branch distinguishing the two clades. Analysis of the entire dataset (474 taxa) by heuristic search using maximum parsimony criteria yields the same major branches depicted in Fig. 2A. For the analysis of consensus sequences shown in Fig. 2B, 382 monomers from the monomeric clade (Fig. 2A) were aligned using ClustalW and a single X monomeric consensus sequence (Xmon) was derived. Bootstrap analysis was used to determine the relationship of Xmon to the consensus monomers established for the four human suprachromosomal families (28), again using AGM as an outgroup (Fig. 2B). Of 1000 bootstrap replicates, 92% support the major branch separating Xmon and AGM from the remaining consensus monomers. Our analysis indicates that Xmon is more closely related to AGM and to the M1 (suprachromosomal family 4) sequence than to any of the other human consensus monomers.

28. I. Alexandrov, A. Kazakov, I. Tumereva, V. Shepelev, Y. Yurov, *Chromosoma* **110**, 253 (2001).
29. S. J. Durfy, H. F. Willard, *Genomics* **5**, 810 (1989).
30. Specific positions within each DXZ1 monomer at which the sequence varies from a consensus DXZ1 monomer are designated signature sites. The analysis of signature sites was automated using a perl script, modifying the consensus identity index used previously (15). The script first identifies and then compares signature sites of test monomers with the specific signature sites of each DXZ1 monomer. Because any two higher-order repeats of typical DXZ1 (type 1) share 98 to 100% identity (31), the number of signature sites for each DXZ1 monomer varies slightly. Type 2, 3, and 4 DXZ1 repeats have numbers of signature sites that are outside the range of those detected in type 1 repeats.
31. Approximately 1000 to 2000 copies of the DXZ1 higher-order repeat exist on each X chromosome (9, 15). Average sequence divergence was estimated from the sequence of seven complete (M. G. Schueler et al., unpublished data) and over 40 partial repeat units (29), from at least nine different X chromosomes.
32. Type 2 and type 3 DXZ1 sequences show approximately the same degree of sequence divergence within each type as they do between types. This indicates that the gradient of divergence from type 1 DXZ1 is very steep and that the diverged types are no longer being homogenized to the same degree as type 1. This is consistent with a model in which the efficiency of homogenization decreases both as a function of physical distance from the type 1 array and as a direct consequence of the increased sequence dissimilarity itself. (Type 3 is represented by only a single repeat and thus cannot be evaluated from this perspective.)
33. D. J. Wolff et al., *Am. J. Hum. Genet.* **58**, 154 (1996).
34. K. Gustashaw et al., data not shown. The most proximal probe tested was 1050H14 (Fig. 1), containing both monomeric alpha satellite and types 3 and 4 DXZ1 sequences.
35. A. W. Higgins, M. G. Schueler, H. F. Willard, *Chromosoma* **108**, 256 (1999).
36. A. W. Higgins et al., data not shown. The Xq contig (54) terminates an estimated few hundred kb from DXZ1. Pulsed-field gel analysis was carried out as in (9). Isochromosomes with breaks within the DXZ1 array showed both missing and altered restriction fragments relative to the parental X chromosome prior to isochromosome formation. Isochromosomes with breaks outside of DXZ1 on proximal Xq showed unaltered DXZ1 restriction patterns. Isochromosomes with breaks in DXZ1 confirm Darlington's original model of centromere misdivision [C. D. Darlington, *J. Genet.* **37**, 341 (1939)].
37. J. J. Harrington, G. Van Bokkelen, R. W. Mays, K. Gustashaw, H. F. Willard, *Nature Genet.* **15**, 345 (1997).
38. M. K. Rudd et al., data not shown. A ~85 kb NotI fragment from BAC (242E23) containing DXZ1 higher-order repeats was isolated and cloned into the pPAC4 vector, containing a blasticidin-resistance gene [T. A. Ebersole et al., *Hum. Mol. Genet.* **9**, 1623 (2000)]. The circular PAC was then transfected into HT1080 cells, as described (37). Blasticidin-resistant colonies were isolated and screened cytogenetically by FISH for DXZ1-containing artificial chromosomes.

Tests of mitotic stability were carried out in the presence and absence of blasticidin. After 30 days, clones containing an artificial chromosome in >60% of cells were scored as stable (37). Indirect immunofluorescence for CENP-E staining was performed as described (37, 39).

39. B. A. Sullivan, S. Schwartz, *Hum. Mol. Genet.* **4**, 2189 (1995).
40. A. M. Laurent et al., *Genomics* **46**, 127 (1997).
41. A. F. Smit, G. Toth, A. D. Riggs, J. Jurka, *J. Mol. Biol.* **246**, 401 (1995).
42. Observed LINE elements (Fig. 5) include those detected through our sequence sampling efforts and from sequence within 344I7 or other sequenced clones available from sequence databases (4, 18, 25). Only L1 elements that were imbedded within alpha satellite were considered. A directed effort was made to identify LINE elements within the DXZ1 array. DNA from 45 BAC stabs (from the RPCI-13 library) positive for DXZ1 by PCR and hybridizing to 348d (type 2 DXZ1; Fig. 1) at high stringency were digested with BamHI, followed by Southern analysis using probes to the 5′ and 3′ ends of L1 at reduced stringency. Two clones yielded both the diagnostic DXZ1 2.0-kb BamHI fragment and hybridization with L1 probes following colony purification. Fragments from both BAC clones that hybridized with L1 probes were subcloned and sequenced. Restriction enzyme digestion patterns and sequence were identical between the subclones originating from the two BAC clones, indicating that these two clones sample the same LINE element.
43. I. G. Goldberg, H. Sawhney, A. F. Pluta, P. E. Warburton, W. C. Earnshaw, *Mol. Cell. Biol.* **16**, 5156 (1996).
44. L. Donehower, D. Gillespie, *J. Mol. Biol.* **134**, 805 (1979).
45. A. M. Laurent, J. Puechberty, G. Roizes, *Chromosome Res.* **7**, 305 (1999).
46. S. Ohno, *Nature* **244**, 259 (1973).
47. S. J. O'Brien et al., *Science* **286**, 458 (1999).
48. M. Ventura, N. Archidiacono, M. Rocchi, *Genome Res.* **11**, 595 (2001).
49. G. Montefalcone, S. Tempesta, M. Rocchi, N. Archidiacono, *Genome Res.* **9**, 1184 (1999).
50. T. D. Murphy, G. H. Karpen, *Cell* **82**, 599 (1995).
51. G. P. Copenhaver et al., *Science* **286**, 2468 (1999).
52. J. A. Bailey, A. M. Yavor, H. F. Massa, B. J. Trask, E. E. Eichler, *Genome Res.* **11**, 1005 (2001).
53. P. E. Warburton, *Trends Genet.* **17**, 243 (2001).
54. M. G. Schueler et al., *Genomics* **66**, 104 (2000).
55. Funded by a March of Dimes Franklin Delano Roosevelt Award and by NIH grants HD32111 and HG00107 (H.F.W.). M.K.R. is supported by NIH training grant HD07518. We thank E. Eichler, B. Grimes, J. Bailey, B. Sullivan, and J. Dunn for helpful discussions and assistance. H.F.W. particularly thanks J. Waye and G. Dover for stimulating thoughts regarding satellite DNA structure, function, and evolution.

Replication Dynamics of the Yeast Genome

M. K. Raghuraman,[1]*† Elizabeth A. Winzeler,[3]* David Collingwood,[2]*
Sonia Hunt,[1] Lisa Wodicka,[4]‡ Andrew Conway,[5]
David J. Lockhart,[4]§ Ronald W. Davis,[6] Bonita J. Brewer,[1]
Walton L. Fangman[1]

Oligonucleotide microarrays were used to map the detailed topography of chromosome replication in the budding yeast *Saccharomyces cerevisiae*. The times of replication of thousands of sites across the genome were determined by hybridizing replicated and unreplicated DNAs, isolated at different times in S phase, to the microarrays. Origin activations take place continuously throughout S phase but with most firings near mid–S phase. Rates of replication fork movement vary greatly from region to region in the genome. The two ends of each of the 16 chromosomes are highly correlated in their times of replication. This microarray approach is readily applicable to other organisms, including humans.

The replication of eukaryotic chromosomes is highly regulated. Replication is limited to the S phase of the cell cycle; and within S phase, initiation of replication is controlled with respect to both location and time. The sites of initiation, called replication origins, have been best characterized in the budding yeast *Saccharomyces cerevisiae*, in which a functional assay based on plasmid maintenance has allowed identification of potential origins of replication [autonomous replication sequence elements (ARSs)]. There are estimated to be ~200 to 400 ARSs in the yeast genome (1, 2), and most, but not all, function as chromosomal origins (3). The few origins investigated at the sequence level usually encompass ~200 base pairs (bp); most contain a perfect match or a one-base mismatch to an 11-bp ARS consensus sequence (ACS) (4, 5). However, the presence of an ACS is not sufficient to predict an origin of replication: There are many more ARS consensus sequences in the genome than origins

[1]Department of Genetics, [2]Department of Mathematics, University of Washington, Seattle, WA 98195, USA. [3]Genomics Institute of the Novartis Research Foundation, 3115 Merryfield Row, San Diego, CA 92121, USA. [4]Affymetrix, 3380 Central Expressway, Santa Clara, CA, USA. [5]Silicon Genetics, 935 Washington Street, San Carlos, CA 94070, USA. [6]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: raghu@u.washington.edu
‡Present address: Aventa Biosciences, 4757 Nexus Centre Drive, Suite 200, San Diego, CA 92121, USA.
§Present address: The Salk Institute for Biological Studies, Laboratory of Genetics, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA.

and the ~200 bp of sequence that flank the ACS, although essential, share no obvious sequence similarities (*3*).

From the analysis of a few origins, the general features of origin control in yeast appear to be as follows (*6–8*). Most (perhaps all) origins are found in intergenic regions, but only a subset are chosen as replication origins in any given S phase. Furthermore, origins range in efficiency from those that are active in almost every cell cycle to those that

are used in only a small proportion of cell cycles. Origins are activated in a reproducible temporal sequence through S phase. The time of origin activation appears not to be an intrinsic property of the origin itself; rather, it is imposed by cis-acting elements that are separable from the origin (*9, 10*). Some transacting factors involved with the temporal program have been identified recently (*11–14*), but how they read and execute the program is still unclear.

How origins are chosen for activation and how timing is determined remain open questions. With the availability of the yeast genome sequence and DNA microarray technologies, it is possible to expand our understanding of origin activity greatly by examining the kinetics of replication across the entire yeast genome. Here we provide the locations and times of activation of the most efficient origins, as well as the directions of replication fork movement and fork migra-
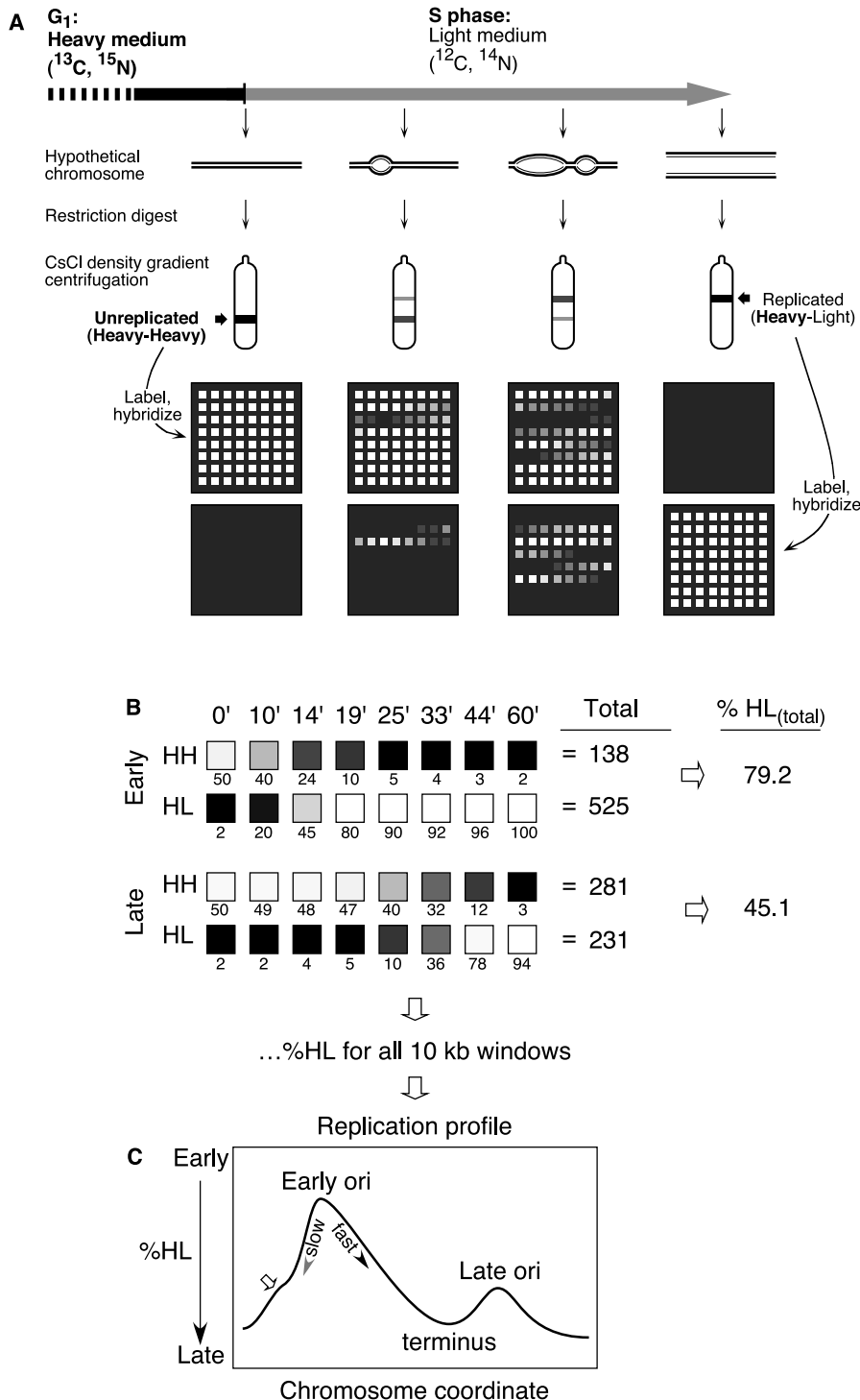
Fig. 1. Outline of the experiment and the data-processing steps. (**A**) *MAT***a** yeast cells carrying a temperature-sensitive lesion in *CDC7* are grown for many generations in medium containing ^{13}C glucose and ^{15}N nitrogen [(^{15}NH$_4$)$_2$SO$_4$] (*28*). The cells are arrested before the onset of DNA replication by the addition of α factor and by shifting the cells to the restrictive temperature, then are washed and resuspended in isotopically light medium. When DNA replication begins again, the newly synthesized DNA is labeled exclusively with the light isotopes. Samples are collected throughout S phase. The DNA is fragmented with a restriction enzyme (Eco RI) and fractionated by cesium chloride density-gradient centrifugation to separate the molecules carrying the two different density labels. These two fractions are then biotinylated and separately hybridized to high-density arrays carrying probes to the entire genome. (**B**) Data processing (*16*) illustrated with a hypothetical data set for one early- and one late-replicating region. The shaded boxes with the numbers below them represent cells of the microarray and the corresponding hybridization values (a black box corresponds to no signal). The sum of the hybridization intensities obtained with unreplicated, fully dense DNA (HH) for the various S phase samples is calculated, as is the sum of the intensities with replicated (HL) DNA. The % HL$_{(total)}$ is then calculated with the equation

$$\%HL_{(total)} = \left(\frac{\sum HL}{\sum HH + \sum HL} \right) 100$$

A DNA fragment that replicates early in S phase will be present in the HL fraction through much of the time course; in contrast, a late-replicating fragment will be represented in the HL fraction only at later times in S phase. Thus, the %HL$_{(total)}$ value is a direct reflection of the time of replication of a fragment, which we have calculated in the past as t_{rep} (*10, 11, 28*) [see (*17*)]. (**C**) The replication profile for a chromosome is constructed by plotting the %HL$_{(total)}$ as a function of the chromosome coordinate. Peaks represent regions that replicate earlier than the neighboring sequences and must therefore correspond to origins of replication; the taller a peak, the earlier the origin fires (ori, origin). Valleys correspond to termination zones. Shoulders along the lines connecting peaks and valleys (open arrow) could result either from inefficient origins or from changes in the fork migration rate. The slope of the line connecting a peak and a valley gives the direction and rate of fork migration through that region: A shallow slope reflects a fast fork, whereas a steep slope reflects a slow fork.

tion rates from these origins. After growth in isotopically dense culture medium, cells are released into S phase, and replicated [heavy-light (HL)] DNAs and unreplicated [heavy-heavy (HH)] DNAs are isolated from samples collected 10, 14, 19, 25, 33, 44, and 60 min later (Fig. 1). Each DNA sample is labeled and hybridized to a high-density oligonucleotide array (*15*). Hybridization reveals the relative abundance of specific genome sequences present in the unreplicated versus replicated DNAs at different times in S phase. Summing the hybridization values over all the samples for both HH and HL DNA at each chromosome coordinate gives an aggregate %HL value for that coordinate, which reflects its relative time of replication [Fig. 1B (*16*)] [see (*17*) for details]. Graphing these aggregate %HL [%HL$_{(total)}$] values ver-

sus chromosome coordinates yields a replication profile for a chromosome (Fig. 1C), with higher values of %HL$_{(total)}$ being indicative of earlier replication.

Interpreting the plots. To allow automated identification of peaks and valleys in the replication profiles, a Fourier convolution algorithm was used to smooth them (*16*, *17*). These smoothed profiles were used to identify potential origins of replication: Because a peak reveals a region that replicates before its neighbors, it must correspond to an origin of replication. Likewise, valleys correspond to termination zones ("terminus" in Fig. 1C). Tall peaks correspond to early-activated origins and shorter peaks to later-activated origins. Overall, we detected 332 origins in the yeast genome. Replication profiles for all chromosomes and the corresponding data in

tabular form can be found in the supplementary material (*17*).

Chromosome VI has been studied extensively (*18*, *19*) and is a good candidate for assessment of the microarray approach. The replication profile deduced from the microarray data for chromosome VI is shown in Fig. 2. Superimposed on this plot are the locations of restriction fragments that are known to contain origins [gray bars; heights of bars correspond to the replication times of the origin fragments themselves or of adjacent restriction fragments, determined previously (*18*)]. The array data clearly show a good match with the known replication characteristics of chromosome VI. For example, the tallest peak in the profile, which predicts the location of the earliest-activated origin on the chromosome, matches the known location of *ARS607*, which has been shown previously to be the earliest-activated origin on the chromosome (*18*). Likewise, the shorter peaks (such as the one centered at ~63 kb) correspond to the locations of origins known to be activated later in S phase.

Predicting origin locations. To further test the efficacy of the array data in predicting origin locations and replication times, we used two approaches. First, we asked how well the peaks in replication profiles reveal the locations of 18 origins known to be efficient and localized to regions of ≤5 kb (*20*). Peaks in the replication profiles occurred close to each of these 18 origins (*21*), with a mean distance of 3.7 kb between predicted origins and centers of known origin windows (Table 1), a match that is significantly better than that expected by chance ($P = 4.8 \times 10^{-17}$).

In a second approach, we examined chromosome X, which had been largely unexamined for replication. Four prominent peaks were selected at random from its replication profile (a, b, d, and e in Fig. 3A). Restriction fragments (~13 kb or smaller) corresponding to the predicted origins were tested by two-dimensional (2D) gel electrophoresis. Abundant bubble structures were detected at each of the four locations (Fig. 3B), confirming that they did correspond to active replication origins, whereas flanking restriction fragments tested did not show bubble structures (*22*). Given that bubble structures can be detected only if replication originates within the central half of a restriction fragment, and assuming that there are ~400 origins in the yeast genome (*2*), the probability of detecting origin activity in any random 13-kb fragment is ~0.22. Therefore, the likelihood of having origin activity in all four fragments just by chance is 0.002. This value is a conservative estimate, as two of the four fragments tested were actually substantially smaller than 13 kb. These results indicate that the replication profiles can indeed predict locations of effi-

Fig. 2. Replication profile for chromosome VI, showing t_{rep} plotted as a function of coordinate. The %HL$_{(total)}$ values were converted to t_{rep} values using the correlation between the set of %HL values and the corresponding set of t_{rep} values calculated for each coordinate (*16*, *17*). The small circle on the X axis marks the location of the centromere. Dark gray bars correspond to restriction fragments known to contain replication origins (*18*, *19*). The heights of the bars correspond to their relative replication times (*18*), scaled to match the t_{rep} values calculated for the microarray experiment. Inverted triangles mark the four origins that are used in ≥50% of the cell cycles. Gaps in the profile correspond to regions of low probe density. Numbers above the peaks indicate the robustness of the peaks on a scale of 1 to 9 [larger numbers are indicative of more robust peaks (*16*)].
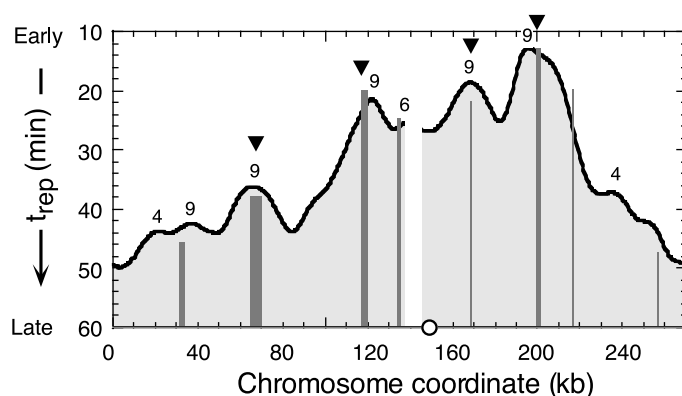

Table 1. Match between known efficient origins and peaks in replication profiles.

Chromosome	Name	Left end*	Right end*	Nearest peak (distance†)
II	SH1022	61.200	61.500	58.273 (3.077)
III	ARS305	39.132	39.682	38.301 (1.106)
III	ARS306	74.428	74.648	77.855 (3.317)
III	ARS307	108.775	108.975	115.906 (7.031)
III	ARS309	131.582	131.928	128.423 (3.332)
III	ARS310	166.211	167.060	174.486 (7.850)
IV	ARS1	462.611	462.648	464.369 (1.740)
IV	SH1023	750.000	752.500	755.444 (4.194)
V	ARS501	549.644	549.680	542.464 (7.198)
VI	ARS603	65.000	70.000	66.346 (1.154)
VI	ARS603.5	117.101	120.038	122.427 (3.858)
VI	ARS606	167.698	167.750	168.994 (1.270)
VI	ARS607	199.051	200.972	196.033 (3.979)
X	ARS121	683.653	683.702	684.266 (0.589)
XIV	ARS1411	168.000	170.100	158.780 (10.270)
XIV	ARS1412	195.652	196.883	201.788 (5.521)
XIV	ARS1413	250.800	250.849	250.798 (0.027)
XIV	ARS1414	280.001	280.049	279.803 (0.222)

*Coordinates are in kb from the left end of each chromosome.   †The distance to the center of the nearest known origin fragment is given in parentheses.
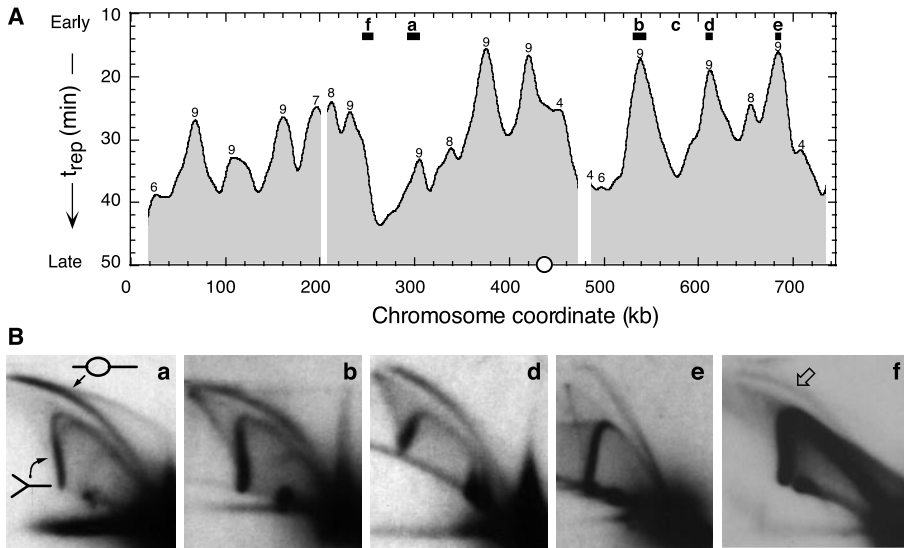
Fig. 3. Replication profiles predict origin locations. (**A**) Replication profile for chromosome X. Numbers above the peaks indicate the robustness of the peaks (*16*). (**B**) Origin activity at five locations on chromosome X. Features marked a, b, d, e, and f in the profile (A) were tested for origin activity by 2D gel electophoresis (*46*). The small horizontal bars under the letters in (A) correspond to the restriction fragments that were tested. The origin marked e corresponds to *ARS121*. Bubbles were detected in each fragment, indicating the presence of active origins. The long exposure required to see the bubble intermediates in region f indicates that the origin at ~250 kb is inefficient.
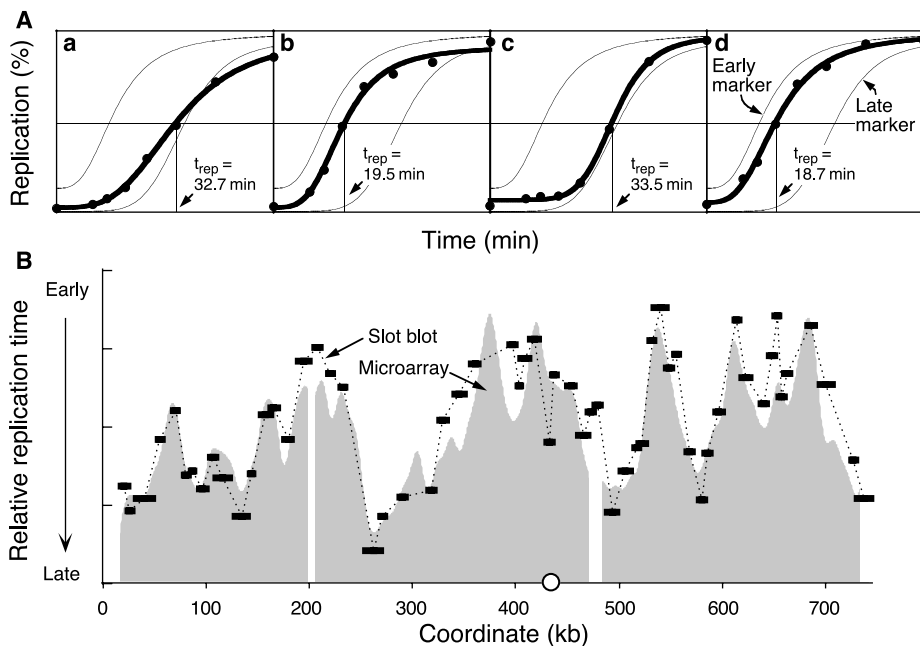
Fig. 4. Profiles predict the temporal program of replication. (**A**) Replication kinetics of fragments corresponding to locations marked a through d in Fig. 3A were determined by slot blot analysis as described (*28*). The DNA samples used were from the same experiment as was used for hybridization to the microarrays. "Early marker" and "late marker" refer to the replication kinetics of *ARS305* (chromosome III) and R11 (chromosome V), known to replicate very early and very late in S phase, respectively (*33*). (**B**) Replication times of 60 restriction fragments on chromosome X from slot blot hybridization (*23*). The profiles deduced from microarray data (gray shaded area; data of Fig. 3A) and from this slot blot experiment (bars connected by dotted lines) are shown. The width of a bar reflects the length of the restriction fragment that was tested; the height of the bar (mean of six replicates) corresponds to the relative %HL$_{(total)}$ for that fragment, and therefore corresponds to replication time: High bars represent early replication, and low bars represent late replication. Because this slot blot method measures the relative times of replication rather than the absolute times, the plot shows just the relative replication times (on the *y* axis), with the microarray and slot blot results scaled to match.

cient origins. The shoulders on the sides of peaks may indicate the locations of inefficient origins; one example of a 2D gel analysis detecting a weak origin from such a shoulder (f in Fig. 3A) is shown (Fig. 3B).

Some discrepancies between the array data and previous reports were observed; for example, the array data predicted an origin at ~235 kb on chromosome VI, where no origin had previously been reported. Although errors arising from poor hybridization signals or from assumptions made in the smoothing process (*21*) may lead to such discrepancies, they could also indicate genuine origins that had been missed previously. In the chromosome VI example, 2D gel analysis of the directions of fork migration supports the prediction of the array data (*22*).

Predicting replication times. Replication times predicted by the microarray analysis for several locations on chromosome X (a through d in Fig. 3A) were tested by slot blot analysis of CsCl density gradient fractions as described previously (*18*). In each case, the kinetics of replication predicted by the microarray data matched those detected by the standard slot blot method (Fig. 4A). As an additional test of the overall replication timing profile, the time of replication was assessed separately for ~60 restriction fragments on chromosome X (*23*), which correspond to annotated open reading frames (ORFs) in the *Saccharomyces* Genome Database. The replication profile deduced from the microarray experiment closely matched the profile obtained from this slot blot experiment (Fig. 4B).

Taken together, the success of the replication profiles in predicting both general origin locations and relative times of replication for different chromosomal segments demonstrates the viability of using microarrays to map the topography of replication. At present, origin activity has been confirmed only for origins predicted with high confidence levels (*16*). More precise assignments of origin locations, especially for less efficient origins, will have to await several repetitions of the experiment. However, the replication profiles will provide useful information even without precise and exhaustive identification of all origins, such as in the comparison of genotypes or culture conditions.

Origin activation times. S phase in the microarray experiment spans an interval of ~55 min. Although we have previously described origins as belonging to distinct "early" or "late" classes, this genome-wide analysis reveals that origins really show a continuum of activation times (Fig. 5A). Most of the origin firings occur near mid-S. This observation is consistent with measurements made on a randomly selected set of 24 origins (*20*) examined by our premicroarray assay

(*22*). Each chromosome shows a range of origin activation times, with the distribution of activation times varying from chromosome to chromosome (Fig. 5B). As deduced earlier from DNA fiber autoradiography for mammals and yeast (*1*, *24*), adjacent origins are generally activated at about the same time (Fig. 5C).

No sequence elements that were absolute predictors of origin location were found; nor have we, so far, uncovered any DNA sequence determinants that allow the prediction of replication time. An obvious feature to examine with respect to origin activation time is base composition. The yeast genome has broad variations in base composition along the length of the chromosomes, with AT-rich and GC-rich isochores occurring with a periodicity of ~50 kb (*25*). Although the 11-bp AT-rich ARS consensus sequence is expected to occur at elevated frequency in AT-rich isochores, there does not appear to be any correlation between base composition and either origin frequency or activation time (Fig. 5, D and E). One measure of how closely the replication profiles are dictated by the underlying DNA sequence is the extent of similarity between profiles for regions believed to have arisen from gene duplications. We were unable to detect any similarity in the replication profiles for the largest reported blocks of duplicated sequence (*26*, *27*). If the underlying DNA sequence does directly dictate the replication profile, the sequences in these blocks must have diverged enough to alter their replication profiles.

Replication of centromeres and telomeres. Consistent with previous observations on a subset of centromeres and telomeres (*28*), we find that centromeres are replicated earlier than subtelomeric regions [the most distal ORF sequences represented on the microarrays (Fig. 6A)]. The average replication time of 10-kb windows containing centromeres is significantly earlier than that of all 10-kb windows ($P = 1.9 \times 10^{-9}$); replication of subtelomeric regions, on the other hand, is much later than the genomic average ($P = 2.4 \times 10^{-16}$). However, subtelomeric regions as a class are not the last sequences to be replicated; for instance, sequences at ~280 kb on chromosome IV are replicated later than many telomeres. The times of replication of the two ends of the same chromosome are highly correlated (Fig. 6B), raising the possibility that intrachromosomal telomere interaction (*29*) may influence the time of origin activation.

The average activation time for the set of most distal predicted origins is ~5 min later than that of all other origins ($P = 2.5 \times 10^{-4}$), consistent with the observation that telomeres cause late activation of origins placed in their vicinity (*9*, *29*, *30*). Although this position effect is known to extend at least

~30 kb from the right telomere of chromosome V, the distance over which telomeres generally exert an effect had not been determined previously. To map the telomere position effect more systematically, the distribution of activation times of all predicted origins located within a sliding 20-kb window was compared to that of all predicted origins in the genome. As the sliding window was moved inward from the chromosome ends, the average activation time of predicted origins within the window approached that of the genomic average (Fig. 6C, top). The difference in average activation time was statistically significant ($P < 0.05$) until the 20-kb window was centered ~45 to 50 kb from the end (Fig. 6C, top), suggesting that the telomere position effect, on average, extends at least ~35 kb.

Because centromeres are replicated earlier than the bulk chromosomal sequences, we did a similar analysis for origins that flank centromeres. The average activation time for the 32 predicted origins flanking centromeres was found to be ~5 min earlier than that of all other origins ($P = 8.0 \times 10^{-5}$). As with the telomeres, this difference in activation

time, when compared to the genomic average, decreased at increasing distances from the centromeres (Fig. 6C, bottom): Predicted origins up to a distance of ~25 kb from the centromeres showed a significantly earlier activation time than the genomic average. These observations raise the possibility that centromeres, like telomeres, may have a position effect on origin activation. Consistent with this idea, introduction of a centromere close to one potential origin on a plasmid containing two identical ARS elements led to preferential activation of the origin closer to the centromere (*31*).

Fork migration rates. The rates of fork migration were measured by taking absolute values of the slopes of the lines connecting peaks and valleys (that is, origins and termini) in the replication profiles, ignoring the region immediately flanking each peak or valley (5 kb on either side), where local flattening of the curve introduces artifacts in the measurement of fork rates. As with origin activation times, a broad range of fork rates was observed (*32*), with a mean of 2.9 kb/min and a median of 2.3 kb/min (Fig. 7). These values are close to a previous estimate of fork
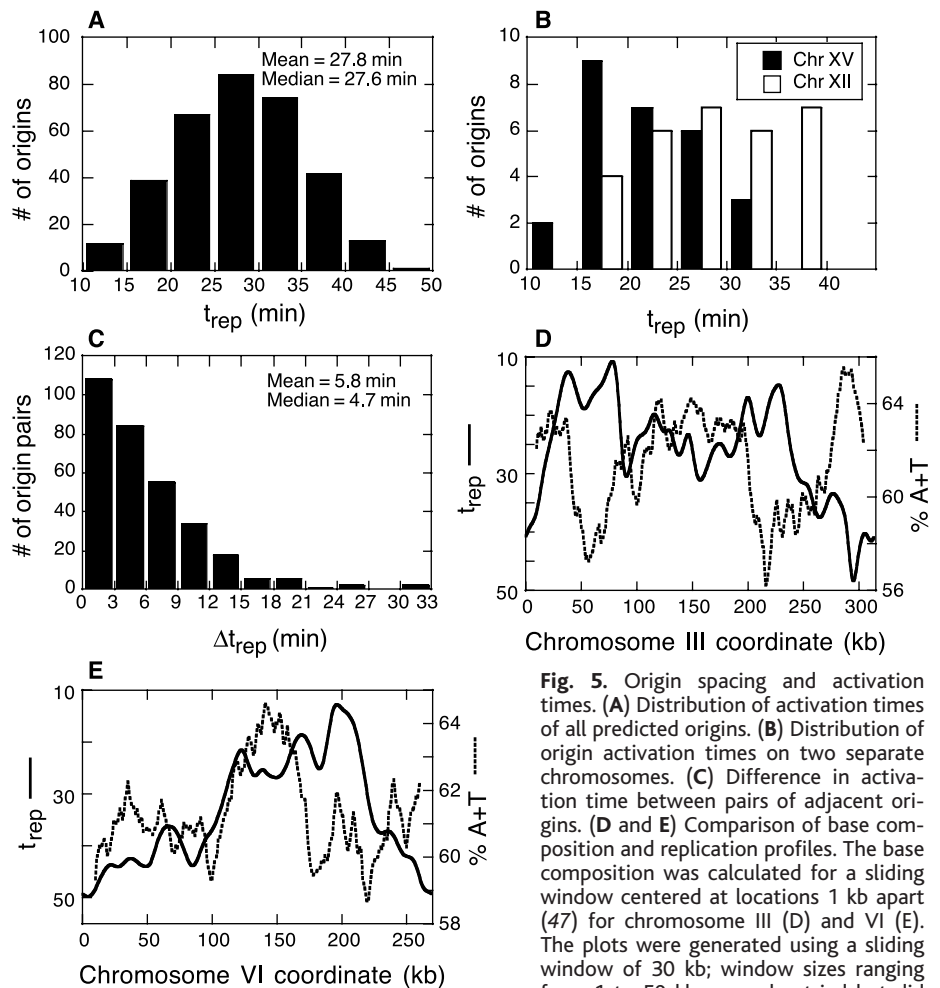
Fig. 5. Origin spacing and activation times. (**A**) Distribution of activation times of all predicted origins. (**B**) Distribution of origin activation times on two separate chromosomes. (**C**) Difference in activation time between pairs of adjacent origins. (**D** and **E**) Comparison of base composition and replication profiles. The base composition was calculated for a sliding window centered at locations 1 kb apart (*47*) for chromosome III (D) and VI (E). The plots were generated using a sliding window of 30 kb; window sizes ranging from 1 to 50 kb were also tried but did not reveal any correlations to origin location or replication time.
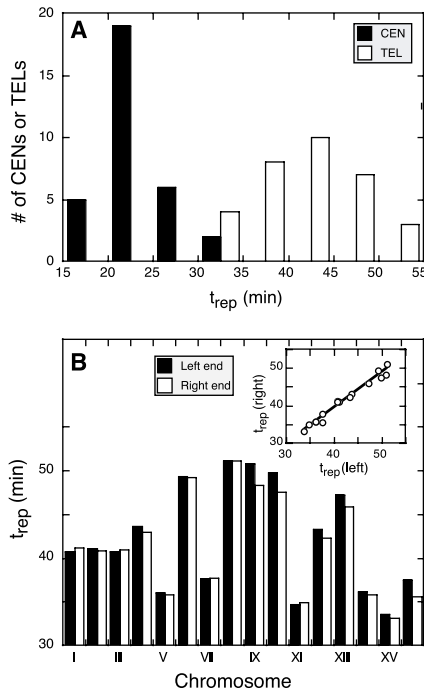
Fig. 6. Time of replication of centromeres and telomeres. (**A**) Distribution of replication times of centromere regions (CEN) and the most distal unique sequences represented on the oligonucleotide microarrays (TEL). (**B**) The times of replication of telomeres on the same chromosome. (Inset) Correlation in replication time between the left end and the right end of each chromosome. (**C**) Position effect on origin activation time. The plots show a comparison of the average activation time of predicted origins within a sliding 20-kb window centered at increments of 5 kb from chromosome ends (top panel) or centromeres (bottom panel). The difference in average t_{rep} for origins within the window when compared to the average of all predicted origins is plotted (Δt_{rep}; open circles) along with the probability in each instance of the observed difference being due to chance (P value; solid circles). The dashed horizontal line shows the threshold P value of $P = 0.05$.

Fig. 7. Replication fork migration rates. Fork migration rates were calculated for regions between peaks and valleys in the replication profiles. The histogram shows the maximum estimate for fork rate within each peak-valley interval.

migration rates made for isogenic cells grown under similar culture conditions [3.7 kb/min (*1*)]. Previously determined slow fork rates in two late-replicating regions [the right end of chromosome V and the left end of chromosome III (*10, 30, 33*)] are confirmed by our microarray data. However, there is no general correlation between fork rate and the time in S phase when the forks are initiated.

Differences in fork rates throughout the genome could result from some local property of the DNA sequence or chromatin structure being replicated by a fork. Alternatively, qualitative differences in proteins assembled at different origins (such as minichromosome maintenance protein composition differences) could result in different rates of fork movement. However, models proposing qualitative differences between forks at different origins would have to accommodate the observation that there are some origins where the leftward and rightward forks move at different rates (for example, compare the left and right slopes at the origin labeled "a" in Fig. 3, located at ~300 kb on chromosome X, where the leftward fork is estimated to be moving at 3.9 kb/min and the rightward fork at 1.4 kb/min). We cannot rule out the possibility that some origins are unidirectional in a

subset of the cells (*32*). If such origins exist, they could account for asymmetry in fork rates between leftward and rightward forks from an origin.

Replication and transcription. A general correlation between transcription and replication has been observed in mammalian cells—genes that are actively transcribed are often replicated early in S phase (*34*)—and has led to the suggestion that replication timing is one way for the cell to control transcription (*35, 36*). With one exception, we found no such correlation between transcription and replication time in yeast (*37*). The lone exception was the family of eight histone genes, which are replicated on average ~10 min earlier than the genome average of 31 min ($P = 6 \times 10^{-6}$). A limitation of our analysis is that we have only examined one growth regimen; it is possible that the program of replication changes when cells modify their transcription profile to adapt to altered growth conditions. We anticipate that some version of the method described here, in combination with single molecule studies (*38*) and genome-wide analysis of transcription, will be especially useful in examining cellular responses to altered growth conditions (including the response to drugs such as
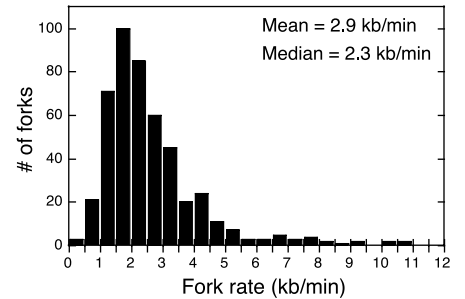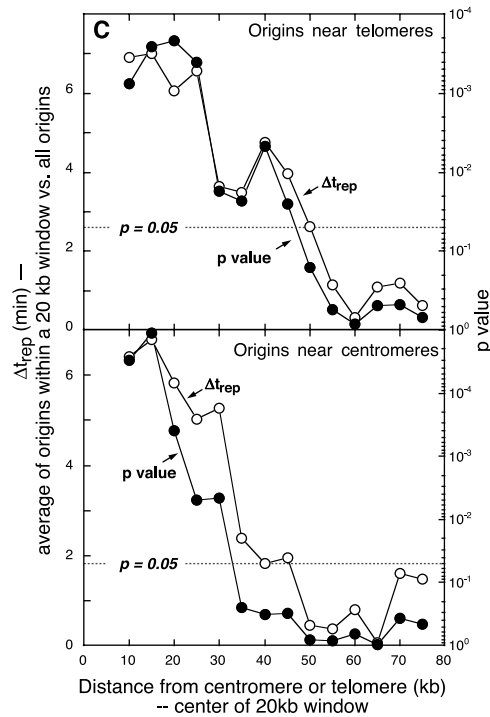
hydroxyurea or DNA-damaging agents such as methylmethane sulfonate).

For the future, we envision at least two immediate benefits of this approach. First, we are compiling a database of origin locations and activation times that will help identify consensus sequence elements affecting origin choice and activation time. Second, this method provides a powerful way of comparing the topography of chromosome replication in different cell cultures: between cells of different genotypes (wild type versus mutant) and for cells of the same genotype under different growth conditions.

We see no reason why this method [or related methods (*39*)] cannot be applied to other organisms, including cultured human cells. There is no requirement for sequenced genomes; all that is required is that an ordered set of unique-sequence clones spaced at reasonable intervals be available for constructing microarrays and that some method of synchronizing cell populations and of density-labeling the DNA be available.

References and Notes
1. C. J. Rivin, W. L. Fangman, *J. Cell. Biol.* **85**, 108 (1980).
2. C. S. Newlon, W. Burke, in *ICN-UCLA Symposium on Molecular and Cellular Biology* (Academic Press, New York, vol. 19, 1980), pp. 399–409.
3. C. S. Newlon, J. F. Theis, *Curr. Opin. Genet. Dev.* **3**, 752 (1993).
4. S. Kearsey, *Cell* **37**, 299 (1984).
5. J. V. Van Houten, C. S. Newlon, *Mol. Cell. Biol.* **10**, 3917 (1990).
6. R. Drouin, N. Lemieux, C. L. Richer, *Chromosoma* **99**, 273 (1990).
7. A. D. Donaldson, J. J. Blow, *Curr. Opin. Genet. Dev.* **9**, 62 (1999).
8. P. Pasero, E. Schwob, *Curr. Opin. Genet. Dev.* **10**, 178 (2000).
9. B. M. Ferguson, W. L. Fangman, *Cell* **68**, 333 (1992).
10. K. L. Friedman *et al.*, *Genes Dev.* **10**, 1595 (1996).
11. A. D. Donaldson *et al.*, *Mol. Cell* **2**, 173 (1998).
12. A. D. Donaldson, W. L. Fangman, B. J. Brewer, *Genes Dev.* **12**, 491 (1998).
13. K. Bousset, J. F. Diffley, *Genes Dev.* **12**, 480 (1998).
14. J. B. Stevenson, D. E. Gottschling, *Genes Dev.* **13**, 146 (1999).
15. The DNA samples were fragmented to an average size of ~50 bp with deoxyribonuclease I and then end-labeled with a biotinylated dideoxynucleotide. Before fragmentation, 100 ng of control DNA from bacterial genes having complementary probes on all

arrays was added as a control for normalization between arrays (*17*). Samples were hybridized sequentially to high-density oligonucleotide arrays containing probes to yeast genomic sequences. After hybridization, the biotinylated target was stained with a streptavidin-phycoerythin conjugate, and then the arrays were scanned. The oligonucleotide arrays (Affymetrix, Santa Clara, CA) are similar to those described elsewhere (*40*). Briefly, they consist of an ordered array of 25-oligomer oligonucleotides, with each annotated ORF in the *Saccharomyces* genome database being represented by 20 oligonucleotides. For every perfect match oligonucleotide, there is a single-base mismatch oligonucleotide on the arrays, serving as controls for the specificity of hybridization.

16. We outline the algorithm here; see (*17*) for details. For each 10-kb window in the genome, spaced at coordinates 0.5 kb apart, we measure the intensity of hybridization to HH and HL DNA; there are ~24,000 such data coordinates. For each data coordinate, the HH hybridization intensities obtained with the eight timed S phase samples [time (t) = 0, 10, 14, 19, 25, 33, 44, and 60 min] are summed, as are the eight HL values. These two sums are used to compute an aggregate or %HL$_{(total)}$ value for that coordinate. The percent representation of a sequence in the HL fraction is a measure of how early it replicates, and it is linearly correlated with t_{rep}, the time of half-maximal replication (*17*). A plot of %HL$_{(total)}$ versus chromosomal coordinates gives a replication profile showing the order of replication of the different coordinates across the genome. Local maxima in the %HL$_{(total)}$ plots reveal potential replication origins. The plots are smoothed by Fourier convolution smoothing (*17*). To estimate the robustness of each peak in this smoothed plot, the data are subjected to nine additional rounds of smoothing; the more robust a peak, the more rounds of smoothing it withstands. Peaks are thus assigned a confidence value of 1 to 9, with 9 representing the most robust peaks, surviving through ≥9 additional rounds of smoothing. Of the 332 putative origins detected, 55% (184 out of 332) had a confidence value of ≥9. All putative origins tested by 2D gel analysis fell in this class. Finally, based on the linear correlation between %HL$_{(total)}$ and t_{rep} values, we can transform the smoothed %HL$_{(total)}$ plots into replication timing profiles.

17. Supplementary Web information is available on *Science* Online at www.sciencemag.org/cgi/content/full/294/5540/115/DC1.

18. K. L. Friedman, B. J. Brewer, W. L. Fangman, *Genes Cells* **2**, 667 (1997).

19. M. Yamashita *et al.*, *Genes Cells* **2**, 655 (1997).

20. Sixteen ARSs have previously been shown to be efficient origins in yeast chromosomes. In addition, SH1022 and SH1023 (Table 1) were identified in our lab during a random search for ARSs (*22*). We identified 22 other origins but excluded them from the comparison of predicted versus known origin loca-

tions, because the regions to which they were mapped were large (~10.5 kb on average).

21. Our analysis is based on the genome sequence in the *Saccharomyces* Genome Database; differences between that sequence and the actual sequence of our strain (such as the positions of Ty elements) could cause some inaccuracy in the assignment of origin locations. Also, one consequence of measuring microarray hybridization intensities over a 10-kb window is that differences in average fork rate on either side of an origin can influence the location of the peak representing that origin. In extreme cases, where an origin is unidirectional, the location of the resulting peak will be shifted by 5 kb in the direction in which the single fork moves. A separate project to map the locations of prereplicative complexes [which would mark potential origins (*41*)] has been under way in parallel with our effort (*42*). That approach, which would identify potential origin locations at relatively high resolution, would be highly complementary to our timing-based approach.

22. S. Hunt, M. Raghuraman, B. Brewer, W. Fangman, unpublished data.

23. An S-phase pooling strategy was used, paralleling the array data-processing method. CsCl density gradient fractions containing HH DNA from all S phase samples were pooled; HL fractions from the same timed samples were likewise combined. Each pooled DNA sample was labeled by random primed synthesis with [^{32}P]-labeled deoxyadenosine triphosphate and excess deoxynucleoside triphosphates, and then separately hybridized to slot blots of 50 to 100 ng of various polymerase chain reaction–amplified ORF DNA fragments. From the hybridization intensities, the percent of labeled DNA that was HL was calculated: The greater the %HL for a fragment, the earlier the replication time. The density transfer experiment that was used for Fig. 4B was different from the one used to probe the microarrays.

24. R. Hand, *J. Cell. Biol.* **64**, 89 (1975).

25. P. M. Sharp, A. T. Lloyd, *Nucleic Acids Res.* **21**, 179 (1993).

26. C. Seoighe, K. H. Wolfe, *Gene* **238**, 253 (1999).

27. K. H. Wolfe, D. C. Shields, *Nature* **387**, 708 (1997).

28. R. M. McCarroll, W. L. Fangman, *Cell* **54**, 505 (1988).

29. R. J. Wellinger, A. J. Wolf, V. A. Zakian, *Cell* **72**, 51 (1993).

30. B. M. Ferguson, B. J. Brewer, A. E. Reynolds, W. L. Fangman, *Cell* **65**, 507 (1991).

31. K. Kolor, B. Brewer, W. Fangman, unpublished data.

32. This analysis disregards the possible occurrence of inefficient origins or of origins that set up unidirectional forks, which would result in fork rate overestimates or underestimates, respectively. Therefore, the values reported here should be considered as rough estimates of the fork rates.

33. A. E. Reynolds, R. M. McCarroll, C. S. Newlon, W. L. Fangman, *Mol. Cell. Biol.* **9**, 4488 (1989).

34. K. S. Hatton *et al.*, *Mol. Cell. Biol.* **8**, 2149 (1988).

35. A. P. Wolffe, *Dev. Genet.* **15**, 463 (1994).

36. R. S. Hansen, T. K. Canfield, A. D. Fjeld, S. M. Gartler, *Hum. Mol. Genet.* **5**, 1345 (1996).

37. The directions of transcription and replication were compared for the 137 ribosomal protein genes, which account for ~50% of the transcription by RNA polymerase II (*43*); no correlation was found. As with the other gene classes tested, the distribution of ribosomal protein gene replication times was not significantly different from that of the genome average. A similar result was obtained for small nucleolar RNA genes (transcribed by RNA polymerase II and III) and for tRNA genes (transcribed by RNA polymerase III). The low transcript levels of the loci of Sir protein–mediated transcriptional silencing (*44*) did not correlate with late replication of these loci; the origins closest to the promoters of genes affected by *sir* mutations showed no difference from the genome average activation time (P = 0.19, 0.39, and 0.70, for *SIR2*, *SIR3*, and *SIR4*, respectively).

38. A. Lengronne, P. Pasero, A. Bensimon, E. Schwob, *Nucleic Acids Res.* **29**, 1433 (2001).

39. A. B. Khodursky *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9419 (2000).

40. E. A. Winzeler *et al.*, *Science* **281**, 1194 (1998).

41. W. M. Toone, B. L. Aerne, B. A. Morgan, L. H. Johnston, *Annu. Rev. Microbiol.* **51**, 125 (1997).

42. O. Aparicio, S. Bell, personal communication.

43. J. R. Warner, *Trends Biochem. Sci.* **24**, 437 (1999).

44. J. J. Wyrick *et al.*, *Nature* **402**, 418 (1999).

45. D. A. Natale, A. E. Schubert, D. Kowalski, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2654 (1992).

46. Electrophoresis in the first dimension was in a 0.3% (w/v) agarose gel in 1× tris-borate EDTA run at 0.67 V/cm for 40 hours, and electrophoresis in the second dimension was in a 0.6% agarose gel run at 1 V/cm for 60 or 72 hours. The DNA fragments tested were as follows: a, a 12.5-kb Nsi I fragment; b, a 13.5-kb Fsp I fragment; d, a 6.5-kb Pst I fragment; e, a 4.8-kb Hpa I fragment; and f, a 10.7-kb Msc I fragment.

47. Base composition was computed for a sliding window, with the use of the program Isachore for Apple Macintosh. The ease of DNA unwinding was calculated with VooDUE, a Macintosh version of the Thermodyn program (*45*). The free energy difference (ΔG) between duplex and single-stranded DNA states was calculated for various window sizes; to look for statistically significant variations in ΔG of unwinding, the mean ΔG for windows over a local region was compared to the mean of all windows in the chromosome or the genome as a whole.

48. We thank D. Richards, D. Kowalski, C. Newlon, Y. Zhou, and R. Abagayan for helpful advice; O. Aparicio for sharing results before publication; and G. Alvino and O. Aparicio for helpful discussions. Partially supported by grants from NIH.

12 July 2001; accepted 11 September 2001