No general pattern was found in the functions of the genes that underwent a fission event. Genes resulting from a fission event are often annotated as hypothetical, because the split forms a problem in annotation of function[11]. The reverse, fusion proteins being annotated as having only one of two functions, has also been observed.

Here for the first time, we have systematically and comprehensively surveyed the occurrence of gene fission and fusion. We find a correlation of fission with thermophily and argue that this lifestyle results in a selective pressure for the split organization of genes. As such, it is an example of the relation between phenotype and its composing parts. Cross-level relations like this stand at the core of genome function and evolution, and we expect that our understanding of them will eventually allow us to elucidate the principles that govern the dynamics of genome evolution.

References

1 Leffers, H. *et al.* (1989) Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaeabacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. *J. Mol. Biol.* 206, 1–17
2 Zakharova, N. *et al.* (1999) Fused and overlapping *rpoB* and *rpoC* genes in helicobacters, campylobacters, and related bacteria. *J. Bacteriol.* 181, 3857–3859
3 Fitch, W.M. (1970) Distinguishing homologous from analogous proteins. *Syst. Zool.* 19, 99–113
4 Smith, T. and Waterman, M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195–197
5 Pearson, W. (1998) Empirical statistical estimates for sequence similarity searches. *J. Mol. Biol.* 276, 71–84
6 Huynen, M.A. and Bork, P. (1998) Measuring genome evolution. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5849–5856
7 Himmelreich, R. *et al.* (1997) Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res.* 25, 701–712
8 Marcotte, E.M. *et al.* (1999) Detecting protein function and protein–protein interactions from genome sequences. *Science* 285, 751–753
9 Jaenicke, R. and Boehm, G. (1998) The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* 8, 738–748
10 Burns, D.M. *et al.* (1995) Silent genes in bacteria: the previously designated 'cryptic' ilvHI locus of '*Salmonella typhorium* LT2' is active in natural isolates. *FEMS Microbial. Lett.* 131, 167–172
11 Bork, P. and Koonin, E.V. (1998) Predicting functions from protein sequences – where are the bottlenecks? *Nat. Genet.* 18, 313–318
12 Thoden, J.B. *et al.* (1999) The structure of carbamoyl phosphate synthase determined to 2.1 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* 55, 8–24
13 Iwabe, N. *et al.* (1989) Evolutionary relationships of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. U. S. A.* 86, 9355–9359
14 Strimmer, K. and von Haeseler, A. (1996) Quartet puzzling: A maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13, 964–969
15 Jeanmougin, F. *et al.* (1998) Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 23, 403–405

The *glnKamtB* operon
a conserved gene pair in prokaryotes

The new subject of comparative genomics has facilitated significant advances in our understanding of genome structure and organization[1]. One notable conclusion is that gene order is not highly conserved and that genomes are highly plastic[2,3]. Conservation of gene order in distantly related eubacteria appears to occur in operons that encode essential cellular apparatus, for example ribosomal and cell division proteins[3], or in operons in which the gene products physically interact[4]. Indeed, the conservation of gene clusters coupled with the rapid increase in the availability of prokaryotic sequence data now allows the prediction of the functional coupling of genes and their assignment to metabolic pathways[5].

Here, we describe a conserved gene pair, *glnK* and *amtB*, that has not been identified in previous studies[3,4]. These genes are not essential for growth in *Escherichia coli*[6,7], which strongly suggests that the reason for their conserved association is that the two proteins interact physically; an interaction that would be between a small cytoplasmic signal transduction protein (GlnK) and a membrane-bound ammonium transport protein (AmtB).

The genetic association of these two genes is found in a diverse range of Eubacteria and Archaea. The operon has been best described in *E. coli*, but homologues of *glnK* and *amtB*, which are co-transcribed, can be found in the completed genome sequences of many other organisms. These include other Gram-negative bacteria (e.g. *Aquifex aeolicus*), Gram-positive bacteria (e.g. *Bacillus subtilis* and *Mycobacterium tuberculosis*), and the Archaea, where the pair are found in multiple copies (e.g. *Archaeoglobus fulgidus*, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*). Linkage of *glnK* and *amtB* genes can also be observed in many of the uncompleted genome sequencing projects (Table 1). It might be of interest to note that, to date, there are no prokaryotic examples of organisms that have *amtB* but no gene that encodes a GlnK-like protein. Examples of the genetic organization of *glnKamtB* operons are presented in Fig. 1, which illustrates some flexibility in the order of the two genes, but not in their genetic linkage. An exception is one set of *glnK* and *amtB* homologues in *M. jannaschii* that are transcribed divergently, although such organization often suggests a related expression pattern (Fig. 1). With one exception described to date, namely in *Azotobacter vinelandii*[8], expression of the *glnKamtB* operon (or its equivalent) is induced specifically during conditions of ammonium limitation, implying that the functions of GlnK and AmtB are specifically required during growth in these conditions[6,9–12].

Homologues of AmtB were originally identified in *Saccharomyces cerevisiae* and *Arabidopsis thaliana*[13,14]. In *S. cerevisiae*, the three AmtB homologues are high-affinity ammonium transporters. The Amt protein family has since been shown to be ubiquitous, and representatives are found in eubacteria, archaea, fungi, plants, protists and lower animals (see http://alize.ulb.ac.be/~bandre/Mep_List.htm). They are all predicted to be membrane-bound, with 10–12 membrane-spanning helices, and they constitute a discrete family that is unrelated to other transport proteins[15]. The precise function of AmtB in bacteria has

Gavin Thomas
gav.thomas@bbsrc.ac.uk
Graham Coutts
graham.coutts@bbsrc.ac.uk
Mike Merrick
mike.merrick@bbsrc.ac.uk

Department of Molecular Microbiology, John Innes Centre, Colney Lane, Norwich, UK NR4 7UH.

0168-9525/00/$ – see front matter © 2000 Elsevier Science Ltd. All rights reserved. PII: S0168-9525(99)01887-9

TABLE 1. The presence or absence of *glnK–amtB* operons and additional *glnK*, *amtB* homologues in completely or partially sequenced genomes

Organism	Presence of *glnKamtB*[a]	Other unlinked *amtB* and *glnK*-like genes
Completed genomes		
Escherichia coli	*glnKamtB*	*glnB*
Aquifex aeolicus	*glnKglnAamtB*	*glnB*
Thermotoga maritima	*amtBglnK*	No
Bacillus subtilis	*amtBglnK*[b]	No
Mycobacterium tuberculosis	*amtBglnKglnD*	No
Archaeoglobus fulgidus	*amtBglnK* (×3)	No
Methanobacterium thermoautotrophicum	*amtBglnK* (×2)	*glnB1, glnB2*
Methanococcus jannaschii	*glnKamtB*	No
	amtB-glnK[c]	
Uncompleted genomes or cloned independently[d]		
Salmonella typhi	*glnKamtB*	*glnB*
Yersinia pestis	*glnKamtB*	*glnB*
Azorhizobium caulinodans	*glnKamtB*	*glnB*
Klebsiella pneumoniae	*glnKamtB*	*glnB*
Rhizobium etli	*glnKamtB*	*glnB*

[a]We list here cases where the *glnK* and *amtB* genes appear in the same operon although not necessarily in this gene order. For the purposes of this discussion, we have defined the *amtB*-linked gene as *glnK*, as this is consistent with independent phylogenetic analysis (see Fig. 2).
[b]Originally designated *nrgAnrgB* (Ref. 9).
[c]Transcribed divergently.
[d]The listing here is a minimal one: a fully referenced Table 1 of *glnK* and *amtB* genes in the prokaryotes is available at http://www.jic.bbsrc.ac.uk/staff/mike-merrick/glnkamtb/table1.html.

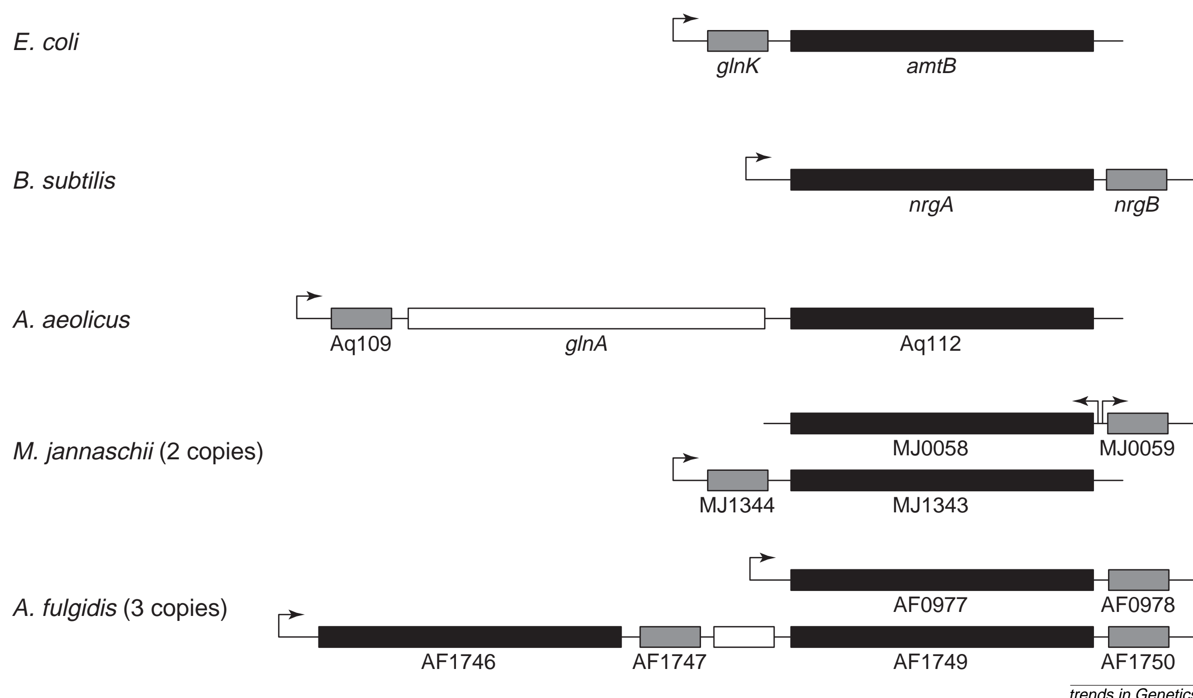
still to be confirmed. Several authors have suggested that, as in *S. cerevisiae*, it serves as a high-affinity ammonium transporter[16,17], whereas Soupene *et al.*[7] conclude that AmtB is an ammonia-facilitator protein. Nevertheless, it is certainly involved in movement of some form of ammonium across the bacterial cytoplasmic membrane.

GlnK is a member of the P_{II} family of bacterial signal transduction proteins, which are almost invariably involved in sensing the nitrogen status of the cell. They transduce this signal, via protein–protein interactions, to a variety of enzymes involved in nitrogen metabolism[16]. In *E. coli* the P_{II} protein interacts with a range of proteins that have quite distinct structures and functions. They include: NtrB, the histidine protein kinase of the nitrogen regulatory system; GlnE, an adenylyltransferase that modifies glutamine synthetase; and GlnD, a uridylyltransferase that modifies P_{II} itself[6,18].

The presence of two P_{II}-like proteins, GlnK and GlnB, is almost universal among the Proteobacteria, and phylogenetic analysis indicates that these are paralogous proteins and are likely to have evolved from a common ancestor (Fig. 2). In *E. coli*, the GlnK and GlnB proteins are 67% identical and show a high degree of structural similarity[19]. However, the precise role of GlnK has still to be determined, although in *E. coli* it can substitute for some functions of P_{II} and in *Klebsiella pneumoniae* it is required for nitrogen regulation of nitrogen fixation (*nif*) gene expression[6,11].

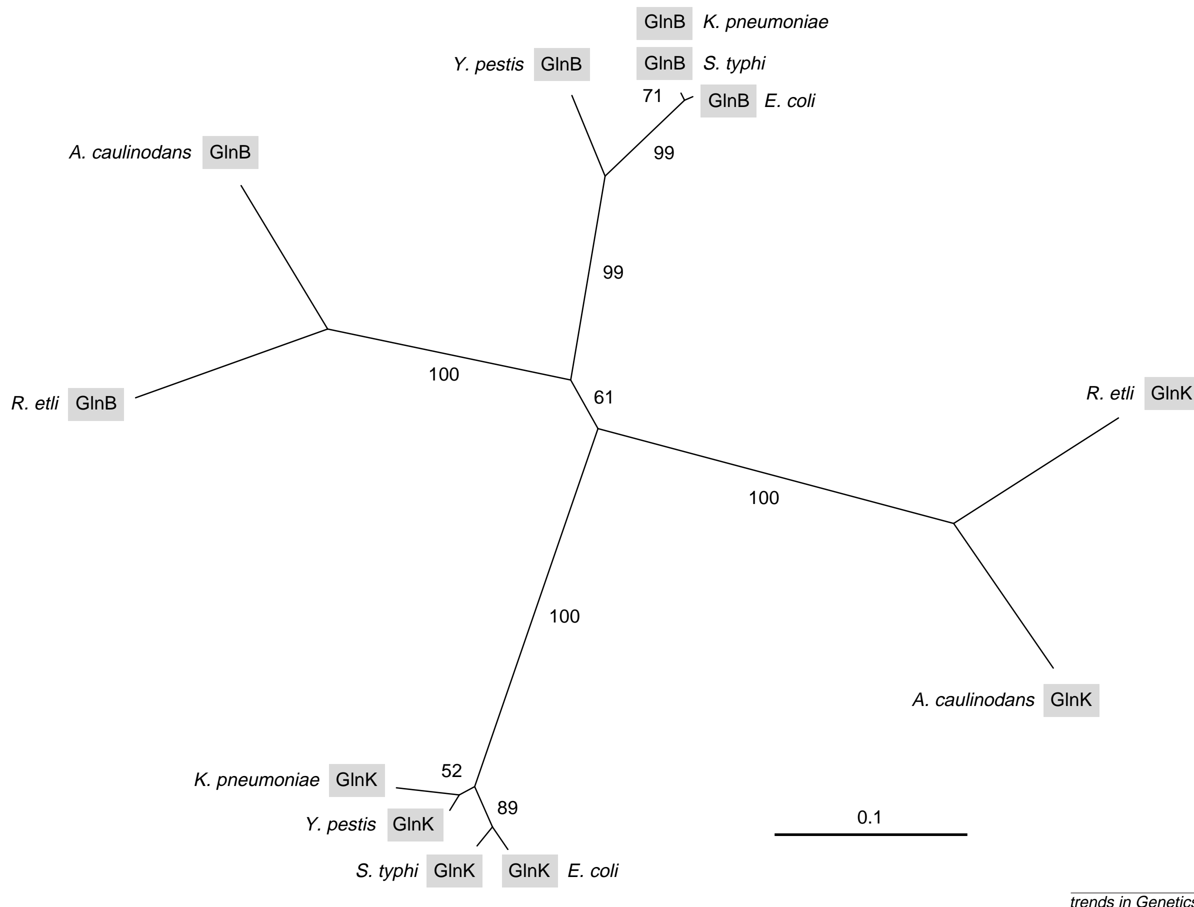
A phylogenetic analysis of the AmtB and GlnK proteins from the Eubacteria and Archaea demonstrated a similar evolutionary pattern for both proteins,

FIGURE 1. Genetic organization of *glnKamtB* genes in a range of Eubacteria and Archaea



Homologues of the *Escherichia coli amtB* gene are presented as filled black boxes, while homologues of *E. coli glnK* are presented as grey boxes. Sequences for *Methanococcus jannaschii* and *Archaeoglobus fulgidus* genes were obtained from their respective genome servers within the TIGR site (http://www.tigr.org). The genetic organization of the genes from *E. coli*, *Bacillus subtilis* and *Aquifex aeolicus* are from Refs 16, 9 and 25, respectively. Full details of all these genes can be found in an extended Table at http://www.jic.bbsrc.ac.uk/staff/mike-merrick/glnkamtb/table1.html.

FIGURE 2. Phylogenetic analysis of the P_{II} proteins from a range of Proteobacteria



Organisms that have a P_{II} protein encoded by a gene linked to *amtB* (called GlnK) and a second, encoded by a gene not linked to *amtB* (called GlnB), are included. The tree was generated in PHYLIP using a distance-matrix analysis and displayed in TREEVIEW with added bootstrap values. The sequences for *Klebsiella pneumoniae*, *Azorhizobium caulinodans* and *Rhizobium etli* used in this analysis were obtained from SWISSPROT; the sequences for *Salmonella typhi* and *Yersinia pestis* were identified in their respective genome sequences by BLAST searches against *Escherichia coli glnK* and *glnB*. Scale bar represents 0.1 substitutions per residue.

suggesting that they each evolved from an ancestral protein (data not shown). As the genetic association of the structural genes for these two proteins is conserved in both the Archaea and the Eubacteria, it suggests that the two genes were associated before the divergence of these two domains of life. In *A. fulgidis*, the evolutionary history of the *glnK–amtB* gene pairs can be deduced from a combination of genetic and phylogenetic evidence. Phylogenetically, the two pairs AF0977/78 and AF1749/50 are very close, but these are not the pairs which are encoded by adjacent genes on the chromosome (see Fig. 1). This suggests that AF1747/47 and AF1749/50 are the most ancient of the gene pairs, one having evolved from a duplication of the other. Much later, AF0977/78 has evolved from a gene duplication of AF1749/50. The demonstration that the GlnK and GlnB proteins form a paralogous pair in the Proteobacteria (see Fig. 2), suggests that it was probably a duplication of an ancient *glnK* gene that gave rise to the *glnB* gene and not vice versa, and that *glnB* then evolved to function in other contexts, most notably in the nitrogen-regulation system.

Therefore, based on their conserved genetic organization, we propose that GlnK and AmtB interact directly and that this could serve one of two functions. Either GlnK could regulate the transport activity of AmtB in response to the ammonium requirements of the cell, or the activity of AmtB could act as a signal of ammonium availability and could function to modulate intracellular nitrogen metabolism through the signal transduction properties of GlnK. P_{II}-like proteins have not previously been shown to interact with membrane-bound proteins, although P_{II} has recently been implicated in the control of nitrate and nitrite uptake in *Synechococcus*[20].

The recent study of Dandekar *et al.*[4] failed to recognize the *glnK* and *amtB* genes as a conserved pair because they are absent in a number of the genomes that they analysed. It is therefore of some interest to examine those completed genome sequences in which the *glnK* and *amtB* genes are absent, as they might offer some clues to the physiological functions of the GlnK and AmtB proteins. Nearly all of the organisms that lack AmtB are pathogenic bacteria, whose genomes have undergone reductive evolution to compete in specific niches within their respective hosts. Intracellular pathogens, such as *Rickettsia prowazekii* and *Chlamydia trachomatis*, rely on their hosts for the provision of all their nitrogen-containing compounds, a lifestyle that has also been adopted by extracellular pathogens,

such as *Borrelia burgdorferi* and *Mycoplasma genitalium*. Correspondingly, these organisms lack not only the *glnK* and *amtB* genes but also other genes required for the assimilation of ammonium.

By contrast, other extracellular pathogens do assimilate ammonium for biosynthesis and so they might be expected to have a mechanism to transport ammonium efficiently into the cell. The pathogen *Helicobacter pylori* synthesizes many of its own nitrogenous compounds, yet its genome lacks the *glnK* and *amtB* genes[21]. However, the organism rapidly metabolizes urea from the host in a process that is thought to aid survival in the low-pH environment of the stomach. As a consequence, it produces copious amounts of intracellular ammonium, which is assimilated to glutamine by glutamine synthetase[22]. Therefore, it is unlikely that the bacterium encounters a situation in which the intracellular concentration of ammonium is limited, and this is probably why it lacks the *glnK* and *amtB* genes.

Similarly, *Haemophilus influenzae* synthesizes most of its amino acids but, again, lacks the *glnK* and *amtB* genes[23]. In a defined growth medium, high concentrations of glutamate (1.3 g l^{-1}) are required for rapid growth[24], and deamination of glutamate is thought to provide carbon skeletons to a truncated tricarboxylic acid cycle[23]. This catabolic reaction releases ammonium, which can be used subsequently by glutamine synthetase to synthesize glutamine from glutamate. Therefore, in the presence of glutamate, sufficient intracellular ammonium is probably liberated to bypass a need for an efficient transport system.

A contrasting case is that of *Aquifex aeolicus*, which, despite being a free-living bacterium, has a relatively small genome (~1.5 Mb)[25]. Unlike the pathogenic bacteria, it must still synthesize most of its metabolites, including its nitrogenous compounds, in an environment where these metabolites are not abundant. Consistent with this, the organism retains *glnK* and *amtB* on the chromosome, although in this case the gene that encodes glutamine synthetase (*glnA*) is between *glnK* and *amtB* (see Fig. 1).

To conclude, we have described a conserved gene pair that has not previously been identified in other studies[3,4]. Consideration of the physiology of the organisms whose genomes are being analysed can provide rational explanations for the absence of the gene pair in certain organisms. Indeed, we would suggest that the presence of the *glnK* and *amtB* genes in distantly related organisms, but not invariably in closely related ones, is purely a reflection of the lifestyle of the organism in question. Future searches for conserved gene sets might reveal further associations if the genomes of organisms that inhabit similar environments are considered together.

Acknowledgements

G.T. and M.M. were supported by EU Contract No. BIO4 CT 972310 and G.C. acknowledges a BBSRC studentship. We thank S. Austin, R. Dixon and G. Sawers for constructive comments on the manuscript.

References

1 Koonin, E.V. and Galperin, M.Y. (1997) Prokaryotic genomes: the emerging paradigm of genome-based microbiology. *Curr. Opin. Genet. Dev.* 7, 757–763
2 Mushegian, A.R. and Koonin, E.V. (1996) Gene order is not conserved in bacterial evolution. *Trends Biochem. Sci.* 12, 289–290
3 Watanabe, H. *et al.* (1997) Genome plasticity as a paradigm of eubacterial evolution. *J. Mol. Evol.* 44 (Suppl. 1), 57–64
4 Dandekar, T. *et al.* (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Genet.* 23, 324–328
5 Overbeek, R. *et al.* (1999) The use of gene clusters to infer functional coupling. *Proc. Natl. Acad. Sci. U. S. A.* 96, 2896–2901
6 Atkinson, M. and Ninfa, A.J. (1998) Role of the GlnK signal transduction protein in the regulation of nitrogen assimilation in *Escherichia coli. Mol. Microbiol.* 29, 431–447
7 Soupene, E. *et al.* (1998) Ammonia acquisition in enteric bacteria: physiological role of the ammonium/methylammonium transport B (AmtB) protein. *Proc. Natl. Acad. Sci. U. S. A.* 95, 7030–7034
8 Meletzus, D. *et al.* (1998) Characterization of the *glnK-amtB* operon of *Azotobacter vinelandii. J. Bacteriol.* 180, 3260–3264
9 Wray, L. *et al.* (1994) The nitrogen-regulated *Bacillus subtilis nrgAB* operon encodes a membrane protein and a protein highly similar to the *E. coli glnB*-encoded P_{II}. *J. Bacteriol.* 176, 108–114
10 Taté, R. *et al.* (1998) The *Rhizobium etli amtB* gene coding for an NH_4^+ transporter is down-regulated early during bacteroid differentiation. *Mol. Plant–Microbe Interact.* 11, 188–198
11 Jack, R. *et al.* (1999) The signal transduction protein GlnK is required for NifL-dependent nitrogen control of *nif* expression in *Klebsiella pneumoniae. J. Bacteriol.* 181, 1156–1162
12 Michel-Reydellet, N. *et al.* (1998) Characterisation of the *glnK-amtB* operon and involvement of AmtB in methylammonium uptake in *Azorhizobium caulinodans. Mol. Gen. Genet.* 258, 671–677
13 Marini, A-M. *et al.* (1994) Cloning and expression of the MEP1 gene encoding an ammonium transporter in *Saccharomyces cerevisiae. EMBO J.* 13, 3456–3463
14 Ninnemann, O. *et al.* (1994) Identification of a high affinity NH_4^+ transporter from plants. *EMBO J.* 13, 3464–3471
15 Saier, M.H., Jr *et al.* (1999) Phylogenetic characterisation of novel transport protein families revealed by genome analysis. *Biochem. Biophys. Acta* 1422, 1–56
16 van Heeswijk, W.C. *et al.* (1996) An alternative P_{II} protein in the regulation of glutamine synthetase in *Escherichia coli. Mol. Microbiol.* 21, 133–146
17 Siewe, R.M. *et al.* (1996) Functional and genetic characterisation of the (methyl)ammonium uptake carrier of *Corynebacterium glutamicum. J. Biol. Chem.* 271, 5398–5403
18 Merrick, M.J. and Edwards, R.A. (1995) Nitrogen control in bacteria. *Microbiol. Rev.* 59, 604–622
19 Xu, Y. *et al.* (1998) GlnK, a P_{II}-homologue: Structure reveals ATP binding site and indicates how the T-loops may be involved in molecular recognition. *J. Mol. Biol.* 282, 149–165
20 Lee, H.M. *et al.* (1998) A role for the signal transduction protein P_{II} in the control of nitrate/nitrite uptake in a cyanobacterium. *FEBS Lett.* 427, 291–295
21 Tomb, J.F. *et al.* (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori. Nature* 388, 539–547
22 Garner, R.M. *et al.* (1998) *Helicobacter pylori* glutamine synthetase lacks features associated with transcriptional and post-translational regulation. *Infect. Immun.* 66, 1839–1847
23 Fleischmann, R.D. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* RD. *Science* 269, 496–512
24 Klein, R.D. and Luginbuhl, G.H. (1979) Simplified media for the growth of *Haemophilus influenzae* from clinical and normal floral sources. *J. Gen. Microbiol.* 113, 409–411
25 Deckert, G. *et al.* (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus. Nature* 392, 353–358

Happy New Year!

All the staff at *Trends in Genetics* wish our readers a happy and prosperous New Year