral values of S_{LR}, we reconstructed the ancestral positions of each column of aligned matches by recursively assigning to each ancestor the average position of the matches in its children.

Construction of GFP Reporters and Localization Assays. We obtained genomic DNA for *C. albicans*, *C. glabrata*, and *K. lactis* from D. Galgoczy (University of California, San Fransico) and A. Johnson (University of California, San Francisco) and for *A. gossypii* from A. Gladfelter (Dartmouth College, Hanover, NH). We amplified the region homologous to the *Sc*-Mcm3-CTR or the entire C terminus by PCR (Phusion; Finnzymes, Espoo, Finland), using primers (IDT Technologies, Coralville, IA) that introduced ClaI or EcoRI restriction sites into the 5′ or 3′ ends of the PCR product. Primer sequences are available on request. These PCR products were inserted between the ClaI and EcoRI sites in the plasmid pML104, a gal inducible *TRP1* integrating plasmid containing the *S. cerevisiae* Mcm2 nuclear localization signal and three tandem copies of GFP (25). All constructs were confirmed by sequencing (MClab, South San Francisco, CA). Plasmids were transformed into YJL310 (67), grown, arrested and photographed as described in figures 4, 5, 6*B*, 8, and 9 of ref. 25. The cell-cycle arrests were confirmed by scoring the fraction budded for >60 cells for each strain under each condition. The GFP localization panels shown were "representative," and observations were confirmed by scoring the fraction showing nuclear staining for >60 cells for each construct under each condition.

We thank Dr. Dave Morgan, Dr. Seth Grant, Dr. Avril Coghlan, and Dr. Jean-Karim Hériché, and Dave Galgoczy for helpful discussions; Dave Galgoczy, Dr. Alexander Johnson, and Dr. Amy Gladfelter for genomic DNA; and Dr. Avril Coghlan for comments on the manuscript. This work was supported by National Institutes of Health Grants 5F31CA110268-03 (to M.E.L.) and R01 GM59704 (to J.J.L.). A.M.M. is a Sanger Postdoctoral Fellow. R.D. and the Wellcome Trust Sanger Institute are funded by the Wellcome Trust.

1. Wilson AC, Maxson LR, Sarich VM (1974) *Proc Natl Acad Sci USA* 71:2843–2847.
2. Tautz D (2000) *Curr Opin Genet Dev* 10:575–579.
3. Levine M, Tjian R (2003) *Nature* 424:147–151.
4. Wray GA (2003) *Int J Dev Biol* 47:675–684.
5. Ronshaugen M, McGinnis N, McGinnis W (2002) *Nature* 415:914–917.
6. Galant R, Carroll, SB (2002) *Nature* 415:910–913.
7. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) *Mol Biol Evol* 20:1377–1419.
8. Johnson SA, Hunter T (2005) *Nat Methods* 2:17–25.
9. Palm D, Goerl R, Burger KJ (1985) *Nature* 313:500–502.
10. Hwang PK, Fletterick RJ (1986) *Nature* 324:80–84.
11. Rittenhouse J, Harrsch PB, Kim JN, Marcus F (1986) *J Biol Chem* 261:3939–3943.
12. Budovskaya YV, Stephan JS, Deminoff SJ, Herman PK (2005) *Proc Natl Acad Sci USA* 102:13933–13938.
13. Jensen LJ, Jensen, TS, de Lichtenberg U, Brunak S, Bork P (2006) *Nature* 443:594–597.
14. Norbury CJ, Nurse P (1989) *Biochim Biophys Acta* 989:85–95.
15. Nasmyth K (1993) *Curr Opin Cell Biol* 5:166–179.
16. Murray AW (1994) *Chem Biol* 1:191–195.
17. Kelly TJ, Brown GW (2000) *Annu Rev Biochem* 69:829–880.
18. Bell SP, Dutta A (2002) *Annu Rev Biochem* 71:333–374.
19. Lee MG, Nurse P (1987) *Nature* 327:31–35.
20. Elledge SJ, Spottswood MR (1991) *EMBO J* 10:2653–2659.
21. Brown NR, Noble ME, Endicott JA, Johnson LN (1999) *Nat Cell Biol* 1:438–443.
22. Vas A, Mok W, Leatherwood J (2001) *Mol Cell Biol* 21:5767–5777.
23. Remus D, Blanchette M, Rio DC, Botchan MR (2005) *J Biol Chem* 280:39740–39751.
24. Serber Z, Ferrell JE, Jr (2007) *Cell* 128:441–444.
25. Liku ME, Nguyen VQ, Rosales AW, Irie K, Li JJ (2005) *Mol Biol Cell* 16:5026–5039.
26. Jallepalli PV, Brown GW, Muzi-Falconi M, Tien D, Kelly TJ (1997) *Genes Dev* 11:2767–2779.
27. Pelizon C, Madine MA, Romanowski P, Laskey RA (2000) *Genes Dev* 14:2526–2533.
28. Herbig U, Griffith JW, Fanning E (2000) *Mol Biol Cell* 11:4117–4130.
29. Perkins G, Drury LS, Diffley JF (2001) *EMBO J* 20:4836–4845.
30. Weinreich M, Liang C, Chen HH, Stillman B (2001) *Proc Natl Acad Sci USA* 98:11211–11217.
31. Jiang W, Wells NJ, Hunter T (1999) *Proc Natl Acad Sci USA* 96:6193–6198.
32. Kim J, Feng H, Kipreos ET (2007) *Curr Biol* 17:966–972.
33. Takeda DY, Parvin JD, Dutta A (2005) *J Biol Chem* 280:23416–23423.
34. Thomer M, May NR, Aggarwal BD, Kwok G, Calvi BR (2004) *Development (Cambridge, UK)* 131:4807–4818.
35. Nguyen VQ, Co C, Li JJ (2001) *Nature* 411:1068–1073.
36. Komamura-Kohno Y, Karasawa-Shimizu K, Saitoh T, Sato M, Hanaoka F, Tanaka S, Ishimi Y (2006) *FEBS J* 273:1224–1239.
37. Montagnoli A, Valsasina B, Brotherton D, Troiani S, Rainoldi S, Tenca P, Molinari A, Santocanale C (2006) *J Biol Chem* 281:10281–10290.
38. Laman H, Peters G, Jones N (2001) *Exp Cell Res* 271:230–237.
39. Moses AM, Heriche JK, Durbin R (2007) *Genome Biol* 8:R23.
40. Mendez J, Zou-Yang XH, Kim SY, Hidaka M, Tansey WP, Stillman B (2002) *Mol Cell* 9:481–491.
41. Li CJ, Vassilev A, DePamphilis ML (2004) *Mol Cell Biol* 24:5875–5886.
42. Mailand N, Diffley JF (2005) *Cell* 122:915–926.
43. Miyake Y, Mizuno T, Yanagi K, Hanaoka F (2005) *J Biol Chem* 280:12643–12652.
44. Yan H, Merchant AM, Tye BK (1993) *Genes Dev* 7:2149–2160.
45. Nguyen VQ, Co C, Irie K, Li JJ (2000) *Curr Biol* 10:195–205.
46. Kearsey SE, Labib K (1998) *Biochim Biophys Acta* 1398:113–136.
47. Mimura S, Seki T, Tanaka S, Diffley JF (2004) *Nature* 431:1118–1123.
48. Tak YS, Tanaka Y, Endo S, Kamimura Y, Araki H (2006) *EMBO J* 25:1987–1996.
49. Strickfaden SC, Winters MJ, Ben-Ari G, Lamson RE, Tyers M, Pryciak PM (2007) *Cell* 128:519–531.
50. Nash P, Tang X, Orlicky S, Chen Q, Gertler FB, Mendenhall MD, Sicheri F, Pawson T, Tyers M (2001) *Nature* 414:51.
51. Lenz P, Swain PS (2006) *Curr Biol* 16:2150–2155.
52. Kim SY, Ferrell JE, Jr. (2007) *Cell* 128:1133–1145.
53. Ludwig MZ, Patel NH, Kreitman M (1998) *Development (Cambridge, UK)* 125:949–958.
54. Ludwig MZ, Bergman C, Patel NH, Kreitman M (2000) *Nature* 403:564–567.
55. Gasch AP, Moses AM, Chiang DY, Fraser HB, Berardini M, Eisen MB (2004) *PLoS Biol* 2:e398.
56. Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB (2005) *Nature* 433:481–487.
57. Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, Osmotherly L, Li R, Liu T, Zhang Z, Bolund L, *et al.* (2006) *Nucleic Acids Res* 34:D572–D880.
58. Byrne KP, Wolfe KH (2005) *Genome Res* 15:1456–1461.
59. Notredame C, Higgins DG, Heringa J (2000) *J Mol Biol* 302:205–217.
60. Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T, Schroeder M, *et al.* (1998) *Nucleic Acids Res* 26:73–80.
61. Yang Z (1997) *Comput Appl Biosci* 13:555–556.
62. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, *et al.* (2002) *Nucleic Acids Res* 30(1):38–41.
63. Stoye J, Evers D, Meyer F (1998) *Bioinformatics* 14:157–163.
64. Durbin R, Eddy S, Krogh A, Mitchison G (1998) *Biological Sequence Analysis: Probabalistic Models of Proteins and Nucleic Acids* (Cambridge Univ Press, Cambridge, UK).
65. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, *et al.* (2006) *Nature* 443:818–822.
66. Fitzpatrick DA, Logue ME, Stajich JE, Butler G (2006) *BMC Evol Biol* 6:99.
67. Detweiler CS, Li JJ (1998) *Proc Natl Acad Sci USA* 95:2384–2389.
68. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL (2000) *Nucleic Acids Res* 28:263–266.