PomBase 2015: updates to the fission yeast database

Mark D. McDowall[1,*], Midori A. Harris[2], Antonia Lock[3], Kim Rutherford[2], Daniel M. Staines[1], Jürg Bähler[3], Paul J. Kersey[1], Stephen G. Oliver[2,*] and Valerie Wood[2,*]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK, [2]Cambridge Systems Biology and Department of Biochemistry, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge, Cambridgeshire CB2 1GA, UK and [3]Research Department of Genetics, Evolution and Environment, and UCL Cancer Institute, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK

ABSTRACT

PomBase (http://www.pombase.org) is the model organism database for the fission yeast *Schizosaccharomyces pombe*. PomBase provides a central hub for the fission yeast community, supporting both exploratory and hypothesis-driven research. It provides users easy access to data ranging from the sequence level, to molecular and phenotypic annotations, through to the display of genome-wide high-throughput studies. Recent improvements to the site extend annotation specificity, improve usability and allow for monthly data updates. Both in-house curators and community researchers provide manually curated data to PomBase. The genome browser provides access to published high-throughput data sets and the genomes of three additional *Schizosaccharomyces* species (*Schizosaccharomyces cryophilus*, *Schizosaccharomyces japonicus* and *Schizosaccharomyces octosporus*).

INTRODUCTION

The fission yeast *Schizosaccharomyces pombe* is a unicellular eukaryote that has been used as a model organism for studying a diverse array of biological processes, from the cell cycle to signaling, for over 60 years (1). It was the sixth eukaryotic organism to have its genome completely sequenced (2). With a thriving community generating data from small and large-scale projects, a central hub to curate and integrate information is vital to facilitate data interpretation and hypothesis generation, and to guide further research.

PomBase (http://www.pombase.org) was launched in 2011 as the model organism database for fission yeast (3). The PomBase portal provides centralized access to gene- and genome-scale information, emphasizing data acquired by manual literature curation. In a novel community cura-

tion initiative, fission yeast researchers now contribute significantly to gene annotation, using the Canto online curation tool (4).

PomBase presents information in gene-specific pages that include summary data on each gene and its product, such as its biological functions, cellular localization, phenotype data, modifications, interactions, regulation and gene expression.

PomBase offers a customized Ensembl Genome browser (5) to provide access to the genome sequence and features, and to visualize high-throughput data sets in a genomic context.

BIOLOGICAL DATA

PomBase curators focus on extracting data from historical papers and on providing help and guidance to researchers who curate their own papers using Canto. The inclusion of genome-scale datasets has resulted in a large increase in the volume of data curated.

High-throughput datasets

PomBase gene pages typically include data from various types of large-scale experiments, such as gene expression data (6,7), phenotypic analysis (8,9) and interaction data (10). Within the genome browser, PomBase hosts sequence-based datasets from a variety of high-throughput experimental techniques, such as nucleosome positioning (11), transcriptomic data (see Figure 1A) (6,11–12), replication profiling (13), polyadenylation sites (14,15) and chromatin binding (16). The datasets included to date are those requested by the fission yeast community, and for which the publication authors have provided data to PomBase.

Additional species

In addition to *S. pombe*, the genomes of *Schizosaccharomyces cryophilus*, *Schizosaccharomyces japonicus* and

---

*To whom correspondence should be addressed. Tel: +44 1223 494589; Fax: +44 1223 494468; Email: mcdowall@ebi.ac.uk
Correspondence may also be addressed to Valerie Wood. Tel: +44 1223 746961; Fax: +44 1223 766002; Email: vw253@cam.ac.uk
Correspondence may also be addressed to Stephen G. Oliver. Tel: +44 1223 333667; Fax: +44 1223 766002; Email: sgo24@cam.ac.uk
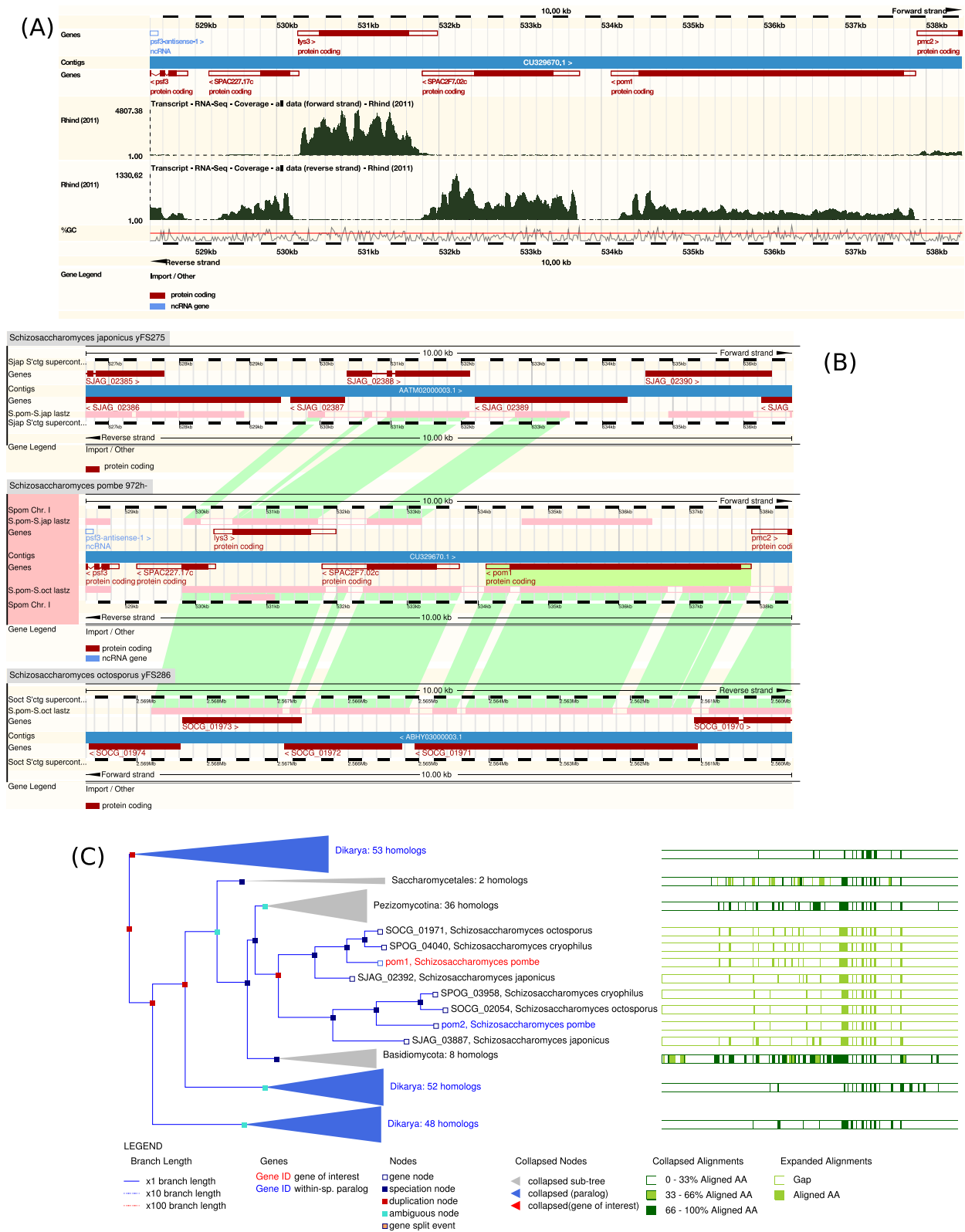
Figure 1. Views available in the genome browser. (**A**) Region display with two tracks enabled, displaying RNA-Seq coverage data ([12]). The two tracks show the forward (top) and reverse (bottom) strand reads along with the genes that are mapped to the region. (**B**) Region comparison display showing the regions of alignment between *Schizosaccharomyces pombe* and *Schizosaccharomyces japonicus* (top) and *Schizosaccharomyces octosporus* (bottom) in green. (**C**) Gene tree view generated using the Compara framework, with the gene of interest, *pom1* (SPAC2F7.03c), highlighted in red.

Schizosaccharomyces octosporus (12) are now accessible via the genome browser, as the result of collaboration with the Ensembl Genomes project (17). All-against-all DNA alignments between the four *Schizosaccharomyces* species can also be displayed in the PomBase browser (see Figure 1B). The new *Schizosaccharomyces* species are among the 52 fungal species included in a protein multiple sequence alignment (see Figure 1C). *S. pombe* also represents the fission yeast clade in protein multiple sequence alignments of a further set of 178 species covering a broad taxonomic range including human and bacteria, which are visible in the genome browser.

INFRASTRUCTURE AND PROCEDURAL IMPROVE- MENTS

Manually curated data is stored within a Chado relational database (18). During the release procedure, a snapshot of the curation database is created and the annotations are imported into an Ensembl schema on a MySQL database. The Ensembl schema provides the back-end architecture for the genome browser and houses the annotations for the PomBase site.

The import pipeline that transfers data from the PomBase Chado curation database to the Ensembl database has been altered to accommodate increasing annotation complexity (as described below). Update procedures have been improved to implement data consistency checks on the database and to use the Selenium testing framework (http://www.seleniumhq.org) on the web interface. This more robust infrastructure has enabled PomBase to implement a monthly release cycle. Improved back-end data storage and retrieval has reduced the gene page loading time, even as the amount of data presented has increased, enabling users to navigate through multiple genes with minimal delays.

INTEGRATING AND VISUALIZING DATA

To maintain a readable display of increasingly complex data, and incorporate new data types, there have been major improvements to the organization and presentation of data on the gene pages. The most extensive changes have affected three key regions of the gene page: the displays of Gene Ontology (GO) (19,20) annotations, Fission Yeast Phenotype Ontology (FYPO) annotations (21) and gene expression data. More subtle changes have also been introduced throughout the gene pages.

Ontology annotations and extensions

The most significant change affecting annotation complexity in PomBase is the introduction of 'annotation extensions' that increase the expressivity of annotations to ontology terms. With active participation by PomBase curators, the GO Consortium introduced annotation extensions in 2013 (22) to enable curators to capture additional contextual details such as effector–target relationships and temporal or spatial aspects of biological processes. Whereas previously each GO annotation combined a single gene product with a single GO term, and was independent from any other

GO annotations, extended annotations can capture interconnections between multiple annotations as well as links to additional ontologies.

Each annotation extension consists of a relation and an identifier that refers to another ontology term (GO, SO (23), ChEBI (24), PSI-MOD (25), etc.) or another gene. An annotation may have one or more extensions, each with its own relationships and sources, and 'compound' extensions can be made by combining single extensions.

To date, PomBase curators have added extensions to over 2000 GO annotations. PomBase has also adopted the annotation extension model for phenotype (FYPO) and gene expression annotations, as described below. Table 1 shows the total number of annotations in PomBase as of August 2014, and the number that have extensions, for four ontologies plus gene expression.

To accommodate annotation extensions, PomBase has adapted its Chado and Ensembl relational database schemata and loading procedures and enhanced the gene page ontology annotation displays. On PomBase gene pages, annotation extensions are shown in rows below the ontology term, with the relevant evidence code and annotation source. Identifiers and relation strings are converted to human-friendly text, such as a gene name or ontology term, wherever possible. For example, Figure 2A shows annotations to GO:0045944 from the *ste11* (SPBC32C12.02) gene page. Annotations without extensions are displayed first, followed by those with extensions, and the bottom row shows a compound annotation extension (26).

Phenotype annotations have also considerably increased both in number and complexity. FYPO annotations include allele details, expression level, experimental conditions, the evidence and source, and annotation extensions that represent penetrance and severity. Phenotype annotation extensions also capture specific genes used in assays for phenotypes such as protein localization or gene expression level. On each gene page, FYPO annotations are grouped by whether the phenotype is relevant at the level of a cell population or an individual cell, and display annotation extensions similarly to the GO tables. Figure 2B shows alleles of *cdc2* (SPBC11B10.09) that have been annotated to 'decreased protein binding' (FYPO:0001645), affecting several other gene products.

Targets

A new table on the gene pages, 'Target Of', reports effects of other genes on the gene of interest, such as modification or regulation. 'Target Of' annotations are the reciprocal of GO and FYPO annotation extensions. The 'Target Of' display includes the relevant gene, a relationship and the annotation source. For example, *cdc2* (SPBC11B10.09) is annotated as the substrate of *csk1* (SPAC1D4.06c) protein kinase activity (27) and *cdc25* (SPAC24H6.05) protein phosphatase activity (28).

Gene expression

Quantitative gene expression data have been imported from two large datasets covering the expression of 3175 (7) and 7016 (6) gene products. Gene expression annotations may

(A)

Term ID	Term Name	Evidence	With/From	Reference	Count
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	IDA		Pereira PS et al. (2001)	81
	regulates mfm2	IMP		Kjaerulff S et al. (1994)	
	regulates mfm3	IMP		Kjaerulff S et al. (1994)	
	regulates mfm1	IMP		Kjaerulff S et al. (1994)	
	during single-celled organism vegetative growth phase, **regulates** mat2-Pc	IMP		Sugimoto A et al. (1991)	

(B)

Term ID	Term Name	Evidence	Allele	Expression	Condition	Reference	Count
FYPO:0001645	decreased protein binding						35
	affecting cdc2 and suc1	Western blot assay	cdc2-33 (A177T)	Not specified		Booher RN et al. (1989)	
	affecting cdc2 and suc1	Western blot assay	cdc2-L7 (P208S)	Not specified		Booher RN et al. (1989)	
	affecting cdc2 and suc1	Western blot assay	cdc2-56 (G183E)	Not specified		Booher RN et al. (1989)	
	affecting cdc2 and cdc13	Western blot assay	cdc2-L7 (P208S)	Not specified		Booher RN et al. (1989)	
	affecting suc1 and cdc2	Western blot assay	cdc2-L7 (P208S)	Endogenous		Brizuela L et al. (1987)	

(C)

▾ Gene Expression

Contact Curators | Help

Qualitative Gene Expression

Description	Level	Evidence	Reference
protein level	fluctuates **during** mitotic cell cycle	mass spectrometry evidence	Carpy A et al. (2014)

Quantitative Gene Expression

Protein Level

Molecules/Cell (average)	During	Condition	Scale	Evidence	Reference
7703	**during** mitotic G1 phase	glucose MM, standard temperature	single_cell	mass spectrometry evidence	Carpy A et al. (2014)
2863	**during** mitotic S phase	glucose MM, standard temperature	single_cell	mass spectrometry evidence	Carpy A et al. (2014)
3924	**during** mitotic G2 phase	glucose MM, standard temperature	single_cell	mass spectrometry evidence	Carpy A et al. (2014)
3105	**during** mitotic M phase	glucose MM, standard temperature	single_cell	mass spectrometry evidence	Carpy A et al. (2014)
3202	**during** single-celled organism vegetative growth phase	glucose MM, standard temperature	single_cell	mass spectrometry evidence	Carpy A et al. (2014)
3058.34	**during** single-celled organism vegetative growth phase	standard temperature, glucose MM	population_wide	experimental evidence	Marguerat S et al. (2012)
817.95	**during** cell quiescence following G1 arrest due to nitrogen limitation	standard temperature, glucose MM, nitrogen absent	population_wide	experimental evidence	Marguerat S et al. (2012)

RNA Level

Molecules/Cell (average)	During	Condition	Scale	Evidence	Reference
2.3	**during** single-celled organism vegetative growth phase	standard temperature, glucose MM	population_wide	experimental evidence	Marguerat S et al. (2012)
0.81	**during** cell quiescence following G1 arrest due to nitrogen limitation	standard temperature, glucose MM, nitrogen absent	population_wide	experimental evidence	Marguerat S et al. (2012)

Figure 2. PomBase gene page views for example annotations: (**A**) GO (gene *ste11* - SPBC32C12.02); (**B**) FYPO (gene *cdc2* - SPBC11B10.09); (**C**) Gene expression (gene *clr3* - SPBC800.03). These examples highlight the display of annotation extensions and their use within the context of different ontologies.

include extensions indicating that the expression level was measured in a specific phase of the cell cycle or under specific growth conditions. Further qualitative data have also been manually imported into PomBase from the literature. When available, this information is displayed on the gene pages, providing details of the experimental conditions, evidence, scale and source. Figure 2C shows the display of gene expression data for *clr3* (SPBC800.03).

Data visualization in genome browser

In the PomBase genome browser, data tracks are presented with curated metadata, links to the relevant publication via the Europe PMC portal (29) and, where appropriate, links to external source databases. Users have the option of viewing their own data privately within the context of the genome browser, or submitting data to be hosted by PomBase for public viewing.

Table 1. Summary of annotations and extensions in PomBase as of September 2014

Curated data type	Annotation count	Annotation extensions	Gene coverage
Gene expression	40 403	40 403	7017
Phenotype (FYPO)	36 382	11 722	4942
Gene Ontolgy (GO)	37 224	2149	5301
Modifications (MOD)	11 265	7255	2009
Protein sequence Features (SO)	943	N/A	764

Annotation Count: total number of annotations of each type, including those with extensions; Annotation extensions: number of annotations that have one or more extensions apiece; Gene coverage: number of genes that have at least one annotation of the given type.

OTHER IMPROVEMENTS

PomBase now offers a motif finder that can retrieve lists of genes that match a particular protein sequence pattern. In the PomBase advanced search, the interfaces for constructing custom queries and retrieving results have been enhanced.

User-experience testing conducted after the initial Pom-Base release identified several opportunities for usability improvements. Accordingly, changes to the navigation and organization of the gene pages, such as collapsible intra-page menus, now make data more intuitively visible. Interfaces requiring user interaction are now also more intuitive.

OUTREACH AND USAGE

PomBase includes documentation for all gene page sections, links to Ensembl documentation for the genome browser and a Frequently Asked Questions section. Various web forms offer convenient links for users to contact curators to ask questions or submit high-throughput datasets to be included in the genome browser or on the gene pages. Pom-Base curators invite all authors of new fission yeast publications to curate their own papers using Canto. Pom-Base also sends announcements and help to a dedicated mailing list and to various social media outlets including Twitter (@PomBase), LinkedIn (http://www.linkedin.com/company/pombase) and Google+ (+PombaseOrg).

FUTURE DIRECTIONS

Canto and the gene pages will be extended to support the curation and display, respectively, of multiple-gene phenotypes (double mutants, triple mutants, etc.). Work has also begun to create pages for non-gene sequence features, such as the centromeres, which at present can only be viewed in the genome browser.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Egel,R. (2000) Fission yeast on the brink of meiosis. *Bioessays*, **22**, 854–860.
2. Wood,V., Gwilliam,R., Rajandream,M., 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.
3. Wood,V., Harris,M.A., McDowall,M.D., Rutherford,K., Vaughan,B.W., Staines,D.M., Aslett,M., Lock,A., Bähler,J., Kersey,P.J. *et al.* (2012) PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res.*, **40**, D695–D699.
4. Rutherford,K.M., Harris,M.A., Lock,A., Oliver,S.G. and Wood,V. (2014) Canto: an online tool for community literature curation. *Bioinformatics*, **30**, 1791–1792.
5. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Billis,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fitzgerald,S. *et al.* (2014) Ensembl 2014. *Nucleic Acids Res.*, **42**, D749–D755.
6. Marguerat,S., Schmidt,A., Codlin,S., Chen,W., Aebersold,R. and Bähler,J. (2012) Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell*, **151**, 671–683.
7. Carpy,A., Krug,K., Graf,S., Koch,A., Popic,S., Hauf,S. and Macek,B. (2014) Absolute proteome and phosphoproteome dynamics during the cell cycle of *Schizosaccharomyces pombe* (Fission Yeast). *Mol. Cell. Proteomics*, **13**, 1925–1936.
8. Hayles,J., Wood,V., Jeffery,L., Hoe,K.-L., Kim,D.-U., Park,H.-O., Salas-Pino,S., Heichinger,C. and Nurse,P. (2013) A genome-wide resource of cell cycle and cell shape genes of fission yeast. *Open Biol.*, **3**, 130053.
9. Sun,L.-L., Li,M., Suo,F., Liu,X.-M., Shen,E.-Z., Yang,B., Dong,M.-Q., He,W.-Z. and Du,L.-L. (2013) Global analysis of fission yeast mating genes reveals new autophagy factors. *PLoS Genet.*, **9**, e1003715.
10. Chatr-Aryamontri,A., Breitkreutz,B.-J., Heinicke,S., Boucher,L., Winter,A., Stark,C., Nixon,J., Ramage,L., Kolas,N., O'Donnell,L. *et al.* (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res.*, **41**, D816–D823.
11. Soriano,I., Quintales,L. and Antequera,F. (2013) Clustered regulatory elements at nucleosome-depleted regions punctuate a constant nucleosomal landscape in *Schizosaccharomyces pombe*. *BMC Genomics*, **14**, 813.
12. Rhind,N., Chen,Z., Yassour,M., Thompson,D.A., Haas,B.J., Habib,N., Wapinski,I., Roy,S., Lin,M.F., Heiman,D.I. *et al.* (2011) Comparative functional genomics of the fission yeasts. *Science*, **332**, 930–936.
13. Xu,J., Yanagisawa,Y., Tsankov,A.M., Hart,C., Aoki,K., Kommajosyula,N., Steinmann,K.E., Bochicchio,J., Russ,C., Regev,A. *et al.* (2012) Genome-wide identification and characterization of replication origins by deep sequencing. *Genome Biol.*, **13**, R27.

14. Mata,J. (2013) Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biol.*, **10**, 1407–1414.

15. Schlackow,M., Marguerat,S., Proudfoot,N.J., Bähler,J., Erban,R. and Gullerova,M. (2013) Genome-wide analysis of poly(A) site selection in *Schizosaccharomyces pombe*. *RNA*, **19**, 1617–1631.

16. Woolcock,K.J., Gaidatzis,D., Punga,T. and Bühler,M. (2011) Dicer associates with chromatin to repress genome activity in *Schizosaccharomyces pombe*. *Nat. Struct. Mol. Biol.*, **18**, 94–99.

17. Kersey,P.J., Allen,J.E., Christensen,M., Davis,P., Falin,L.J., Grabmueller,C., Hughes,D. S.T., Humphrey,J., Kerhornou,A., Khobova,J. *et al.* (2014) Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Res.*, **42**, D546–D552.

18. Mungall,C.J., Emmert,D.B. and FlyBase Consortium (2007) A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics*, **23**, i337–i346.

19. Gene Ontology Consortium, Blake,J., Dolan,M., Drabkin,H., Hill,D., Li,N., Sitnikov,D., Bridges,S., Burgess,S., Buza,T. *et al.* (2013) Gene Ontology annotations and resources. *Nucleic Acids Res.*, **41**, D530–D535.

20. Ashburner,M., Ball,C., Blake,J., Botstein,D., Butler,H., Cherry,J., Davis,A., Dolinski,K., Dwight,S., Eppig,J. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

21. Harris,M.A., Lock,A., Bähler,J., Oliver,S.G. and Wood,V. (2013) FYPO: the fission yeast phenotype ontology. *Bioinformatics*, **29**, 1671–1678.

22. Huntley,R.P., Harris,M.A., Alam-Faruque,Y., Blake,J.A., Carbon,S., Dietze,H., Dimmer,E.C., Foulger,R.E., Hill,D.P., Khodiyar,V.K. *et al.* (2014) A method for increasing expressivity of Gene Ontology annotations using a compositional approach. *BMC Bioinformatics*, **15**, 155.

23. Eilbeck,K., Lewis,S.E., Mungall,C.J., Yandell,M., Stein,L., Durbin,R. and Ashburner,M. (2005) The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.*, **6**, R44.

24. Hastings,J., de Matos,P., Dekker,A., Ennis,M., Harsha,B., Kale,N., Muthukrishnan,V., Owen,G., Turner,S., Williams,M. *et al.* (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.*, **41**, D456–D463.

25. Montecchi-Palazzi,L., Beavis,R., Binz,P.-A., Chalkley,R.J., Cottrell,J., Creasy,D., Shofstahl,J., Seymour,S.L. and Garavelli,J.S. (2008) The PSI-MOD community standard for representation of protein modification data. *Nat. Biotechnol.*, **26**, 864–866.

26. Sugimoto,A., Iino,Y., Maeda,T., Watanabe,Y. and Yamamoto,M. (1991) *Schizosaccharomyces pombe* ste11+ encodes a transcription factor with an HMG motif that is a critical regulator of sexual development. *Genes Dev.*, **5**, 1990–1999.

27. Lee,K., Saiz,J., Barton,W. and Fisher,R. (1999) Cdc2 activation in fission yeast depends on Mcs6 and Csk1, two partially redundant Cdk-activating kinases (CAKs). *Curr. Biol.*, **9**, 441–444.

28. Moreno,S., Hayles,J. and Nurse,P. (1989) Regulation of p34cdc2 protein kinase during mitosis. *Cell*, **58**, 361–372.

29. McEntyre,J.R., Ananiadou,S., Andrews,S., Black,W.J., Boulderstone,R., Buttery,P., Chaplin,D., Chevuru,S., Cobley,N., Coleman,L.-A. *et al.* (2011) UKPMC: a full text article resource for the life sciences. *Nucleic Acids Res.*, **39**, D58–D65.