WormBook: the online review of *Caenorhabditis elegans* biology

Lisa R. Girard[1],*, Tristan J. Fiedler[3], Todd W. Harris[3], Felicia Carvalho[4],
Igor Antoshechkin[1], Michael Han[5], Paul W. Sternberg[1,2], Lincoln D. Stein[3]
and Martin Chalfie[6]

[1]Division of Biology, 156-29 and [2]Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA 91125, USA, [3]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA, [4]Department of Pharmaceutical Sciences, St John's University, Queens, NY 11439, USA, [5]Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK and [6]Department of Biological Sciences, Columbia University, NY 10027, USA

ABSTRACT

WormBook (www.wormbook.org) is an open-access, online collection of original, peer-reviewed chapters on the biology of *Caenorhabditis elegans* and related nematodes. Since WormBook was launched in June 2005 with 12 chapters, it has grown to over 100 chapters, covering nearly every aspect of *C.elegans* research, from Cell Biology and Neurobiology to Evolution and Ecology. WormBook also serves as the text companion to WormBase, the *C.elegans* model organism database. Objects such as genes, proteins and cells are linked to the relevant pages in WormBase, providing easily accessible background information. Additionally, WormBook chapters contain links to other relevant topics in WormBook, and the in-text citations are linked to their abstracts in PubMed and full-text references, if available. Since WormBook is online, its chapters are able to contain movies and complex images that would not be possible in a print version. WormBook is designed to keep up with the rapid pace of discovery in the field of *C.elegans* research and continues to grow. WormBook represents a generic publishing infrastructure that is easily adaptable to other research communities to facilitate the dissemination of knowledge in the field.

INTRODUCTION

WormBook is a comprehensive online review of *Caenorhabditis elegans* biology, containing over 100 original, peer-reviewed chapters on a wide range of topics related to the biology of *C.elegans* and related nematodes; as well as WormMethods, a collection of laboratory methods and protocols useful for nematode researchers.

Information about *C.elegans* biology had been freely distributed for nearly two decades as part of the Worm Breeder's Gazette, a now defunct newsletter, which provided an informal arena for descriptions of new techniques and preliminary results. In the mid-1980s the worm community decided a review volume that provided a narrative on a wide range of topics would be useful for new researchers in the field, other scientists, and interested lay people, as well as serve as a useful reference for *C.elegans* biologists. This effort culminated in the publication of *The Nematode C.elegans* (1). However, by the mid-1990s a more up-to-date resource was needed. The resulting volume, *C.elegans II* (2), had 30 chapters (more than twice as many as *The Nematode C.elegans*), a reflection of the growing breadth and complexity of research in the field.

By 2003, this second volume was itself out-of-date and the Worm Breeder's Gazette was no longer being published. WormBook was created in response to the need to develop a publishing model that could keep pace with a rapidly growing knowledge base and provide a central repository for methods and protocols. Because of the accelerated pace of discovery in *C.elegans* biology and the greatly increased number of research areas, the community viewed that another print review volume could no longer be reasonably comprehensive. Instead, we decided to utilize an online format for WormBook. This format provides numerous advantages:

(i) The ability to support a wider range and more extensive use of media than print, including movies and complex images [see Figure 1 in (3), dx.doi.org/

*To whom correspondence should be addressed at 1200 E, California Blvd, Pasadena, CA 91125, USA. Tel: +1 603 646 3501; Fax: +1 603 646 1347; Email: lgirard@caltech.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

10.1895/wormbook. 1.41.1 and Movie 1 in (4), dx.doi.org/10.1895/wormbook.1.53.1 for examples].

 (ii) The use of in-text hyperlinks to other web resources [see Table 1 in (5), dx.doi.org/10.1895/wormbook.1.72.1, for example].

(iii) The ability to make easier and more regular updates. Revisions can be made rapidly without having to wait for a next edition to be printed, a great assistance in keeping the resource accurate and up-to-date.

(iv) The retention of control over the content. A print publication often involves surrendering ownership of content to a publisher. This transfer may compromise access to, and freedom with, the material. The accessibility of an online publishing model allows WormBook to come to fruition independent of a publisher.

 (v) The freedom from page limits. Currently, WormBook contains the equivalent of over 1800 printed pages. This is three times as many pages as *The Nematode C.elegans* and over one-third more pages than *C.elegans II*.

WormBook is an open-access publication and its original content is freely available. Our contributors retain the copyright for their material and agree to the attachment of a Creative Commons Attribution License, which allows others to copy, distribute or display the WormBook material, as well as derivative works based upon it, if proper credit is given (see www.creativecommons.org for more information). Under this license, schools are allowed to include original WormBook contributions in course readings and WormBook chapters may be more easily archived by libraries.

Since its inception in June 2005, usage rates have steadily risen to the current rate of ∼150 000 pages served monthly. WormBook has enjoyed the enthusiastic support of the *C.elegans* community and has already been cited in such influential publications as Science, Genes and Development and Nature.

WormBook provides a generic, easily adaptable infrastructure for text companions to model organism databases, whose numbers are increasing in the post-genomic era. It also serves as a tightly linked companion to WormBase [(6,7), www.wormbase.org], the *C.elegans* model organism database, providing a biological context for the facts WormBase presents. Genes, proteins and cells named in the text of WormBook are linked to the relevant sections of WormBase. More recently, links from several types of WormBase pages, including gene, protein and phenotype pages into WormBook have been implemented. Approaches similar to WormBook could be employed at other model organism databases.

WormBook is staffed primarily by *C.elegans* biologists. WormBook's editor-in-chief is responsible for overseeing the editorial board. The editor of WormBook, working with members of the editorial board, is responsible for ensuring the quality and flow of chapters from their inception, through their commissioning, review, revision and publication. The editorial board has been instrumental in ensuring the quality of published material. Each editorial board member, or a team of two members, is associated with a particular section of WormBook, based upon their area of expertise. The editorial board works with the editor and editor-in-chief to establish a table of contents for WormBook, commission

authors for contributions and contribute to the peer review process. WormBook's software developer's primary accomplishments include the WormBook chapter production pipeline, database architecture and user interface. WormBook has also benefited from the assistance of its technical production editor, as well as the expertise of the WormBase programming teams on refining the web interface and search capabilities.

CONTENT

The content of WormBook is divided into 10 sections: Genetics and Genomics, Molecular Biology, Cell Biology, Sex Determination, The Germ Line, Developmental Control, Signal Transduction, Neurobiology and Behavior, Evolution and Ecology, and WormMethods. Each section contains ∼10 chapters covering topics related to that subject area written by experts in the field. WormMethods includes a range of methods and protocols, accompanied by commentary, relevant to a range of studies including sections on Behavior, Genetic mapping, Cell Biology and Biochemistry. The chapters are geared toward a wide audience ranging from high school students and interested laypeople to *C.elegans* researchers.

FEATURES AND USER INTERFACE

The homepage of WormBook provides a launching off point from which readers can explore its contents, find links to widely used *C.elegans* resources including WormBase and Wormatlas, links to individual section contents, News and notes, a link to the WormBook mirror site located at the Wellcome Trust Sanger Institute in Cambridge, UK, and a search engine that allows readers to quickly locate desired information throughout the site. From the homepage, readers may also download the entire PDF version of WormBook (currently a 380 MB zip archive), as well as the entire set of WormBook references into bibliographic management software such as Endnote. Instructions to authors and a variety of mailing lists increase communication between the WormBook staff and the community. Readers are encouraged to stay abreast of new chapters coming online by visiting the e-Alerts section on the header of each page.

The chapter pages (see Figure 1) offer readers the choice of reading an HTML version of a chapter or downloading it as a PDF in order to print it more easily. WormBook chapters are extensively hyperlinked to other online resources. Chapters contain links to other WormBook chapters to provide background information on a particular topic. Genes, cells, proteins and other objects are linked to the relevant pages in WormBase and in-text references are linked to their abstracts in PubMed, and the full-text version of the reference, if available. Each chapter has a clickable outline visible to the reader, allowing instant navigation throughout the text. Capitalizing on its online capabilities, it also contains numerous movies and complex images that would not be possible in a print version.

WORMBOOK PRODUCTION PIPELINE

Each WormBook chapter is an invited contribution. Once a chapter is submitted, it is assessed by the editor and the editorial

Figure 1. WormBook chapters are extensively connected to other web resources. Chapters contain links to WormBase, PubMed abstracts, full-text journal articles, as well as other WormBook chapters.

board member associated with the particular section, and sent to two outside experts for review. The chapter is returned to the author with the reviewer's comments for revision. Once a chapter is accepted, the manuscript is converted into XML, and subsequently HTML and PDF file formats. The production pipeline was developed in-house using open-source software. Scaling of the WormBook project to the current level has been made possible by working with a digital publishing firm that has the capability to handle production capacity workloads for the XML markup process. The authors are sent proofs of their chapter to check for errors, and a final version of their chapter is then published on the WormBook website.

WORMBOOK SOFTWARE

DocBook markup language [(8), www.docbook.org], a type of XML designed for authoring technical documentation and publications, facilitates the creation of WormBook articles suitable for online and print display. DocBook markup establishes the logical structure of the document's content that can then be output in a variety of formats without any modification to the source. As a publishing method, DocBook offers other advantages including the ability to handle large quantities of structured content and tools for automated batch processing.

The Textpresso text mining system [(9), www.textpresso.org] identifies any biological information in the DocBook files through the use of a two-component ontology trained on the WormBase content. Keywords identified in the Worm-Book text via Textpresso are linked to entries at external web-accessible databases, such as WormBase. Similarly,

valid bibliographic reference elements are automatically inserted based on author-supplied PubMed identification data, establishing links to both the citation's abstract in PubMed and, through the resolution of the citation's Digital Object Identifier (DOI; www.doi.org), the full-text reference when available.

WormBook publishing takes advantage of the Extensible Style Language (XSL) to control the presentation of specific DocBook elements. XSL is an XML-based formatting object language possessing a rich set of tools for specifying the typeset and layout of all components of the DocBook element set (10). Being template-driven rather than a procedural language, XSL specifies a customizable output sample for each element type. XSL invokes two additional tools, namely XSL transformation (XSLT) for producing HTML or text output and the XML path language (Xpath) for addressing parts of the source XML document. Rearrangement of the source document into the final output sequence of elements is also controlled by XSLT transformation. Saxon is the open-source java XSLT processor (saxon.sourceforge.net) generating the HTML output format for all WormBook articles. Production of WormBook's print quality PDF output is done with FOP (xml.apache.org/fop), a freely available java-based formatting objects processor from the Apache XML project.

The WormBook chapters have been structured to allow external resources to link directly to specific sections, figures, tables and media objects within each html format chapter. These links have recently been implemented for WormBase gene and cell pages. As new WormBook chapters come online or existing chapters are revised, new WormBase paper objects are created. These database objects contain

the updated citation information, DOI and URL for each WormBook chapter. Candidate object names in WormBook are subsequently revised using regular expressions, then matched with the corresponding object identifiers in the current release of WormBase including gene, variation, clone, rearrangement, transgene and cell. Pre-formatted queries for the WormBook search engine also allow users to search the WormBook index for all instances of a particular object. This index is updated weekly to reflect any changes in content.

FUTURE DIRECTIONS

WormBook's content and features continue to expand. Currently, three new sections of WormBook are in progress: Biochemistry, Post-embryonic Development, and Disease Models and Drug Discovery in *C.elegans*. In addition to expanding WormBook content, all current chapters are scheduled to be updated every three years. Some authors, however, have already updated their chapters as new information becomes available. Readers will have access not only to the most current version of a chapter, but also to archived earlier versions.

WormBook provides a generic infrastructure for a model organism database text companion. Although the utility of community-based resources such as WormBook has been demonstrated, a funding niche, public or private, that will enable the model to be easily adapted by other research communities needs to be delineated. We hope that the extensible nature of WormBook will encourage other research communities to adopt similar projects to create an up-to-date resource, facilitating the dissemination of information and fostering the growth of knowledge.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Wood,W.B. and the Community of *C. elegans* researchers (eds) (1988) *The Nematode Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Riddle,D.L., Blumenthal,T., Meyer,B.J. and Priess,J.R. (eds) (1997) *C. elegans* II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3. De Ley,P. (2006) A quick tour of nematode diversity and the backbone of nematode phylogeny (January 25, 2006). The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.41.1. .
4. Greenstein,D. (2005) Control of oocyte meiotic maturation and fertilization (December 28, 2005). The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.53.1.
5. Oegema,K. and Hyman,A.A. (2006) Cell division (January 19, 2006). The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.72.1.
6. Schwarz,E.M., Antoshechkin,I., Bastiani,C., Bieri,T., Blasiar,D., Canaran,P., Chan,J., Chen,N., Chen,W.J., Davis,P. *et al.* (2006) WormBase: better software, richer content. *Nucleic Acids Res.*, **34**, D475–D478.
7. Stein,L., Sternberg,P., Durbin,R., Thierry-Mieg,J. and Spieth,J. (2001) WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res.*, **29**, 82–86.
8. Walsh,N. and Muellner,L. (1999) *DocBook: The Definitive Guide*. O'Reilly and Associates, Inc., Sebastopol, CA.
9. Müller,H., Kenny,E.E. and Sternberg,P.W. (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol.*, **2**, e309.
10. Stayton,R. (2005) *DocBook XSL: The Complete Guide, 3rd edn.* Sagehill Enterprises, Santa Cruz, CA.