The difficulty of annotating genes: the case of putrescine carbamoyltransferase

Incorrect annotation of genes and enzymes is an important source of confusion in phylogenetic studies and analysis of metabolic networks.

We have recently identified as putrescine carbamoyltransferases (PTCases; EC

2.1.3.6, GO: 0050231) a set of enzymes that have been erroneously annotated as ornithine carbamoyltransferases (EC 2.1.3.3, GO: 0009348) on the basis of genomic data (Naumoff *et al.*, 2004). These PTCases are involved in the catabolism of the polyamine agmatine via the agmatine deiminase system (Simon & Stalon, 1982) present in several Gram-positive bacteria. Accordingly, we have been able to re-annotate the function of the corresponding encoding genes in those bacteria (*Enterococcus faecalis*, *Listeria monocytogenes*, *Streptococcus mutans* and *Lactococcus lactis*) for which the genomes have been entirely sequenced (Naumoff

et al., 2004). This is summarized in Fig. 1 (see upper part, Cluster A) which also displays more recent data confirming and extending our re-annotation process to other bacteria including one distantly related proteobacterium, *Photobacterium profundum* [re-annotation of the hypothetical ornithine carbamoyltransferase PBPRB1921 (TrEMBL Q6LG11) as a PTCase].

While our article (Naumoff *et al.*, 2004) was being reviewed, Griswold *et al.* (2004a) identified independently the SMU.262 gene in *S. mutans* as encoding a PTCase and re-designated it with the name *aguB*

(Fig. 1). We would like to draw attention to the fact that this renaming is inappropriate and misleading since *aguB* has been already used by Haas *et al.* (1984) and more recently by Nakada *et al.* (2001) and Nakada & Itoh (2003) to designate the gene encoding the *N*-carbamoylputrescine amidohydrolase (*N*-carbamoylputrescine amidase, EC 3.5.1.53, GO: 0050126), an enzyme involved in agmatine utilization by *Pseudomonas aeruginosa* (Mercenier *et al.*, 1980) as summarized in Fig. 1 (see lower part, Cluster B). As a matter of fact, the two enzymes are neither homologous nor analogous. Indeed, PTCase converts

Fig. 1. Contradictory use of the name *aguB*. The re-annotation we have proposed (Naumoff *et al.*, 2004) for the set of genes encoding the proteins involved in the agmatine deiminase pathway in four Gram-positive bacteria (*Enterococcus faecalis*, GenBank/EMBL/DDBJ accession no. AE016830; *Listeria monocytogenes*, AL591824; *Streptococcus mutans*, AE014133; *Lactococcus lactis*, AE005176) has been summarized here and extended to another, more distant, Gram-positive bacterium, *Mycoplasma mycoides* (GenBank/EMBL/DDBJ accession no. BX293980). The re-annotation proposed by Griswold *et al.* (2004a) is also indicated in the case of *S. mutans*. This set of data is framed to indicate the homogeneity of these respective sets of genes (Cluster A). Below this frame the homologues of those genes that we have detected in the newly sequenced genome of the γ -proteobacterium *Photobacterium profundum* are indicated using the same motif for the respective homologous genes. Finally, we show the cluster of genes present in *Pseudomonas aeruginosa* (GenBank/EMBL/DDBJ accession no. AE004091) that is involved in agmatine utilization. This unrelated Cluster B contains the actual *aguB* gene (see text). Note also the homology between the *Pseudomonas aeruginosa* *aguR* gene and a previously unannotated *Photobacterium profundum* gene (PBPRB1919).

Table 1. Summary of the matches found in TrEMBL when querying the UniProt knowledgebase with the text search ‘aguB’

There were no matches in SwissProt.

Accession no.	Enzyme name	Gene name	Species	Entered	Annotation is:
Q9I6J8	N-Carbamoylputrescine amidohydrolase	<i>aguB</i>	<i>Pseudomonas aeruginosa</i>	March 2001	Correct
Q7M114	Putrescine carbamoyltransferase	<i>aguB</i> (fragment)	<i>Enterococcus faecalis</i>	March 2004	Incorrect (contradicts Tricot <i>et al.</i> , 1989)
Q6IFZ2	Putrescine carbamoyltransferase	<i>aguB</i>	<i>Streptococcus mutans</i> UA159	July 2004	Incorrect (see text)
Q6XR80	Agarase	<i>aguB</i> (synonyms <i>aguD</i> , <i>aguH</i> and <i>aguK</i>)	Uncultured bacterium (soil metagenome)	July 2004	Incorrect (see text)

carbamoylputrescine to putrescine and carbamoylphosphate, whereas *N*-carbamoylputrescine amidohydrolase converts *N*-carbamoylputrescine to putrescine, CO₂ and NH₃, and their respective amino acid sequences are completely unrelated. Moreover, the cognate genes belong to different genetic contexts (Naumoff *et al.*, 2004, and Fig. 1).

Furthermore, the name *aguB* has also been used by Voget *et al.* (2003) to designate a gene encoding a putative β-agarase (see Table 1) identified from a soil metagenome. This adds to the confusion and should be considered an inappropriate designation also.

Consequently, we propose the name *ptcA* to designate the gene encoding PTCase.

It should be stressed that the uncontrolled misuse of the name *aguB* has created an intolerable chaos in databases, as shown by the following excerpts from the latest version of TrEMBL (release 27.4, 31 August 2004), which is part of the UniProt knowledgebase (Apweiler *et al.*, 2004). This confusion is clearly apparent in Table 1, which was obtained by querying UniProt with the text search ‘aguB’.

In order to prevent people obtaining such misleading and worthless results where three unrelated enzymic activities are recovered when querying with a unique gene name descriptor, we strongly suggest to use the name *aguB* to designate exclusively genes encoding *N*-carbamoylputrescine amidohydrolase as we have specified in the column ‘Annotation is:’ of Table 1 since both gene *aguB* and protein AguB have been well studied for a long time and unambiguously characterized as playing a major role in

putrescine biosynthesis as well as in agmatine catabolism (Nakada & Itoh, 2003, and references therein).

With regard to *aguA*, the use of this name by Griswold *et al.* (2004a) to designate agmatine deiminase respects the nomenclature proposed by Haas *et al.* (1984) and more recently by Nakada *et al.* (2001) and Nakada & Itoh (2003) and is thus correct. By contrast, using *aguC* for carbamate kinase instead of *arcC* is again not appropriate since the name *arcC* is universally recognized as designating carbamate kinases. It must be noted that, in a subsequent article studying in detail the arginine deiminase operon in another *Streptococcus* species, the same authors went back to using *arcC* in the case of the corresponding carbamate kinase (Griswold *et al.*, 2004b).

Interestingly, Fig. 1 further shows that a gene homologous to the transcriptional regulator *aguR* which has been characterized by Nakada *et al.* (2001) and Nakada & Itoh (2003) as negatively regulating the expression of the *aguBA* operon in *Pseudomonas aeruginosa* (cluster B in Fig. 1) is also found in *Photobacterium profundum*. However, in *Photobacterium profundum*, this *aguR*-like gene is inserted in a set of genes that appear to be homologous to the agmatine deiminase gene cluster (cluster A in Fig. 1) found in Gram-positive bacteria and encoding PTCase, amino acid permease, agmatine deiminase and carbamate kinase, respectively (Griswold *et al.*, 2004a; Naumoff *et al.*, 2004), but with a different gene order. This *aguR*-like gene does not appear to be evolutionarily related to the putative regulator SMU.261c of the agmatine deiminase pathway in *S. mutans* (Griswold *et al.*, 2004a) nor to its

homologues EF0731 (*E. faecalis*) and *llrH* (*Lactococcus lactis*). This would appear to be an interesting instance of recruitment of a regulatory gene from one pathway to another and deserves further study.

Acknowledgements

This work was supported by the CNRS (UMR 8621) and by the Flanders Foundation for Joint and Fundamental Research. D. G. N. was supported by a postdoctoral grant from the French Ministère de la Recherche.

Daniil G. Naumoff,^{1†} Ying Xu,² Victor Stalon,² Nicolas Glansdorff² and Bernard Labedan¹

¹Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris Sud, Bâtiment 409, 91405 Orsay Cedex, France

²Laboratoire de Microbiologie, Université Libre de Bruxelles and Institut de Recherches Microbiologiques J.M. Wiame, 1, ave E. Gryzon, B-1070 Brussels, Belgium

[†]Present address: State Institute for Genetics and Selection of Industrial Micro-organisms, I-Dorozhny proezd, 1, Moscow 117545, Russia

Correspondence: Bernard Labedan (bernard.labeledan@igmors.u-psud.fr)

Apweiler, R., Bairoch, A., Wu, C. H. & 12 other authors (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32 Database issue**, D115–D119.

Griswold, A. R., Chen, Y. Y. & Burne, R. A. (2004a). Analysis of an agmatine deiminase gene cluster in *Streptococcus mutans* UA159. *J Bacteriol* **186**, 1902–1904.

Griswold, A. R., Chen, Y. Y., Snyder, J. A. & Burne, R. A. (2004b). Characterization of the arginine deiminase operon of *Streptococcus*

rattus FA-1. *Appl Environ Microbiol* **70**, 1321–1327.

Haas, D., Matsumoto, H., Moretti, P., Stalon, V. & Mercenier, A. (1984).

Arginine degradation in *Pseudomonas aeruginosa* mutants blocked in two arginine catabolic pathways. *Mol Gen Genet* **193**, 437–444.

Mercenier, A., Simon, J. P., Haas, D. & Stalon, V. (1980). Catabolism of L-arginine by *Pseudomonas aeruginosa*. *J Gen Microbiol* **116**, 381–389.

Nakada, Y. & Itoh, Y. (2003). Identification of the putrescine biosynthetic genes in *Pseudomonas aeruginosa* and characterization of agmatine deiminase

and N-carbamoylputrescine amidohydrolase of the arginine decarboxylase pathway. *Microbiology* **149**, 707–714.

Nakada, Y., Jiang, Y., Nishijyo, T., Itoh, Y. & Lu, C. D. (2001). Molecular characterization and regulation of the *aguBA* operon, responsible for agmatine utilization in *Pseudomonas aeruginosa* PAO1. *J Bacteriol* **183**, 6517–6524.

Naumoff, D. G., Xu, Y., Glansdorff, N. & Labedan, B. (2004). Retrieving sequences of enzymes experimentally characterized but erroneously annotated: the case of the putrescine carbamoyltransferase. *BMC Genomics* **5**, 52.

Simon, J. P. & Stalon, V. (1982). Enzymes of agmatine degradation and the control of their synthesis in *Streptococcus faecalis*. *J Bacteriol* **152**, 676–681.

Tricot, C., De Coen, J. L., Momin, P., Falmagne, P. & Stalon, V. (1989). Evolutionary relationships among bacterial carbamoyltransferases. *J Gen Microbiol* **135**, 2453–2464.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K. E. & Streit, W. R. (2003). Prospecting for novel biocatalysts in a soil metagenome. *Appl Environ Microbiol* **69**, 6235–6242.

DOI 10.1099/mic.0.27640-0