Letter to the Editor

Inference from Proteobacterial Operons Shows Piecewise Organization: A Reply to Price et al.

R. Fani,[1] M. Brilli,[1] P. Liò[2]

[1] Department of Animal Biology and Genetics, University of Florence, Via Romana 17-19, I-50125 Florence, Italy
[2] Computer Laboratory, Cambridge University, 15 JJ Thomson Avenue, Cambridge CB3 0FD, UK

Introduction

Price and collaborators (2006; preceding article) have recently reported an interesting and exhaustive analysis of *his* gene orders in a large number of bacterial genomes deposited in the MicrobesOnline database (http://www.microbesonline.org/). This, coupled with a phylogenetic analysis using distance-based methods of the His concatamers, showed that a *unified his-GDC(NB)BHAF(IE)* operon is present in many lineages. This result has led the authors to hypothesize that this operon is ancient, an assumption that is in partial disagreement with our previous results (Fani et al. 2005). However, our work did not focus on the evolution of *his* operons in general (i.e., in all completed genomes) but on the construction of the proteobacterial operons. Our main conclusion was that the *E. coli*-type operon was constructed piecewise during proteobacterial evolution starting from a proteobacterial ancestor whose genome probably contained a set of scattered *his* genes (or only partially clustered; see below the "*his* core," a cluster of four *his* genes arranged in the order *hisBHAF* [Fani et al. 1995]. This does not per se imply that the Last Universal Common Ancestor possessed scattered *his* genes.

The analysis of gene order is of great importance to understand the evolution of operon structures; however, it should be integrated with other information, such as gene regulative controls and the presence of fused genes. In this sense, histidine biosynthesis represents an excellent study model since it has been investigated for over 40 years and a wealth of experi-

Correspondence to: R. Fani, e-mail: renato.fani@unifi.it

mental and bioinformatics results is available. Therefore, we did not limit our study to the analysis of strings of sequences, but also considered other important aspects of *his* genes (e.g., presence/absence of given gene fusions, type of transcriptional regulation).

If we look in detail at the results of Price et al. (2006), we see that they support that *his* core is present in the proteobacterial ancestor, but not that the *E. coli*-type operon is ancient and common in other bacterial branches, because it has peculiar features.

The term *unified* operon introduced by Price et al. seems to indicate very similar structures in different organisms, but *his* operons may have very peculiar features in different taxonomic lineages, and these have to be taken into account when deriving the evolutionary history of these biosynthetic operons.

Gene Order

Price et al. propose that the unified *E. coli* type *his* operon is present in the Euryarchaeota; this is not true, since these archaeal operons lack the *hisNB* fusion and include *hisZ*, a gene which is not present in the *E. coli* genome (see below). In addition, there are hints concerning the *Pyrococcus furiosus/Thermococcus kodakarensis* and *Ferroplasma acidarmanus/Picrophilus torridus* operons suggesting that they were acquired by lateral gene transfer from a bacterial donor (Brilli and Fant, manuscript in preparation). Actually, these operons show complete synteny conservation with some bacterial operons, parallel with an evident divergence from all the other Archaea (Fig. 1). This is in agreement with data from Huynen
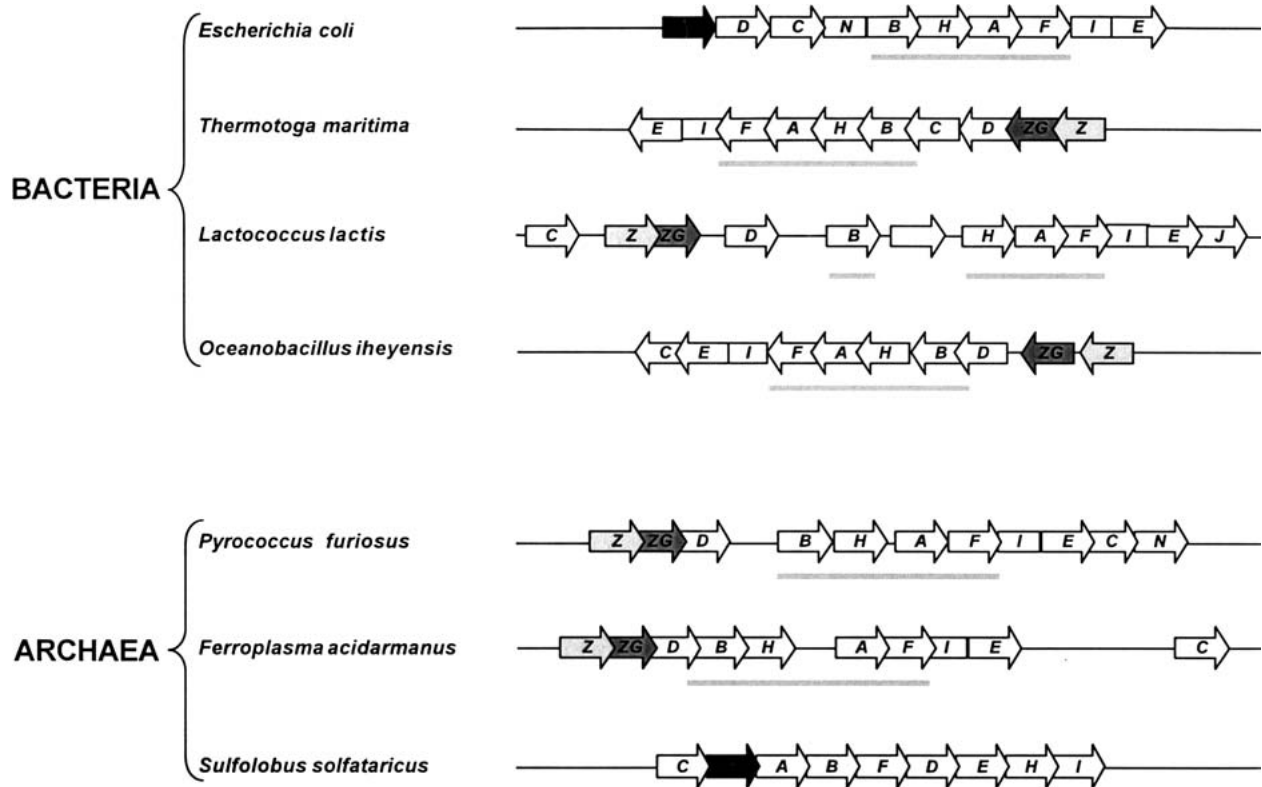
Fig. 1. Organization of *his* genes in some representatives of Archaea and Bacteria. The *E. coli*-like *hisG* gene is in black, the *hisZG* and *hisZ* genes are shown in dark and light gray, respectively. A gray bar indicates the genes forming the core of histidine biosynthesis.

and Bork (1998), who stated that "gene order is rarely conserved in evolution," hence "the presence in two distant evolutionary branches of the same order of genes, combined with the absence of this gene order in other more closely related branches, can point to horizontal transfer." Moreover, the *his* operon of these Euryarchaeota includes both a bifunctional *hisIE* gene and *hisZ*, the latter coding for a regulatory protein. Both of them are present in these four species (of two taxonomic groups) but never in other archaeal species, while both are common in Bacteria (Sissler et al. 1999; Vega et al. 2005). Finally, all the closest relatives of the above-mentioned archaeal species lack the entire set of histidine biosynthetic genes. The crenarchaeal *Sulfolobales*-like operon *(hisCGABFDEHI)* shows a gene order that is completely different from the bacterial type and it lacks the *his* core. The gene order of partial *his* clusters in other Archaea often resembles the *Sulfolobales* operon gene order but not bacterial gene orders (e.g., *hisAG* in *Methanopyrus kandleri, hisGAB* in *Methanosarcina mazei, hisAB* in *Archaeoglobus fulgidus;* these genes have never been found in these orders on bacterial chromosomes). These features, along with phylogenetic results, seem to indicate that the *resident* archaeal *his* operon is represented by the *Sulfolobus* type, while the *core* containing operons appeared in Bacteria and were later transferred to some Archaea.

Gene order differences in operons also occur in Bacteria, but a constant theme across different taxonomic groups is the presence of the *his* core, hence a construction of the *his* core outside, and before, proteobacterial divergence is possible (Brilli and Fani 2004; Fani et al. 2005). In our paper (Fani et al. 2005) we did not completely rule out this possibility and the new available genomes support this view. Nonetheless, *his* operons contain more genes than those in the core and their gene order is not as constant as the analysis by Price et al. (2006) seems to indicate: e.g., in some Firmicutes the *hisC* gene is located outside the operon and belongs to a different transcriptional unit (e.g., *Bacillus subtilis* and *Zymomonas mobilis* [Gu et al. 1995], containing *trp, aro,* and *tyr* genes, which are involved in other biosynthetic pathways HisC is involved in, in addition to histidine biosynthesis; on the contrary, in other Firmicutes, the *hisC* gene is the first of the operon and it is not involved in other pathways (e.g., *Lactococcus lactis* [Delorme et al. 1992].

Gene Fusions

E. coli-like operons contain the *hisNB* gene fusion, which is peculiar to γ-proteobacteria, except for documented cases of lateral gene transfer, such as *Campylobacter jejuni* (Brilli and Fani 2004). Thus, it is

quite unlikely that this fusion was present in the last common ancestor of Bacteria (or Proteobacteria) and then lost (often independently; see the $\alpha/\beta/\delta/\epsilon$-proteobacterial branches) in all but one (the γ-proteobacteria) taxonomic branch (see Brilli and Fani 2004).

The HOL-Pase Coding Gene

The *hisN* gene codes for a histidinol-phosphate phosphatase (HOL-Pase), and this activity has been ascribed to at least three different (and phylogenetically unrelated) types of enzymes, depending on the (group of) organisms considered: in at least some Firmicutes, the budding and fission yeasts, this enzyme belongs to the PHP (polymerases and histidinol phosphatases) family (le Coq et al. 1999; German and Hu 1969; Millay and Huston 1973). ln the γ-proteobacteria (and probably in other proteobacterial branches) it is part of the DDDD superfamily (Brilli and Fani 2004 and references therein). Moreover, in *Neurospora crassa* this activity is probably carried out by an aspecific alkaline phosphatase (Morales et al. 2000). ln most organisms the gene coding for this enzyme is completely unknown and often the sequence corresponding to known types of HOL-Pases cannot be retrieved in complete genomes. As previously suggested (Brilli and Fani 2004), this gene was very likely recruited in a later evolutionary stage of histidine biosynthesis evolution. Different types of phosphatases are devoted to histidinol dephosphorylation in different groups of organisms, and the position of the corresponding genes *in his* operons is highly variable, even between closely related organisms (i.e., Gram positives).

Feedback Inhibition

In *E. coli* and the other γ-proteobacteria the HisG enzyme is able to autoregulate its activity by binding histidine when the intracellular amino acid concentration exceeds a threshold level. A few years ago a different feedback inhibitory mechanism of histidine biosynthesis was disclosed in *L. lactis* (Sissler et al. 1999) and then found in an increasing number of microorganisms. This mechanism is based on the interaction between HisZ and a shorter version of the HisG protein, which is *per se* unable to bind histidine, a function performed by HisZ, which then turns off HisG. Thus, HisZ regulates HisG and it is not present in γ-proteobacteria (Kleeman and Parsons 1977; Boveè et al. 2002; Vega et al. 2005; Lohkamp et al. 2004; Fani et al. manuscript in preparation), representing a remarkable difference in the feedback regulatory mechanism of histidine biosynthesis that impacted on operon evolution too: if present, the *hisZ* and *hisG* genes are very often contiguous (and overlapping) on the genome. If clustered in the operon, the two genes occupy the first two positions.

Transcriptional Regulation

The ways in which *his* operon transcription is regulated can also vary among different organisms: several Firmicutes and γ-Proteobacteria possess an attenuator sequence upstream of *his* operons. However, the mechanisms are different in the two groups: the *E. coll* attenuator region contains seven *his* codons, which are directly involved in the attenuation mechanism (Alifano et al. 1997 and references therein), while the *Lactococcus lactis his* operon promoter lacks significant pairings with the anticodon of the tRNA[his] (Delorme et al. 1999) and histidine codons are not retrieved in its sequence.

All these differences indicate that also starting from partial clusters in the proteobacterial (or the LC) ancestor, the evolutionary history of the *E. coli*-like operons appears to be unique, as are many of its peculiar features.

Phylogenetic Analysis

Price et al. performed a phylogeny of the concatenated His proteins using a distance method. This approach has the main disadvantage, with respect to maximum likelihood (ML), that concatenated sequences are transformed into a single distance matrix, losing the information on each protein's effective contribution to the phylogeny. Moreover, the information in alignment columns containing gaps is not used. It is noteworthy that distance methods are more susceptible to long-branch attraction (Felsenstein 2004) than ML; this may happen with diverging amino acid sequences. Several studies show that ML methods generally outperform distance methods over a broad range of realistic conditions (Kuhner and Felsenstein 1994; Gaut and Lewis 1995; Huelsenbeck 1995).

Finally, we note that species such as *Aquifex*, the cyanobacteria and the T/D group have scattered genes. If the complete operon is really ancient, we have to postulate several independent partial or total destructions of the operon in these species and differential maintenance in those which differentiated in a later stage.

Conclusions

The scenario of a proteobacterial ancestor harboring a *his* core *(hisBHAF)*, and of different genes being joined to this aggregation nucleus at different times,

seems to be better supported by sequence and other functional data than other hypotheses.

The presence of a completely formed *E. coli*-like *his* operon very early in evolution does not appear to be supported by the comparative analysis of genomes and by the analysis of gene structure, organization, and regulation. This indicates that if this view is correct, then a great number of independent rearrangements would be necessary to explain the extant situation.

We cannot a priori exclude—and it was beyond the scope of the paper by Fani et al. (2005)—the possibility that a partial clustering of *his* genes occurred in the early stages of bacterial evolution. However, the *E. coli*-type operons containing all *his* genes in the order *hisGDC(NB)HAF(IE)* emerged later during evolution, as did the fine regulatory mechanisms which control transcription today in many species and the gene fusions found in the γ branch.

By this we can explain the present-day differences as different evolutionary outcomes which all started from a common set of histidine biosynthetic genes, but which happened in independent ways in the different lineages, in conformity with the different organisms' needs, which are directly related to the ecological niche the organism itself lives in.

References

Boveè ML, Champagne KS, Demeler B, Francklyn CS (2002) The quaternary structure of the HisZ-HisG N-1-(5′-phosphoribosyl)-ATP transferase from *Lactococcus lactis*. Biochemistry 41(39):11838–11846

Brilli M, Fan R (2004) Molecular evolution of *hisB* genes. J Mol Evol 58(2):225–237

Charlebois RL, Sensen CW, Doolittle WF, Brown JR (1997) Evolutionary analysis of the *hisCGABdFDEHI* gene cluster from the archaeon *Sulfolobus solfataricus* P2. J Bacteriol 179(13):4429–4432

Delorme C, Ehrlich SD, Renault P (1992) Histidine biosynthesis genes in *Lactococcus lactis* subsp. *lactis*. J Bacteriol 174(20):6571–6579

Delorme C, Ehrlich SD, Renault P (1999) Regulation of expression of the *Lactococcus lactis* histidine operorn. J Bacteriol 181(7):2026–2037

Fani R, Liò P, Lazcano A (1995) Molecular evolution of the histidine biosynthetic pathway. J Mol Evol 41(6):760–774

Fani R, Brilli M, Lio P (2005) The origin and evolution of operons: The piecewise building of the proteobacterial histidine operon. J Mol Evol 60:370–390

Felsenstein J (2004) Inferring priylogenies. Sinauer Associates, Sunderland, MA

Gaut BS, Lewis PO (1995) Success of maximum-likelihood phytogeny inference in the 4-taxon case. Mol Biol Evol 12:152–162

Gorman JA, Hu AS (1969) The separation and partial characterization of L-histidinol phosphatase and an alkaline phosphatase of *Saccharomyces cerevlsiae*. J Biol Chem 244 (6):1645–1650

Gu W, Zhao G, Eddy C, Jensen RA (1995) Imidazole acetol phosphate aminotransferase in *Zymomonas mobilis*: molecular genetic, biochemical, and evolutionary analyses. J Bacteriol 177(6):1576–1584

Kuhner MK, Felsentein J (1994) Simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol 11:459–468. See also Erratum. Mol Biol Evo (12:525, 1995)

Huelsenbeck JP (1995) Performance of phylogenetic methods in simulation. Syst Biol 44:17–48

Huynen MA, Bork P (1998) Measuring genome evolution. Proc Natl Acad Sci USA 95(11):5849–5856

Kleeman JE, Parsons SM (1977) Inhibition of histidyl-tRNA-adenosine triphosphate phosphoribosyltransferase complex formation by histidine and by guanosine tetraphosphate. Proc Natl Acad Sci USA 74(4):1535–l537

Le Coq D, Fillinger S, Aymerich S (1999) Histidinol phosphate phosphatase, catalyzing the penultimate step of the histidine biosynthesis pathway, is encoded by *ytvP (hisJ)* in *Bacillus subtilis*. J Bacteriol 181(10):3277–3280

Lohkamp B, McDermott G, Campbell SA, Coggins JR, Lapthorn AJ (2004) The structure of Escherichia coli ATP-phosphoribosyltransferase: identification of substrate binding sites and mode of AMP inhibition. J Mol Biol 336(1):131–144

Millay RH Jr, Houston LL (1973) Purification and properties of yeast histidinol phosphate phosphatase. Biochemistry. 12(14):259l–2596

Moore CM, Gaballa A, Hui M, Ye RW, Helmann JD (2005) Genetic and physiological responses of *Bacillus subtilis* to metal ion stress. Mol Microbiol 57(1):27–40

Morales AC, Nozawa SR, Thedei G Jr, Maccheroni W Jr, Rossi A (2000) Properties of a constitutive alkaline phosphatase from strain 74A of the mold *Neurospora crassa*. Braz J Med Biol 33(8):905–912

Price MN, Alm LJ, Arkin AP (2006) The histidine operon is ancient. J Mol Evol 62:807–808

Riggs DL, Mueller RD, Kwan HS, Artz SW (1986) Promoter domain mediates guanosine tetraphosphate activation of the histidine operon. Proc Natl Acad Sci USA 83(24):9333–9337

Shand RF, Blum PH, Holzschu DL, Urdea MS, Artz SW (1999) Mutational analysis of the histidine operon promoter of Salmonella typhimurium. J Bacteriol 171:6330–6337

Sissler M, Delorme C, Bond J, Ehrlich SD, Renault P, Francklyn C (1999) An aminoacyl-tRNA synthetase paralog with a catalytic role in histidine biosynthesis. Proc Natl Acad Sci USA 96(16):8985–8990

Stephens JC, Artz SW, Ames BN (1975) Guanosine 5′-diphosphate 3′-di phosphate (ppGpp): positive effector for histidine operon transcription and general signal for amino acid deficiency. Proc Natl Acad Sci USA 72(11):4389–4393

Vega MC, Zou P, Fernandez FJ, Murphy GE, Sterner R, Popov A, Wilmanns M (2005) Regulation of the hetero-octameric ATP phosphoribosyl transferase complex from *Thermotoga maritima* by a tRNA synthetase-like subunit. Mol Microbiol 55(3):675–686

Whelan S, Liò P, Goldman N (2001) Molecular phylogenetics: state-of-the art methods for looking into the past. Trends Genet 17(5):262–272

Winkler ME, Zawodny RV, Hartman PE (1979) Mutation *spoT* of *Escherichia coli* increases expression of the histidine operon deleted for the attenuator. J Bacteriol 139(3):993–1000