BuchneraBASE: a post-genomic resource for Buchnera sp. APS

M. Dennis Pickett, Matt Page†, Angela E. Douglas and Gavin H. Thomas*

Department of Biology, University of York, PO Box 373, York YO10 5YW, UK

Received on October 19, 2005; revised and accepted on December 22, 2005
Advance Access publication January 5, 2006
Associate Editor: Nikolaus Rajewsky

ABSTRACT

Summary: BuchneraBASE is a bioinformatic research tool for the genome of the symbiotic bacterium Buchnera sp. APS that includes an improved genome annotation, comparative information about related insect symbiont genomes and a complete mapping of metabolic reactions to an Escherichia coli in silico model. The database is designed to accommodate genome-wide post-genomic datasets that are becoming available for this organism.

Availability: BuchneraBASE is available at http://www.buchnera.org/
Contact: ght2@york.ac.uk

INTRODUCTION

The first symbiotic bacterium of insects for which a full genome sequence was obtained is the γ-proteobacterium Buchnera sp. APS in pea aphids (Shigenobu et al., 2000). Buchnera sp. APS has a tiny genome, just 0.64 Mb, and the original annotation identified 583 coding sequences (CDSs), all but four of which had unambiguous sequence similarity to CDSs in the related bacterium Escherichia coli (Shigenobu et al., 2000). In other words, the Buchnera sp. APS genome approximates to a subset of the E.coli genome and, by transfer of annotations from the well-studied E.coli, the function of ~90% of putative gene products of Buchnera sp. APS could be predicted. Subsequently, complete genome sequences have become available for Buchnera sp. in two other aphids, Schizaphis graminum (SG) (Tamas et al., 2002) and Baizongia pistacea (BP) (van Ham et al., 2003), and for two further symbiotic bacteria considered to be closely related to Buchnera sp. by some authorities: Wigglesworthia glossinidiae brevipalis in the tsetse fly Glossina brevipalpis (Akman et al., 2002) and Blochmannia floridanus in the carpenter ant Camponotus floridanus (Gil et al., 2003). The predicted gene contents of these bacteria are subsets of the E.coli genome and overlap with each other and with Buchnera sp. APS. In this paper, we describe a new database for Buchnera sp. and allied symbiotic bacteria based around our successful post-genomic database for E.coli, EchoBASE (Misra et al., 2005). During the creation of this database, we reannotated portions of the Buchnera sp. APS genome and key changes are summarized.

CREATION AND ORGANIZATION OF BUCHNERABASE

The database was constructed with MySQL using the schema developed for EchoBASE and Macromedia Coldfusion to create the WWW interface. It holds information on Buchnera genes and their products originally obtained from the GenBank entry for Buchnera aphidicola str. APS (BA000003), and includes the sequences of the two Buchnera sp. APS plasmids, pLeu and pTrp. Each gene can be viewed on a ‘gene page’ that displays basic features and annotation information. Gene pages displaying information on CDSs display the top three BLASTP hits in E.coli, and each CDS has been mapped to its E. coli orthologue and links to the respective gene page in EchoBASE. The gene page provides links to the nucleotide and protein sequence of each gene and to other useful resources. Synonyms for gene names were added from EchoBASE. MultiFun codes have been imported from E. coli and the gene products can be browsed using this protein functional classification scheme (Serres and Riley, 2000). The database has been designed to be flexible and allow the simple addition and integration of additional γ-proteobacterial symbiont genomes as they are completed. It also builds on a structure created in EchoBASE that will allow the connection of experimental information to individual gene products as these become available.

SUMMARY OF ANNOTATION CHANGES TO BUCHNERA SP. APS GENOME SEQUENCE

In the first stage of the annotation check, each CDS from Buchnera sp. APS was used as the query sequence in a BLAST search against E. coli K-12 MG1655 genome sequence. The gene name of the most significant match was checked against the name of the gene encoding that CDS in Buchnera sp. APS. There was generally agreement with the original annotation of Shigenobu et al. (2000). However, the few errors in the original annotation noted by us and others (van Ham et al., 2003) are corrected, including the identification of the yba1 gene (originally annotated as unique to Buchnera) as the 5′-region of a truncated flIK gene. We have updated the functions of several gene products, e.g. the gene for the elusive 6-phosphogluconolactonase recently characterized in E.coli as the ybbE gene product (Thomason et al., 2004). Also, gene names have been changed, where appropriate, from the ybXXX nomenclature to the y gene nomenclature now preferred in E.coli. To facilitate easy querying of genes, a synonym table of gene names taken from our sister database EchoBASE is provided. We have added annotations for obvious pseudogenes, of interest in tracing the recent evolutionary history of Buchnera, e.g. metR, disrupted relatively recently in the history of Buchnera sp. APS. The lengths of some gene have been changed where comparison with E.coli orthologues suggests an incorrect initiator methionine was chosen during the original annotation.

*To whom correspondence should be addressed.
†Current address: UCB, 216 Bath Road, Slough, Berkshire SL1 4EN, UK
Using our knowledge of the *E.coli* genome annotation, we have identified two previously unrecognized small RNA-encoding genes. *Buchnera* sp. APS was originally annotated to have *ffh* and *ftsY*, which encode the protein components of the signal recognition particle (SRP) (Shigenobu *et al.*, 2000). We found the gene encoding the RNA component of the SRP, *ffs*, between *mdlB* and *dnaX* and have defined the CDS based on the coordinates provided by SRPDB (Rosenblad *et al.*, 2003). Also, based on the presence of *smpB*, we identified the gene for the tmRNA, *ssrA*, and confirmed this using the tmRNA website (Gueneau and Williams, 2004).

### COMPARATIVE ANALYSIS OF INSECT SYMBIONT GENOMES

To aid understanding of the APS genome and its products, we have included the genomes of four additional insect symbionts, *B.aphidica* str. SG (Schizaphis graminum) (Tamas *et al.*, 2002), *B.aphidica* str. BP (Baizonga pistaciaceae) (van Ham *et al.*, 2003), *Wigglesworthia glossinidia* (Akman *et al.*, 2002) and *Blochmannia floridanus* NL (Gil *et al.*, 2003), which were parsed and added to the database as for *Buchnera* sp. APS. A manually curated non-redundant table of genes was created for the products of the five genomes with mapping of all orthologous genes across the genomes. This allows rapid assessment of the distribution of individual genes within the sequenced insect symbionts, displayed on the genes page. These data can also be viewed in a summary table providing referencing across the five genomes to allow rapid determination of genes shared between or particular to certain symbionts genomes. More limited gene pages are available for genes for all four of the additional symbionts.

We have updated the function lines of the *Wigglesworthia* gene products for the 618 genes that have clear orthologues in *E.coli*, based on the current *E.coli* function lines. This improves the usefulness of the annotation, which included non-specific descriptions of many gene products in the original submission (Akman *et al.*, 2002).

### MAPPING OF BUCHNERA SP. APS GENE PRODUCTS TO AN E.COLI METABOLIC MODEL (IJR904)

We have mapped all metabolic reactions from *Buchnera* sp. APS to a metabolic model of *E.coli* (IJR904) created for flux-balance analysis (Reed *et al.*, 2003). A total of 233 gene/enzyme relationships were mapped by matching of gene names and then by manual checks. The mappings are shown on the gene pages for the genes that encode these activities, displaying the reaction equation, reaction name (and synonyms) and reaction pathway. This includes 182 distinct biochemical reactions encoded by the products of 193 distinct APS genes. This will allow examination of the properties of the metabolic network with graph analysis tools and facilitate use of the mapping in flux-balance modeling of *Buchnera* metabolism.

### NOTE ADDED IN PROOF

While this manuscript was being reviewed the genome sequence of *Blochmannia pensyvallanica* was published [Degnan, Lazarns and Wernegreen (2005) Genome Research, 15 1023–1033], which has been added to *BuchneraBASE*.

### ACKNOWLEDGEMENTS

We thank Raju Misra and Richard Horler for advice in setting up the database.

**Conflict of Interest:** none declared.

### REFERENCES


