# A hidden metabolic pathway exposed

**Andrei Osterman***

*Burnham Institute for Medical Research, La Jolla, CA 92037; and Fellowship for Interpretation of Genomes, Burr Ridge, IL 60527*

Before reading the report by Sydney Kustu's research group published in a recent issue of PNAS (1), I would have seriously doubted that discovery of a novel pathway for pyrimidine catabolism in *Escherichia coli* K12 was something that was still possible today. After all of the years of exhaustive genetic and biochemical studies, almost a decade past completion of the genome (2), after innumerable annotation and reannotation efforts, metabolic modeling (3), and systematic functional genomics studies (4–6), a completely novel pathway was found? Although functional roles of many *E. coli* genes (at least 20%) are still unknown, these remaining *y* genes are largely perceived as potential off-mainstream players in some obscure areas of metabolism or, even more likely, in nonmetabolic processes. Therefore, a string of eight *y* genes encoding an entire pathway in the very center of the *E. coli* metabolic map felt like it was too big a skeleton to be hidden in a closet.

How could not only the specific genes, but the very existence of this pathway have eluded earlier studies? One possible explanation is that "pyrimidine rings can be used as sole N source at room temperature but not at 37°C" (1). The question still remains: why among the many successful hunts for missing genes (7) did all of the powerful bioinformatics techniques fail to spot this missing pathway? A likely explanation (in one of its many versions) is that "to find a black cat in the dark room we should at least know that it is hiding there." Indeed, early physiological evidence (8), as long as it was not captured by public archives, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genome.ad.jp/kegg) or EcoCyc, the Encyclopedia of *Escherichia coli* K-12 Genes and Metabolism (http://ecocyc.org), remained inaccessible to genome annotators, preventing them from asking the right questions. On the other hand, experts in a field are often unaware of bioinformatics tools that would help them address this question. A tale of a hidden pyrimidine catabolism pathway (1) illustrates the power of combining both worlds. I will return to this topic after a brief discussion of specific implications of this study.

An elaborate machinery of pyrimidine catabolism revealed by Kustu and coworkers (1) is an important addition to the N-utilization network in *E. coli*. This machinery, a classical operon (now termed *rut* operon, see Fig. 1*A*), is under regulation of a new repressor RutR (former *ycdC* gene) competing with NtrC, whose role as a global "nitrogen dispatcher" was elucidated by Kustu's group earlier (9). The main physiological role of this pathway is not obvious, at least to members of that group, who suspect that it may contribute more to global cellular regulation than to mere nutrition (1). Nevertheless, in the best catabolic traditions, the *rut* operon is equipped with a transporter RutG (former *ycdG* gene). A final product of this pathway, hydroxypropionate (or 2-methylhydroxypropionate) is another novelty, distinguishing it from both "legacy" pathways described in other species and captured in the KEGG map (www.genome.jp/dbget-bin/www_bget?path:rn00240). The distinction from the reductive pathway is also supported by gene comparison, whereas no genetic data are available for the oxidative pathway described in early studies of Corynebacteria and other species (10–12).

**Table 1. Occurrence of genes involved in pyrimidine catabolism in completely sequenced genomes**

| | Organism | Novel Pathway, Enzymes | | | | | | Permease | Regulator | Reductive Pathway, Enzymes | | | PydABC transporter | | | | Permease |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | RutA | RutB | RutC | RutD | RutE | RutF | RutG | RutR | PydA | PydB | PydC | A | B1 | B2 | C | PydP |
| γ | *Escherichia coli (6 strains)* | + | + | + | + | + | + | + | + | | | | | | | | |
| γ | *Shigella flexneri (2 strains) + 2 ssp* | + | + | + | + | + | + | + | + | | | | | | | | |
| γ | *Yersinia enterocolitica (only)* | + | + | + | + | + | + | + | + | | | | | | | | |
| γ | *Acinetobacter sp. ADP1* | + | + | + | + | | + | + | + | | | | | | | | |
| γ | *Pseudomonas syringae (3 strains)* | + | + | + | + | | + | | + | | | | | | | | |
| α | *Caulobacter crescentus* | + | + | + | + | | | | + | | | | | | | | |
| α | *Agrobacterium tumefaciens (2 strains)* | + | + | + | + | + | + | | + | ? | + | + | + | + | + | + | |
| γ | *Pseudomonas aeruginosa (2 strains)* | | | | | | | | + | + | + | + | | | | | + |
| γ | *Pseudomonas fluoresens (3 strains)* | | | | | | | | + | + | + | + | | | | | + |
| γ | *Pseudomonas putida* | | | | | | | | + | + | + | + | | | | | + |
| β | *Burkholderia cepacia (2 strains)+2 ssp* | | | | | | | | + | + | + | + | | | | | + |
| α | *Sulfitobacter sp (2 strains)* | | | | | | | | + | + | + | + | + | + | + | + | |
| α | *Silicibacter pomeroy + 1 ssp* | | | | | | | | + | + | + | + | + | + | + | + | |
| α | *Loktanella vestfoldensis* | | | | | | | | + | + | + | + | + | + | + | + | |
| α | *Roseobacter sp.* | | | | | | | | + | + | + | + | + | + | + | + | |
| α | *Roseovarius nubinhibens + 1 ssp* | | | | | | | | + | + | + | + | + | + | + | + | |
| α | *Rhodobacterales bacterium* | | | | | | | | + | + | + | + | + | + | + | + | |
| α | *Rhodobacter sphaeroides* | | | | | | | | + | + | + | + | + | + | + | + | |
| α | *Sinorhizobium meliloti* | | | | | | | | + | + | + | + | + | + | + | + | |
| α | *Rhizobium leguminosarum* | | | | | | | | + | + | + | + | + | + | + | + | |
| α | *Mesorhizobium loti + 1 ssp* | | | | | | | | + | + | + | + | + | + | + | + | |
| α | *Brucella suis + 2 ssp* | | | | | | | | + | + | + | + | | | | | |
| F | *Bacillus clausii* | | | | | | | | | + | + | + | | | | | |
| | *Homo sapiens* | | | | | | | | | + | + | + | | | | | |
| | *Drosophila melanogaster* | | | | | | | | | + | + | + | | | | | |
| | *Caenorhabditis elegans* | | | | | | | | | + | + | + | | | | | |

Modified from The SEED (http://theseed.uchicago.edu/FIG/index.cgi) display. Columns are functional roles, and rows correspond to species (or groups of related species) with complete genomes. Taxonomic groups are marked: α, alphaproteobacteria; β, betaproteobacteria; γ, gammaproteobacteria; and F, firmicutes. The presence of gene orthologs is marked by +. Actual gene IDs are provided in the online version of pyrimidine utilization subsystem at: http://theseed.uchicago.edu/FIG/subsys.cgi?ssa_name=Pyrimidine_utilization&request=show_ssa. Genes that occur close to each other on the chromosome are highlighted by matching background colors.

Comparative genome analysis allowed Kustu and coworkers (1) to expand the impact of their findings beyond the model system of *E. coli*, as illustrated in Table 1, a condensed form of a pyrimidine utilization subsystem compiled with the SEED platform (http://theseed.uchicago.edu/FIG/index.cgi). The subsystem approach to genome annotation and pathway analysis allows us to capture the existing knowledge of biological processes and reliably project it across the entire collection of diverse species with completely sequenced genomes (13). In addition to establishing which organisms implement one or the other functional variants of a subsystem (e.g., reductive pathway or *rut* pathway), this approach helps to reveal gaps in knowledge (missing genes) and potential new players (predicted genes).

Conflict of interest statement: No conflicts declared.

See companion article on page 5114 in issue 13 of volume 103.

*E-mail: osterman@burnham.org.

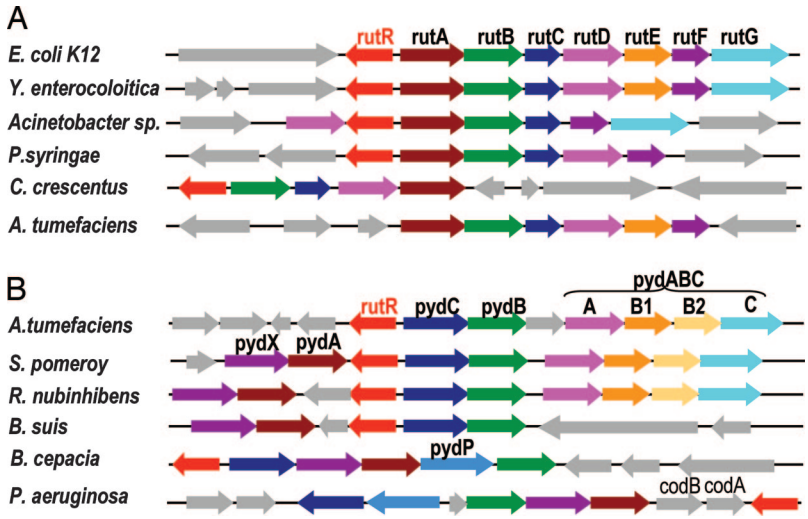


**Fig. 1.** Clustering on the chromosome of the genes involved in the novel pathway (*A*, *rut* genes) and in the previously characterized reductive pathway (*B*, *pyd* genes) of pyrimidine catabolism. Gene names are as in Table 1. Matching colors (within each panel) correspond to gene orthologs. A unique feature of the *pyd* operon in *P. aeruginosa* (but not in other Pseudomonadales) is the presence of genes encoding cytosine permease (*codB*, a homolog of *PydP*) and cytosine deaminase (*codA*), which would feed cytosine to the same utilization pathway.

Genes of the *rut* pathway are clustered on the chromosome in all species (see Table 1 and Fig. 1*A*), providing strong evidence of their functional coupling (14). Regulator RutR and most enzymatic components are conserved, except for putative reductase RutE, a possible subject of nonorthologous gene displacement in several species beyond Enterobacteria. An absence of RutG orthologs in some species may reflect an alternative transport system or an exclusive role for the *rut* pathway in controlling the internal pyrimidine pool.

The core of the reductive pathway is comprised of three enzymatic steps: (*i*) dihydropyrimidine dehydrogenase, (*ii*) dihydropyrimidinase, and (*iii*) β-ureidopropionase, historically characterized in a context of human metabolic disorders (15). Bacterial genes encoding functional homologs of these enzymes [*pydA*, *pydB*, and *pydC*, characterized in *Brevibacillus agri* (16)] form a chromosomal cluster conserved in a variety of species (see Fig. 1*B*). The most striking observation in the comparison of the two pathways is that the only gene shared by them is the regulator *rutR* (17), which is conserved and tightly clustered with both operons. Although this conservation may be rationalized by sharing of effectors (pyrimidines), an evolutionary scenario explaining this observation is not obvious. A single copy of *rutR* in *Agrobacterium tumefaciens* carrying both operons is clustered with *pyd* operon (see Fig. 1), which is also the only example of an incomplete reductive pathway. Although the absence of *pydA* in this organism could be explained by a nonorthologous displacement, the unique coexistence of the two pathways may invoke other explanations.

Transport mechanisms vary between pathways and between groups of species. Two types of transport systems can be inferred for the reductive pathway from chromosomal clustering (see Fig. 1*B*). A permease PydP characteristic of β- and γ-proteobacteria (e.g., PA0443 in *Pseudomonas aeruginosa*) appears to be functionally replaced by an ABC cassette in α-proteobacteria. An additional component of reductive pathway, termed here PydX (e.g., PA0440 in *P. aeruginosa*) can be inferred based on chromosomal clustering (14) and protein fusion (18) evidence. The latter is provided by an observed homology between PydX and the N-terminal domain of the human dihydropyrimidine dehydrogenase known to supply an electron transfer chain for a catalytic C-terminal domain (19), a homolog of bacterial PydA.

Notwithstanding the importance of specific findings discussed above, an impact of the published study (1) extends beyond this interesting subsystem providing us with a remarkable illustration of the power of combining functional genomics techniques, such as microarray analysis and phenotype screening, with comparative genomics for mapping of completely novel pathways. Analyzing the *rut* pathway in retrospect, one may notice a number of important clues provided just by bioinformatics: (*i*) chromosomal clustering and co-occurrence of *rut* genes indicated that they could belong to the same pathway, (*ii*) a homology-based annotation of RutG as nucleobase transporter hinted at a possible role of this pathway in utilization of pyrimidines or purines, and (*iii*) sharing of RutR orthologs between *rut* and *pyd* operons pointed to a possible involvement with pyrimidine catabolism. These clues, combined with sparse evidence in the literature, could induce a hypothesis amenable to focused experimental testing. This analysis (no matter how speculative it is with respect to an already accomplished study) is aimed to emphasize the tremendous opportunities provided by comparative analysis of multiple sequenced genomes. A growing number of successful examples (among which the tale of the hidden *rut* pathway is one of the most striking) and a growing awareness of comparative genomics tools will prompt many research groups to explore these opportunities, thus accelerating the pace of gene and pathway discovery.

I thank Dr. Sydney Kustu for her kind permission to include the results for the pyrimidine utilization subsystem into the SEED database (http://theseed.uchicago.edu/FIG/subsys.cgi?ssa_name=Pyrimidine_utilization&request=show_ssa) before their publication. I am grateful to Dr. Veronika Vonstein for discussion and help with this manuscript.

1. Loh, K. D., Gyaneshwar, P., Papadimitriou, E. M., Fong, R., Kim, K.-S., Parales, R., Zhou, Z., Inwood, W. & Kustu, S. (2006) *Proc. Natl. Acad. Sci. USA* **103,** 5114–5119.
2. Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997) *Science* **277,** 1453–1474.
3. Reed, J. L. & Palsson, B. O. (2004) *Genome Res.* **14,** 1797–1805.
4. Kang, Y., Durfee, T., Glasner, J. D., Qiu, Y., Frisch, D., Winterberg, K. M. & Blattner, F. R. (2004) *J. Bacteriol.* **186,** 4921–4930.
5. Mori, H. (2004) *J. Biochem. Mol. Biol.* **37,** 83–92.
6. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. & Mori, H. (February 21, 2006) *Mol. Syst. Biol.*, 10.1038/msb4100050.
7. Osterman, A. & Overbeek, R. (2003) *Curr. Opin. Chem. Biol.* **7,** 238–251.
8. Simaga, S. & Kos, E. (1981) *Int. J. Biochem.* **13,** 615–619.
9. Zimmer, D. P., Soupene, E., Lee, H. L., Wendisch, V. F., Khodursky, A. B., Peter, B. J., Bender, R. A. & Kustu, S. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 14674–14679.
10. Hayaishi, O. & Kornberg, A. (1952) *J. Biol. Chem.* **197,** 717–732.
11. Lara, F. J. (1952) *J. Bacteriol.* **64,** 271–277.
12. Wang, T. P. & Lampen, J. O. (1952) *J. Biol. Chem.* **194,** 775–783.
13. Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards,R., *et al.* (2005) *Nucleic Acids Res.* **33,** 5691–5702.
14. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 2896–2901.
15. Nyhan, W. L. (2005) *Mol. Genet. Metab.* **86,** 25–33.
16. Kao, C. H. & Hsu, W. H. (2003) *Biochem. Biophys. Res. Commun.* **303,** 848–854.
17. Low, L. Y., Yang, C., Perego, M., Osterman, A. & Liddington, R. C. (2005) *J. Biol. Chem.* **280,** 35433–35439.
18. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg, D. (1999) *Nature* **402,** 83–86.
19. Dobritzsch, D., Ricagno, S., Schneider, G., Schnackerz, K. D. & Lindqvist, Y. (2002) *J. Biol. Chem.* **277,** 13155–13166.