[Fly 1:3, e1-e9, EPUB Ahead of Print: http://www.landesbioscience.com/journals/fly/article/4599; May/June 2007]; ©2007 Landes Bioscience

Research Paper

Molecular Evolution of Drosophila *Cdc6*, an Essential DNA Replication Licensing Gene, Suggests an Adaptive Choice of Replication Origins

Benjamin L. Wiggins[1,2]
Harmit S. Malik[2,*]

[1]Molecular and Cellular Biology Program; University of Washington; Seattle, Washington USA

[2]Division of Basic Sciences; Fred Hutchinson Cancer Research Center; Seattle, Washington USA

*Correspondence to: Harmit S. Malik; Division of Basic Sciences; Fred Hutchinson Cancer Research Center; 1100 Fairview Avenue N. A1 162; Seattle, Washington 98109 USA; Tel.: 206.667.5204; Fax: 206.667.6497; Email: hsmalik@fhcrc.org

Original manuscript submitted: 05/25/07
Revised manuscript submitted: 06/18/07
Manuscript accepted: 06/19/07

This manuscript has been published online, prior to printing for Fly, Volume 1, Issue 3. Definitive page numbers have not been assigned. The current citation is: Fly 2007; 1(3):
http://www.landesbioscience.com/journals/fly/article/4599
Once the issue is complete and page numbers have been assigned, the citation will change accordingly.

KEYWORDS

McDonald-Kreitman test, non-coding DNA, replication origins, origin 'choice', *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Drosophila pseudoobscura*

ACKNOWLEDGEMENTS

We thank Josh Bayes, Nels Elde, Mosur Raghuraman, Sara Sawyer, Danielle Vermaak, Giovanni Bosco and one anonymous referee for their comments on the manuscript. We also wish to thank Steve Henikoff and Teri Markow for their encouragement, Corbin Jones and Daven Presgraves for their advice on codon bias and MK tests, and the Drosophila Species Center and our Drosophila colleagues (in particular Soojin Yi and Mohammed Noor) for the kind gift of Drosophila strains. This work was supported by a National Institute of Health grant GM074108 and a Searle Scholar Award (to Harmit S. Malik).

NOTE

Supplementary figures can be found at www.landesbioscience.com/supplement/WigginsFLY1-3-sup.pdf

ABSTRACT

Increased size of eukaryotic genomes necessitated the use of multiple origins of DNA replication, and presumably selected for their efficient spacing to ensure rapid DNA replication. The sequence of these origins remains undetermined in metazoan genomes, leaving important questions about the selective constraints acting on replication origins unanswered. We have chosen to study the evolution of proteins that recognize and define these origins every cell cycle, as a surrogate to the direct analysis of replication origins. Among these DNA replication proteins is the essential Cdc6 protein, which acts to license origins for replication. We find that two different species pairs of Drosophila show evidence of positive selection in *Cdc6* in their highly conserved C-terminal AAA-ATPase domain. We also identified amino acid segments that are highly conserved in the N-terminal tail of Cdc6 proteins from various Drosophila species, but are not conserved even in closely related insect species. Instead, we find that the N-terminal tails of Cdc6 proteins vary extensively in size and sequence across different eukaryotic lineages. Our results suggest that choice of origin firing may be significantly altered in closely related species, as each set of replication proteins optimizes to its own genomic landscape.

©2007 LANDES BIOSCIENCE. DO NOT DISTRIBUTE.

INTRODUCTION

Precise transfer of genetic information is a basic necessity for life. Genomic DNA must be replicated accurately and packaged into daughter cells at each cell division. In eukaryotes, this is accomplished by duplication of genomic DNA in S-phase and subsequent segregation during M-phase. Considering the importance of fidelity of DNA replication, it is not surprising that the molecular components of replication systems are highly conserved from most archaea to all eukaryotes.[1] The transition to larger genome sizes and multiple chromosomes in eukaryotes also required several adaptations in DNA replication, including the coordination of multiple origins of replication to ensure both faithful and efficient duplication of eukaryotic genomes.[2,3] Pressure to replicate efficiently dictates that these origins cannot be spaced randomly because the largest continuous replicated regions of a chromosome ('replicon') will be rate-limiting for completion of replication. This origin spacing problem necessitates a more regular spacing of origins by some means.[4] Among eukaryotes studied so far, the sequence of genomic origins of replication is well-defined only in yeasts. An 11-bp consensus binding site embedded in 200 bp suffices as an autonomous origin of replication in *S. cerevisiae*[5,6] and 300 such origins are spaced throughout the *S. cerevisiae* genome, although these do not all fire every cell cycle.[7] A similar origin-like element has been identified in *Schizosaccharomyces pombe* (100–500bp, A/T rich, multiple domains),[8] but it appears that multiple sequences from the *S. pombe* genome bearing similar characteristics can also act as origins.[9]

Generally speaking, most origins of more complex eukaryotic genomes demonstrate only weak sequence preferences[10] and it has been suggested that they can be defined in a sequence-independent manner.[11,12] Studies of origin sequence specificity in some eukaryotes suggest that while most origins have distinct chromosomal locations, precisely which sequences have this activity are undefined, and much or all euchromatin is competent to initiate replication if necessary.[13-16] How complex eukaryotic genomes accomplish this task in the absence of *cis*-acting sequence information is largely unknown.[17] This is in contrast to cases where DNA replication is developmentally programmed, for instance in the specific amplification of chorion genes in insects, where the origins of DNA replication are well-defined by site-specific binding of proteins;[18,19] these origins are highly conserved across broad phylogenetic ranges. In addition, in early embryonic

stages of Drosophila development, rapid cycles of DNA replication ensue and it is estimated that DNA replication is completed every 3.4 minutes;[2] at this developmental stage, origins are spaced regularly and more closely together than in somatic cells in the same organism, which accomplish DNA replication 200-fold slower.

Another important factor in coordinating multiple origins of replication is ensuring that each replicon be replicated only once per cell cycle, a process carried out by the DNA replication licensing complex. DNA licensing requires the ordered recruitment of a few highly conserved proteins at origins of replication to form the pre-replication complex. This is followed by rapid removal of key components after initiation, thereby ensuring that each origin 'fires' once and only once per cell cycle. According to prevalent models of DNA replication, origins of replication are first bound by Origin Recognition Complex (ORC) proteins,[20] and these origins are 'licensed' by the binding of Cdc6 and Cdt1 proteins at the ORC.[12] This licensing allows the recruitment of the minichromosome maintenance (MCM) complex.[1,21] The MCM complex likely acts as a helicase, opening the DNA for the initial steps of replication,[22] and proceeding along the DNA after initiation as bidirectional DNA replication forks.

Cdc6/Cdt1 binding is the primary control point in licensing the replication of genomic DNA. Besides enhancing the interactions between ORC and chromatin[23] and between Cdt1 and the MCM complex, Cdc6 is stringently regulated to insure exactly one replicative firing per origin.[12] Overexpression of Cdc6 results in continuous replication in the absence of mitosis in *S. pombe*, even in G_2-arrested cells.[24] Therefore, limiting Cdc6 availability is an important step in preventing aberrant initiation. This is accomplished via proteolytic degradation in yeast and vertebrates.[1] The important control function for Cdc6 is also evident from its involvement in crucial developmental steps. In Xenopus, oocytes acquire the ability to replicate DNA during maturation upon breakdown of the nuclear envelope, and Cdc6 translation provides the sole missing factor needed to allow DNA replication,[25] a property apparently conserved in metazoans.[26] While traditional models suggest that Cdc6 acts downstream of ORC binding, recent studies have shown that the site-specific recruitment of Cdc6 to genomic DNA is sufficient to create an artificial origin of replication in mammalian cells.[27] Furthermore, recent findings suggest that the Cdc6 ATPase activity may directly regulate the stability of the ORC-Cdc6 complex; Cdc6's specificity of interaction with DNA sequences may therefore directly translate into specifying which origins will successfully 'fire'.[28] These results suggest that Cdc6 binding can influence origin choice directly.

A previous screen had suggested *Cdc6* from *S. cerevisiae* as a candidate gene that may be evolving under positive selection.[29] Positive selection is typically seen in instances where protein sequence innovation has selectively beneficial consequences. For instance, in the screen in which *Cdc6* was putatively identified, most other genes were involved in some host-pathogen conflict, where amino acid substitutions would directly lead to an evolutionary advantage (to evade host defenses, for instance). Because *Cdc6* is essential, regulatory and single-copy, the possibility of positive selection led us to investigate the molecular evolution of *Cdc6* in the Drosophila lineage, which presents more statistical power to detect positive selection because a large number of strains are available from multiple, closely related species. We found an unambiguous signature of positive selection acting on the *Cdc6* gene in two independent species pairs of Drosophila. Intriguingly, this signature appears restricted to the highly conserved C-terminal domain, but not in the more variable N-terminal tail. Analysis of nearly 30 additional species reveals several protein segments in the N-terminal tail that have been highly conserved in the ~60 million years of Drosophila evolution. Surprisingly, these highly conserved protein motifs are absent from any of the other characterized Cdc6 proteins in other eukaryotes, even in other Diptera. Rapid evolution of such an essential DNA replication gene is very unusual. We discuss possible models that may have led to this unusual selective regime.

MATERIALS AND METHODS

Fly strains. Strains from *D. melanogaster* and *D. simulans* were obtained from the National Drosophila Species Resource Center (Tucson, Arizona) and have been previously described.[30] Various Drosophila species were obtained from the Tucson Drosophila Species Stock Center. *D. trilutea* was a generous gift from Dr. MT Kimura (Hokkaido University, Japan). The *D. miranda* flies were a generous gift from Soojin Yi (then at the Univ. of Chicago, presently at Georgia Tech University) and the *D. pseudoobscura* strains were a gift from Dr. Mohammed Noor (Duke University).

PCR and DNA sequencing. Primers were designed to the *D. melanogaster Cdc6* gene (Primer Table included in Supplementary Information) and PCR was carried out using PCR Supermix Hifi (Invitrogen). Both strands of PCR products were directly sequenced throughout the coding region using Big Dye sequencing (Applied Biosystems) and sequences were analyzed using Vector NTI software. Sequencing of Cdc6 from a species panel was performed using primers constructed to regions of conservation between *D. melanogaster* and *D. pseudoobscura*. In most instances, species subgroup specific primers had to be employed to finish PCR and sequencing. The sequences of 12 diverse Drosophila species were used as a guide to design primers to finish sequencing. However, the N- and C-termini of *Cdc6* could not be completely sequenced in some species.

Evolutionary and bioinformatic analyses. Sequences from strain and species panels were aligned using CLUSTAL_X[31] and manually adjusted for gaps. dN/dS sliding window tests were performed using alignments from pairwise alignments created in CLUSTAL_X and analyzed using Kestimator.[32] Evolutionary tests for deviations from neutrality (McDonald-Kreitman tests) were performed using DnaSP.[33] From the multiple alignment of the various Cdc6 proteins, BLOCKS of protein conservation were identified using the multiple alignment processor (http://blocks.fhcrc.org[34]). These position-specific scoring matrices (PSSMs) were displayed as Logos,[35] a graphical representation of aligned sequences where at each position the size of each letter is proportional to the frequency of that particular residue in that position and the total height of all of the letters in the position is proportional to the conservation (information content) of the position. Letters in the logo are colored according to the physical and chemical characteristics of the amino acid residues they specify. PSSMs generated were also used to search both the non-redundant database (protein and nucleotide) and a collection of known Cdc6 proteins using the Motif Alignment Search Tool (MAST)[36,37] using the same methodology as previously described for Drosophila centromeric histones.[38] Sequences obtained by this study are deposited in Genbank under the accession nos. DQ839668-DQ839694, DQ839711-DQ839733, DQ839735-DQ839743.

Maximum likelihood analysis was performed with codeml in the PAML 3.14.1 software package. To detect selection, multiple alignments were fitted to the NSsites models that disallow positive

Table 1 **McDonald-Kreitman tests for positive selection in *Cdc6*, comparing *D. melanogaster* versus *D. simulans***

	Fixed (Interspecies)	Polymorphic (Intraspecies)	Ratios (Rf:Sf::Rp:Sp)	P-Value*
Pooled				
Replacement	10 (N) + 2 (C)	23 (N) + 8 (C)	10:9::23:22 (N)	0.911 (N)
Synonymous	9 (N) + 11 (C+I)	22 (N) + 114 (C+I)	2:11::8:114 (C+I)	0.248 (C+I)
			12:20::31:136 (W)	**0.017 (W)***
D. melanogaster alone				
Replacement	6 (N) + 1 (C)	7 (N) + 3 (C)	6:6::7:5 (N)	0.682 (N)
Synonymous	6 (N) + 4 (C+I)	5 (N) + 35 (C+I)	1:5::3:35 (C+I)	0.381 (C+I)
			7:10::10:40 (W)	0.083 (W)
D. simulans alone				
Replacement	4 (N) + 1 (C)	16 (N) + 5 (C)	3:3::16:17 (N)	0.677 (N)
Synonymous	3 (N) + 7 (C+I)	17 (N) + 87 (C+I)	1:7::5:87 (C+I)	0.474 (C+I)
			5:10::21:104 (W)	0.157 (W)

N, C, I and W refer to the N-terminal tail, C-terminal domain, intronic regions and the whole gene respectively. *p-value calculated by the Chi-square test although similar results are obtained by the Fisher's exact test.

selection (M1, M7) or to models that permit positive selection (M2, M8 respectively) assuming the f61 model of codon frequencies.[39] Simulations were run with multiple starting values for dN/dS. Likelihood ratio tests were performed to assess whether permitting codons to evolve under positive selection gives a significantly better fit to the data. Similarly, multiple alignment input files were used to detect codons evolving under positive selection using the Random Effects Likelihood method (REL), in which statistical significance is measured in terms of Bayes factor (a BF > 50 is considered highly unlikely by chance).[40]

RESULTS

Molecular evolution of *Cdc6* in the melanogaster subgroup. The essential, single-copy gene *Cdc6* in *D. melanogaster* (*CG5971*) is approximately 2070 nucleotides long with one intron (spanning nucleotides 1570-1650). The highly conserved C-terminus is encoded by bases 750-2070, while the N-terminal tail consists of approximately the first 750 nucleotides. We sequenced *Cdc6* from a strain panel of *D. melanogaster* and *D. simulans*, two species that diverged 2.5 million years ago.[41] We determined the sequence of *Cdc6* from 11 strains of *D. melanogaster*, 15 strains of *D. simulans* and several outgroup species by PCR and direct sequencing. For both *D. melanogaster* ($\pi = 0.00826$, Tajima's D = -0.267) and *D. simulans* ($\pi = 0.01712$, Tajima's D = -0.429), there is a slight tendency towards rare, unshared polymorphisms. However, this is not significant, suggesting that no recent adaptive sweeps have occurred in either species.

We tabulated the sequencing data from the two species (Table 1), classifying each change as either a fixed change between species (f) or a polymorphic change between strains of the same species (p). We further classified these as either a replacement change (R) that alters the encoded amino acid residue, or a synonymous change (S) that does not alter the encoded amino acid. In the absence of adaptive evolution, the Rf:Sf ratio between species would be expected to be approximately equal to the Rp:Sp ratio between strains of the same species.[42]

Two cases which would deviate from this expectation are: an excess of replacement polymorphisms within species which denotes the circulation of deleterious mutations,[43] or an excess of replacement fixed changes between species, indicating positive selection.[42] The *melanogaster-simulans* comparison displays an excess of fixed replacement changes (Table 1). The Rf:Sf ratio between *melanogaster* and *simulans* is 12:20, while the Rp:Sp ratio within both species is 30:136 (intronic changes are assumed to be synonymous[44,45]). The number of fixed replacement changes deviates from the expected number at a statistically significant level ($p < 0.02$). This suggests that *Cdc6*, a single copy gene encoding an essential DNA replication licensing protein has evolved under positive selection between *D. melanogaster* and *D. simulans*. We can also estimate the rate at which adaptive changes were fixed within this gene over these 5 million years of divergence. In the absence of positive selection, we might expect only ~ 5 fixed replacement changes (30/136 x 20). Even under a statistical cutoff of $p < 0.05$, non-adaptive evolution is not sufficient to explain at least two of the fixed changes between these species (i.e., an Rf:Sf of 10:20 is not significantly different from 30:136). Such a low rate of adaptive changes (a minimum of two replacements in five million years of species divergence) in *Cdc6* suggests that there is a subtle selective advantage associated with adaptive changes in DNA replication.

Which domains has the adaptive evolution affected? An HKA test comparing the N-terminal tail and the C-terminal domain did not find significant differences in the polymorphism spectrum for either of the two species ($p > 0.1$). There appears to be virtually no evidence of higher than expected Rf (suggestive of adaptive evolution) in the N-terminal tail (10:9::23:22) which shows a high Rf but also tolerating more Rp, suggesting a lower degree of selective constraint (Table 1). However, the C-terminal domain shows a slightly higher than expected Rf (2:11::8:114) although this is not significant overall because of the low numbers of Rf (Table 1). There is similarly not enough statistical power to determine whether positive selection happened predominantly along the lineage leading to *D. melanogaster* or *D. simulans* although Rf:Sf ratios are comparable across the two lineages in both domains (Table 1).

It has been suggested previously that codon bias might affect the interpretation of McDonald-Kreitman tests because this would impose stringent selection on synonymous sites.[46] To account for this possibility, we compared the codon bias acting on *Cdc6* in both species. We find that the degree of codon bias in *Cdc6* is not unusual (Effective number of codons: 46.37 and 46.24[47] and Codon Bias index of 0.41 & 0.42[48] in *D. melanogaster* and *D. simulans* respectively) when compared to other genes.[49] Thus, we do not find

Table 2 **McDonald-Kreitman tests for positive selection in *Cdc6*, comparing *D. pseudoobscura* versus *D. miranda***

	Fixed (Interspecies)	Polymorphic (Intraspecies)	Ratios (Rf:Sf::Rp:Sp)	P-Value
Pooled				
Replacement	1 (N) + 2 (C)	8 (N) + 1 (C)	1:0::8:11 (N)	0.257 (N)
Synonymous	0 (N) + 2 (C+I)	11 (N) + 45 (C+I)	2:2::1:45 (C+I)	**0.0001 (C+I)***
			3:2::9:56 (W)	**0.0083 (W)***
D. pseudoobscura alone				
Replacement	1 (N) + 1 (C)	5 (N) + 1 (C)	1:0::5:6 (N)	0.296 (N)
Synonymous	0 (N) + 0 (C+I)	6 (N) + 29 (C+I)	1:0::1:29 (C+I)	**0.0001 (C+I)***
			2:0::6:35 (W)	**0.0024 (W)***
D. miranda alone				
Replacement	0 (N) + 1 (C)	3 (N) + 0 (C)	0:0::3:5 (N)	cannot be calculated
Synonymous	0 (N) + 2 (C+I)	5 (N) + 16 (C+I)	1:2::0:16 (C+I)	**0.018 (C+I)***
			1:2::3:21 (W)	0.338 (W)

N, C, I and W refer to the N-terminal tail, C-terminal domain, intronic regions and the whole gene respectively. *p-value calculated by the Chi-square test although similar results are obtained by the Fisher's exact test.

any evidence of high codon bias acting on *Cdc6*. Nonetheless, one approach to minimize the confounding effects of weakly deleterious synonymous mutations on the McDonald-Kreitman tests is to avoid the use of unpreferred synonymous mutations in both the fixed or polymorphic categories.[49] When we compare ratios of Rf: Sf(preferred)+i and Rp:Sp (preferred)+i, we get a ratio of 12:4 versus 30:44 ($p < 0.02$), which is still significant (compare to Table 1). This suggests that a confounding effect of codon bias is not likely to have given a false signature of adaptive evolution.

Another important caveat to the use of McDonald-Kreitman tests is that even transient relaxation of selective constraints during species' separation followed by a reimposition of selective constraints within extant populations can lead to the appearance of an elevated Rf, which might be interpreted as positive selection. This situation can be encountered either when the individual gene function is not essential (transient "pseudogene-like" phase), or in case of a gene duplication, where constraints on one particular copy are relaxed,[50] or if the population expands up significantly and selective constraints become more stringent.[42,51] However, we can argue that all these possibilities do not apply to our current analyses. First, *Cdc6* is essential, and presumed to be so throughout the evolutionary distances we have analyzed. Second, we have no evidence to suggest that *Cdc6* has undergone any gene duplications in this period. Third, at least in the *D. melanogaster- D. simulans* pair, a significant expansion of effective population size does not appear to have affected the efficacy of the McDonald-Kreitman test.[51] It thus appears that even essential DNA replication proteins are not immune from the torrid rate of adaptive evolution seen in *D. melanogaster*.[52]

***Cdc6* evolution in the pseudoobscura subgroup.** We wanted to investigate whether the positive selection signature we had observed in the melanogaster subgroup was typical of *Cdc6* evolution more generally in Drosophila. To this end, we chose to investigate polymorphism and divergence patterns in an additional pair of species from a distant part of the Drosophila species phylogeny, the pseudoobscura subgroup, which diverged from the melanogaster subgroup approximately 25 mya. We chose *D. miranda* and *D. pseudoobscura*, two species that diverged about two million years ago in the pseudoobscura subgroup.[53] We sequenced eight strains each of *D. pseudoobscura* and *D. miranda*. The polymorphism maps and McDonald-Kreitman data are shown in Table 2. For both *D. pseudoobscura* (π= 0.00923, Tajima's D = -0.335) and *D. miranda* (π = 0.00531, Tajima's D = -0.209), we found no evidence of a recent adaptive sweep.

The *D. pseudoobscura*/ *D. miranda* species showed an Rf:Sf::Rp:Sp ratio of 3:2::9:56. Like in the melanogaster subgroup, *Cdc6* evolution in this species pair also rejects neutrality in the McDonald-Kreitman test ($p < 0.01$), which also indicates that at least 1 of the fixed changes between *D. miranda* and *D. pseudoobscura* cannot be explained in the absence of adaptive evolution. The pace of adaptive evolution is similar to that seen in the melanogaster subgroup i.e., one adaptive replacement change in approximately four million years of divergence. However, in this comparison, we were able to conclusively determine that the adaptive evolution happened in the C-terminal conserved domain along both the *D. pseudoobscura* and *D. miranda* lineages (Table 2). Thus, in both the melanogaster and pseudoobscura subgroups, the signature of adaptive evolution is seen in the C-terminal domain of *Cdc6*. This is intriguing because the C-terminal half of *Cdc6* encodes the AAA+ ATPase domain, which is highly conserved amongst all eukaryotes in both length and sequence, while it is the N-terminal tail that appears to be much more variable.

Molecular evolution of *Cdc6* in the Drosophila lineage. To investigate the selective constraints that shape the evolution of the more variable N-terminal tail, we determined the sequence of the *Cdc6* gene from 34 species within the Drosophila lineage (including seven species whose genome sequences were released before or during this study) and constructed a neighbor-joining tree with the amino acid alignments. Since *Cdc6* is a single-copy, essential gene, Figure 1 provides valuable phylogenetic detail in the evolutionary analysis of Drosophila species; our phylogeny is largely consistent with previous analyses on this subject from a variety of nuclear and mitochondrial genes.[54-56]

Maximum likelihood methods based on multiple alignments are a powerful means of identifying codons that are repeatedly subject to positive selection;[39,40] identification of such codons allows the identification of functional 'surfaces' in proteins.[57,58] However, in the case of the Cdc6 proteins, some of the fixed amino acid changes between *D. melanogaster*/*D. simulans* and *D. pseudoobscura*/*D. miranda* occur in positions that are otherwise quite well conserved in all other species. We therefore do not expect to identify these particular

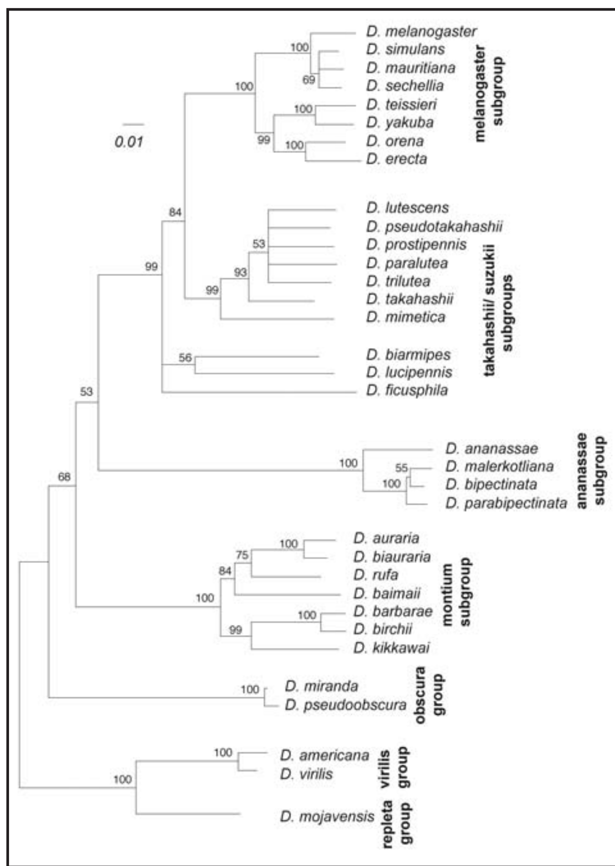

Figure 1. Cdc6 phylogeny. A neighbor-joining tree based on an amino acid alignment of the Cdc6 proteins in Drosophila is presented. Bootstrap support (percentage of 1,000 trials) is shown next to each node. This phylogeny is largely congruent with that of Drosophila species.

amino acid positions as being especially prone to positive selection. In order to determine whether any positions do show a signature of recurrent positive selection in the C-terminal domain, we used the codeml program and compared models of codon evolution (NSites M7) in which positive selection was prohibited, to models in which positive selection was allowed (M8).[39] We found no strong evidence for positive selection ($p > 0.05$). We also used the Random Effects Likelihood method[40] against this dataset and found two positions that showed suggestive evidence of positive selection (corresponding to amino acids 315 and 524 in the full length *D. melanogaster* Cdc6 protein, with Bayes factors of 54 and 75 respectively). This translates to a rather weak signature of recurrent positive selection,[59] suggesting that positive selection on *Cdc6* is episodic and likely involves different amino acid residues in different lineages.

We also identify many highly conserved blocks in the N-terminal tail of Cdc6 (Fig. 2), using methods previously applied to an analysis of Drosophila centromeric histones.[38] The high degree of conservation of these Cdc6 blocks suggest that they carry out important roles in Cdc6 function and may mediate important protein-protein interactions, as has been shown for human Cdc6.[60] We used these highly conserved Drosophila blocks and performed a search against the non-redundant database of proteins as well as a database of selected Cdc6 proteins from various eukaryotes. To our surprise, we were unable to find significant matches to any other protein in the database, including many other arthropod *Cdc6* genes (*Anopheles gambiae, Aedes aegyptii, Tribolium castaneum, Bombyx mori*) while blocks constructed from the C-terminal conserved domain accurately identified *Cdc6* orthologs from all eukaryotes. Thus, in spite of the high degree of conservation of the protein blocks outlined in Figure 2 (many of which are as highly conserved as the C-terminal domain within Drosophila species), they are apparently not conserved even in other Dipteran species. We have found that highly conserved segments of the vertebrate Cdc6 proteins are similarly not found in other lineages (data not shown). This suggests an intriguing short-term conservation but long-term turnover of these motifs, which might contribute to the general lack of homology between Cdc6 N-termini. Indeed, when we compared the Cdc6 proteins encoded in a variety of eukaryotic genomes, we found considerable variation in length and sequence in the N-terminal tail of Cdc6 where N-termini from different Cdc6 proteins are not even homologous to each other (no significant BLAST e-values).

The variability we see in the N-terminal tail could indicate a region largely devoid of function and undergoing little selection apart from maintaining an open reading frame. This alternative is belied by the important roles that the N-terminal tails of Cdc6 are known to carry out in a variety of eukaryotes. For instance, this N-terminal tail is known to be essential in humans and likely encodes the interaction interface between Cdc6 and Cdt1.[60] Phosphorylation sites in the N-terminus appear to be important for regulation of Cdc6 degradation and chromatin binding[61] and removal of a 47-amino acid segment of the N-terminus from yeast Cdc6 prevents degradation.[62] Similarly, a conserved RXL motif in the N-terminus of *Xenopus* Cdc6 is required for Cyclin E binding, which may facilitate initiation and/or prevent repeated initiations.[63] Instead, our analysis suggests that Cdc6 N-terminal tails might perform different functions in different eukaryotic lineages. This might be because specific functions in DNA replication licensing may be partitioned differently between various proteins in separate taxa, as has been recently suggested for Cdc6 and ORC1 in Drosophila.[64]

Revisiting *Cdc6* evolution in Saccharomyces species. Since the original observation on possible positive selection on *Cdc6* was made using sliding window dN/dS analyses in *S. cerevisiae*,[29] considerably more sequence information has become available. We wanted to follow up on this initial finding using the complete genome sequences of a number of species that are now available.[65] We therefore performed sliding window dN/dS analyses on Cdc6 in all pairwise combinations in the sensu stricto group of Saccharomyces, but found no indication of positive selection. Thus, we were unable to recreate the original observation of positive selection[29] but we must point out that we do not know which species pairs were compared in this earlier study, and what window sizes were used in the dN/dS study. We also sequenced eight strains each of the closely related species *S. cerevisiae* and *S. paradoxus* to perform a McDonald-Kreitman test, which did not find any evidence for positive selection because there were too few intraspecies polymorphisms to have any statistical power for this test (data not shown).

Since the *Cdc6* orthologs from the various sensu stricto species were well aligned throughout their length with very few indels, we also employed multiple alignment methodology using maximum likelihood techniques to ascertain whether any amino acid positions showed a consistent signature to detect positive selection. To this end, we also sequenced the complete *Cdc6* gene from *S. pastorianus* and *S. cariocanus*, two additional species from the sensu stricto group, and performed REL and PAML analyses[39,40] to detect positive selection. We found no evidence for positive selection in any of our comparisons (no positively selected sites with Bayes factor > 50 in REL analysis),

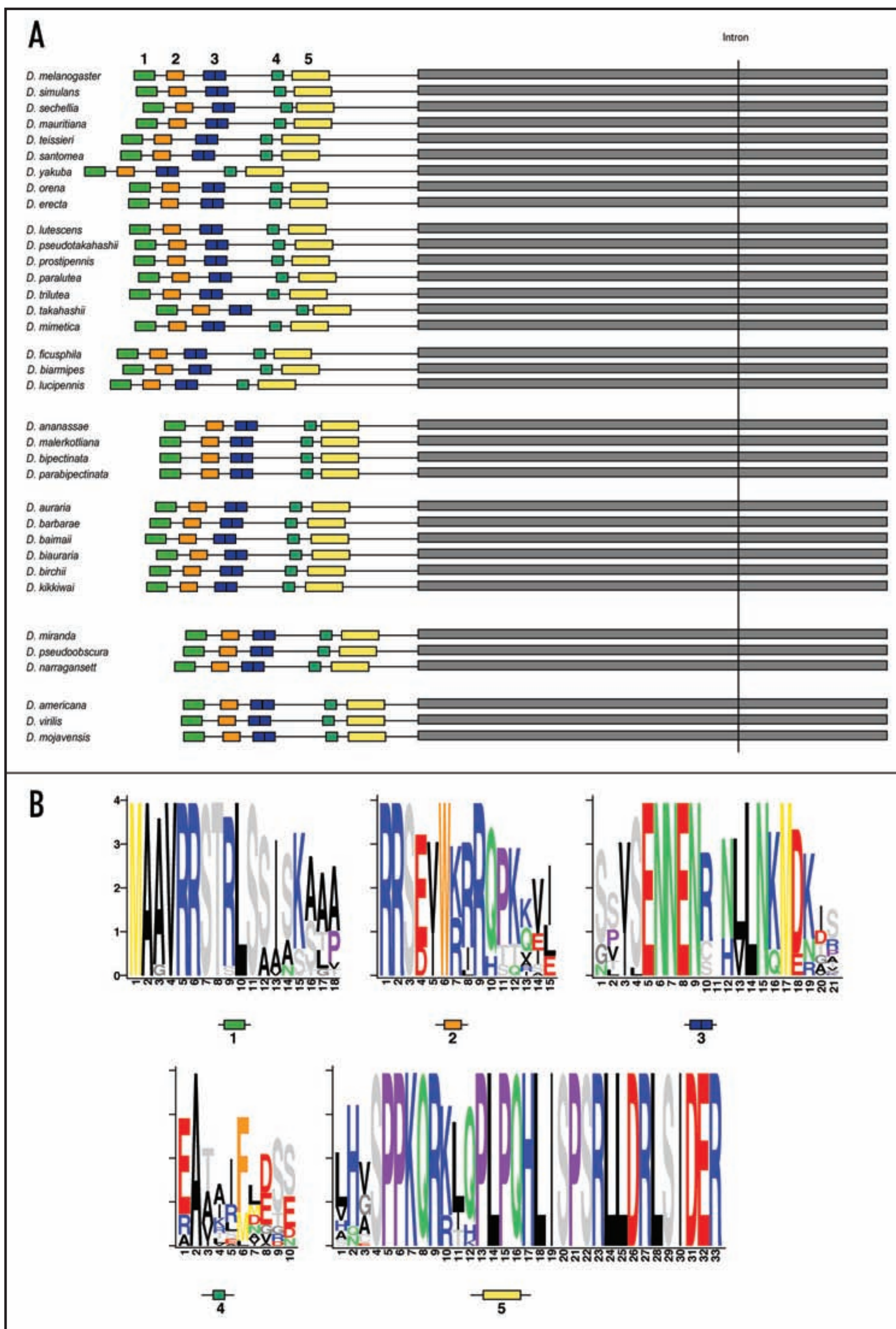

Figure 2. Evolution of Cdc6's N-terminal tail. (A) Schematic representation of *Cdc6* from different Drosophila species. The N-terminal tail is shown as a series of conserved sequence blocks (numbered 1 through 5) and a thin line, while the C-terminal domain (consisting of the AAA+ ATPase and winged helix-like DNA binding domains) is shown in a gray box. The position of the intron conserved through all the Drosophila species is shown with a thin vertical line. (B) The 5 conserved blocks schematized in (B) are presented in Logos format. Note that position 11 in Block 3 is a gapped position.

which all overwhelmingly suggested purifying selection as the overall theme in yeast *Cdc6* evolution.

It is disappointing not to have found a signature of positive selection in yeast *Cdc6* because we know of the compendium of replication origins only in *S. cerevisiae*,[7] and a functional analysis to evaluate the significance of *Cdc6*'s positive selection would only currently be possible in this species. It is also surprising considering that the original suggestion that *Cdc6* is under positive selection was made in yeast[29] and because this result has been widely cited in most recent reviews on which genes are shaped by positive selection.[66] Instead, we discuss the possibility that the *Cdc6* genes in Drosophila and yeast operate under different selective regimes because while Drosophila has large blocks of late-replicating heterochromatin, *S. cerevisiae* does not.

DISCUSSION

What might drive positive selection in *Cdc6*? What force could be acting on the DNA replication licensing system that might drive adaptive evolution in Drosophila? We consider several possibilities. One possibility is that Drosophila Cdc6 proteins function very differently than what has been proposed for yeast and vertebrate Cdc6s. However, recent studies in *D. melanogaster* indicate that Cdc6 appears to function in a similar manner, at least in tissue culture cells.[67] *Cdc6* participates in biological processes other than DNA replication in some lineages,[68] which might impose specific biological pressures. For instance, Cdc6 serves to tightly link cell cycle and gamete surface properties in some metazoans,[69] therefore it is formally possible that sexual selection plays a role in its evolution, as has been seen extensively in sperm-egg interactions.[70,71]

Drosophila species also go through a period of rapid replication during early embryonic development (wherein the entire genome is estimated to be replicated in 3.4 minutes[2,72]), during which delayed replication could be a significant selective detriment.[73,74] Moreover, Drosophila have specialized tissues which uncouple DNA replication and cell division, for the purpose of ramping up gene copy number for transcriptional purposes (e.g., chorion genes). It is conceivable that pressures associated with this selective DNA amplification could provide the impetus for positive selection of *Cdc6* exclusively in Drosophila.[3,75]

Another impetus for positive selection of Cdc6 could be a host-pathogen interaction. Many DNA viruses rely on host replication machineries to replicate themselves as episomes in host cells. Many others also rely on shutting down host cells at various stages of the cell cycle to ensure virus budding. In both these instances, it is conceivable that interactions with host DNA replication proteins may lead to genetic conflicts that cause the adaptive evolution of *Cdc6*. The selfish element need not be

episomal. Selfish DNA-based transposable elements (P, mariner) use a 'cut-and-paste' mechanism for increasing their copy number. Because this mechanism does not lead to an increase in copy number, it is in the element's best interests to excise from the 'donor site' only after the replication machinery has already duplicated this site, so that the excising element can be replenished by repair machineries using the sister chromatid.[76] For DNA-mediated transposons, insertion into regions proximal to origin sequences therefore helps ensure a greater opportunity for copy number increase. One means of accomplishing this would be to use DNA replication proteins assembled on origins as a 'beacon' to guide insertion sites, perhaps by protein-protein interactions between the transposition machinery and replication proteins. Intriguingly several DNA-based transposons have 'hotspots' for insertion in the Drosophila genome,[77] and avoid inserting into heterochromatic sequences.[78] Whether this correlates with DNA replication, or transcriptional status (these two features themselves appear strongly correlated[10,79]), or yet another feature of chromatin, is currently an open question. If it were replication, a genetic conflict with transposons could explain adaptive changes in *Cdc6*, as host genomes evolve away from this interaction.

Recent findings suggest the possibility that changes in Cdc6 protein sequence might influence the pattern of replication initiation timing[27] by affecting the subset of origins that successfully fire during replication initiation.[28] Under this model, the ATPase activity of Cdc6 modulates the stability of the Cdc6-ORC complex specifically on certain DNAs, and thereby determines which DNA sequences will successfully act as origins of replication. Amino acid replacements in Cdc6's ATPase domain may therefore alter the 'preference' of Cdc6 for certain DNAs over others. The positive selection we have observed may be a result of selection for altering that pattern, presumably to optimize the time required to finish DNA replication. Changes in the replication pattern would be especially necessary in the case of large-scale changes in the genome, like large expansions or deletions, such as those that happen in heterochromatin.[2] Heterochromatic regions of the genome are especially noteworthy because they are devoid of origins of replication,[80-82] easily imposing a rate-limiting step in the replication of complex genomes like Drosophila.[2] Large-scale changes in heterochromatin, by transposition, rearrangements or unequal crossing over in satellite repeats thus provide a subtle impetus to reorganize the landscape of replication origin firing. These adaptive changes could be due to selection of newer origins of DNA replication, either by direct DNA binding preferences[28] or by protein-protein interactions with heterochromatin proteins. A large-scale two-hybrid screen of interactions among Drosophila proteins[83] suggests that an interaction does exist between Cdc6 and a testis-specific heterochromatin protein, HP1E[84] which may be similar to the interaction between the ORC proteins, Cdc6 and HP1 in metazoans.[85,86] Quite presciently, exactly this scenario was originally imagined by Hogness and colleagues[2]: "...We imagine that it is the heterochromatization of the regions containing the satellite DNAs that provides the mechanism for controlling the duration of S phase in somatic cells. Perhaps it is the kind, amount and distribution of the various satellite DNAs in the centromeric DNA that determines the nature of this structural control."

Selective constraints on origins of DNA replication. What selective constraints act on origins of replication? Computational analyses of putative replication origins in mammalian and yeast genomes suggest that the majority of 'predicted' origins of replication can be preserved for tens of millions of years in yeast[87] and mammals.[88] However, in a complex eukaryotic genome with thousands of origins of DNA replication, each individual origin of replication is likely to be only minimally preserved by purifying selection. This important issue is unlikely to be suitably addressed in an appropriate manner by comparative genomics, due to the large amount of effort required to identify the compendium of replication origins in even a relatively simple eukaryotic genome.[7] In this context, our discovery of adaptive evolution in proteins involved in determining choice of origins of replication is one of the strongest pieces of evidence that they are acted upon by natural selection, despite the fact that the selective coefficient of these adaptive changes may be relatively small. Under the model whereby DNA replication origins are defined epigenetically by the binding of replication proteins, this may help explain the lack of any global sequence conservation of replication origins in higher eukaryotes.

Similar 'surrogate' approaches using proteins bound to other non-coding DNA like centromeres[30] and heterochromatin[84] have also previously revealed that, like DNA replication origins, these important components of eukaryotic genomes are being actively shaped by natural selection. Indeed, there are several striking parallels between our analyses of Cdc6 and centromeric histone proteins in Drosophila; both evolve under positive selection,[30] have well defined C-terminal domains, but have N-terminal tails that evolve so rapidly that they cannot be compared across different eukaryotic taxa.[38,89] In addition, both centromeres and DNA replication origins are genetically defined by conserved DNA segments in *S. cerevisiae* but are poorly defined in metazoan genomes. Since non-coding DNA like origins of replication are hard to define genetically and appear to be seemingly tolerant of a variety of mutational insults, this might tempt one to conclude that changes in these are likely of little selective consequence, by and large. Our findings challenge this view and validate the utility of our surrogate approaches to interrogate essential biological processes like DNA replication to establish the role that natural selection plays in shaping these important non-coding DNA, that are not active targets of study in comparative genomics.

References

1. Bell SP, Dutta A. DNA replication in eukaryotic cells. Annu Rev Biochem 2002; 71:333-74.
2. Blumenthal AB, Kriegstein HJ, Hogness DS. The units of DNA replication in *Drosophila melanogaster* chromosomes. Cold Spring Harb Symp Quant Biol 1974; 38:205-23.
3. Spradling AC. *ORC* binding, gene amplification, and the nature of metazoan replication origins. Genes Dev 1999; 13:2619-23.
4. Blow JJ, Gillespie PJ, Francis D, Jackson DA. Replication origins in *Xenopus* egg extract Are 5-15 kilobases apart and are activated in clusters that fire at different times. J Cell Biol 2001; 152:15-25.
5. Fangman WL, Hice RH, Chlebowicz-Sledziewska E. ARS replication during the yeast S phase. Cell 1983; 32:831-8.
6. Brewer BJ, Fangman WL. The localization of replication origins on ARS plasmids in *S. cerevisiae*. Cell 1987; 51:463-71.
7. Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL. Replication dynamics of the yeast genome. Science 2001; 294:115-21.
8. Dubey DD, Kim SM, Todorov IT, Huberman JA. Large, complex modular structure of a fission yeast DNA replication origin. Curr Biol 1996; 6:467-73.
9. Dai J, Chuang RY, Kelly TJ. DNA replication origins in the *Schizosaccharomyces pombe* genome. Proc Natl Acad Sci USA 2005; 102:337-42.
10. MacAlpine DM, Rodriguez HK, Bell SP. Coordination of replication and transcription along a Drosophila chromosome. Genes Dev 2004; 18:3094-105.
11. MacAlpine DM, Bell SP. A genomic view of eukaryotic DNA replication. Chromosome Res 2005; 13:309-26.
12. Takeda DY, Dutta A. DNA replication and progression through S phase. Oncogene 2005; 24:2827-43.
13. Mello CC, Kramer JM, Stinchcomb D, Ambros V. Efficient gene transfer in *C.elegans*: Extrachromosomal maintenance and integration of transforming sequences. Embo J 1991; 10:3959-70.
14. Kim CS, Preer Jr JR, Polisky B. Bacteriophage lambda DNA fragments replicate in the Paramecium macronucleus: Absence of active copy number control. Dev Genet 1992; 13:97-102.

15. Smith JG, Calos MP. Autonomous replication in *Drosophila melanogaster* tissue culture cells. Chromosoma 1995; 103:597-605.
16. Coverley D, Laskey RA. Regulation of eukaryotic DNA replication. Annu Rev Biochem 1994; 63:745-76.
17. Cvetic C, Walter JC. Eukaryotic origins of DNA replication: Could you please be more specific? Semin Cell Dev Biol 2005; 16:343-53.
18. Austin RJ, Orr-Weaver TL, Bell SP. Drosophila ORC specifically binds to ACE3, an origin of DNA replication control element. Genes Dev 1999; 13:2639-49.
19. Beall EL, Manak JR, Zhou S, Bell M, Lipsick JS, Botchan MR. Role for a Drosophila Myb-containing protein complex in site-specific DNA replication. Nature 2002; 420:833-7.
20. Mendez J, Zou-Yang XH, Kim SY, Hidaka M, Tansey WP, Stillman B. Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. Mol Cell 2002; 9:481-91.
21. Tye BK. MCM proteins in DNA replication. Annu Rev Biochem 1999; 68:649-86.
22. Diffley JF. Regulation of early events in chromosome replication. Curr Biol 2004; 14:R778-86.
23. Harvey KJ, Newport J. Metazoan origin selection: Origin recognition complex chromatin binding is regulated by CDC6 recruitment and ATP hydrolysis. J Biol Chem 2003; 278:48524-8.
24. Nishitani H, Lygerou Z, Nishimoto T, Nurse P. The Cdt1 protein is required to license DNA for replication in fission yeast. Nature 2000; 404:625-8.
25. Lemaitre JM, Bocquet S, Mechali M. Competence to replicate in the unfertilized egg is conferred by Cdc6 during meiotic maturation. Nature 2002; 419:718-22.
26. Lemaitre JM, Bocquet S, Terret ME, Namdar M, Ait-Ahmed O, Kearsey S, Verlhac MH, Mechali M. The regulation of competence to replicate in meiosis by Cdc6 is conserved during evolution. Mol Reprod Dev 2004; 69:94-100.
27. Takeda DY, Shibata Y, Parvin JD, Dutta A. Recruitment of ORC or CDC6 to DNA is sufficient to create an artificial origin of replication in mammalian cells. Genes Dev 2005; 19:2827-36.
28. Speck C, Stillman B. Cdc6 ATPase activity regulates ORC x Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication. J Biol Chem 2007; 282:11705-14.
29. Endo T, Ikeo K, Gojobori T. Large-scale search for genes on which positive selection may operate. Mol Biol Evol 1996; 13:685-90.
30. Malik HS, Henikoff S. Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics 2001; 157:1293-8.
31. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997; 25:4876-82.
32. Comeron JM. K-Estimator: Calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics 1999; 15:763-4.
33. Rozas J, Rozas R. DnaSP version 2.0: A novel software package for extensive molecular population genetics analysis. Comput Appl Biosci 1997; 13:307-11.
34. Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. Blocks-based methods for detecting protein homology. Electrophoresis 2000; 21:1700-6.
35. Schneider TD, Stephens RM. Sequence logos: A new way to display consensus sequences. Nucleic Acids Res 1990; 18:6097-100.
36. Bailey TL, Gribskov M. Concerning the accuracy of MAST E-values. Bioinformatics 2000; 16:488-9.
37. Bailey TL, Gribskov M. Methods and statistics for combining motif match scores. J Comput Biol 1998; 5:211-21.
38. Malik HS, Vermaak D, Henikoff S. Recurrent evolution of DNA-binding motifs in the *Drosophila* centromeric histone. Proc Natl Acad Sci USA 2002; 99:1449-54.
39. Wong WS, Yang Z, Goldman N, Nielsen R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 2004; 168:1041-51.
40. Pond SL, Frost SD. Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 2005; 21:2531-3.
41. Li YJ, Satta Y, Takahata N. Paleo-demography of the *Drosophila melanogaster* subgroup: Application of the maximum likelihood method. Genes Genet Syst 1999; 74:117-27.
42. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature 1991; 351:652-4.
43. Nachman MW, Brown WM, Stoneking M, Aquadro CF. Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics 1996; 142:953-63.
44. Rooney AP, Zhang J. Rapid evolution of a primate sperm protein: Relaxation of functional constraint or positive Darwinian selection? Mol Biol Evol 1999; 16:706-10.
45. Wyckoff GJ, Wang W, Wu CI. Rapid evolution of male reproductive genes in the descent of man. Nature 2000; 403:304-9.
46. Akashi H. Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA. Genetics 1995; 139:1067-76.
47. Wright F. The 'effective number of codons' used in a gene. Gene 1990; 87:23-9.
48. Morton BR. Chloroplast DNA codon use: Evidence for selection at the psb A locus based on tRNA availability. J Mol Evol 1993; 37:273-80.
49. Schlenke TA, Begun DJ. Natural selection drives Drosophila immune system evolution. Genetics 2003; 164:1471-80.
50. Zhang J, Rosenberg HF, Nei M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA 1998; 95:3708-13.
51. Eyre-Walker A. Changing effective population size and the McDonald-Kreitman test. Genetics 2002; 162:2017-24.
52. Smith NG, Eyre-Walker A. Adaptive protein evolution in Drosophila. Nature 2002; 415:1022-4.
53. Wang RL, Hey J. The speciation history of *Drosophila pseudoobscura* and close relatives: Inferences from DNA sequence variation at the period locus. Genetics 1996; 144:1113-26.
54. Goto SG, Kimura MT. Phylogenetic utility of mitochondrial COI and nuclear *Gpdh* genes in *Drosophila*. Mol Phylogenet Evol 2001; 18:404-22.
55. Malik HS, Henikoff S. Positive selection of Iris, a retroviral envelope-derived host gene in *Drosophila melanogaster*. PLoS Genet 2005; 1:e44.
56. Kopp A. Basal relationships in the *Drosophila melanogaster* species group. Mol Phylogenet Evol 2006; 39:787-98.
57. Sawyer SL, Wu LI, Emerman M, Malik HS. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proc Natl Acad Sci USA 2005; 102:2832-7.
58. Yang Z. The power of phylogenetic comparison in revealing protein function. Proc Natl Acad Sci USA 2005; 102:3179-80.
59. Kosakovsky Pond SL, Frost SD. Not so different after all: A comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 2005; 22:1208-22.
60. Cook JG, Chasse DA, Nevins JR. The regulated association of Cdt1 with minichromosome maintenance proteins and Cdc6 in mammalian cells. J Biol Chem 2004; 279:9625-33.
61. Sanchez M, Calzada A, Bueno A. The Cdc6 protein is ubiquitinated in vivo for proteolysis in *Saccharomyces cerevisiae*. J Biol Chem 1999; 274:9092-7.
62. Drury LS, Perkins G, Diffley JF. The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. Embo J 1997; 16:5966-76.
63. Furstenthal L, Kaiser BK, Swanson C, Jackson PK. Cyclin E uses Cdc6 as a chromatin-associated receptor required for DNA replication. J Cell Biol 2001; 152:1267-78.
64. Araki M, Wharton RP, Tang Z, Yu H, Asano M. Degradation of origin recognition complex large subunit by the anaphase-promoting complex in Drosophila. Embo J 2003; 22:6115-26.
65. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 2003; 423:241-54.
66. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends in Ecology and Evolution 2000; 15:496-503.
67. Crevel G, Mathe E, Cotterill S. The *Drosophila* Cdc6/18 protein has functions in both early and late S phase in S2 cells. J Cell Sci 2005; 118:2451-9.
68. Gonzalez S, Klatt P, Delgado S, Conde E, Lopez-Rios F, Sanchez-Cespedes M, Mendez J, Antequera F, Serrano M. Oncogenic activity of Cdc6 through repression of the *INK4/ARF* locus. Nature 2006; 440:702-6.
69. Tian J, Thomsen GH, Gong H, Lennarz WJ. *Xenopus* Cdc6 confers sperm binding competence to oocytes without inducing their maturation. Proc Natl Acad Sci USA 1997; 94:10729-34.
70. Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proc Natl Acad Sci USA 2001; 98:7375-9.
71. Swanson WJ, Wong A, Wolfner MF, Aquadro CF. Evolutionary expressed sequence tag analysis of Drosophila female reproductive tracts identifies genes subjected to positive selection. Genetics 2004; 168:1457-65.
72. Rabinowitz M. Studies on the cytology and early embryology of the egg of *Drosophila melanogaster*. J Morphol 1941; 69:1.
73. Edgar BA, Lehner CF. Developmental control of cell cycle regulators: A fly's perspective. Science 1996; 274:1646-52.
74. Lee LA, Orr-Weaver TL. Regulation of cell cycles in Drosophila development: Intrinsic and extrinsic cues. Annu Rev Genet 2003; 37:545-78.
75. Calvi BR, Spradling AC. Chorion gene amplification in Drosophila: A model for metazoan origins of DNA replication and S-phase control. Methods 1999; 18:407-17.
76. Engels WR, Johnson-Schlitz DM, Eggleston WB, Sved J. High-frequency P element loss in Drosophila is homolog dependent. Cell 1990; 62:515-25.
77. Green MM. Genetic instability in *Drosophila melanogaster*: De novo induction of putative insertion mutations. Proc Natl Acad Sci USA 1977; 74:3490-3.
78. Berg CA, Spradling AC. Studies on the rate and site-specificity of P element transposition. Genetics 1991; 127:515-24.
79. Schubeler D, Scalzo D, Kooperberg C, van Steensel B, Delrow J, Groudine M. Genome-wide DNA replication profile for *Drosophila melanogaster*: A link between transcription and replication timing. Nat Genet 2002; 32:438-42.
80. Barigozzi C, Dolfini S, Fraccaro M, Raimondi GR, Tiepolo L. In vitro study of the DNA replication patterns of somatic chromosomes of *Drosophila melanogaster*. Exp Cell Res 1966; 43:231-4.
81. Dolfini S. Karyotype polymorphism in a cell population of *Drosophila melanogaster* cultured in vitro. Chromosoma 1971; 33:196-208.
82. Ahmad K, Henikoff S. Centromeres are specialized replication domains in heterochromatin. J Cell Biol 2001; 153:101-10.
83. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley Jr RL, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM. A protein interaction map of *Drosophila melanogaster*. Science 2003; 302:1727-36.

84. Vermaak D, Henikoff S, Malik HS. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. PLoS Genet 2005; 1:96-108.
85. Auth T, Kunkel E, Grummt F. Interaction between HP1alpha and replication proteins in mammalian cells. Exp Cell Res 2006; 312:3349-59.
86. Pak DT, Pflumm M, Chesnokov I, Huang DW, Kellum R, Marr J, Romanowski P, Botchan MR. Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. Cell 1997; 91:311-23.
87. Nieduszynski CA, Knox Y, Donaldson AD. Genome-wide identification of replication origins in yeast by comparative genomics. Genes Dev 2006; 20:1874-9.
88. Touchon M, Nicolay S, Audit B, Brodie of Brodie EB, d'Aubenton-Carafa Y, Arneodo A, Thermes C. Replication-associated strand asymmetries in mammalian genomes: Toward detection of replication origins. Proc Natl Acad Sci USA 2005; 102:9836-41.
89. Malik HS, Henikoff S. Phylogenomics of the nucleosome. Nat Struct Biol 2003; 10:882-91.