

PomBase Community Curation: A Fast Track to Capture Expert Knowledge

Midori A. Harris¹, Kim Rutherford¹, Jürg Bähler², Paul Kersey³, Stephen G. Oliver⁴, Valerie Wood¹

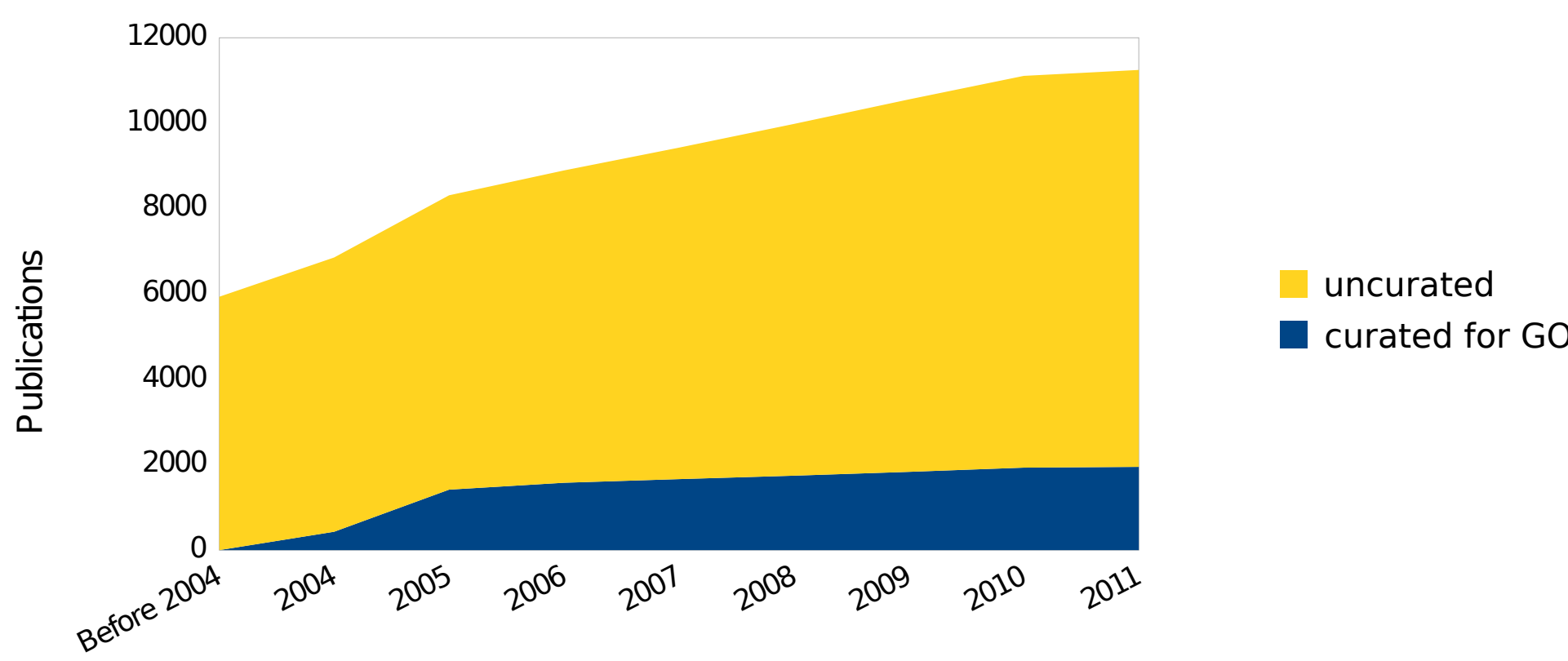
¹Department of Biochemistry, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, UK; ²Department of Genetics, Evolution and Environment, and UCL Cancer Institute, University College London, London WC1E 6BT, UK; ³EMBL Outstation —European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SD UK; ⁴Cambridge Systems Biology Centre and Department of Biochemistry, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, UK

Abstract:

PomBase, scheduled for launch in summer 2011, will be a new, comprehensive online resource for researchers working on *Schizosaccharomyces pombe*, combining structural and functional annotation, literature curation and access to large-scale data sets. Contributions from the *S. pombe* research community will be the primary source of curated data in PomBase. Following a successful pilot annotation project in which several laboratories contributed annotations — including GO annotations, phenotypes, genetic and physical interactions, and domain data — based on their recent publications, we are now developing a web-based tool to allow the entire *S. pombe* community to participate in PomBase curation. This tool for community annotation is being piloted by *S. pombe* community, but will also be made available to other communities of researchers working with model organisms. Thus, by participating in this programme, *S. pombe* researchers will not only benefit their own community, but also further the study of fundamental aspects of biology with other model organisms. We present the prototype curation interface, supporting GO annotations. By streamlining community contributions to PomBase and establishing pipelines to collaborating databases, we will rapidly work towards comprehensive, high-quality coverage of the *S. pombe* literature, reflect the best, most up-to-date knowledge available for this key model organism, and make *S. pombe* research globally visible.

Why Add Community Curation?

Accumulating Literature



In-depth curation of the entire body of *S. pombe* scientific literature is a primary goal of PomBase, but at present about 7000 publications remain uncured and the backlog continues to grow.

Active Community

Community Curation Annotation Summary							
PMID	Laboratory	Submitter if different	GO terms	Interactions	Phenotypes	Other curation	Total
PHASE 1							
1910154	Peter Folk	Martin P_evorovsk	19	0	15	3	37 4/109
19098712	Juan Mata		3	7	0	1	11 4/109
19076239	Paul Russell		5	0	2	2	9 4/109
18675827	Per Sønnerhagen		2	1	3	0	6 6/109
17353264	Chris Norbury		2	0	0	0	2 6/109
19057642	Luis Roekach						4/109
19056896	Fred Winston	Dom Helmlinger					4/109
19041767	Peter Espenshade	John Burg	4	2	2	1	9 4/109
18556659	Kevin Hardwick		16	16	2	0	34 6/109
18430926	Charlie Hoffman		4	0	2	0	6 6/109
<p><i>S. pombe</i> is ideal for introducing community curation: researchers contributed annotations based on ~90 papers in a successful pilot project.</p>							
18725026	Takashi Tada		2	3	3	0	8 6/109
19109429	Stephen Kearsey		9	2	2	0	13 6/109
17552403	Georgios Kouras						6/109

S. pombe is ideal for introducing community curation: researchers contributed annotations based on ~90 papers in a successful pilot project

A New Tool for Community Curation

To complement the efforts of PomBase's dedicated database curators, we are developing a web-based curation tool that will allow researchers to contribute annotations directly to PomBase.

GO annotation in the curation tool

PomBase CURS v1.26
Pombe community curation application

Postreplication repair and PCNA modification in Schizosaccharomyces pombe.

Submitter update

Please enter your name and email address to continue:

Name | Midori Harris
Email | mah79@hermes.cam.ac.uk

Start curation

Gene upload

Please enter the systematic identifier (eg. SPCC1739.10) or the primary identifier (eg. cdc11) of the genes referred to in PMID:16641370

abc1
mms2
rad8
rhp18

List all genes described in the paper

Choose what to curate

Curating ubc13 from PMID:16641370

Please read the definition of the currently selected term to ensure that it accurately describes your gene:

Find a term

GO:0070534
GO biological process
protein K63-linked ubiquitination
A protein ubiquitination process in which a polymer of ubiquitin, formed by linkages between lysine residues at position 63 of the ubiquitin molecule, is added to a protein. K63-linked ubiquitination does not target substrates for degradation, but is involved in several pathways, notably as a signal to promote error-free DNA post-replication repair.

Annotate!

Pick an evidence code

Curating ubc13 from PMID:16641370

Please read the definition of the currently selected term to ensure that it accurately describes your gene:

GO:0070534
GO biological process
protein K63-linked ubiquitination
A protein ubiquitination process in which a polymer of ubiquitin, formed by linkages between lysine residues at position 63 of the ubiquitin molecule, is added to a protein. K63-linked ubiquitination does not target substrates for degradation, but is involved in several pathways, notably as a signal to promote error-free DNA post-replication repair.

Table of all annotations from the paper

Systematic identifier	Gene name	Annotation type	Term ID	Term name	Evidence code	With	Note
SPAC11E3.04c	ubc13	GO biological process	GO:0071479	cellular response to ionizing radiation	IDA		✗
SPAC11E3.04c	ubc13	GO biological process	GO:0071479	cellular response to ionizing radiation	IMP		✗
SPAC11E3.04c	ubc13	GO biological process	GO:0071479	cellular response to ionizing radiation	IGI	mms2	✗
SPAC11E3.04c	ubc13	GO biological process	GO:0071479	cellular response to ionizing radiation	IGI	rad8	✗
SPAC11E3.04c	ubc13	GO biological process	GO:0071479	cellular response to ionizing radiation	IGI	rhp18	✗
SPAC11E3.04c	ubc13	GO biological process	GO:0070534	protein K63-linked ubiquitination	IDA		✗
SPCC338.05c	mms2	GO biological process	GO:0070534	protein K63-linked ubiquitination	IDA		✗
SPCC338.05c	mms2	GO biological process	GO:0071479	cellular response to ionizing radiation	IDA		✗
SPCC338.05c	mms2	GO biological process	GO:0071479	cellular response to ionizing radiation	IMP		✗
SPCC338.05c	mms2	GO biological process	GO:0071479	cellular response to ionizing radiation	IGI	rhp18	✗

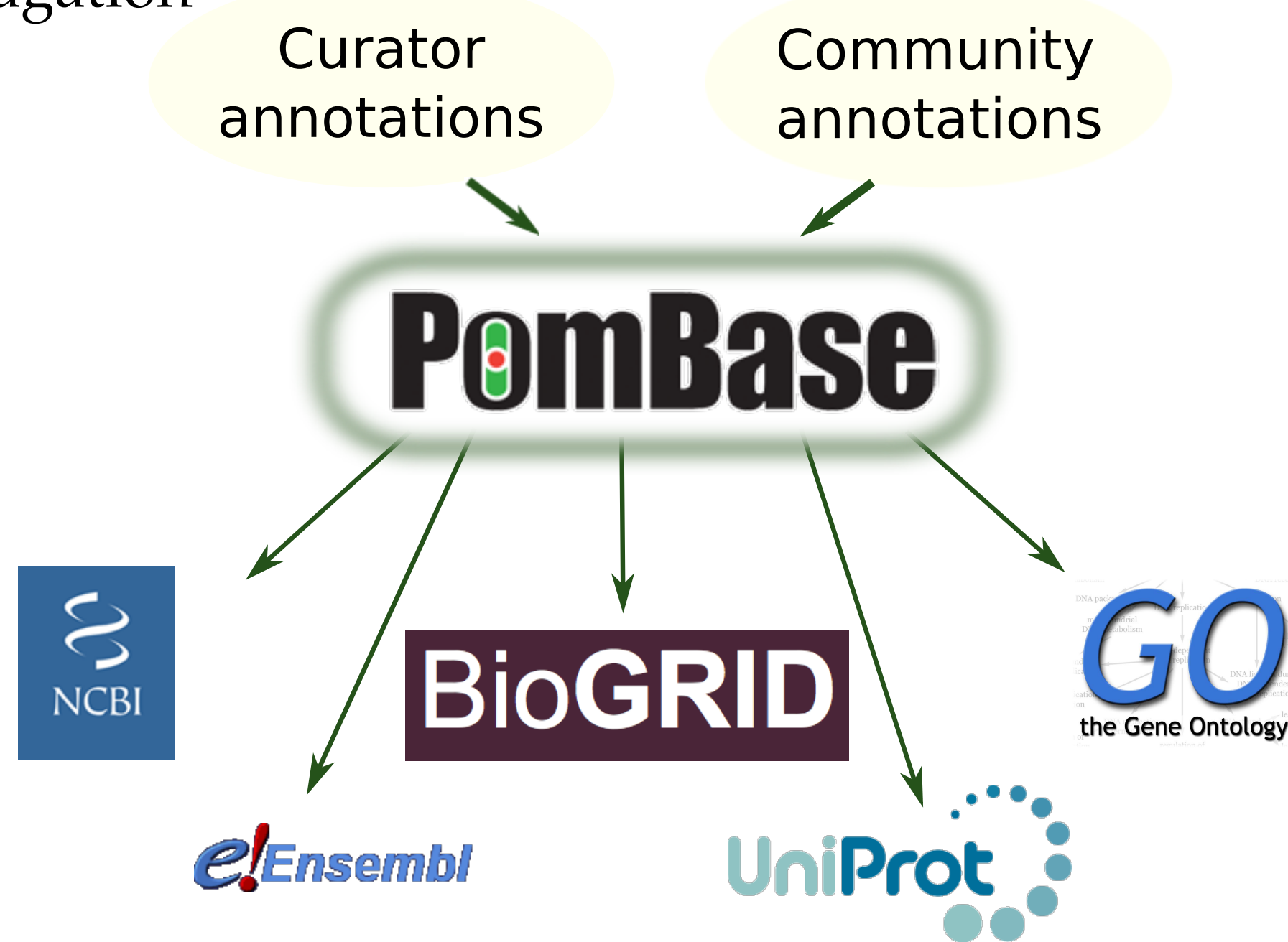
Download table ...

Tool benefits:

- easy-to-use interface, accessible on web
- capture expert knowledge direct from community & convert into specific, accurate annotations
- uses controlled vocabularies (ontologies) to ensure consistency
- automated error checking/quality control
- extensible — can add more data types, more controlled vocabularies, etc.
- co-ordinate curation with publication

Beyond PomBase

Data Propagation



Annotation data go into PomBase via the curation tool, and are then automatically made available to several other databases, ensuring that *S. pombe* research results are visible throughout the online biomedical data sphere.

Expanded Community Curation

- More data types will be available for *S. pombe*
 - Phenotypes
 - Modifications
 - Interactions
- The curation tool itself will be made available to researchers working on other model organisms, such as nematodes, fungi and plants.

Acknowledgements: PomBase is funded by the Wellcome Trust, and run by a consortium comprising the University of Cambridge, the European Bioinformatics Institute, and University College London.