D682–D686 *Nucleic Acids Research, 2012, Vol. 40, Database issue* *Published online 24 November 2011*
doi:10.1093/nar/gkr1091

OriDB, the DNA replication origin database updated and extended

Cheuk C. Siow, Sian R. Nieduszynska, Carolin A. Müller and Conrad A. Nieduszynski*

Centre for Genetics and Genomics, The University of Nottingham, Queen's Medical Centre, Nottingham, NG7 2UH, UK

Received October 3, 2011; Revised October 28, 2011; Accepted November 1, 2011

ABSTRACT

OriDB (http://www.oridb.org/) is a database containing collated genome-wide mapping studies of confirmed and predicted replication origin sites. The original database collated and curated *Saccharomyces cerevisiae* origin mapping studies. Here, we report that the OriDB database and web site have been revamped to improve user accessibility to curated data sets, to greatly increase the number of curated origin mapping studies, and to include the collation of replication origin sites in the fission yeast *Schizosaccharomyces pombe*. The revised database structure underlies these improvements and will facilitate further expansion in the future. The updated OriDB for *S. cerevisiae* is available at http://cerevisiae.oridb.org/ and for *S. pombe* at http://pombe.oridb.org/.

INTRODUCTION

Complete, accurate replication of the genome is crucial for life. Chromosomes must be precisely copied exactly once, a process that takes place during S phase. To complete DNA replication within S phase, replication of eukaryotic genomes is initiated at multiple discrete chromosomal sites called replication origins. Appropriate distribution of the origin sites is important to ensure that every sequence is replicated. However, not every origin site is used in every cell cycle; that is replication origins differ in their efficiency. Furthermore, origins activate at characteristic times during S phase, with some origins activating early in S phase and others later.

Replication origins are best characterized in the budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe*. In both organisms, origin sequences have been isolated through their ability to support plasmid replication (called Autonomously Replicating Sequences or ARS) (1,2). Chromosomal origin activity has been assayed using two-dimensional (2D) gel electrophoresis to detect replication intermediates in both *S. cerevisiae* and *S. pombe* (3,4). *Saccharomyces cerevisiae* origins contain an essential sequence element called the ARS consensus sequence (ACS) (5). In contrast, *S. pombe* origins feature AT-rich sequences, but no specific sequence motif (6). Origin sites in both yeasts are bound by the Origin Recognition Complex (ORC), which in turn recruits Cdc6 and Cdt1 to load Mcm2-7 double hexamers and form a pre-replication complex (pre-RC). Assembly of the pre-RC 'licenses' the origin for activation in the subsequent S phase.

Saccharomyces cerevisiae ORC binds to the ACS, however the ACS alone is not sufficient for origin function. Indeed, there are approximately 12 000 matches to the ACS in the genome, but only approximately 500 of these are functional replication origins. Consequently, there must be additional mechanisms to specify replication origin sites. These are thought to include transcription that ablates origin function (7,8), chromatin structure that can aid ORC recruitment (9,10) and secondary sequence motifs (11,12). The *S. pombe* Orc4 protein contains AT-hook domains that recognize and bind AT-rich origin sequences (13). The high AT content of *S. pombe* replication origins has allowed their identification, genome wide, as AT-rich islands (14).

Genome-wide approaches to identify and characterize replication origin locations rely on detecting either the origin-associated proteins or the DNA synthesis at active origin sites. Chromatin-immunoprecipitation (ChIP) of ORC and/or MCM proteins have been used to isolate origin sites (15–17). In *S. cerevisiae*, this has been combined with motif searches or phylogenetic footprinting to predict the location of the ACS (5,9,18). Active replication origins have been identified as local points of the earliest replicating sequence in genome-wide measures of when each sequence in the yeast genome replicates (19–22). Origin sites have also been identified as sites of BrdU incorporation or accumulation of single-stranded DNA when cells are challenged with hydroxyurea (16,23,24).

*To whom correspondence should be addressed. Tel: +44 115 823 0352; Fax: +44 115 823 0338; Email: conrad.nieduszynski@nottingham.ac.uk

Previously, we collated the proposed location of *S. cerevisiae* origin sites from the available genome-wide mapping studies and presented the results in a web-accessible database, OriDB (25). This collated data set has facilitated comparisons with a range of other chromosomal features including transcription (26), genomic rearrangements (27) and fragile sites (28,29). Furthermore, the comprehensive origin data sets and the underlying data have permitted mathematical approaches to investigate genome replication (30–32). Now we present a major update to OriDB. We have completely restructured the underlying database tables to enable the incorporation of many additional data sets, improvements in user access to the raw data and expansion to a second model system, *S. pombe*.

RESULTS

Revised database structure

The original release of OriDB implemented a simple, but limited table structure. The majority of data was stored in a single table. This has made updating the database time consuming and has risked the introduction of errors. The rapid growth in replication origin studies necessitated a complete restructuring of the underlying database.

We have replaced the original OriDB database tables with a large number of non-redundant tables with defined relationships (Figure 1). Four primary tables define the database and the relationship between all the tables. First, the table 'sc_ori' contains the list of confirmed, likely and dubious origin sites collated from published studies as described previously (25). Second, the table 'sc_ori_studies' lists the studies that have published lists of origin locations. Third, the table 'sc_repl_data' lists the studies for which OriDB has stored the experimental data from which origin predictions have been made. Fourth, the table 'sc_elements_studies' lists the studies that have proposed origin sequence elements. Each of these primary tables defines the relationship with further tables that store the data from each of these studies. These tables of published data are from genome-wide studies and are supplemented with additional tables that have been collated by OriDB from the literature: origin sites confirmed by 2D gel electrophoresis (sc_2D_gel), origin sites confirmed by ARS assay (cloned_ori) and confirmed origin sequence elements (sc_confirmed_ACS).

All collated data sets and chromosomal coordinates are presented relative to the 1 October 2003 release of the *S. cerevisiae* genome (referred to as sacCer1 at the UCSC genome browser) (33,34). To convert between the various sequence releases, we have used the liftOver tool from UCSC (35) with custom generated parameter (over.chain) files. Members of the yeast community can use this tool through a web interface at: http://www.nieduszynski.org/liftover/.

The restructuring of the OriDB database tables necessitated a complete re-writing of the web pages. The resulting changes offer a number of significant benefits for

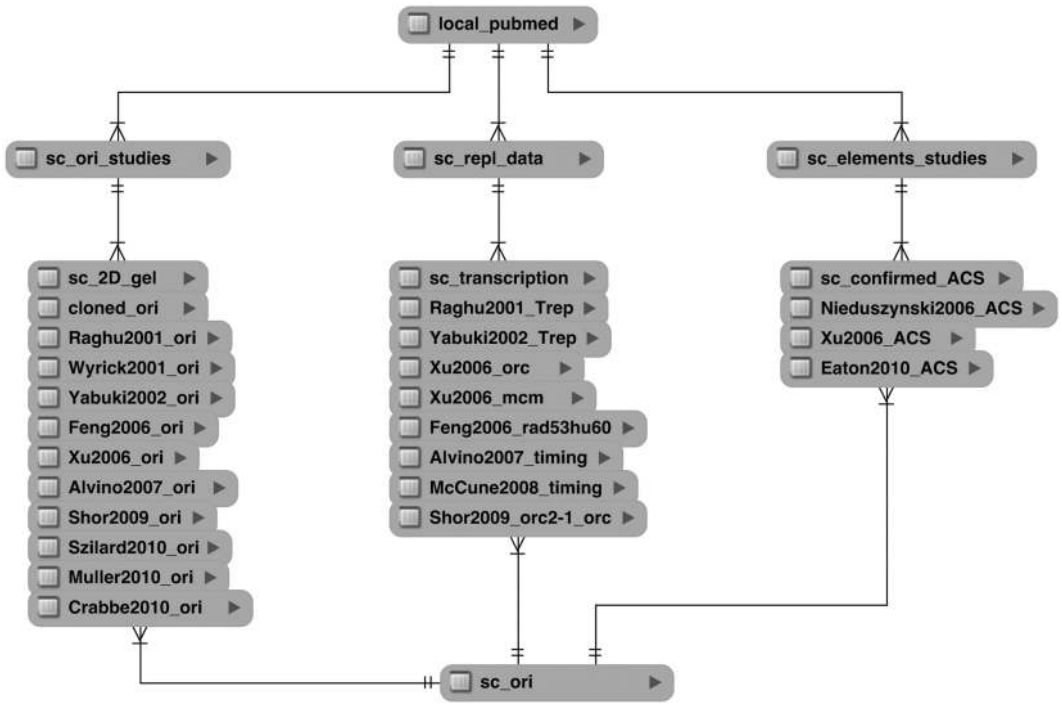

Figure 1. Four primary tables define the database structure for *S. cerevisiae* OriDB. (Left-hand side) The table 'sc_ori_studies' describes the curated studies that have reported replication origin sites, each of which is represented by a further table. (Middle) The table 'sc_repl_data' describes additional tables that contain the experimental data from origin mapping studies. (Right-hand side) The table 'sc_elements_studies' describes the curated studies that reported sequence elements at replication origins; each of these studies is represented by a table. (Bottom) The table 'sc_ori' contains the collated list of all reported replication origin sites from those studies listed in 'sc_ori_studies'. Finally, each table is linked to the appropriate PubMed record in a locally stored table ('local_pubmed') that retrieves data directly from PubMed.

users. The origin details pages now load all tabs concurrently, but only display the user-selected tab; this allows rapid switching between tabs. Furthermore, the new data structures allow improved user access to the underlying data, making it straightforward to include many additional origin mapping data sets and allow for the expansion of OriDB to include the fission yeast *S. pombe*. These pages are available at http://cerevisiae.oridb.org/ and http://pombe.oridb.org/ (with backup sites available at http://www.nottingham.ac.uk/plzcnlab/oridb/cerevisiae/ and at http://www.nottingham.ac.uk/plzcnlab/oridb/pombe/).

Improved user access to data

The most frequent user request is to retrieve data from an OriDB curated study in a user-specified format. The new database structure, described above, allows us to implement a straightforward yet powerful route to the underlying data. A 'download' link present in the top bar of every OriDB page allows access to all the datasets curated at OriDB, including those tables curated from the literature. Data sets are grouped first by the data type (e.g. predictions of origin location) and then by the original study. Access to the underlying data is also available from links on the pages that summarize the findings of individual studies. The user has a choice of appropriate formats for downloading the data (including the raw data in a tab or comma separated format, BED or WIG formats for display in genome browsers, and FASTA for sequence download). These pages and links are generated from the underlying database tables and therefore will automatically update to include new studies, as they are included in OriDB.

Expanded data coverage

The original OriDB database collated four genome-wide (microarray) data sets (15,19,20,23) and our phylogenetic footprinting of origin sequence elements (5). The availability of high-resolution microarrays (18) and more recently, deep-sequencing technologies (9) have led to a large increase in the number of studies proposing origin locations. The new database structure has allowed us to integrate many additional data sets, so that at the time of writing, *S. cerevisiae* OriDB includes 10 genome-wide data sets and has the capability to include an effectively unlimited number in the future. The data from these studies are presented to the user through the details page for each origin. As in the previous version of OriDB, the details page includes an 'Origin Location Assignments' tab which lists all the studies that identify the particular origin [as described previously this is based upon the proposed resolution of the study in question (25)]. The 'Origin Location Assignments' tab also has the capability to display additional information from each study for each origin location. For example, a recent study mapped the activity of origins in different mutant cells subjected to the drug hydroxyurea (24); OriDB includes the details of which mutants the origin was reported to be active in.

Collation of *S. pombe* replication origin sites

The mapping of replication origin sites in *S. pombe* has drawn on a similar range of experimental techniques as used in *S. cerevisiae*, including ARS assays and 2D gels. Although *S. pombe* replication origins do not contain a discrete sequence motif for ORC recruitment, the replication origins have a characteristic AT composition, called AT islands. The computational identification of AT islands allowed the accurate predication of replication origin sites throughout the *S. pombe* genome (14). More recently, genome-wide studies have employed microarray technologies to identify origins based upon the location of pre-RC proteins, newly synthesized DNA (16), the increase in DNA copy number as a sequence replicates (21) or the single-stranded DNA that accumulates at stalled replication forks (23). Each of these studies produced a genome-wide list of replication origin sites. To facilitate access to these data sets and allow comparison between them, we generated a single collated list of replication origin sites presented through a web-accessible database, which includes text and graphic representations of the data (Figure 2). The independent studies that identified *S. pombe* replication origins have used a range of naming conventions that have resulted in different names being assigned to the same origin. To consolidate replication origin naming in *S. pombe*, we have assigned each *S. pombe* replication origin site a systematic name based upon the chromosome number (in roman numerals) and the chromosomal coordinate. Hence the origin on chromosome 1 at 3060 kb is named *ori-I-3060* [other names for this origin are *ars1119* (16), *ori1095* (21), AT1098 (14) and *ars766* (1)]. Our collated *S. pombe* replication origin data is presented relative to the current genome sequence, downloaded on 1 October 2011 (36). The *S. pombe* replication origin database can be accessed at http://pombe.oridb.org/.

DISCUSSION

In the era of high-throughput genome-wide data generation, it is essential that the scientific community can access the data and the conclusions drawn from these data. For replication origin mapping studies, this means access to microarray (or deep sequencing) data and the inferred origin locations. OriDB aims to provide access to exactly these data types, presenting them through a user-friendly interface. In this update, we improve user access to the underlying data (now available for download), extend the number of studies collated and for the first time collate origin sites from *S. pombe*.

ACKNOWLEDGEMENTS

OriDB has been built on a large body of work only a fraction of which we have been able to mention here. We apologize to colleagues whose work has not been cited and remind readers that OriDB contains a more extensive and expanding list of cross-referenced citations. We thank members of the Nieduszynski laboratory for helpful discussions and authors from the curated studies

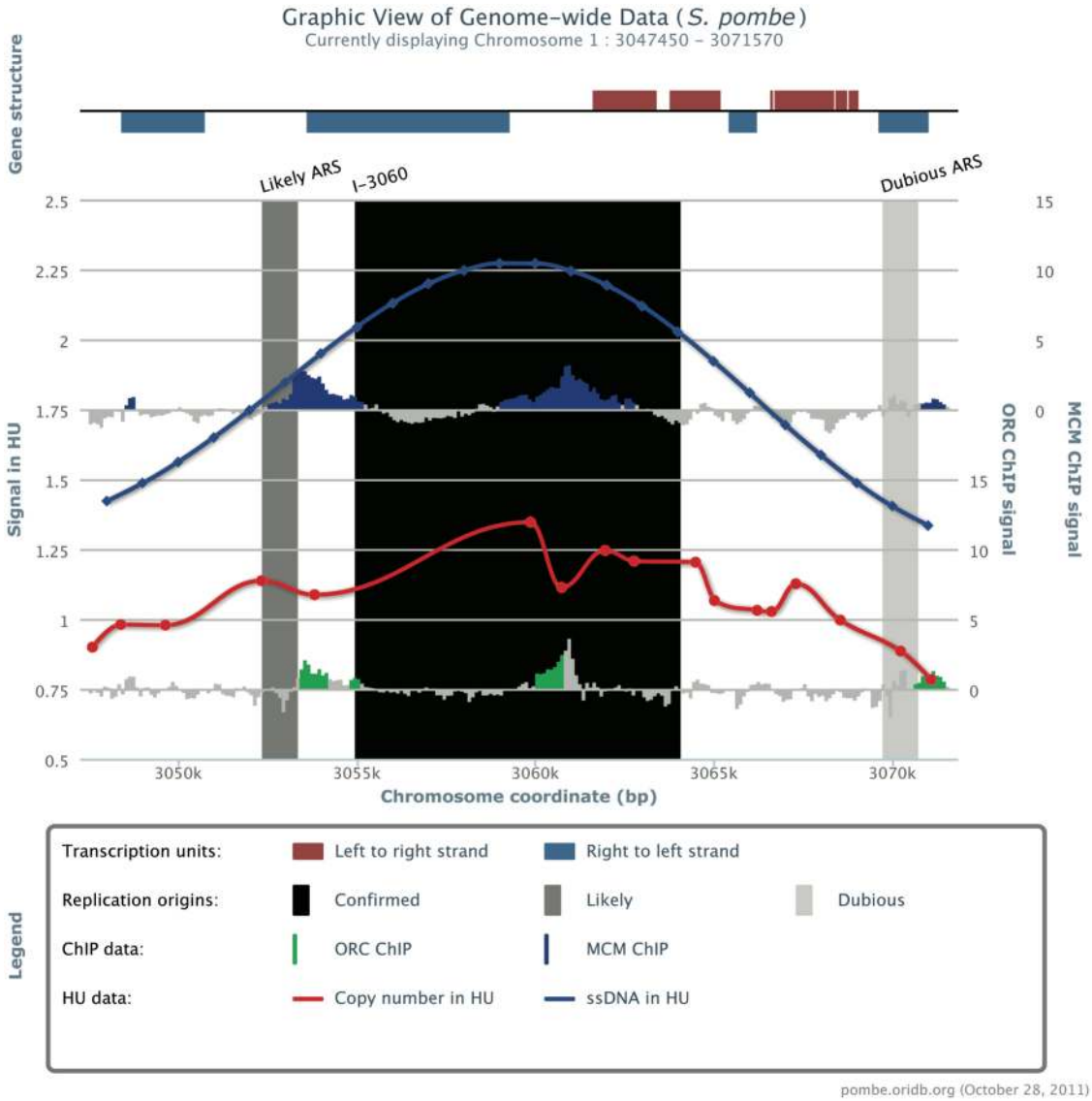

Figure 2. Screen shot from *S. pombe* OriDB showing the Origin Summary Graphic tab for *ori-I-3060*. A window of the *S. pombe* genome is shown centred upon the origin of interest. (Top) the gene structure is shown ('mouse over' displays the name of the each gene). (Main plot) Vertical bars show the replication origin sites (black for confirmed; dark grey for likely; light grey for dubious). Blue and green bars illustrate the location of signal from ChIP of Mcm6 and Orc1, respectively (16). The red curve gives the increase in DNA content during DNA replication in the presence of hydroxyurea (21). The blue curve shows the accumulation of single-stranded DNA during DNA replication in Δ*cds1* cells exposed to hydroxyurea (23).

for making data available. Particular thanks are due to Prof. Joel Huberman for help and advice on *S. pombe* replication origin sites.

FUNDING

The Royal Society, The University of Nottingham and the Biotechnology and Biological Sciences Research Council (grant numbers BB/E023754/1, BB/G001596/1); David Phillips Fellowship (to C.A.N.). Funding for open access charge: Biotechnology and Biological Sciences Research Council.

Conflict of interest statement. None declared.

REFERENCES

1. Maundrell,K., Hutchison,A. and Shall,S. (1988) Sequence analysis of ARS elements in fission yeast. *EMBO J.*, **7**, 2203–2209.
2. Stinchcomb,D.T., Struhl,K. and Davis,R.W. (1979) Isolation and characterisation of a yeast chromosomal replicator. *Nature*, **282**, 39–43.
3. Zhu,J., Brun,C., Kurooka,H., Yanagida,M. and Huberman,J.A. (1992) Identification and characterization of a complex chromosomal replication origin in Schizosaccharomyces pombe. *Chromosoma*, **102**, S7–S16.
4. Huberman,J.A., Zhu,J.G., Davis,L.R. and Newlon,C.S. (1988) Close association of a DNA replication origin and an ARS element on chromosome III of the yeast, Saccharomyces cerevisiae. *Nucleic Acids Res.*, **16**, 6373–6384.
5. Nieduszynski,C.A., Knox,Y. and Donaldson,A.D. (2006) Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev.*, **20**, 1874–1879.

6. Clyne,R.K. and Kelly,T.J. (1995) Genetic analysis of an ARS element from the fission yeast Schizosaccharomyces pombe. *EMBO J.*, **14**, 6348–6357.
7. Snyder,M., Sapolsky,R.J. and Davis,R.W. (1988) Transcription interferes with elements important for chromosome maintenance in Saccharomyces cerevisiae. *Mol. Cell Biol.*, **8**, 2184–2194.
8. Nieduszynski,C.A., Blow,J.J. and Donaldson,A.D. (2005) The requirement of yeast replication origins for pre-replication complex proteins is modulated by transcription. *Nucleic Acids Res.*, **33**, 2410–2420.
9. Eaton,M.L., Galani,K., Kang,S., Bell,S.P. and MacAlpine,D.M. (2010) Conserved nucleosome positioning defines replication origins. *Genes Dev.*, **24**, 748–753.
10. Muller,P., Park,S., Shor,E., Huebert,D.J., Warren,C.L., Ansari,A.Z., Weinreich,M., Eaton,M.L., MacAlpine,D.M. and Fox,C.A. (2010) The conserved bromo-adjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin. *Genes Dev.*, **24**, 1418–1433.
11. Chang,F., Theis,J.F., Miller,J., Nieduszynski,C.A., Newlon,C.S. and Weinreich,M. (2008) Analysis of chromosome III replicators reveals an unusual structure for the ARS318 silencer origin and a conserved WTW sequence within the origin recognition complex binding site. *Mol. Cell Biol.*, **28**, 5071–5081.
12. Chang,F., May,C.D., Hoggard,T., Miller,J., Fox,C.A. and Weinreich,M. (2011) High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. *Nucleic Acids Res.*, **39**, 6523–6535.
13. Kong,D. and DePamphilis,M.L. (2001) Site-specific DNA binding of the Schizosaccharomyces pombe origin recognition complex is determined by the Orc4 subunit. *Mol. Cell Biol.*, **21**, 8095–8103.
14. Segurado,M., de Luis,A. and Antequera,F. (2003) Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe. *EMBO Rep.*, **4**, 1048–1053.
15. Wyrick,J.J., Aparicio,J.G., Chen,T., Barnett,J.D., Jennings,E.G., Young,R.A., Bell,S.P. and Aparicio,O.M. (2001) Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. *Science*, **294**, 2357–2360.
16. Hayashi,M., Katou,Y., Itoh,T., Tazumi,A., Yamada,Y., Takahashi,T., Nakagawa,T., Shirahige,K. and Masukata,H. (2007) Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *EMBO J.*, **26**, 1327–1339.
17. Shor,E., Warren,C.L., Tietjen,J., Hou,Z., Muller,U., Alborelli,I., Gohard,F.H., Yemm,A.I., Borisov,L., Broach,J.R. *et al.* (2009) The origin recognition complex interacts with a subset of metabolic genes tightly linked to origins of replication. *PLoS Genet.*, **5**, e1000755.
18. Xu,W., Aparicio,J.G., Aparicio,O.M. and Tavare,S. (2006) Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. *BMC Genomics*, **7**, 276.
19. Raghuraman,M.K., Winzeler,E.A., Collingwood,D., Hunt,S., Wodicka,L., Conway,A., Lockhart,D.J., Davis,R.W., Brewer,B.J. and Fangman,W.L. (2001) Replication dynamics of the yeast genome. *Science*, **294**, 115–121.
20. Yabuki,N., Terashima,H. and Kitada,K. (2002) Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells*, **7**, 781–789.
21. Heichinger,C., Penkett,C.J., Bahler,J. and Nurse,P. (2006) Genome-wide characterization of fission yeast DNA replication origins. *EMBO J.*, **25**, 5171–5179.
22. Alvino,G.M., Collingwood,D., Murphy,J.M., Delrow,J., Brewer,B.J. and Raghuraman,M.K. (2007) Replication in Hydroxyurea: It's a Matter of Time. *Mol. Cell Biol.*, **27**, 6396–6406.
23. Feng,W., Collingwood,D., Boeck,M.E., Fox,L.A., Alvino,G.M., Fangman,W.L., Raghuraman,M.K. and Brewer,B.J. (2006) Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat. Cell Biol.*, **8**, 148–155.
24. Crabbe,L., Thomas,A., Pantesco,V., De Vos,J., Pasero,P. and Lengronne,A. (2010) Analysis of replication profiles reveals key role of RFC-Ctf18 in yeast replication stress response. *Nature Struct. Mol. Biol.*, **17**, 1391–1397.
25. Nieduszynski,C.A., Hiraga,S., Ak,P., Benham,C.J. and Donaldson,A.D. (2007) OriDB: a DNA replication origin database. *Nucleic Acids Res.*, **35**, D40–D46.
26. Omberg,L., Meyerson,J.R., Kobayashi,K., Drury,L.S., Diffley,J.F. and Alter,O. (2009) Global effects of DNA replication and DNA replication origin activity on eukaryotic gene expression. *Mol. Syst. Biol.*, **5**, 312.
27. Gordon,J.L., Byrne,K.P. and Wolfe,K.H. (2009) Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern Saccharomyces cerevisiae genome. *PLoS Genet.*, **5**, e1000485.
28. Di Rienzi,S.C., Collingwood,D., Raghuraman,M.K. and Brewer,B.J. (2009) Fragile genomic sites are associated with origins of replication. *Genome Biol. Evol.*, **1**, 350–363.
29. Szilard,R.K., Jacques,P.E., Laramee,L., Cheng,B., Galicia,S., Bataille,A.R., Yeung,M., Mendez,M., Bergeron,M., Robert,F. *et al.* (2010) Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX. *Nat. Struct. Mol. Biol.*, **17**, 299–305.
30. de Moura,A.P., Retkute,R., Hawkins,M. and Nieduszynski,C.A. (2010) Mathematical modelling of whole chromosome replication. *Nucleic Acids Res.*, **38**, 5623–5633.
31. Sekedat,M.D., Fenyo,D., Rogers,R.S., Tackett,A.J., Aitchison,J.D. and Chait,B.T. (2010) GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol. Syst. Biol.*, **6**, 353.
32. Yang,S.C., Rhind,N. and Bechhoefer,J. (2010) Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol. Syst. Biol.*, **6**, 404.
33. Engel,S.R., Balakrishnan,R., Binkley,G., Christie,K.R., Costanzo,M.C., Dwight,S.S., Fisk,D.G., Hirschman,J.E., Hitz,B.C., Hong,E.L. *et al.* (2010) Saccharomyces Genome Database provides mutant phenotype data. *Nucleic Acids Res.*, **38**, D433–436.
34. Fujita,P.A., Rhead,B., Zweig,A.S., Hinrichs,A.S., Karolchik,D., Cline,M.S., Goldman,M., Barber,G.P., Clawson,H., Coelho,A. *et al.* (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.*, **39**, D876–882.
35. Hinrichs,A.S., Karolchik,D., Baertsch,R., Barber,G.P., Bejerano,G., Clawson,H., Diekhans,M., Furey,T.S., Harte,R.A., Hsu,F. *et al.* (2006) The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.*, **34**, D590–598.
36. Wood,V., Gwilliam,R., Rajandream,M.A., Lyne,M., Lyne,R., Stewart,A., Sgouros,J., Peat,N., Hayles,J., Baker,S. *et al.* (2002) The genome sequence of Schizosaccharomyces pombe. *Nature*, **415**, 871–880.