RiceWiki: a wiki-based database for community curation of rice genes

Zhang Zhang[1],*, Jian Sang[2], Lina Ma[1], Gang Wu[1], Hao Wu[1], Dawei Huang[1], Dong Zou[1], Siqi Liu[1], Ang Li[1], Lili Hao[1], Ming Tian[3], Chao Xu[3], Xumin Wang[1], Jiayan Wu[1], Jingfa Xiao[1], Lin Dai[3], Ling-Ling Chen[4], Songnian Hu[1] and Jun Yu[1]

[1]CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China, [2]Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Fuyang, Zhejiang 311400, China, [3]School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China and [4]College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

ABSTRACT

Rice is the most important staple food for a large part of the world's human population and also a key model organism for biological studies of crops as well as other related plants. Here we present RiceWiki (http://ricewiki.big.ac.cn), a wiki-based, publicly editable and open-content platform for community curation of rice genes. Most existing related biological databases are based on expert curation; with the exponentially exploding volume of rice knowledge and other relevant data, however, expert curation becomes increasingly laborious and time-consuming to keep knowledge up-to-date, accurate and comprehensive, struggling with the flood of data and requiring a large number of people getting involved in rice knowledge curation. Unlike extant relevant databases, RiceWiki features harnessing collective intelligence in community curation of rice genes, quantifying users' contributions in each curated gene and providing explicit authorship for each contributor in any given gene, with the aim to exploit the full potential of the scientific community for rice knowledge curation. Based on community curation, RiceWiki bears the potential to make it possible to build a rice encyclopedia by and for the scientific community that harnesses community intelligence for collaborative knowledge curation, covers all aspects of biological knowledge and keeps evolving with novel knowledge.

INTRODUCTION

Rice (*Oryza sativa*) is not only the most important staple food feeding a large part of the world population, but also an important model organism for biological studies of crops as well as other related plants (1,2). For this reason, rice was chosen as the first crop for whole genome sequencing. The availability of genome sequences of the two most common cultivated rice subspecies, *indica* 93-11 and *japonica* Nipponbare (3,4), along with the resequencing of cultivated and wild rice accessions (5), enables the in-depth characterization of rice genes (6), exploration of agronomically important traits (7) and investigation of rice diversity and domestication (8).

Multiple databases (9–13) have been developed for rice. However, building standardized rice reference genomes with comprehensive and accurate annotations remains a formidable challenge. Extant related databases are most based on expert curation, viz., conducted manually by dedicated experts. To perform curation, expert curators often administer raw biological data, conduct a thorough literature search, extract essential information from multiple publications, curate the information using structured and controlled vocabularies and then submit the information to a knowledge database [e.g., the Reference Sequence database (11) at the National Center for Biotechnology Information] to make it public. However, with the exponentially exploding volume of rice knowledge and other relevant data, expert curation is becoming more laborious and time-consuming. Keeping biological knowledge in expert-curated databases comprehensive, up-to-date and accurate is increasingly lagging behind knowledge creation, or worse, not being done at

all in fields where insufficient funds can be allocated to curation (14). Simply put, although expert-curated databases have traditionally proven important for scientific research, they are struggling with the flood of knowledge. Explosive growth in the volume of rice data requires a large number of people getting involved in knowledge curation, viz., community curation.

Community curation exploits the whole power of the scientific community for knowledge curation (15–19). A case in point that harnesses community intelligence in knowledge integration is Wikipedia (http://www.wikipedia.org). Wikipedia is an online encyclopedia, allowing any user to create/edit any content. It features collaborative knowledge curation, up-to-date content, huge coverage and low cost for maintenance (20). Despite fears that the openness of editorial capacity could lead to incorporation of significant flawed content (21), it is reported that Wikipedia rivals the traditional encyclopedia in accuracy (22). Broad participation in Wikipedia not only increases knowledge coverage and keeps knowledge up-to-date, but also improves knowledge accuracy ('with enough eyeballs, all bugs are shallow'; http://oreilly.com/web2/archive/what-is-web-20.html). Owing to the extraordinary success of Wikipedia, it has been advocated that biological databases go wiki (23). As a consequence, more than a dozen biological wikis (24–42) have been constructed to exploit the full potential of the scientific community for knowledge curation, such as EcoliWiki for the model bacteria *Escherichia coli* (27), Gene Wiki for human gene annotation (28) and PDBWiki for protein structures (31). To date, however, there is no community-curated resource for rice.

To establish a platform for community integration of rice data and to facilitate efficient management of community knowledge on rice, here we develop RiceWiki (http://ricewiki.big.ac.cn), a wiki-based, publicly editable and open-content platform for community curation of rice genes. Unlike extant relevant databases, RiceWiki features harnessing community intelligence in curation of rice genes, quantifying users' contributions in each curated gene and providing explicit authorship for each contributor in any given gene, with the aim to attract more participation from the scientific community in collaborative and collective curation of rice genes.

IMPLEMENTATION

RiceWiki has been implemented using MediaWiki (http://www.mediawiki.org; a free and open source wiki engine; version 1.18.4), MySQL (http://www.mysql.org; a free and popular relational database management system; version 5.1.58) and PHP (http://www.php.net; a widely used general-purpose scripting language; version 5.2.17) on a Red Hat Enterprise Linux Server. The wide adoption of community intelligence in knowledge curation is primarily attributable to free wiki software such as MediaWiki that provides a collaborative framework for knowledge collection, management and dissemination (that powers Wikipedia). MediaWiki allows any user to add, modify or delete any content (with customized permission

control for editing) via a web browser without any extra add-ons and thus enables web content to be edited easily, swiftly and collaboratively by multiple different users. Every page in MediaWiki has an associated page named 'History'; every change made to a page can be stored; the user responsible for every change can be identified and every history revision can be reviewed or recovered. Built on MediaWiki, RiceWiki enables users to be involved in an ongoing process of creation and collaboration that constantly changes the contents. Thus, RiceWiki can significantly ease the process of knowledge collection, curation and sharing, befitting the exploding volume of biological data.

In addition, MediaWiki allows any user to develop customized functionalities by packaging a bunch of codes as MediaWiki extensions. We installed our newly developed extension named 'AuthorReward' (http://www.mediawiki.org/wiki/Extension:AuthorReward) in RiceWiki, with the aim to attract more participation from the scientific community for collaborative curation of rice genes. A wiki page can be collaboratively curated by multiple users and thus may have different edit versions. For each version that is contributed by a specific person, 'AuthorReward' quantifies his/her contribution by factoring edit quality as well as edit quantity; the edit quantity amounts to the edit distance in comparison with its previous version (i.e. the minimum number of edit operations required to transform one string into the other), and the edit quality corresponds to whether the edit persists in comparison with the last version, ranging from -1, when the edit is entirely reverted (short-lived), to 1, indicating that the edit is totally preserved in the last version (long-lived). Because one person may perform many discontinuous edits for a wiki page, his/her contribution score in this page is the sum of quantified contributions over all contributed edits. 'AuthorReward' provides RiceWiki with an authorship metric; it quantifies users' contributions and yields explicit authorship for each wiki page according to their quantitative contributions (described later in the text). Thus, 'AuthorReward' bears the potential to increase community participation in RiceWiki and to achieve community curation of massive biological knowledge. All extensions as well as software installed in RiceWiki can be found at http://ricewiki.big.ac.cn/index.php/Special:Version.

In RiceWiki, users can access any content, but only registered users can perform the edits. This restriction is due to a trade-off between simplicity for users to make contributions and reliability of the edits provided. Open identity provided by registration not only improves content reliability and increases users' collaborations and communications, but it is also supportive to reward community-curated efforts by giving explicit authorship. Thus, although the requirement of registration poses an unpleasant obstacle to community curation, it is of crucial significance for bio-wikis that would like to give credit to all contributors in reward for community-provided content. Additionally, it is beneficial to greatly avoid vandalism or spammed content in bio-wikis, albeit, of course, there are multiple solutions for vandalism detection (43).

DATABASE CONTENT

RiceWiki incorporates the two common cultivated rice subspecies (*O. sativa indica* 93–11 and *O. sativa japonica* Nipponbare) and covers >66 000 rice genes. As implemented on MediaWiki, RiceWiki inherits its features: each content page is associated with a discussion page (where users can discuss content or leave a comment), a history page (where revision as well as its contributor can be recognized) and category terms (that increase the usability for information management).

The central objects of RiceWiki are rice genes and thus each gene corresponds to a specific wiki page. To provide easy access to gene-specific pages, they are built based on gene identifiers. Although they can also be based on gene names, it is noted that only a small part of rice genes has been well studied. With the ongoing studies on the biological functions of rice genes, a gene name often comes and goes as our knowledge about that entity increases, which would bring uncertainty of designations and synonyms of rice genes. Thus, gene identifiers are relatively stable to refer to specific genes, as they do not change with the accumulation of novel information. The gene information in RiceWiki was initially seeded from NCBI RefSeq (11), Ensembl (44), RAP-DB (12) and MSU Rice Genome Annotation Project (http://rice.plantbiology.msu.edu).

The content of every gene in RiceWiki is structured into multiple sections, namely, 'Annotated Information', 'Structured Information', 'Labs Working on This Gene' and 'References', as well as a one-sentence summary for gene description at the top of each page (Figure 1). 'Annotated Information' is organized as free text and is helpful to users who share their knowledge and contribute edits without training in the curation or wiki techniques, significantly simplifying edits' provision and lowering technological entrance barriers for wider community participation in curation. It can also fall into several sub-sections, such as 'Function', 'Evolution' and 'Expression', making it convenient to direct users to the sub-section(s) of interest. Although these sub-sections are preset, new sub-sections can be easily added and irrelevant sub-section(s) can be deleted. Such arrangement with multiple sub-sections enables not only automatic entry of information via application programming interface but also facilitates users to intuitively edit the information by clicking an 'edit' link available to each sub-section. On the contrary, 'Structured Information' is organized structurally in the form of a table, including gene symbol, gene description, sequence information, expression profile, external links to other related databases and a set of images from GBrowse (45) showing the genomic context of the gene and the gene structure. Albeit the structure of this table is preset, it also allows users to provide updates and additions. 'Labs Working on This Gene' is a list of laboratories throughout the world working on this gene, facilitating communication and collaboration in curation of this gene. 'References' are publications closely related to this gene and automatically formatted using the 'Cite' extension (http://www.mediawiki.org/wiki/Extension:Cite).

The major focus of RiceWiki is to exploit the full potential of the scientific community in collaborative curation of rice genes. It should be noted, however, that one crucial essential to making knowledge aggregation successful in RiceWiki and also in other bio-wikis is sufficient participation (19), which requires a large number of people from the scientific community to curate biological knowledge collectively and collaboratively. Unfortunately, the current status is that, despite the presence of well-constructed bio-wikis, most researchers seldom make direct contributions. The reason is well known: most people in areas covered by bio-wikis are in academic research fields, and the currency of academic research careers is authorship, and bio-wikis do not offer attributable authorship. It is an especially severe issue for young researchers, who are most open to new technologies such as wikis, but for whom one's publication record is the significant determinant to his/her academic professional success (41). It is inevitably time-consuming and poorly motivating for researchers to spend time performing curation without career-advancing credit (46) as opposed to writing their next paper–although, arguably, contributing to a common and open resource such as a bio-wiki may have much broader scientific impacts. There is no mechanism to reward people who perform knowledge curation in bio-wikis. It has been recognized that the major limitation deterring researchers from active participation in bio-wikis is the lack of explicit authorship and thus no credit for their contributions (17,19,41). It should be also noted, however, that not all people care about authorship, e.g., long-term Wikipedia editors. Despite this, authorship might be essential to academic-based researchers.

To attract more participation from the scientific community for RiceWiki and to make it a vivid platform for community curation of rice genes, we installed 'AuthorReward' (47), an extension to MediaWiki, which provides a standard practice to reward community-provided contents in bio-wikis by quantifying researchers' contributions and providing explicit authorship according to their quantitative contributions. For each gene-centric page in RiceWiki, community curation is quantified as a contribution score for each contributor in which both edit quantity and quality are taken into account (Figure 1). Accordingly, authorship in RiceWiki is awarded only to a contributor whose contribution score is >1 (by default; this cut-off score is configurable). At the top of each page, we provide the authorship information, displaying contributor name(s), gene ID, hyperlink to this gene and last update time, which is aimed to encourage community participation and to show clearly how to cite community-curated efforts. Additionally, at the bottom of each page, we present the detailed information, including a pie chart to visualize quantified contributions of multiple contributors who were involved in curating this gene, a histogram to depict the edit quantity and quality for each contributor and a table summarizing contributor name, contribution score, edit count, edit quantity, edit quality, last edit time and edit details. As one person may perform curations for multiple different pages, we define the total contribution of this person to RiceWiki as the sum of multiple
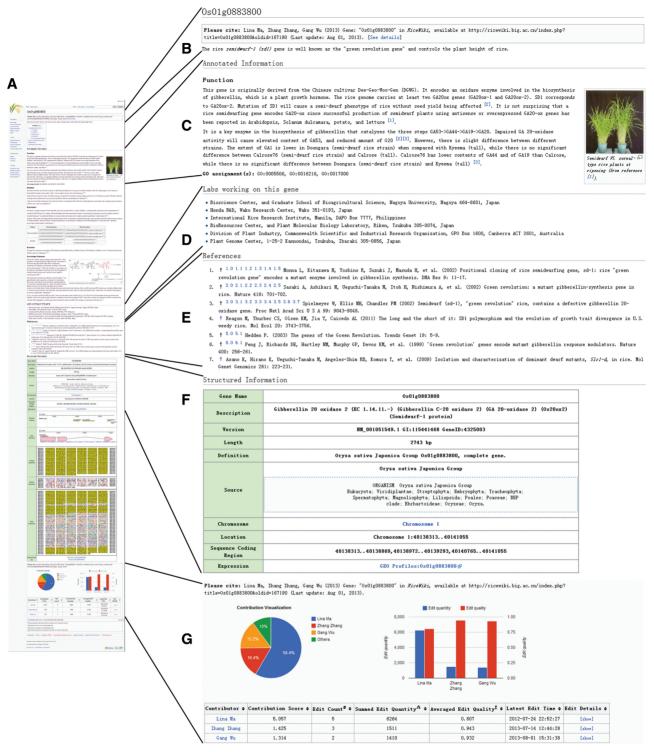
Figure 1. Screenshots of the RiceWiki page for the rice semidwarf-1 (*sd1*) gene, available at http://ricewiki.big.ac.cn/index.php/Os01g0883800. This gene was collaboratively curated by nine researchers, yielding 89 versions as of 1 August 2013. (**A**) Whole page, containing multiple different sections. (**B**) Brief authorship information in reward for community curation and a one-sentence summary for description of this gene. (**C**) Annotated information. It is organized as free text and contains several sub-sections. (**D**) Labs working on this gene. It is a list of laboratories working on this gene derived from references and provided by the community. (**E**) References, which are automatically generated and formatted with the help of the 'Cite' extension. (**F**) Structured information. It is organized structurally in the form of a table, including gene symbol, gene description, sequence information, gene structure, etc. (**G**) Authorship details generated by the 'AuthorReward' extension that quantifies researchers' contributions and provides explicit authorship according to their quantitative contributions. The cutoff score for awarding authorship is configurable and set to 1 (by default) in RiceWiki.

contribution scores in all participated pages. With explicit authorship as a reward for community curation, RiceWiki has the potential to attract more people to share their expertise and to provide edits on genes of their interest.

DISCUSSION AND FUTURE DIRECTIONS

The explosion of multiple different bio-wikis and RiceWiki shows the potential of 'wikiomics' in action (48). Considering the ever-growing volume of rice data and related literature and contrastingly the relatively small number of expert curators working on rice, it should be noted that community curation is an important complement to expert curation, and community-based bio-wikis, like RiceWiki, are not aimed for replacement of traditionally expert-driven databases. RiceWiki harnesses community intelligence in curating a wide range of rice-related topics and thus can save considerable time and efforts of expert curators. To encourage collaborative curations of rice genes in RiceWiki, expert curators can conduct quality control for community-contributed information and provide a variety of training (e.g., webinars, online tutorials and open discussion) for the community on how to perform curation in an accurate and standard manner. Meanwhile, journals can be also involved in curation by building a mechanism to regard community curation as a compulsory post-publication process and by providing obligations or incentives to authors submitting relevant information to RiceWiki. Such a mechanism has already been put into practice in the journal Plant Physiology, partnering with The Arabidopsis Information Resource to increase the curation of Arabidopsis. Wider adoption of this practice would be enormously beneficial to curatorial efficiency, accuracy and reliability (as testified by the admirable annotations for Arabidopsis). Therefore, the community as well as expert curators, authors, and journals should collaborate together to make RiceWiki more influential and to achieve community curation of massive rice knowledge.

Future directions for RiceWiki include establishment of close collaborations with laboratories in the world working on rice. 'Nothing great is ever accomplished in isolation'—Yo-Yo Ma. Rice knowledge and related data developed by many laboratories and researchers should be added to RiceWiki and shared with the whole scientific community. We will also promote, as best as we can partnership with journals to require community curation as a compulsory post-publication. In addition, we encourage investigators/teachers to incorporate community curation of rice genes in RiceWiki as student assignments. For example, N students collaborate to curate N rice genes, where $N \geq 3$, and contribution score for each student should be >1. RiceWiki will continue to integrate more types of data (e.g., mutants, repetitive elements, expression and phenotyping) from different resources and improve the connections with existing relevant databases. We have collected ~40 000 rice-related publications from PubMed (http://www.ncbi.nlm.nih.gov/pubmed/), including title, author(s), affiliation(s), abstract and hyperlinks to the full text (if available), and currently are attempting to dig out the 'treasure' from the flood of literature. Therefore, RiceWiki will also integrate tools particularly for literature mining (49,50) and incorporate literature-based curated annotation, to realize automatic information retrieval and improve the credibility of community-provided contents. In response to an update of literature collection, RiceWiki will build a mechanism to send an invitation to authors with recent publications for curation of specific genes (51).

In sum, RiceWiki serves as a community-curated knowledgebase for the rice research community. It exploits the whole power of the scientific community in collaborative curation of rice genes by rewarding community-provided content through contribution quantification and explicit authorship. Such a collaborative community-contributed and contribution-rewarded resource would make it possible to build a rice encyclopedia by and for the scientific community (52) that harnesses collective intelligence for collaborative knowledge curation, covers all aspects of biological knowledge and keeps evolving with novel knowledge.

DATABASE AVAILABILITY

RiceWiki is freely available at http://ricewiki.big.ac.cn.

REFERENCES

1. Shimamoto,K. and Kyozuka,J. (2002) Rice as a model for comparative genomics of plants. *Annu. Rev. Plant. Biol.*, **53**, 399–419.
2. Paterson,A.H., Freeling,M. and Sasaki,T. (2005) Grains of knowledge: genomics of model cereals. *Genome Res.*, **15**, 1643–1650.
3. Yu,J., Hu,S., Wang,J., Wong,G.K., Li,S., Liu,B., Deng,Y., Dai,L., Zhou,Y., Zhang,X. *et al.* (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science*, **296**, 79–92.

4. Goff,S.A., Ricke,D., Lan,T.H., Presting,G., Wang,R., Dunn,M., Glazebrook,J., Sessions,A., Oeller,P., Varma,H. *et al.* (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science*, **296**, 92–100.

5. Xu,X., Liu,X., Ge,S., Jensen,J.D., Hu,F., Li,X., Dong,Y., Gutenkunst,R.N., Fang,L., Huang,L. *et al.* (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.*, **30**, 105–111.

6. Zhang,G., Guo,G., Hu,X., Zhang,Y., Li,Q., Li,R., Zhuang,R., Lu,Z., He,Z., Fang,X. *et al.* (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.*, **20**, 646–654.

7. Huang,X., Wei,X., Sang,T., Zhao,Q., Feng,Q., Zhao,Y., Li,C., Zhu,C., Lu,T., Zhang,Z. *et al.* (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.*, **42**, 961–967.

8. Huang,X., Kurata,N., Wei,X., Wang,Z.X., Wang,A., Zhao,Q., Zhao,Y., Liu,K., Lu,H., Li,W. *et al.* (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature*, **490**, 497–501.

9. Zhao,W., Wang,J., He,X., Huang,X., Jiao,Y., Dai,M., Wei,S., Fu,J., Chen,Y., Ren,X. *et al.* (2004) BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Res.*, **32**, D377–D382.

10. Ouyang,S., Zhu,W., Hamilton,J., Lin,H., Campbell,M., Childs,K., Thibaud-Nissen,F., Malek,R.L., Lee,Y., Zheng,L. *et al.* (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.*, **35**, D883–D887.

11. Pruitt,K.D., Tatusova,T., Brown,G.R. and Maglott,D.R. (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.*, **40**, D130–D135.

12. Sakai,H., Lee,S.S., Tanaka,T., Numa,H., Kim,J., Kawahara,Y., Wakimoto,H., Yang,C.C., Iwamoto,M., Abe,T. *et al.* (2013) Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.*, **54**, e6.

13. Wang,D., Xia,Y., Li,X., Hou,L. and Yu,J. (2013) The Rice Genome Knowledgebase (RGKbase): an annotation database for rice comparative genomics and evolutionary biology. *Nucleic Acids Res.*, **41**, D1199–1205.

14. Baker,M. (2012) Databases fight funding cuts. *Nature*, **489**, 19.

15. Salzberg,S.L. (2007) Genome re-annotation: a wiki solution? *Genome Biol.*, **8**, 102.

16. Hu,J.C., Aramayo,R., Bolser,D., Conway,T., Elsik,C.G., Gribskov,M., Kelder,T., Kihara,D., Knight,T.F. Jr, Pico,A.R. *et al.* (2008) The emerging world of wikis. *Science*, **320**, 1289–1290.

17. Howe,D., Costanzo,M., Fey,P., Gojobori,T., Hannick,L., Hide,W., Hill,D.P., Kania,R., Schaeffer,M., St Pierre,S. *et al.* (2008) Big data: The future of biocuration. *Nature*, **455**, 47–50.

18. Zhang,Z., Bajic,V.B., Yu,J., Cheung,K.-H. and Townsend,J.P. (2011) In: Mahdavi,M.A. (ed.), *Bioinformatics - Trends and Methodologies*, Vol. 1. InTech, Rijeka, Croatia, pp. 41–56.

19. Finn,R.D., Gardner,P.P. and Bateman,A. (2012) Making your database available through Wikipedia: the pros and cons. *Nucleic Acids Res.*, **40**, D9–D12.

20. Zhang,Z., Cheung,K.H. and Townsend,J.P. (2009) Bringing Web 2.0 to bioinformatics. *Brief. Bioinformatics*, **10**, 1–10.

21. Bidartondo,M.I. (2008) Preserving accuracy in GenBank. *Science*, **319**, 1616.

22. Giles,J. (2005) Internet encyclopaedias go head to head. *Nature*, **438**, 900–901.

23. Giles,J. (2007) Key biology databases go wiki. *Nature*, **445**, 691.

24. Kumar,S., Schiffer,P.H. and Blaxter,M. (2012) 959 Nematode Genomes: a semantic wiki for coordinating sequencing projects. *Nucleic Acids Res.*, **40**, D1295–D300.

25. Stokes,T.H., Torrance,J.T., Li,H. and Wang,M.D. (2008) ArrayWiki: an enabling technology for sharing public microarray data repositories and meta-analyses. *BMC Bioinformatics*, **9(Suppl. 6)**, S18.

26. Hoehndorf,R., Bacher,J., Backhaus,M., Gregorio,S.E. Jr, Loebe,F., Prufer,K., Uciteli,A., Visagie,J., Herre,H. and Kelso,J. (2009) BOWiki: an ontology-based wiki for annotation of data and integration of knowledge in biology. *BMC Bioinformatics*, **10(Suppl. 5)**, S5.

27. McIntosh,B.K., Renfro,D.P., Knapp,G.S., Lairikyengbam,C.R., Liles,N.M., Niu,L., Supak,A.M., Venkatraman,A., Zweifel,A.E., Siegele,D.A. *et al.* (2012) EcoliWiki: a wiki-based community resource for *Escherichia coli*. *Nucleic Acids Res.*, **40**, D1270–D1277.

28. Good,B.M., Clarke,E.L., de Alfaro,L. and Su,A.I. (2012) The Gene Wiki in 2011: community intelligence applied to human gene annotation. *Nucleic Acids Res.*, **40**, D1255–D1261.

29. Renfro,D.P., McIntosh,B.K., Venkatraman,A., Siegele,D.A. and Hu,J.C. (2012) GONUTS: the gene ontology normal usage tracking system. *Nucleic Acids Res.*, **40**, D1262–D1269.

30. Bolser,D.M., Chibon,P.Y., Palopoli,N., Gong,S., Jacob,D., Del Angel,V.D., Swan,D., Bassi,S., Gonzalez,V., Suravajhala,P. *et al.* (2012) MetaBase—the wiki-database of biological databases. *Nucleic Acids Res.*, **40**, D1250–D1254.

31. Stehr,H., Duarte,J.M., Lappe,M., Bhak,J. and Bolser,D.M. (2010) PDBWiki: added value through community annotation of the Protein Data Bank. *Database (Oxford)*, **2010**, baq009.

32. Hodis,E., Prilusky,J., Martz,E., Silman,I., Moult,J. and Sussman,J.L. (2008) Proteopedia - a scientific 'wiki' bridging the rift between three-dimensional structure and function of biomacromolecules. *Genome Biol.*, **9**, R121.

33. Gardner,P.P., Daub,J., Tate,J., Moore,B.L., Osuch,I.H., Griffiths-Jones,S., Finn,R.D., Nawrocki,E.P., Kolbe,D.L., Eddy,S.R. *et al.* (2011) Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Res.*, **39**, D141–D145.

34. Li,J.W., Robison,K., Martin,M., Sjodin,A., Usadel,B., Young,M., Olivares,E.C. and Bolser,D.M. (2012) The SEQanswers wiki: a wiki database of tools for high-throughput sequencing analysis. *Nucleic Acids Res.*, **40**, D1313–D1317.

35. Cariaso,M. and Lennon,G. (2012) SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Res.*, **40**, D1308–D1312.

36. Mader,U., Schmeisky,A.G., Florez,L.A. and Stulke,J. (2012) SubtiWiki–a comprehensive community resource for the model organism Bacillus subtilis. *Nucleic Acids Res.*, **40**, D1278–D1287.

37. Csosz,E., Mesko,B. and Fesus,L. (2009) Transdab wiki: the interactive transglutaminase substrate database on web 2.0 surface. *Amino Acids*, **36**, 615–617.

38. Zhao,D., Wu,J., Zhou,Y., Gong,W., Xiao,J. and Yu,J. (2012) WikiCell: a unified resource platform for human transcriptomics research. *OMICS*, **16**, 357–362.

39. Kelder,T., van Iersel,M.P., Hanspers,K., Kutmon,M., Conklin,B.R., Evelo,C.T. and Pico,A.R. (2012) WikiPathways: building research communities on biological pathways. *Nucleic Acids Res.*, **40**, D1301–D1307.

40. Mons,B., Ashburner,M., Chichester,C., van Mulligen,E., Weeber,M., den Dunnen,J., van Ommen,G.J., Musen,M., Cockerill,M., Hermjakob,H. *et al.* (2008) Calling on a million minds for community annotation in WikiProteins. *Genome Biol.*, **9**, R89.

41. Hoffmann,R. (2008) A wiki for the life sciences where authorship matters. *Nat. Genet.*, **40**, 1047–1051.

42. Dai,L., Xu,C., Tian,M., Sang,J., Zou,D., Li,A., Liu,G., Chen,F., Wu,J., Xiao,J. *et al.* (2013) Community intelligence in knowledge curation: an application to managing scientific nomenclature. *PLoS One*, **8**, e56961.

43. Potthast,M., Stein,B. and Gerling,R. (2008) Automatic vandalism detection in Wikipedia. *Adv. Inform. Retr.*, **4956**, 663–668.

44. Flicek,P., Ahmed,I., Amode,M.R., Barrell,D., Beal,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fairley,S. *et al.* (2013) Ensembl 2013. *Nucleic Acids Res.*, **41**, D48–D55.

45. Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W., Arva,A. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.

46. Callaway,E. (2010) No rest for the bio-wikis. *Nature*, **468**, 359–360.

47. Dai,L., Tian,M., Wu,J., Xiao,J., Wang,X., Townsend,J.P. and Zhang,Z. (2013) AuthorReward: increasing community curation in biological knowledge wikis through automated authorship quantification. *Bioinformatics*, **29**, 1837–1839.

48. Waldrop,M. (2008) Big data: wikiomics. *Nature*, **455**, 22–25.
49. Muller,H.M., Kenny,E.E. and Sternberg,P.W. (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol.*, **2**, e309.
50. Wei,C.H., Kao,H.Y. and Lu,Z. (2013) PubTator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Res.*, **41**, W518–W522.
51. Bunt,S.M., Grumbling,G.B., Field,H.I., Marygold,S.J., Brown,N.H. and Millburn,G.H. (2012) Directly e-mailing authors of newly published papers encourages community curation. *Database (Oxford)*, **2012**, bas024.
52. Zhang,Q., Li,J., Xue,Y., Han,B. and Deng,X.W. (2008) Rice 2020: a call for an international coordinated effort in rice functional genomics. *Mol. Plant*, **1**, 715–719.