



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib



Data Article

Draft genome sequence of *Trametes villosa* (Sw.) Kreisel CCMB561, a tropical white-rot Basidiomycota from the semiarid region of Brazil

Dalila Souza Santos Ferreira^a, Rodrigo Bentes Kato^b,
Fábio Malcher Miranda^{b,f}, Kenny da Costa Pinheiro^e,
Paula Luize Camargos Fonseca^b, Luiz Marcelo Ribeiro Tomé^b,
Aline Bruna Martins Vaz^b, Fernanda Badotti^c,
Rommel Thiago Jucá Ramos^e, Bertram Brenig^d,
Vasco Ariston de Carvalho Azevedo^b,
Raquel Guimarães Benevides^a, Aristóteles Góes-Neto^{a,b,*}

^a State University of Feira de Santana, Department of Biological Science, Feira de Santana, BA 44036-900, Brazil

^b Federal University of Minas Gerais, Institute of Biological Sciences, Belo Horizonte, MG 31270-901, Brazil

^c Federal Center of Technological Education of Minas Gerais (CEFET-MG), Belo Horizonte, MG 30421-169, Brazil

^d University of Göttingen, Institute of Veterinary Medicine, Burckhardtweg 2, D-37077 Göttingen, Germany

^e Federal University of Pará, Institute of Biological Sciences, Belém, PA 66075-110, Brazil

^f Federal University of Pará, Computer Science Graduate Program, Belém, PA 66075-110, Brazil

ARTICLE INFO

Article history:

Received 24 February 2018

Received in revised form

12 April 2018

Accepted 19 April 2018

ABSTRACT

Herein, we present the draft genome of *Trametes villosa* isolate CCMB561, a wood-decaying Basidiomycota commonly found in tropical semiarid climate. The genome assembly was 57.98 Mb in size with an L50 of 691. A total of 16,711 putative protein-encoding genes was predicted, including 590 genes coding for carbohydrate-active enzymes (CAZy), directly involved in the decomposition of lignocellulosic materials. This is the first genome of this species of high interest in bioenergy research. The draft genome of *Trametes*

* Corresponding author at: Federal University of Minas Gerais, Institute of Biological Sciences, Belo Horizonte - MG, 31270-901, Brazil.

E-mail address: arigoesneto@icb.ufmg.br (A. Góes-Neto).

<https://doi.org/10.1016/j.dib.2018.04.074>

2352-3409/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

villosa isolate CCMB561 will provide an important resource for future investigations in biofuel production, bioremediation and other green technologies.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Biology
More specific subject area	Mycology, Genomics, Biotechnology
Type of data	Genomic sequence, gene prediction and annotation of <i>Trametes villosa</i> isolate CCMB561
How data was acquired	The whole genome was sequenced with an Illumina Hi-Seq. 2500
Data format	Draft genome assembly and gene annotation
Experimental factors	The mycelium derived from field-collected basidiomata was cultured on potato dextrose agar (PDA) medium at room temperature and DNA was extracted with a FastDNA™ Soil kit (MPBio)
Experimental features	The genome was assembled with SPAdes version 3.11.1 and annotated with MAKER version 2.31.9
Data source location	Basidiomata of <i>T. villosa</i> were collected in decaying wood (fallen branch) of an unidentified angiosperm in the Brazilian semiarid region (Serra das Candeias, Quijingue, Bahia, Brazil; Lat: 39°04'30"W and Long: 10°55'16"S).
Data accessibility	This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number PUDQ00000000 (https://www.ncbi.nlm.nih.gov/nucleotide/PUDQ00000000). The short reads have been deposited at SRA under the accession number SRR6763787_ (https://www.ncbi.nlm.nih.gov/sra/?term=SRR6763787).

Value of the data

- It is the first draft genome of *Trametes villosa*, a tropical white-rot Basidiomycota from the semiarid region of Brazil, promising for its production of ligninolytic enzymes.
- *T. villosa* isolate CCMB561 is a good producer of lignin peroxidase, manganese peroxidase and laccase, enzymes considered crucial for lignin degradation, providing a major advantage for its use in bioenergy research.
- The draft genome will accelerate functional genomics research, helping to understand the molecular basis of lignin decay by this fungus as well as advancing its enzymatic applications.

1. Data

The genus *Trametes* Fr. (Polyporaceae, Basidiomycota) comprises 20 species usually growing on decaying wood of angiosperms [1]. *Trametes* is morphologically characterized by its pileate basidiomata with a trimitic hyphal system and non-amyloid, non-dextrinoid and thin-walled spores, without hymenial cystidia [2].

Table 1Comparison of the genomic features of *Trametes villosa* isolate CCMB561 with *Trametes hirsuta* strain 072 [13].

Organism	DB accession number	Isolation source	Contigs/Scaffolds	Genome size (Mb)	G + C (%)	CDSs/ORFs
<i>Trametes villosa</i> CCMB561	PUDQ000000000	Decaying Wood	10,323	57.98	59.34	16,711
<i>Trametes hirsuta</i> 072	LIYB000000000	Soil	141	33.62	57.6	14,598

Trametes villosa (Sw.) Kreisel is a common species in the Brazilian semiarid region [3]. It is a good producer of the three important ligninolytic enzymes: Laccase (Lac) [4], Manganese Peroxidase (MnP) [5] and Lignin Peroxidase (LiP) [6], demonstrating its high potential for biotechnological applications. However, little is known about the function and structure of *T. villosa* genes, which requires detailed investigation.

White-rot basidiomycotan fungi are the main producers of ligninases that substantially contribute to lignin decay of wood [7,8]. Nowadays, ligninolytic enzymes of white-rot fungi have been broadly studied for their potential applications in a wide range of industrial bioprocesses such as decolorization of industrial dyes, the pulp bleaching of paper, textile industry and the degradation of organopollutants [9]. Furthermore, *T. villosa* simultaneously produces LiP, MnP and Lac [5,6,10] whereas other lignin decay fungi produce only one or two of these ligninolytic enzymes simultaneously [11,12]. Thus, a species able to produce the three ligninolytic enzymes in the same bath culture is highly desirable for biotechnological applications [6].

In order to accelerate the studies on functional genomics and elucidate molecular processes of lignin decay in this species, the genome of *T. villosa* CCMB561 was sequenced and assembled. Sequencing was performed using the paired-end method with the Illumina HiSeq. 2500, which generated 25,034,256 reads with a mean read length of 151 bp and a total of 7.5 Gbp of data. The resulting genome assembly of *T. villosa* CCMB561 contained 57.98 Mb, which was larger than the 33.6 Mb genome of *Trametes hirsuta* [13], which is the phylogenetically closest species with available complete genome based on a five-marker dataset [14]. According to QUAST version 4.4 [15], the assembled draft genome of *T. villosa* CCMB561 consisted of 10,323 contigs (6161 longer than 1 kbp), with N50 of 16.5 kb and L50 of 691, while the largest contig spanned 647,839 bp, and the GC content of the genome was predicted as 59.34% (Table 1). A total of 16,711 genes were predicted, encoding proteins with an average length of 496 amino acids. The CAZymes analysis identified 590 these genes encoding carbohydrate-active enzymes (CAZymes), which included 237 glycoside hydrolases (GHs), 78 glycosyltransferases (GTs), 12 polysaccharide lyases (PLs), 69 carbohydrate esterases (CEs) and 112 auxiliary activities (AAs). Additionally, the genome of *T. villosa* CCMB561 was predicted to contain 820 proteins with oxidoreductase activity and 45 with peroxidase activity.

Although *T. villosa* CCMB561 commonly exists as an efficient wood decomposer, with a lignocellulolytic enzyme system mainly comprising laccases, lignin peroxidases and Mn-dependent peroxidases as well as a series of CAZymes [5,6,16], there are limited data about their synthesis, genetic coding and regulation. In the context of basic research, the genome sequencing of *T. villosa* CCMB561 presented herein will enrich the portfolio of potential genes, enzymes and pathways involved in the lignin degradation processes of the white-rot fungi. Additionally, in an applied context, the draft genome assembly of *T. villosa* CCMB561 will facilitate the development of ligninolytic enzyme production for biotechnological applications. Altogether, these efforts will ultimately provide important management tools to be used in industry, especially for lignocellulosic waste management.

2. Experimental design, materials and methods

2.1. Genomic DNA extraction and sequencing

The mycelium of *T. villosa* was grown on PDA medium for 5–7 days, at room temperature and after covering the superficial area of a 9-mm diam. Petri dish, it was scrapped. Genomic DNA was extracted

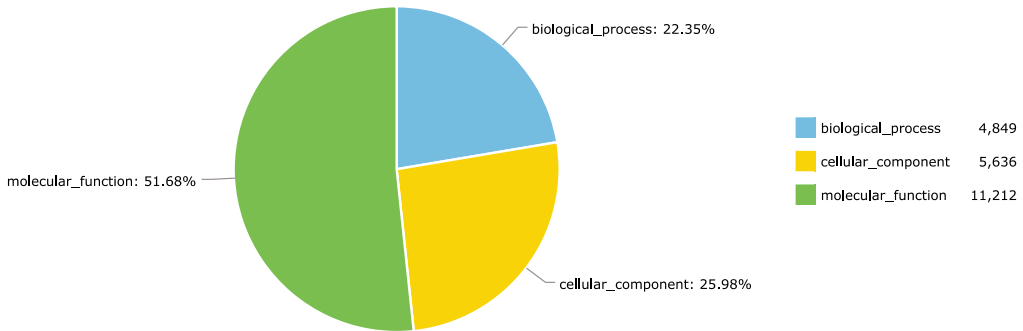


Fig. 1. Groups of ontology.

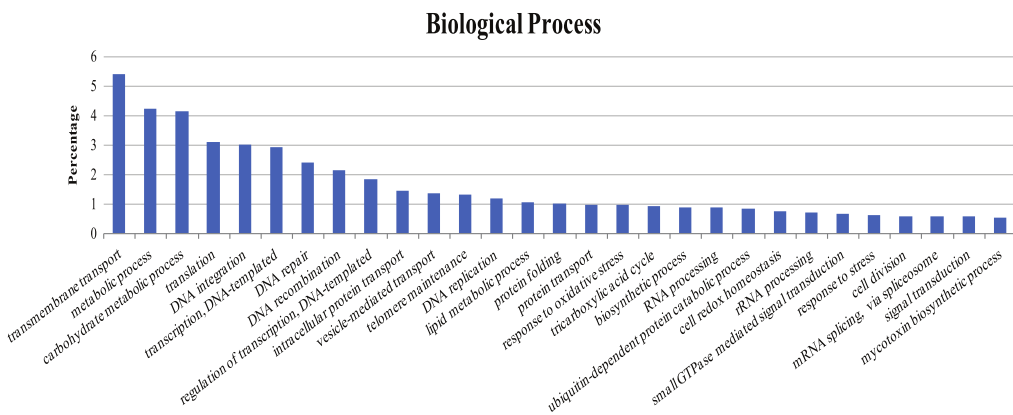


Fig. 2. The histogram shows the percentage of GOs in biological process group.

with a FastDNA™ Soil kit (MPBio). The quality and quantity of the genomic DNA were assessed by agarose gel electrophoresis and fluorometric analysis, respectively. A 450 bp library was prepared from genomic DNA with the NEBNext Fast DNA Fragmentation and Library Preparation Kit (New England Biolabs, Ipswich, NE, USA) following the manufacturer's instructions. Library quality was evaluated with Agilent 2100 Bioanalyzer. Whole genome sequencing was performed using an Illumina HiSeq. 2500.

2.2. Genome assembly and annotation

Sequence read quality was assessed using FastQC v0.11.5 [17], while low quality bases were trimmed (Phred < 20) and overlapping sequences were collapsed with AdapterRemoval v2 [18]. We assembled the genome using SPAdes version 3.11.1 [19] with k -mers 49, 51, 53, 55, 57, 59, 61, 63, 65 and 67, which were estimated by KmerStream version 1.1 [20]. We identified 16,711 protein-coding genes that were predicted using MAKER2 version 2.31.9 [21], with support from *ab initio* predictors SNAP version 2006-07-28 [22] and gene prediction program. Fungal proteins, especially of the order Polyporales were also used to support gene prediction, by providing protein homology evidence. The contigs predicted by MAKER2 were analyzed with GoFeat [23] using the following databases: UNI-PROT [24], INTERPRO [25], PFAM [26], SEED [27], NCBI [28], EMBL [29], KEGG [30]. CAZymes were identified with dbCAN version 5.0 [31]. The GoFeat analysis classified *T. villosa* 16,711 predicted genes in three groups of ontologies (biological process, cellular component and molecular function): 22.35% in biological process, 25.98% in cellular component and 51.68% in molecular function (Fig. 1). GO terms in biological process group are molecular events related to cell functioning, in cellular

Transparency document. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2018.04.074>.

References

- [1] A. Justo, O. Miettinen, D. Floudas, B. Ortiz Santana, E. Sjökvist, D. Lindner, K. Nakasone, T. Niemelä, K.-H. Larsson, L. Ryvarden, D.S. Hibbett, A revised family-level classification of the Polyporales (Basidiomycota), *Fungal Biol.* 121 (9) (2017) 798–824. <http://dx.doi.org/10.1016/j.funbio.2017.05.010>.
- [2] A. Carlson, A. Justo, D.S. Hibbett, Species delimitation in *Trametes*: a comparison of ITS, RPB1, RPB2 and TEF1 gene phylogenies, *Mycologia* 106 (4) (2014) 735–745. <http://dx.doi.org/10.3852/13-275>.
- [3] M.A. Neves, I.G. Baseia, E.R. Drechsler-Santos, A. Goês-Neto, *Guide to the Common Fungi of the Semiarid Region of Brazil*, Tecc Editora Ltda, Florianópolis, 2013.
- [4] M.L.C. Silva, V.B. Souza, V.S. Santos, H.M. Kamida, J.R.T. Vasconcellos-Neto, A. Góes-Neto, M.G.B. Koblitz, Production of manganese peroxidase by *Trametes villosa* on inexpensive substrate and its application in the removal of lignin from agricultural wastes, *Adv. Biosci. Biotechnol.* 5 (2014) 1067–1077. <http://dx.doi.org/10.4236/abb.2014.514122>.
- [5] C. Bukh, M. Lund, M.J. Bjerrum, Kinetic studies on the reaction between *Trametes villosa* laccase and dioxygen, *J. Inorg. Biochem.* 100 (2006) 1547–1557. <http://dx.doi.org/10.1016/j.jinorgbio.2006.05.007>.
- [6] R.T. Carneiro, M.A. Lopes, M.L. Silva, V.D. Santos, V.B. Souza, A.O. Sousa, C.P. Pirovani, M.G. Koblitz, R.G. Benevides, A. Góes-Neto, *Trametes villosa* lignin peroxidase (TvLiP): genetic and molecular characterization, *J. Microbiol. Biotechnol.* 27 (1) (2017) 179–188. <http://dx.doi.org/10.4014/jmb.1606.06055>.
- [7] M. Dashtban, H. Schraft, T.A. Syed, W. Qin, Fungal biodegradation and enzymatic modification of lignin, *Int. J. Biochem. Mol. Biol.* 1 (2019) 36–50 (<http://europepmc.org/articles/pmc3180040>).
- [8] D. Floudas, M. Binder, R. Riley, K. Barry, R.A. Blanchette, B. Henrissat, A.T. Martínez, R. Otillar, J.W. Spatafora, J.S. Yadav, A. Aerts, I. Benoit, A. Boyd, A. Carlson, A. Copeland, P.M. Coutinho, R.P. deVries, P. Ferreira, K. Findley, B. Foster, J. Gaskell, D. Glotzer, P. Görecki, J. Heitman, C. Hesse, C. Hori, K. Igarashi, J.A. Jurgens, N. Kallen, P. Kersten, A. Kohler, U. Kues, T.K. A. Kumar, A. Kuo, K. LaButti, L.F. Larrondo, E. Lindquist, A. Ling, V. Lombard, S. Lucas, T. Lundell, R. Martin, D.J. McLaughlin, I. Morgenstern, E. Morin, C. Murat, L.G. Nagy, M. Nolan, R.A. Ohm, A. Patyshakuliyeva, A. Rokas, F.J. Ruiz-Dueñas, G. Sabat, A. Salamov, M. Samejima, J. Schmutz, J.C. Slot, F. St. John, J. Stenlid, H. Sun, S. Sun, K. Syed, A. Tsang, A. Wiebenga, D. Young, A. Pisabarro, D.C. Eastwood, F. Martin, D. Cullen, I.V. Grigoriev, D.S. Hibbett, The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes, *Science* 336 (2012) 1715–1719. <http://dx.doi.org/10.1126/science.1221748>.
- [9] G. Janusz, K.H. Kucharzyk, A. Pawlik, M. Staszczak, A.J. Paszczynski, Fungal laccase, manganese peroxidase and lignin peroxidase: gene expression and regulation, *Enzym. Microb. Technol.* 52 (2013) 1–12. <http://dx.doi.org/10.1016/j.enzmictec.2012.10.003>.
- [10] M.J. Dinis, R.M. Bezerra, F. Nunes, A.A. Dias, C.V. Guedes, L.M. Ferreira, J.W. Cone, G.S.M. Marques, A.R.N. Barros, M.A. M. Rodrigues, Modification of wheat straw lignin a bysolid state fermentation with white-rot fungi, *Bioresour. Technol.* 100 (2009) 4829–4835. <http://dx.doi.org/10.1016/j.biortech.2009.04.036>.
- [11] V. Elisashvili, E. Kachlishvili, M. Penninckx, Effect of growth substrate, method of fermentation, and nitrogen source on lignocellulose-degrading enzymes production by white-rot basidiomycetes, *J. Ind. Microbiol. Biotechnol.* 35 (2008) 1531–1538. <http://dx.doi.org/10.1007/s10295-008-0454-2>.
- [12] R.D. Koyani, K.S. Rajput, Solid state fermentation: comprehensive tool for utilization of lignocellulosic through biotechnology, *J. Bioprocess. Biotech.* 5 (2) (2015). <http://dx.doi.org/10.4172/2155-9821.1000258>.
- [13] A.R. Pavlov, T.V. Tyazhelova, K.V. Moiseenko, D.V. Vasina, O.V. Mosunova, T.V. Fedorova, L.G. Maloshenok, E.O. Landesman, S.A. Bruskin, N.V. Psurtseva, A.I. Slesarev, S.A. Kozyavkin, O.V. Koroleva, Draft genome sequence of the fungus *Trametes hirsuta* 072, *Genome Announc.* 19 (3(6)) (2015). <http://dx.doi.org/10.1128/genomeA.01287-15>.
- [14] A. Justo, D.S. Hibbett, Phylogenetic classification of *Trametes* (Basidiomycota, Polyporales) based on a five-marker dataset, *Taxon* 60 (6) (2011) 1567–1583 (1).
- [15] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, *Bioinformatics* 29 (2013) 1072–1075. <http://dx.doi.org/10.1093/bioinformatics/btt086>.
- [16] A. Knežević, M. Ivan, S. Mirjana, V. Jelena, Potential of *Trametes* species to degrade lignin, *Int. Biodeterior. Biodegrad.* 85 (2013) 52–56. <http://dx.doi.org/10.1016/j.ibiod.2013.06.017>.
- [17] Babraham Bioinformatics, FastQC: A Quality Control Tool for High Throughput Sequence Data, Babraham Institute, Cambridge, UK, 2011.
- [18] M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification and read merging, *BMC Res. Notes* 9 (2016) 88. <http://dx.doi.org/10.1186/s13104-016-1900-2>.
- [19] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A. D. Pribelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (2012) 455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
- [20] P. Melsted, B.V. Halldórsson, KmerStream: streaming algorithms for k-mer abundance estimation, *Bioinformatics* 30 (2014) 3541–3547. <http://dx.doi.org/10.1093/bioinformatics/btu713>.
- [21] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects, *BMC Bioinform.* 12.1 (2012) 491. <http://dx.doi.org/10.1186/1471-2105-12-491>.

- [22] J. Leskovec, R. Sosič, SNAP: a general-purpose network analysis and graph-mining library, *ACM Trans. Intell. Syst. Technol.* (2016), <http://dx.doi.org/10.1145/2898361>.
- [23] F.A. Araujo, D. Barh, A. Silva, L. Guimarães, R.T.J. Ramos, GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data, *Sci. Rep.* (2018), <http://dx.doi.org/10.1038/s41598-018-20211-9>.
- [24] UniProt Consortium, UniProt: the universal protein knowledgebase, *Nucleic Acids Res.* 45 (D1) (2016) D158–D169. <http://dx.doi.org/10.1093/nar/gkw1099>.
- [25] R.D. Finn, T.K. Attwood, P.C. Babbitt, A. Bateman, P. Bork, A.J. Bridge, H.-Y. Chang, Z. Dosztányi, S. El-Gebali, M. Fraser, J. Gough, D. Haft, G.L. Holliday, H. Huang, X. Huang, I. Letunic, R. Lopez, S. Lu, A. Marchler-Bauer, H. Mi, J. Mistry, D. A. Natale, M. Necci, G. Nuka, C.A. Orengo, Y. Park, S. Pesseat, D. Piovesan, S.C. Potter, N.D. Rawlings, N. Redaschi, L. Richardson, C. Rivoire, A. Sangrador-Vegas, C. Sigrist, I. Sillitoe, B. Smithers, S. Squizzato, G. Sutton, N. Thanki, P. D. Thomas, S.C.E. Tosatto, C.H. Wu, I. Xenarios, L.-S. Yeh, S.-Y. Young, A.L. Mitchell, InterPro in 2017 – beyond protein family and domain annotations, *Nucleic Acids Res.* (2017), <http://dx.doi.org/10.1093/nar/gkw1107>.
- [26] R.D. Finn, P. Coghill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: towards a more sustainable future, *Nucleic Acids Res.* 44 (2016) D279–D285. <http://dx.doi.org/10.1093/nar/gkv1344>.
- [27] R. Overbeek, D. Terry, R. Stevens, The SEED: a peer-to-peer environment for genome annotation, *Commun. ACM* 47 (11) (2004) 46–51. <http://dx.doi.org/10.1145/1029496.1029525>.
- [28] NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information, *Nucleic Acids Res.* 44 (2016) D7. <http://dx.doi.org/10.1093/nar/gkv1290>.
- [29] R. Lopez, A. Cowley, W. Li, H. McWilliam, Using embl-ebi services via web interface and programmatically via web services, *Curr. Protoc. Bioinform.* (2014), <http://dx.doi.org/10.1002/0471250953.bi0312s48>.
- [30] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, *Nucleic Acids Res.* 44 (2016) D457–D462. <http://dx.doi.org/10.1093/nar/gkv1070>.
- [31] Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao, Y. Xu, dbCAN: a web resource for automated carbohydrate-active enzyme annotation, *Nucleic Acids Res.* 40 (2012) W445–W451. <http://dx.doi.org/10.1093/nar/gks479>.