An updated metabolic view of the *Bacillus subtilis* 168 genome

Eugeni Belda,1,2,3 Agnieszka Sekowska,4 François Le Fèvre,1,2,3 Anne Morgat,5 Damien Mornico,1,2,3 Christos Ouzounis,6,7 David Vallenet,1,2,3 Claudine Médigue1,2,3 and Antoine Danchin4,8

Correspondence
Eugeni Belda
ebelda@genoscope.cns.fr
Antoine Danchin
antoine.danchin@normalesup.org

1CEA, Institut de Génomique, Génoscale Laboratoire d’Analyse Bioinformatique en Génomique et Métabolisme, 2 rue Gaston Crémieux, 91087 Evry, France
2CNRS-UMR 8030, 2 rue Gaston Crémieux, 91057 Evry, France
3UEVE, Université d’Evry, boulevard François Mitterrand, 91025 Evry, France
4AMAbiotics SAS, Bldg G1, 2 rue Gaston Crémieux, 91000 Evry, France
5Swiss Institute of Bioinformatics, CMU, 1 Michel-Servet, CH-1211 Genève 4, Switzerland
6Institute of Applied Biosciences, Centre for Research and Technology Hellas (CERTH), Thessaloniki, Greece
7Donnelly Centre for Cellular and Biomolecular Research University of Toronto, Toronto, ON, Canada
8Department of Biochemistry, Li KaShing Faculty of Medicine, The University of Hong Kong, 21, Sassoon Road, Hong Kong SAR, China

Continuous updating of the genome sequence of *Bacillus subtilis*, the model of the Firmicutes, is a basic requirement needed by the biology community. In this work new genomic objects have been included (toxin/antitoxin genes and small RNA genes) and the metabolic network has been entirely updated. The curated view of the validated metabolic pathways present in the organism as of 2012 shows several significant differences from pathways present in the other bacterial reference, *Escherichia coli*: variants in synthesis of cofactors (thiamine, biotin, bacillithiol), amino acids (lysine, methionine), branched-chain fatty acids, tRNA modification and RNA degradation. In this new version, gene products that are enzymes or transporters are explicitly linked to the biochemical reactions of the RHEA reaction resource (http://www.ebi.ac.uk/rhea/), while novel compound entries have been created in the database Chemical Entities of Biological Interest (http://www.ebi.ac.uk/chebi/). The newly annotated sequence is deposited at the International Nucleotide Sequence Data Collaboration with accession number AL009126.4.

INTRODUCTION

Thousands of genome sequences have now been published. New ones appear daily. It is no longer possible to annotate manually even the restricted set of the major features of the genomic objects they encode. Yet, it is essential that the work pursued by a great many groups worldwide is continuously imported into ongoing genome annotations (Ouzounis & Karp, 2002), keeping them as accurate as possible. During this process proper validation is essential as percolation of annotations [in particular wrong annotations (Gilks et al., 2002)] is unavoidable, affecting the way investigators build up their experiments in the future. This need implies that annotation must be automatically transferred from reliable sources. Model organisms play a central role in this process. *Bacillus subtilis* is the only model organism for the Firmicutes clade that has been continuously annotated in depth. Its genome has been resequenced and reannotated by Barbe et al. (2009), and it is now time to deliver to the international community an updated annotation that can be used both for specific work on this organism and for making inferences not only for

**Abbreviations:** BBH, bidirectional best hits; CanOE, candidate genes for orphan enzymes; CDS, predicted protein-coding sequence; ChEBI, chemical entities of biological interest; EC, Enzyme Commission; GR, gene–reaction; INSDC, International Nucleotide Sequence Database Collaboration; PGDB, Pathway Genome DataBase; PkGDB, Prokaryotic Genome DataBase.

The GenBank/EMBL/DDBJ accession number for the annotated genome sequence of *B. subtilis* is AL009126.4.

Three supplementary tables are available with the online version of this paper.
bacteria belonging to the Firmicutes, but also for many other organisms. At the time of the last update, the *B. subtilis* genome was proposed to harbour 4244 predicted protein-coding sequences (CDSs), 48% with identified functions. Since then, a considerable number of studies have allowed investigators to complete, or at least improve, specific annotations, including identification of new genes. Sporulation was the most frequently recognized source of biological interest for this organism. Recent work showed that a specific biologically relevant feature of this organism was that a significant portion of the genome sequence is allocated to metabolic interactions with plants (Barbe et al., 2009). It is now widely recognized that *B. subtilis* is an epiphyte [it can even be found within plants (Huang et al., 2011)]. More and more *B. subtilis* functions are found to be typically relevant to plant-associated metabolism (this is also in line with the traditional use of this organism in food processing), including coping with surfaces by swelling and making and remodelling biofilms (Chen et al., 2012; Diethmaier et al., 2011; Ostrowski et al., 2011; Sekowska et al., 2009). A major interest in the knowledge derived from genome sequencing is the identification of active metabolic pathways. This is particularly important for modelling the physiology of the cell and for industrial applications. Beside plant-related carbohydrate metabolism, a significant number of new enzyme activities were discovered in *B. subtilis*, both in vivo and in vitro, and it is now time to construct a comprehensive metabolic chart of the organism. The present work collected the annotations from publications that appeared after the 2009 sequence release, as well as annotations of novel genomic objects such as small untranslated RNAs, riboswitches and small CDSs, with emphasis on metabolic discoveries that were experimentally validated. The annotated sequence is available to investigators as release AL009126.4 of the corresponding entry at the International Nucleotide Sequence Database Collaboration repositories (INSDC: DDBJ/EMBL-EBI/GenBank) (Karsch-Mizrachi et al., 2012). Additionally, the annotated genome sequence together with an updated reconstruction of the metabolic network of *B. subtilis* 168 are also available in the MicroScope platform, which is based on the Prokaryotic Genome Database (PkGDB) (https://www.genoscope.cns.fr/agc/microscope/home/index.php) (Vallenet et al., 2009).

### METHODS

#### General annotation procedures.

As described by Barbe et al. (2009) genome prediction was performed using the AMIGene software (Bocs et al., 2003), using the gene models built with the previously published version of *B. subtilis* annotations. The CDSs were assigned a locus_tag similar to that of the previous annotations, i.e. 'BSUxxxxx'. In the case of additional gene predictions, we used a label corresponding to the position of the gene in the genome sequence. Sequence data for comparative analyses were obtained from the National Center for Biotechnology Information database (ReSeq section, http://www.ncbi.nlm.nih.gov/RefSeq). Putative orthologues and synteny groups (i.e. displaying conservation of the chromosomal co-localization between pairs of orthologous genes from different genomes) were computed between the newly sequenced genome and all the other complete genomes as described by Vallenet et al. (2006). A further literature-screening procedure for previously characterized CDSs involved searching for its name (and the name of synonyms) in the PubMed and PubMed Central data libraries, as well as via Google Web. Relevant PubMed identifiers were included in the genomic object record when this information could be used to validate the status of the annotation, with a particular emphasis on experimental validations.

Newly annotated non protein-coding RNA genomic objects were given a gene name (when possible) and labelled with the locus_tag ‘BSUmisc_{RNA}_xxx’, (xxx is a digital number indicating the start position of the RNA). For example fsrA labels a regulatory RNA controlling iron-dependent metabolism (Smaldone et al., 2012) and its locus_tag is BSUmisc_RNA1483460D.

#### Automatic genome-scale metabolic network reconstruction.

The metabolic network of *B. subtilis* was generated by combining the automatic and manually curated annotations extracted from PkGDB using Pathway Tools (version 16.0), the BioCyc pathways reconstruction software (Latendresse et al., 2012). Starting with reference metabolic pathways stored in the MetaCyc database (Caspi et al., 2012), together with the set of genome annotations, the Pathway Tools built up a specific Pathway Genome Database (PGDB) in a two-step process: the Reactome projection and subsequently the Pathway projection, as described by Karp et al. (2011). Reactome is an open-source, open access, manually curated and peer-reviewed pathway database. Pathway annotations are authored by expert biologists in collaboration with Reactome editorial staff and cross-referenced to many bioinformatics databases. Metabolic pathways in MetaCyc are defined as linked sets of reaction steps that participate in a single biological process in a single organism, considered as active in a sequence without lag or delay (Green & Karp, 2006). The Reactome projection is performed by inferring associations between genes and metabolic reactions extracted from MetaCyc. To this purpose, the Enzyme Commission (EC) numbers linked to gene names are first examined, followed by the gene product names and the gene ontology terms found in the gene records of the genome. The corresponding reactions are deemed ‘matched reactions’. However, this procedure may lead to erroneous results because some records may be wrongly formatted or may be linked to unspecified EC numbers or to inaccurate or unclear textual annotations. This leads the Pathway Tools either to overpredict reactions or to miss relevant enzymatic reactions. We thus enhanced the matching procedure by directly using the regular MetaCyc reactions’s frame identifier as the functional annotation. PathoLogic is a component of the Pathway Tools suite of MetaCyc that creates a new pathway database containing the predicted metabolic pathways of an organism. Our export procedure from PkGDB directly linked the genes to manually validated MetaCyc reaction identifiers in the PathoLogic input format. Alternatively, if no MetaCyc reaction has been validated and associated to a particular gene, the procedure exports the partial EC number and/or the gene product common name.

In the second step, the Pathway Tools software was used to infer complete metabolic pathways through a pathway projection process (Karp et al., 2011) that takes as the input the initial set of catalysed reactions and the set of reference pathways available in MetaCyc. This required the development of an in-house metabolic database having a data structure identical to that of MetaCyc: this database was named MicroReCyc. It behaves as a repository for existing pathways that are absent from MetaCyc. MicroReCyc pathways are submitted on a regular basis to both the external Microme repository (http://www.microme.eu/) and the MetaCyc pathway resource (http://metacyc.org).
Bacillus subtilis metabolic pathways 2012

org/). Pathways extracted from MetaCyc and MicroRefCyc were merged together and used as the set of reference pathways to initiate the PathoLogic pathway projection process. Finally, some reactions lacking an associated gene (orphan reactions) were retrieved on the basis of their presence in an inferred pathway. The reconstructed metabolic genome scale network is included in the MicroCyc repository available at http://www.genoscope.cns.fr/agc/microcyc.

Metabolic network curation process

The process of metabolic data curation is based on the specifications of the MicroScope platform (Vallenet et al., 2009). It can be divided into three main parts: first, the manual curation of individual gene–reaction (GR) associations; second, the curation of automatically generated metabolic pathway projections and the definition of new metabolic pathways and pathway variants; third, the projection of GR associations in two additional public resources for chemical compounds [Chemical Entities of Biological Interest ontology (ChEBI) (de Matos et al., 2012)] and reactions (RHEA) (Alcántara et al., 2012). A schematic representation of the curation strategy is provided in Fig. 1.

Curation of individual GR associations. As a starting point, the annotation of the B. subtilis 168 genome stored in the MicroScope platform at the end of December 2011 [it was continuously annotated since the previous release of the B. subtilis genome sequence at the INSDC (Barbe et al., 2009)] was used as the input data of the PathoLogic algorithm, and a first set of MetaCyc reactions that can be catalysed by the enzymes encoded in the genome was automatically inferred (Karp et al., 2011). This step produced a set of automatically retrieved GR associations that was subsequently divided into two main subsets. This implied creating groups of similar genes, which we identified as similar using BLASTP best hits of the corresponding proteins. The best hit of a particular protein to all the proteins coded by a target genome is the protein in that proteome that represents a best BLASTP match. With complete genomes the procedure can be symmetrical. The match is bidirectional when the two proteins are best hits of each other. Accordingly, a bidirectional best hit (BBH) indicates a very strong similarity between two proteins. It is used as in silico evidence for orthology (Overbeek et al., 1999; Tatusov et al., 2000). The first subset contained B. subtilis CDSs sharing BBH relationships with E. coli K-12 counterparts, having the same gene name or the same complete EC number annotation, which supports a common functional role. For these CDSs, GR associations were automatically transferred from the corresponding E. coli K-12 BBH coming from curated EcoCyc data (Keseler et al., 2011). The second subset comprised the remaining B. subtilis 168 CDSs having predicted GR associations (with and without BBH relationships with E. coli K-12); this set was manually curated to avoid overassignments of biochemical reactions associated to CDSs with generic functional annotations, and to reconcile B. subtilis 168 predictions with E. coli K-12 and International Union of Biochemistry and Molecular Biology-compliant annotations of enzymatic activities.

Manual curation of GR associations was also extended to B. subtilis 168 CDSs annotated as enzymes and putative enzymes, and for which no MetaCyc reaction had been predicted by PathoLogic. The MicroScope platform gathers results from several computational methods that can assist biologists in the curation process: this includes BLASTP similarity profiles of each CDS against the UniProt resource (Dimmer et al., 2012), PRIAM (an automatic profiling software for automatic annotation of metabolism) profiles for EC number predictions (Claudel-Renard et al., 2003), protein domain profiling based on InterPro domains (Hunter et al., 2012) and protein family assignments based on clusters of orthologues (Tatusov et al., 2003), Fellowship for Interpretation of Genomes families (Meyer et al., 2009) and high-quality automated and manual annotation of protein family repositories (Lima et al., 2009). The outcome of all these methods was further improved by extensive manual literature searches to update B. subtilis 168 gene annotations and to broaden our previous knowledge of biochemical reactions that can be catalysed by enzymes encoded in the genome.

BsubCyc integration. Finally, GR associations coming from the BsubCyc database (http://bsubcyc.org/) were also integrated into MicroScope. BsubCyc is a Tier 2 BioCyc PGDB of B. subtilis 168 based on the reannotation of 2009 (Barbe et al., 2009) that contains curated data of B. subtilis 168 metabolism absent in MetaCyc repository (GR associations and metabolic pathways) together with automatic predictions based on the output of the PathoLogic algorithm (Latendresse et al., 2012). Both B. subtilis 168 PGDBs (BsubCyc and MicroScope) were compared at the level of individual GR associations, and BsubCyc-specific associations supported by experimental evidence (not coming from automatic PathoLogic predictions) were integrated in MicroScope.

Curation of automatically generated pathways projection. Manual curation was also extended to the set of automatically projected metabolic pathways generated by the PathoLogic algorithm (Latendresse et al., 2012), based on the updated genome annotation that includes automatic and manually validated GR associations (see above). MetaCyc V16.0 was used as the reference repository for the pathway projection procedure. This frequently produced incomplete pathways, i.e. pathways with holes (enzymatic reactions not linked to any gene). This situation may come from missing enzymatic activities in annotations, presence of alternative pathway variants different from those defined in MetaCyc, or non-functionality of the projected MetaCyc pathway in the organism of interest. This problem prompted us to curate manually the pathways that were incomplete, using MicroScope. Several pathway-hole filling strategies have been integrated in this platform, including CanOE (Smith et al., 2012) and PhyloProfile (Engelen et al., 2012). CanOE is software that takes advantage of the genomic and metabolic context, the similarity profiles of the B. subtilis 168 proteome against UniProt entries validated by experimental evidence of the missing enzymatic activity, to identify missing activities in comparison with other genes of the pathway. PhyloProfile analyses the co-evolution between phylogenetic profiles of candidate genes. Finally, extensive manual literature searches were performed to find experimental evidence of the missing enzymatic activities in B. subtilis 168. When a candidate gene was found, its functional annotation was updated accordingly (see Fig. 2 for an example of pathway hole filling).

In addition, when the curation process allowed identification of a novel metabolic pathway or a pathway variant not currently present in the MetaCyc resource, the Pathway Tools were used to design novel pathways and collect them in the MicroRefCyc database. In a similar manner, BsubCyc-specific pathway entities absent in MetaCyc repository and supported by experimental evidence were also integrated into the MicroRefCyc database.

Curation of chemical compounds and reactions in ChEBI and RHEA resources. The process of manual curation of GR associations and pathway projections allowed us to identify several B. subtilis enzyme activities that were not currently present in the MetaCyc reaction repository. In order to deal with these situations, the MicroScope platform allows the user to validate biochemical reactions using two different reaction resources, mainly MetaCyc (Caspi et al., 2012) and RHEA (Alcántara et al., 2012). RHEA is a European resource of expert-curated biochemical reactions that are defined based on chemical compounds from the ChEBI, and that are stoichiometrically balanced for mass and charge at pH 7.3 (Alcántara et al., 2012; de Matos et al., 2012). All small molecules of the MetaCyc reactions associated to B. subtilis 168 validated genes were automatically mapped into RHEA, using the MetaCyc-RHEA cross-references
Pathway-hole filling

Pathway projection (PathoLogic)

CanOE module
Similarity module
Phylogenetic module

Pathway validation

“Non-”/“Functional”
“Variant-needed”
“False-positive”

MicroRefCyc metabolic pathways

B. subtilis 168 PKGDB vN+1

Automatic validation

Manual curation

B. subtilis 168 PKGDB vN+2

Automatic mapping

RHEA-MetaCyc cross-reference

B. subtilis 168 PKGDB vN+3

New RHEA reactions

New ChEBI compounds

Fig. 1. Schematic representation of the curation workflow of *B. subtilis* 168 metabolic data. The curation process can be divided into three main parts. First, the curation of individual GR associations. Second, the curation of MetaCyc pathway projections and the creation of new metabolic pathways and pathway variants identified during the curation process and absent in MetaCyc pathway repository. Third, projection of MetaCyc reactions and *B. subtilis* 168 specific reactions on ChEBI and RHEA resources for compounds and reactions. Metabolic data for GR associations and metabolic pathways from BsubCyc database (http://www.bsubcyc.org/) with supporting experimental evidence were also integrated. All annotation updates, GR associations and metabolic pathway projections and reconstructions are stored in the specific *B. subtilis* entry of PKGDB in the MicroScope database environment (Vallenet et al., 2009).
MetaCyc reactions involving small molecules that do not have RHEA cross-references were also manually curated: the corresponding MetaCyc cross-references were added if the reactions were already present in RHEA, or the reactions were created de-novo if they were absent. For those cases of missing compounds in ChEBI with correct 2D structure at pH 7.3, the corresponding compounds were created de novo in ChEBI. The Marvin suite of tools from ChemAxon (http://www.chemaxon.com) was used to finalize the 2D structure of these new compounds at pH 7.3 so that they are now available for reaction creation into RHEA. The same procedure was followed for specific biochemical reactions characterized during the manual curation of GR associations that are absent in the MetaCyc reaction repository.

The resulting curated B. subtilis metabolic network is available at the BacilluScope project website (https://www.genoscope.cns.fr/agc/microscope/home/index.php) and is included in the MicroCyc repository (http://www.genoscope.cns.fr/agc/microcyc).

Experimental validation of annotations. In some cases, when there was ambiguity in the function ascribed to a particular gene we performed experiments designed to validate the proposed annotation. The strain used was strain 168, resequenced in 2009 and deposited at the BacillusScope project website (https://www.genoscope.cns.fr/agc/microscope/home/index.php) for details of the algorithm, see Engelen et al. (2012). The best co-evolved genes with the candidate gene BSU30810 are two neighbour genes involved in other steps of the pathway (BSU30800 and BSU30820). The BLASTP profiles of BSU30810 against protein entries of SwissProt. The best hit corresponds to the BSU30810 protein entry in SwissProt, where it appears annotated as a putative esterase, with an incomplete EC number association. The next best hits show significant similarities with the orphan activity EC 4.2.99.20 in initial pathway projections, confirming the initial predictions of CANOE.
the DSMZ under accession number DSM 23778 (Anagnostopoulos & Spizizen, 1961; Barbe et al., 2009). Experiments were performed by growing relevant B. subtilis strain 168 mutants in liquid or on solid media, as described previously (Barbe et al., 2009).

Cells were grown either in Luria–Bertani (LB) medium (Bertani, 1951) or in ED minimal medium: K₂HPO₄, 8 mM; KH₂PO₄, 4.4 mM; glucose, 27 mM; Na₂-citrate, 0.3 mM; L-glutamine, 15 mM; L-tryptophan, 0.244 mM; ferric citrate, 33.5 mM; MgSO₄, 2 mM; MgCl₂...0.61 mM; CaCl₂, 49.5 mM; FeCl₃, 49.9 mM; MnCl₂, 5.05 mM; ZnCl₂, 12.4 mM; CuCl₂, 2.52 mM; CoCl₂, 2.5 mM; Na₂MoO₄, 2.48 mM, with the relevant carbon, nitrogen or sulfur source when needed, omitting the counterpart (e.g. MgCl₂ replaced MgSO₄ when cells were assayed for a sulfur source).

LB and ED plates were prepared by addition of 17 g Bacto agar or Agar Noble (Difco) l⁻¹, respectively, to the medium. Bacteria were grown at 37 °C. The optical density of bacterial cultures was measured at 600 nm. For assaying growth on sulfur compounds on plates, either the MgSO₄-containing medium or the sulfur-free basal medium was used (MgSO₄ was replaced by MgCl₂ as described above). In the latter case, 10 µl of the sulfur source under investigation was applied onto paper discs (generally 100 mM stock solution) deposited on the plate, after the bacteria had been uniformly spread on the surface of the plate, and growth was measured around the disc. A similar procedure was used for assaying nitrogen sources, omitting glutamine from the medium and placing the putative nitrogen source on the disc.

RESULTS AND DISCUSSION

General features of reannotation

The B. subtilis sequence now comprises 4458 genomic objects (4422 in the preceding release); 1217 genes have been reannotated and updated references have been linked to genes (3644 PubMed references in the current annotation in comparison with 3376 references in the preceding release). A significant effort has been made in the curation of the associations between genes encoding enzymes and the biochemical reactions catalysed by these enzymes in order to provide a curated and comprehensive reconstruction of the global metabolic map of B. subtilis. In the current release, 1083 CDSs have been associated to 1097 chemical reactions (a total of 1751 GR associations). Among those, eight are now complete, whereas 14 remain incomplete. Pathways that are functional according to experimental evidence but where some reaction steps do not yet have an identified gene in B. subtilis remain incomplete in terms of gene identification but are nevertheless deemed ‘validated’. This status could be used to define new targets for experimental studies meant to identify the corresponding catalyst (see Table S1, available in the online version of this paper). The full list of the 283 MetaCyc pathways with the corresponding functional status according to MicroScope pathway curation tool is provided in Table S2.

Finally, manual curation of GR associations and automatic pathway projections resulted in the characterization of 27 new metabolic pathways or pathway variants of B. subtilis that are not represented in MetaCyc V16.0 pathway repository (Table S3). These metabolic pathways have been instantiated within the internal PGDB MicroRefCyc by using the Pathway tools software package. The complete metabolic network of B. subtilis 168 comprising manually curated and automatically validated GR associations together with full set of MetaCyc V16.0 and MicroRefCyc metabolic pathways are freely available in a specific MicroCyc PGDB in MicroScope environment at the following URL http://microcyc.genoscope.cns.fr/BACSU7/organism-summary.

Beside emphasis on metabolic pathways, discussed below, several new features have now been included in the genome annotation.

Specific genomic objects

Toxin/antitoxin genes. Association of gene products that combine a deleterious activity together with a control
inhibitor are widespread in all domains of life. A case in point is the association between restriction endonucleases and associated anti-restriction methylases. This type of association has evolved and is nowadays generally identified under the ‘toxin–antitoxin’ generic name (Hayes & Van Melderen, 2011). Many physico-chemical processes lie under this banner: the toxin can be a protein destabilizing membranes, a protease, a restriction endonuclease or, most often, an RNAse that may have a variety of targets, sometimes with broad specificity, sometimes with extremely narrow sites of action. Most of these systems, restriction endonucleases aside, belong to two classes.

Type I toxins are often short proteins, disrupting membranes. The antitoxin of type I systems is a small RNA which base pairs with the toxin mRNA to prevent protein synthesis (Fozo et al., 2010). The TxpA/RatA with its RNA was already identified in the previous release. In phage SPbeta the BsrG/RtaD (SR4) is expressed in a thermosensitive fashion and controls cell lysis (Jahn et al., 2012). We now included a collection of several related peptides that are probably type I toxins: YoyG, YosA, YjcZ, YczM, YksV, YuzJ and SscA, for which the putative downstream RNA has not yet been identified. They all contain variants of the peptide motif FALLVVLFILLIIVG. The latter, SscA, displayed a poor germination phenotype when disrupted (Kodama et al., 2011).

The antitoxin of the type II systems is a protein that binds to and inhibits the toxin protein. Several such systems have been documented in B. subtilis 168: NdoA/NdoAI modulates the effect of lethal stresses, via the action of a sequence-specific RNAse (Wu et al., 2011); SpoIISA/SpoISB, inactivation of SpoIISA has no effect on sporulation but it fully restores sporulation of a spoISB null mutant, indicating that SpoISB is required only to counteract the negative effect of SpoIISA on sporulation (Florek et al., 2011); YxiD (now RtbD) was annotated as a putative transposase, experiments have recently demonstrated that it is in fact the antitoxin of toxic RNAse of the RtbD(YxiD)/RtbE(YxiE) system. The same is true for the couples RtbG/RtbF and RtbI(YocI)/RtbJ(YocJ), with SsrSA small RNA in between (Beckmann et al., 2012) RtbL/RtbM and RtbN(YwqJ)/RtbO(YwqK) (Holberger et al., 2012).

Small untranslated RNAs and riboswitches. Several small untranslated RNAs have been identified (Irnov et al., 2010), many without known function, while some are involved in the control of iron homeostasis (Gaballa et al., 2008). SrrA (SR1) controls expression of a transcriptional activator of arginine catabolism, AhrC (Heidrich et al., 2006), while it also encodes RgpA, which regulates expression of gapA (Gimpel et al., 2010), and FsrA, which controls iron-responsive genes (sdhCAB, citB, yufW, leuCD) (Gaballa et al., 2008). Apart from the previously identified S-boxes and T-boxes, several novel riboswitches have been identified (Breaker, 2011).

Information transfer. Besides the previously identified features of macromolecule modifications, the biosynthesis of the modified nucleotide 2-methylthio-N6-threocarbamoyladenosine, a modified residue at position 37, adjacent to the 3’-end of the anticodon of several tRNAs has been characterized
The Firmicute degradosome and nanoRNases. The existence of a variety of small untranslated RNAs as well as regulatory regions in translated RNAs has placed RNA degradation in the limelight. In *E. coli* a structure named the degradosome is organized around RNase E (Carpousis, 2007). There is no structural counterpart of this RNase in *B. subtilis*, and a remarkable feature of this organism is that the way it degrades RNA appears to be completely different to that found in *E. coli*. In particular, the counterpart of the degradosome is almost entirely different, being built up around RNase J (A and B) (Deikus & Bechhofer, 2009; degradosome is almost entirely different, being built up around RNase J (A and B) (Deikus & Bechhofer, 2009; Engelen et al., 2008, 2010; Lauhon, 2012).

Another essential function of the degradosome is that of nanoRNases, degrading oligonucleotides shorter than 5 nt. Orn in *E. coli* and at least two counterparts of different descent in *B. subtilis*, NrnA and NrnB (Fang et al., 2009). Remarkably, NrnA displays 3′,5′-adenosine bisphosphate (pAp) 3′-phosphatase activity, coupling RNA degradation to sulfur and lipid metabolism (pAp is the by-product of sulfate reduction, as well as of 4-phosphopantetheine synthesis). This type of coupling between small molecules and macromolecule degradation will have to be taken into account in further developments of metabolic pathway tools and repositories. A further substantiation of this widespread category of pathways is seen with a similar double function, found for NrnC in alpha-proteobacteria (Liu et al., 2012), suggesting that this is an important metabolic feature of RNA catabolism. Another exonuclease of unknown specificity, YhaM, may also participate in the activity of the degradosome. It belongs to a cluster of co-evolving proteins comprising YmIF and YmFG, two peptidyl hydroxylases of unknown substrates, perhaps associated to transport of peptidyl siderophores and connected to YheA [related to metabolism of aromatics (YwbD, AroC, TyrA, AroF)], proteins such as YqA (a putative membrane-bound protein) and YacP (possibly involved in rRNA maturation), to a network with tRNA synthetases CygS and GltX and with sigma factor SigH (Engelen et al., 2012). Nevertheless, the lack of further characterization prevented us from changing its status of ‘Y’ (unknown) protein.

**Metabolic pathways**

The present update of the *B. subtilis* 168 genome annotation provides us with a fairly comprehensive view of its metabolism. We highlight below some of the novel features that have been introduced in the present updated annotation of the genome sequence, with emphasis on metabolism.

**Carbon metabolism.** The pathway of myo-inositol degradation to the glycolytic intermediate dihydroxycetone phosphate has been updated, including the activities of IolD and IolB, catalysing the hydrolysis of the pathway intermediate 3β-(3,5/4)-trihydroxycyclohexane-1,2-dione to yield 5-deoxy-D-glucuronic acid and the subsequent isomerization to 2-deoxy-5-keto-D-gluconic acid (Yoshida et al., 2008). The activity of IolL, the inosine isomerase that catalyses the isomerization of 2-keto-myoinositol intermediate to 1-keto-D-chiro-inositol, which is subsequently transformed to D-chiro-D-inositol by the inositol dehydrogenase IolG, has also been included (Yoshida et al., 2006). These activities have been experimentally validated by the inactivation of IolE, leading to overproduction of D-chiro-inositol, a possibly interesting metabolite as a drug candidate as it improves the efficiency with which the body uses insulin and also promotes ovulation, being a possible candidate for the treatment of type 2 diabetes and polycystic ovary syndrome (Yoshida et al., 2006).

The fermentative metabolism of *B. subtilis* 168 has been further documented based on the experimental evidence reviewed by Nakano & Zuber (1998) and Nakano et al. (1997). In the absence of external electron acceptors, *B. subtilis* 168 still grows anaerobically by fermentation, provided glucose and pyruvate are supplied, with ethanol, lactate, acetate and 2,3-butanediol as major fermentative products detected by nuclear magnetic resonance and HPLC experiments on anaerobic cell cultures with glucose products detected by nuclear magnetic resonance and HPLC experiments on anaerobic cell cultures with glucose and pyruvate as carbon sources and ammonia as a nitrogen source (Nakano & Zuber, 1998; Nakano et al., 1997). Production of lactate and 2,3-butanediol continues even in the presence of the anaerobic electron acceptors nitrate and nitrite, pointing out the importance of these fermentative processes for NADH reoxidation even for anaerobic respiratory growth (Nakano & Zuber, 1998).

The annotation of the gene cluster yngJHGFIE has also been updated to reflect its functional role, encoding enzymes responsible for the degradation of L-leucine to acetyl-CoA. Experimental inactivation of each gene and analysis of the corresponding cell cultures in medium with 13C-labelled L-leucine confirm that the pathway is functional during sporulation in *B. subtilis* 168 and that individual knockouts of this gene cluster block L-leucine degradation (Hsiao et al., 2010). The pathway might also...
be used in isoleucine and valine degradation during sporulation.

The substrate specificity of carbohydrate-degrading enzymes has to be carefully established. Beta-galactosidase GanA from *B. subtilis* is very different from LacZ of *E. coli*. It is an arabinogalactan hydrolase (Shipkowski & Brenchley, 2006), which is restricted in its substrates. For example, in contrast with LacZ its presence does not allow growth on phenyl-galactoside, showing also that its paralogue RhgZ (a possible galacturonisase) does not confer this ability either (Table 1). D-Rhamnose is a fairly rare component of the surface of many pathogens. Because *B. subtilis* has been shown to counteract many actions of pathogens on plants (Choudhary & Johri, 2009), it was interesting to see whether it could use this carbon source. Contrary to the situation with L-rhamnose (involving *rhaA*, *rhaB*, *rhaU*), *B. subtilis* does not grow on D-rhamnose (Table 1).

### Energy metabolism.

Three different metabolic pathways (MicorefCyc: PWYIPS-66, 67 and 95) have been created, reflecting the two main branches of aerobic respiration existing in *B. subtilis* 168, the quinol oxidase branch and cytochrome oxidase branch, according to the presence of three different types of terminal oxidases (Winstedt & von Wachenfeldt, 2000). In the quinol oxidase branch, menaquinol is directly reoxidized to menaquinone using molecular oxygen as an electron acceptor by two different types of quinol oxidases: the cytochrome aa3 oxidase complex encoded by the *qoxABCD* gene cluster (MicorefCyc PWYIPS-66), which belongs to the haem-copper superfamily of respiratory oxidases, coupling menaquinol reoxidation with proton translocation across the membrane (Santana *et al.*, 1992); and two different cytochrome bd oxidases encoded by the *cydABCD* or *ythAB* gene clusters (MicorefCyc PWYIPS-67), which do not pump protons (Winstedt *et al.*, 1998). CydABCD has been shown to be a cytochrome *bb* oxidase rather than a cytochrome bd oxidase with haem D being substituted by high spin haem B at the oxygen reactive site, i.e.

### Table 1. Growth properties of *Bacillus subtilis* 168

Growth is measured on minimal medium plates spread with a liquid culture of *B. subtilis* (see Methods) onto which a 6 mm paper disk is deposited, containing 10 μl of a 100 mM solution of the assayed compound. In the case of serine a complete inhibition zone around the disk is then followed by a zone of growth. Growth was also monitored in liquid medium supplemented with 10 mM (final concentration) of the compound of interest (except for arginine, where 5 mM was used as arginine acts as a chaotropic agent, leading to *B. subtilis* lysis). +, Growth; +/−, weak growth; −, no growth; NA, not applicable.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Carbon</th>
<th>Nitrogen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Arginine*</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Asparagine</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>+/−</td>
<td>+</td>
</tr>
<tr>
<td>Cysteine</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>+/−</td>
<td>+</td>
</tr>
<tr>
<td>Glutamine</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Glycine</td>
<td>−</td>
<td>+/−</td>
</tr>
<tr>
<td>Histidine</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Leucine</td>
<td>+/−</td>
<td>+</td>
</tr>
<tr>
<td>Lysine</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Methionine</td>
<td>+/−</td>
<td>−</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>+/−</td>
<td>−</td>
</tr>
<tr>
<td>Proline</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Serine</td>
<td>−</td>
<td>+†</td>
</tr>
<tr>
<td>Threonine</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Valine</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Beta-alanine</td>
<td>−‡</td>
<td>−‡</td>
</tr>
<tr>
<td>1,5-Anhydro-D-sorbitol</td>
<td>−‡</td>
<td>NA</td>
</tr>
<tr>
<td>Glutarate</td>
<td>−‡</td>
<td>NA</td>
</tr>
<tr>
<td>Phenylgalactoside</td>
<td>−‡</td>
<td>NA</td>
</tr>
<tr>
<td>D-Rhamnose</td>
<td>−‡</td>
<td>NA</td>
</tr>
<tr>
<td>S-Methyl-cysteine</td>
<td>−‡</td>
<td>+‡</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>Sulfur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetylmethionine</td>
<td>+</td>
</tr>
<tr>
<td>Benzensulfonic acid</td>
<td>−</td>
</tr>
<tr>
<td>Cholesterol sulfate</td>
<td>−</td>
</tr>
<tr>
<td>Coenzyme M</td>
<td>+</td>
</tr>
<tr>
<td>Dithiobutyric acid</td>
<td>−</td>
</tr>
<tr>
<td>Dithiodipropionic acid</td>
<td>−</td>
</tr>
<tr>
<td>Formyl-methionine</td>
<td>+</td>
</tr>
<tr>
<td>Gluthathione</td>
<td>+/−</td>
</tr>
<tr>
<td>Isethionate</td>
<td>+</td>
</tr>
<tr>
<td>Methanesulfonate</td>
<td>+</td>
</tr>
<tr>
<td>Methanethiol</td>
<td>−</td>
</tr>
<tr>
<td>Orthanilic acid</td>
<td>−</td>
</tr>
<tr>
<td>Propane sulfonate</td>
<td>+</td>
</tr>
<tr>
<td>Sinigrine</td>
<td>−</td>
</tr>
<tr>
<td>S-Methylmethionine</td>
<td>−</td>
</tr>
<tr>
<td>S-Methylcysteine</td>
<td>+</td>
</tr>
<tr>
<td>S-Methyl-galactoside (TMG)</td>
<td>−</td>
</tr>
<tr>
<td>Sulfanilamide</td>
<td>−</td>
</tr>
</tbody>
</table>

### Table 1. cont.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Sulfur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulfobenzoic acid</td>
<td>−</td>
</tr>
<tr>
<td>Sulfur (S⁰)</td>
<td>+</td>
</tr>
<tr>
<td>Taurine</td>
<td>+/−</td>
</tr>
<tr>
<td>Thialysine</td>
<td>+/−</td>
</tr>
<tr>
<td>Thiaproline</td>
<td>+</td>
</tr>
<tr>
<td>Thio-dipropionic acid</td>
<td>−</td>
</tr>
<tr>
<td>Thiophene</td>
<td>−</td>
</tr>
<tr>
<td>Toluenesulfonic acid</td>
<td>−</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>Sulfur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulfobenzoic acid</td>
<td>−</td>
</tr>
<tr>
<td>Sulfur (S⁰)</td>
<td>+</td>
</tr>
<tr>
<td>Taurine</td>
<td>+/−</td>
</tr>
<tr>
<td>Thialysine</td>
<td>+/−</td>
</tr>
<tr>
<td>Thiaproline</td>
<td>+</td>
</tr>
<tr>
<td>Thio-dipropionic acid</td>
<td>−</td>
</tr>
<tr>
<td>Thiophene</td>
<td>−</td>
</tr>
<tr>
<td>Toluenesulfonic acid</td>
<td>−</td>
</tr>
</tbody>
</table>

*5 mM in liquid conditions.
†Inhibition of 1 cm followed by growth of 0.5 cm.
‡Not assayed in liquid medium.
cytochrome $b(558)b(595)b'$ (Azarkina et al., 1999). The exact role of YthAB is unknown as neither deletion nor overproduction alter the expression of the $b'$ oxidase (Azarkina et al., 1999). In the cytochrome oxidase branch (MicrorefCyc PWYIPS-71), a cytochrome $bc_1$ complex encoded by genes $gcrCAB$ catalyses the reoxidation of menaquinol intermediates by electron transfer to a $C$-type cytochrome, which is subsequently reoxidized by a cytochrome $caaa_3$ complex encoded by the genes $ctaCDEF$, reducing molecular oxygen to water (Yu et al., 1995).

**Anabolism of small molecules.** A new biotin biosynthesis pathway variant has been defined, including the specific activity of BioA and Biol enzymes in *B. subtilis* 168. BioA is lysine-8-amino-7-oxononanoate aminotransferase that uses specifically L-lysine as an amino group donor instead of $S$-adenosylmethionine, which is commonly used by BioA homologues in other bacterial species (Van Arsdell et al., 2005), whereas Biol encodes a cytochrome $p450$ protein that catalyses the two-step oxidative cleavage of long-chain acyl-ACP and long-chain fatty acids to yield pimeloyl-ACP and pimelate, respectively, which are the starting substrates of the pathway (Green et al., 2001; Stok & De Voss, 2000). This activity allows the connection of fatty acid metabolism with biotin biosynthesis, accounting for the unexplained auxotrophy for biotin observed in currently available genome-scale metabolic models of *B. subtilis* 168 (Henry et al., 2009).

The lipoate biosynthesis pathway of *B. subtilis* 168 also proceeds in a slightly different way when compared with that of *E. coli* K-12, with the presence of octanoyl-[(GcvH):protein N-octanoyltransferase LipL, which catalyses the two-step transfer of octanoyl moiety from octanoyl-GcvH complex to the E2 subunit of pyruvate dehydrogenase (Christensen et al., 2011).

The pathway of branched-chain fatty acid biosynthesis has also been updated. This includes the initial two-step degradation of branched-chain amino acids ($\alpha$-valine, $\alpha$-isoleucine and $\alpha$-leucine) to the corresponding 2-ketoacyl-CoA intermediates (isobutyryl-CoA, 2-methylbutanoyl-CoA and isovaleryl-CoA, respectively) by the combined action of the branched-chain amino acid dehydrogenase Bcd and the branched-chain keto-acid dehydrogenase enzyme complex (BkdAA + BkdAB + BkdBB). This degradation is coupled with the successive elongation cycles by using malonyl-ACP as an extender substrate carried out the universal dissociated type II fatty acid synthase system encoded by $\text{fab}$ genes (Choi et al., 2000; Debarbouille et al., 1999; Kaneda, 1991; Oku & Kaneda, 1988). These elongation cycles are equivalent to those involved in the classical pathway of straight-chain fatty acid biosynthesis present in *E. coli* K-12 (with acetyl-CoA as a primer substrate).

In the same way, the metabolism of methionine has been updated with a novel pathway representation (MicrorefCyc RXNIPS-71). This covers the reverse transulfuration pathway from methionine to cysteine existing in *B. subtilis* 168 that allows the bacterium to grow with methionine as a sulphur source, and that includes the specific cystathionine $\beta$-synthase MccA that uses O-acetyl-serine instead of serine as the substrate to produce cystathionine from homocysteine, and MccB, which encodes an enzyme with cystathionine $\gamma$-lyase, homocysteine $\gamma$-lyase and cysteine desulphydrase activities (Hullo et al., 2007).

It is known that *B. subtilis* does not synthesize glutathione, and it took a long time to identify the molecule that plays the role of this reduced-sulfur-buffering molecule. It has now been established that *B. subtilis* synthesizes bacillithiol, the alpha-anomeric glycoside of L-cysteinyl-D-glucosamine with L-malic acid. The whole pathway (encoded by the genes $bshBAbshA$, $bshBB$ and $bshC$) is now present and documented in the genome sequence annotations (Gaballa et al., 2010; Newton et al., 2009, 2011; Upton et al., 2012).

It has been recently reported that *B. subtilis* 168 is able to synthesize *in vivo* an archaeal-type ether lipid by the action of glycerol-1-phosphate (G1P) dehydrogenase AraM, which catalyses the NADH-dependent reduction of dihydroxyacetone phosphate to G1P, followed by the condensation with the heptaprenyl moiety of heptaprenyl diphosphate catalysed by heptaprenylglyceryl-phosphate synthase PcrB (Guldan et al., 2008, 2011). This represents the first experimental evidence of the presence of an archaeal-type ether lipid derived from G1P in bacteria, a remarkable feature from an evolutionary point of view, since one of the main differences between Bacteria and Archaea lies in the chemical composition of their membrane phospholipids as well as the chirality of the glycerol–phosphate head. Bacteria and Eukarya have ester-linked fatty acid phospholipids based on glycerol-3-phosphate (G3P), whereas Archaea have ether-linked isoprenoid phospholipids based on G1P (Lombard et al., 2012). Phylogenomic approaches have revealed the presence of G1P and G3P dehydrogenases in the three domains of life, strongly supporting the view of the ubiquitous presence of phospholipid membranes in the common ancestor population of Bacteria and Archaea despite their different composition (Peretò et al., 2004). Various analyses of metabolic pathways have long suggested a significant proximity between Archaea and Firmicutes (Bright et al., 1993; Jensen et al., 1988; Ruepp et al., 1995; Sekowska et al., 2000) and, although horizontal gene transfer events cannot be excluded, the presence of this archaeal-type ether lipid biosynthesis pathway in *B. subtilis* 168 substantiates the idea of a very-early presence of phospholipids of various types in the ancestral populations of primitive cells.

**Secondary metabolism.** A novel aromatic polyketide biosynthesis pathway has been entered (MicrorefCyc PWYIPS-48) that includes the polyketide synthase BcsA, which catalyses the biosynthesis of tri- and tetraketide
pyrones by the condensation of long-chain fatty acyl-CoA thioesters (primer substrates) with malonyl-CoA (extender substrates), and methyltransferase BcsB, which methylates alklypyrones produced by BcsA-yielding alklyprone methyl ethers. YpbQ represents the first enzyme found to methylate alklypyrones (Nakano et al., 2009).

Finally, annotation of the bacABHF genes has been updated in order to reflect their roles in tetrahydrotryosine biosynthesis, a non-proteogenic amino acid that is synthesized by the products of the genes bacABHF, from which anticapsin, the key residue essential for the antibiotic activity of bacilisin, could be derived (Shomura et al., 2010).

**Other idiosyncratic features.** *B. subtilis* has a complete methionine salvage pathway. While the pathway accounts for salvage of the thiomethyl group under oxidative conditions (with a dioxygenase, MtnD, used at the penultimate step of the pathway), the last transamination step unexpectedly uses glutamine as the alpha-amino donor group (Sekowska et al., 2004). This reaction is surprising because it generates alpha-ketoglutaramate, a reactive molecule that cyclizes and has toxic properties. This entails the existence of a further metabolic step that deamidates this toxic compound. MtnU, previously annotated as a nitrilase, is in fact highly similar to known omega amidases. As it has all the characteristic features of these enzymes, we can be confident that this is indeed its activity in *B. subtilis*. This step introduces an irreversible reaction that prevents back-transamination of methionine. This is supported by the observation that, while being a sulfur source, methionine cannot be used as a carbon source (Table 1).

The pathway of purine nucleotide degradation has been updated by including the specific activity of PucG, which encodes a specific pyridoxal-phosphate-dependent (S)-ureidoglycine:glyoxylate aminotransferase that catalyses the amino group transfer from the purine degradation intermediate (S)-ureidoglycine to glyoxylate, yielding the amino acid glycine and oxalurate (Ramazzina et al., 2010). In this context, (S)-ureidoglycine can also deaminate possibly spontaneously to glyoxylate by releasing urea and subsequently ammonia, representing a novel mechanism by which the same metabolite [(S)-ureidoglycine] is able to act as an amino group donor and acceptor under conditions where the amino group acceptor (glyoxylate) is absent (Ramazzina et al., 2010).

The presence of a specific lysine aminomutase (KamA) suggested that this enzyme might be an intermediate in lysine degradation. However, we found that lysine could not be used as a carbon source under vegetative growth (Table 1). In *B. subtilis*, yodD and kamA belong to a transcriptional unit whose expression is upregulated during sporulation, and, in fact, N(ε)-acetyl-β-lysine has not been detected in vegetatively growing or osmotically stressed *B. subtilis* cells (Müller et al., 2011), pointing out a possible role of N(ε)-acetyl-β-lysine under sporulation. Glutarate, an intermediate in several lysine degradation pathways (MetaCyc pathway III, IV, V and IX), cannot be used either (Table 1). The annotation of the kamA and kamB (yodP) genes has been updated, reflecting their roles encoding an AbI-related lysine-2,3-aminomutase and an AbI-related acetyltransferase, respectively, involved in the biosynthesis of the archaean-type osmolite N(ε)-acetyl-β-lysine (Müller et al., 2011). N(ε)-acetyl-β-lysine is characteristic of methanogenic Archaea, allowing cellular adaptation to high-osmotic conditions, and also having other stabilizing properties for protein, membranes and even entire cells (Pfüger et al., 2003). It is a compatible solute that may protect bacteria against some stressful conditions such as low temperature or extreme pH. Alternatively, a putative lysine degradation pathway might be triggered only under specific developmental conditions. We note that the yodT upstream promoter of the region encompassing kamBA is indeed under the control of sigma E, a mother cell-specific promoter during sporulation. It may be that lysine degradation is triggered only in the mother cell when it undergoes lysis.

**Conclusion**

With the data explosion that parallels the development of next-generation sequencing techniques it is no longer possible to annotate most genomes in-depth. This makes the role of model bacteria as essential partners in defining the reference knowledge that is subsequently automatically propagated in genome annotations all the more important. *B. subtilis* 168 is the reference organism for the Firmicutes and it is particularly useful that its genome sequence annotation is updated in a regular fashion. More than 5000 new references connected to this organism have been included in PubMed since the last annotation of the genome, many of them corresponding to the identification of gene functions. The present work upgraded the overall identification of gene functions, while focusing on creating a consistent picture of the intermediary metabolism of the organism, a feature that will be essential for work on systems and synthetic biology. It will be pursued over the next few years with the aim of having a fairly complete picture of the gene set and organization of this model organism.

**ACKNOWLEDGEMENTS**

This work was supported by European Union’s 7th Framework Programme Microme FP7-KBBE-2007-3-2-08-222886.

**REFERENCES**


Edited by: W. J. Quax