

# Enzyme evolution beyond gene duplication

## A model for incorporating horizontal gene transfer

Lianet Noda-García<sup>†</sup> and Francisco Barona-Gómez\*

Evolution of Metabolic Diversity Laboratory; Laboratorio Nacional de Genómica para la Biodiversidad (Langebio); Cinvestav-IPN; Irapuato, México

<sup>†</sup>Current affiliation: Department of Biological Chemistry; Weizmann Institute of Science; Rehovot, Israel

**Keywords:** enzyme evolution, horizontal gene transfer, substrate specificity, subHisA and PriA, tryptophan and histidine biosynthesis

**Abbreviations:** HGT, horizontal gene transfer; EAC, escape from adaptive conflict; DDC, duplication, degeneration and complementation; ProFAR, *N*'-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide; PRA, *N*'-(5'-phosphoribosyl)anthranilate

Submitted: 08/13/2013

Revised: 09/09/2013

Accepted: 09/10/2013

Citation: Noda-García L, Barona-Gómez F. Enzyme evolution beyond gene duplication: A model for incorporating horizontal gene transfer. *Mobile Genetic Elements* 2013; 3:e26439; <http://dx.doi.org/10.4161/mge.26439>

\*Correspondence to: Francisco Barona-Gómez; Email: [fbarona@langebio.cinvestav.mx](mailto:fbarona@langebio.cinvestav.mx)

Commentary to: Noda-García L, Camacho-Zarco AR, Medina-Ruiz S, Gaytán P, Carrillo-Tripp M, Fülöp V, Barona-Gómez F. Evolution of substrate specificity in a recipient's enzyme following Horizontal Gene Transfer. *Mol Biol Evol* 2013; 30:2024–34; PMID:23800623; <http://dx.doi.org/10.1093/molbev/mst115>

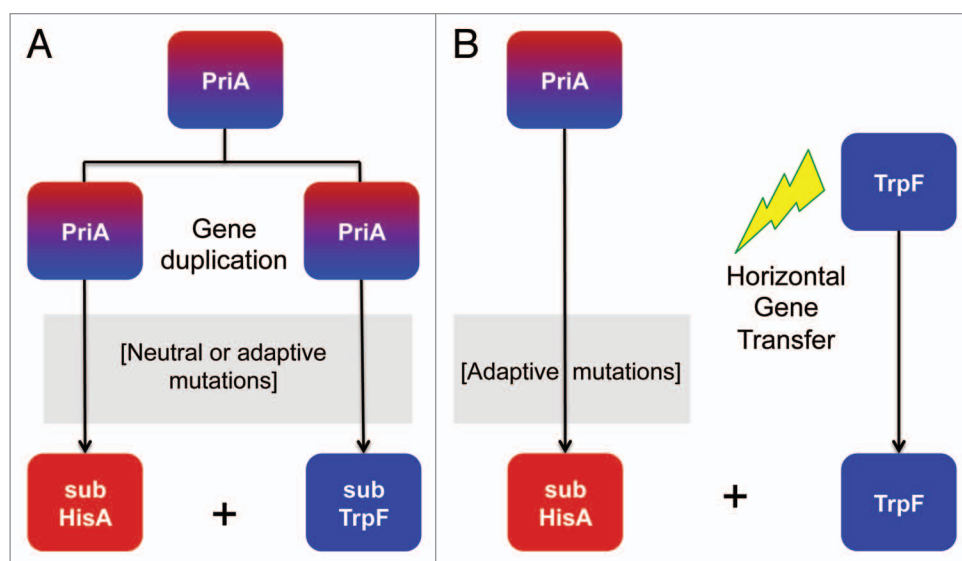
Understanding the evolution of enzyme function after gene duplication has been a major goal of molecular biologists, biochemists and evolutionary biologists alike, for almost half a century. In contrast, the impact that horizontal gene transfer (HGT) has had on the evolution of enzyme specialization and the assembly of metabolic networks has just started to being investigated. Traditionally, evolutionary studies of enzymes have been limited to either the function of enzymes *in vitro*, or to sequence variability at the population level, where in almost all cases the starting conceptual framework embraces gene duplication as the mechanism responsible for the appearance of genetic redundancy. Very recently, we merged comparative phylogenomics, detection of selection signals, enzyme kinetics, X-ray crystallography and computational molecular dynamics, to characterize the sub-functionalization process of an amino acid biosynthetic enzyme prompted by an episode of HGT in bacteria. Some of the evolutionary implications of these functional studies, including a proposed model of enzyme specialization independent of gene duplication, are developed in this commentary.

The core view of enzyme evolution is that gene duplication, followed by functionalization, is the primary mechanism by which novel enzymes have evolved.<sup>1-5</sup> Since it is assumed that at the dawn of life genetic resources were scarce, it is commonly believed that ancestral enzymes lacked specificity, and that this allowed

them to engage in many surrogate activities. This view led to the seminal proposal that modern enzymes are the result of substrate specialization processes prompted by gene duplication, and that current organisms evolved regulation of metabolism in a pathway-specific manner, only possible in the absence of substrate ambiguity.<sup>6</sup> Indeed, it is now broadly believed that functional divergence prior to duplication is a fundamental driving force of enzyme evolution.<sup>4,7</sup>

The most accepted models of enzyme evolution are sub-functionalization involving either escape from adaptive conflict (EAC) or duplication, degeneration and complementation (DDC). In the EAC model, gene duplication has been proposed to allow organisms to escape from adaptive conflicts imposed by 'generalist' or promiscuous enzymes that result from divergence prior to duplication.<sup>1,8,9</sup> During EAC the duplicated copies acquire adaptive mutations and become specialized toward one of the original functions. On the contrary, in the DDC model, the mutations that lead to sub-functionalization are neutral. One of the duplicated gene copies retains the original function, while the other evolves by acquiring mutations that render its product into a sub-specialized enzyme.<sup>1-5</sup>

These hypotheses might explain primitive mechanisms of enzyme evolution, mostly in the absence of available genetic resources; however, it has been proposed that horizontal gene transfer (HGT), and not gene duplication, is the main mechanism by which prokaryotes expand their gene and protein families.<sup>10</sup> What is the



**Figure 1.** Models of enzyme sub-functionalization. When loss of a *trpF* gene occurs evolution of a proficient dual-substrate PriA enzyme takes place after positive selection. From that stage onwards, two possible scenarios can be envisaged. **(A)** The most accepted model of enzyme evolution implies gene duplication followed by narrowing of substrate specificity (subHisA and subTrpF), which is expected to occur after relaxation of the purifying selection. **(B)** Alternatively, differential gain of a *trpF* gene after HGT (marked with a lightning bolt), which does not imply the genetic and functional redundancy that is generated after gene duplication, could also take place. Thus, the sub-functionalization of PriA is constrained since purifying selection for the original function, i.e., HisA, is still in place. Moreover, the need for integrating the HGT-acquired whole-pathway *trp* operon into the receiving metabolic network, by means of altering substrate specificity, implies positive selection.

effect of HGT upon the topology of metabolic networks and enzyme features such as substrate specificity within the receiving organism is a question that remains unanswered. Indeed, although many reports show the pervasive occurrence of HGT in the prokaryotic kingdom, likely to be promoted by the social behavior of bacteria within the communities in which they evolve, insights into the mechanisms allowing the receiving organisms to adapt to the consequences of HGT are scarce.

Multi-specific enzymes from prokaryotic central metabolism, as they are well-integrated within the metabolic network, may advance our current understanding of enzyme evolution beyond what we have learnt using enzymes involved in peripheral pathways.<sup>11</sup> Ideally, these models should be phylogenetically traceable at a whole genome level. In our previous study we adopted and further developed, within a phylogenomics framework, an enzyme model from amino acid biosynthesis.<sup>12</sup> This model has been postulated and adopted as well by others as a suitable system for testing evolutionary hypotheses.<sup>11,13,14</sup> At the center of this model stands a dual-substrate ( $\beta\alpha$ )<sub>8</sub>-barrel phosphoribosyl isomerase (*priA* gene), which is a

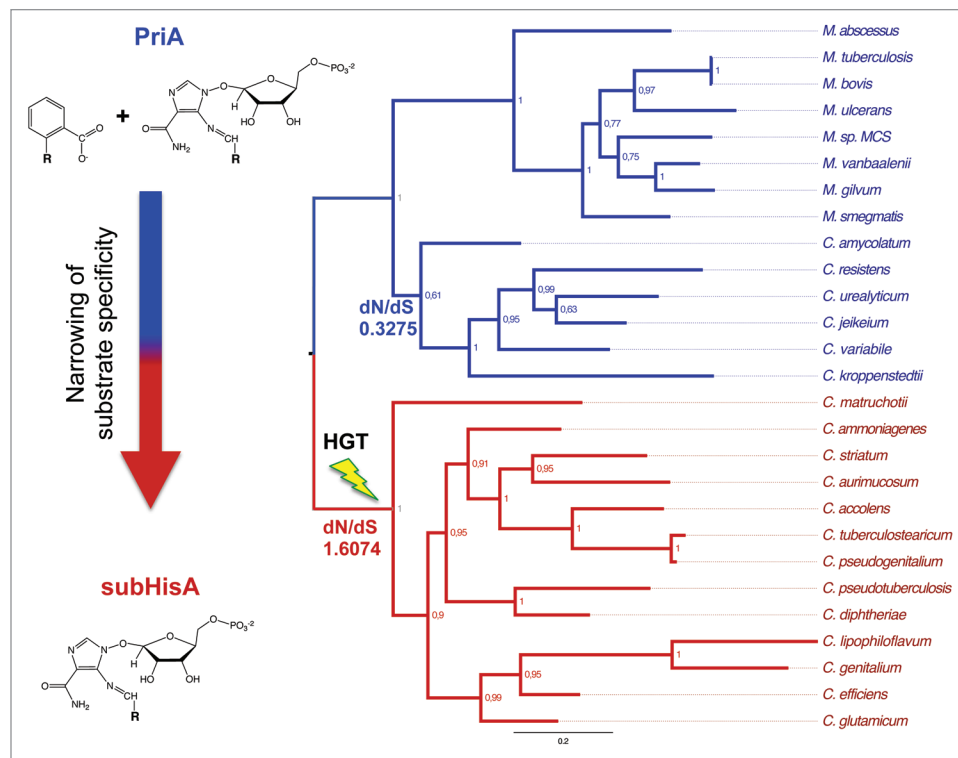
close homolog (~50% ID) of the histidine biosynthetic *bisA* gene.<sup>12</sup>

The product of the *priA* gene has been shown to participate in both the biosynthesis of L-histidine [HisA, *N'*-(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase] and L-tryptophan [TrpF, *N'*-(5'-phosphoribosyl)anthranilate (PRA) isomerase] in the model actinobacteria *Streptomyces coelicolor* and *Mycobacterium tuberculosis*, Gram positive bacteria with high (G + C) content.<sup>12,15</sup> The discovery of PriA stemmed from the realization that after whole genome sequencing of these organisms the *trpF* gene could not be found.

Given that the existence of a generalist enzyme like PriA could impose a metabolic disadvantage, we initially hypothesized the occurrence of gene duplication events that could lead to pathway-specific PriA enzymes with narrow substrate specificities for ProFAR and PRA substrates (Fig. 1A). Thus, we searched for *priA* duplicates, with an emphasis on actinobacterial genome sequences, including closely related organisms as out-groups. Despite it was found that the majority of the organisms analyzed lack a *trpF* gene,

these analyses failed to identify any sign of *priA* paralogy. In contrast, based in co-occurrence of *trpF* and *priA* genes, scenarios involving HGT or differential gene loss, could be identified (Fig. 1B).

Indeed, and in accordance with previous reports,<sup>16,17</sup> we found *trpF* homologs in several species from the genus *Corynebacterium*, which is closely related to the genus *Mycobacterium* (Fig. 2). These organisms, such as *C. glutamicum*, have acquired the *trpF* gene by HGT as part of a whole-pathway *trp* operon from a member of the *Gammaproteobacteria*.<sup>16,17</sup> After this HGT event, genomic and enzyme evolution events occurred, including (1) partial homologous displacement of the original corynebacterial *trp* genes and (2) complete loss of PRA isomerase activity of their cognate PriA homologs, which we explained after identification of mutations within the active sites of these enzymes and loss of protein conformational diversity.<sup>18</sup> Thus, the PriA homologs from this sub-clade of the genus *Corynebacterium* were renamed as subHisA, in order to reflect the sub-functionalization process—involving HGT and not gene duplication—that renders these enzymes with narrow substrate specificity with metabolic implications.



**Figure 2.** Positive selection during evolution of subHisA from PriA after HGT. PriA/subHisA phylogeny was constructed with MrBayes (posterior probability at each node is indicated). PriA and subHisA are shown with blue and red branches, respectively. The tree is rooted using the genus *Mycobacterium*. The proposed HGT event of *trpF* together with other *trp* genes, which coincides with speciation of one of the *Corynebacterium* lineages, is marked with a lightning bolt. The dN/dS values for ancestral PriA (blue) and subHisA (red) nodes are specified. The common phosphoribosyl moiety of ProFAR and PRA, whose chemical structures are shown, is denoted with an R.

As mentioned, the subHisA sequences are limited to the genus *Corynebacterium*. Interestingly, however, this genus also includes species with PriA enzymes, sharing up to a ~70% sequence identity (Fig. 2). Inspection of this enzyme sub-family within the phylogenetic tree suggests sequence divergence, which could be related to the evolution of enzyme substrate specificity. Indeed, sequence analyses revealed that the shape parameters of PriA ( $\alpha = 1.6631$ ) and subHisA ( $\alpha = 0.8311$ ) independent sub-trees are different and show higher rate heterogeneity for subHisA. Moreover, and in disagreement with the DDC model of enzyme sub-functionalization,<sup>1-3</sup> preliminary dN/dS ( ) analysis showed signs of positive selection within the subHisA branch, which correlates with acquisition of the whole-pathway *trp* operon by HGT (Fig. 2; see ref. 19).

Narrowing of enzyme substrate specificity in subHisA shows that, even within a highly constrained active site, loss of one of the ancestral activities can occur

without compromising the catalytic efficiency of the remaining enzyme function. Thus, given that most random mutations are deleterious to protein stability, which implies a trade-off during evolution<sup>20</sup> and that in the DDC sub-functionalization model beneficial mutations need not be invoked,<sup>2</sup> detection of positive selection in the branch where subHisA has evolved, may question the universality of this model to include enzyme substrate sub-functionalization in purely biochemical grounds. Indeed, the DDC model calls upon evolution of gene duplicates in eukaryotes, mainly through tissue-specific gene regulation,<sup>2</sup> and thus this may be the origin of this discrepancy.

Moreover, the occurrence of positive selection, originally implied by the EAC model<sup>1</sup> but which still lacks 'smoking gun' evidence involving a generalist ancestral enzyme,<sup>8,21</sup> can be explained if the mutations identified in the active site of subHisA were positively selected to accommodate the feedback regulated HGT-acquired whole-pathway *trp* operon

within the metabolism of the cognate *Corynebacterium* species. Interestingly, in *C. glutamicum*, where we confirmed the existence of a subHisA enzyme, it has been shown that feedback gene regulation of L-tryptophan<sup>16,17,22,23</sup> and L-histidine<sup>24</sup> biosynthesis has recently evolved. The implication of this observation is that multi-specific enzymes connecting central metabolic pathways will impose a conflict to the cell if pathway-specific regulatory regimens are needed.

Therefore, our model of enzyme evolution independent of gene duplication proposes that during dynamic genome processes involving HGT strong positive selection is needed to drive both (1) narrowing of enzyme substrate specificity from a generalist enzyme; and (2) efficient assembly of HGT-acquired biosynthetic pathways within the receiving metabolic network. In the presence of a PRA isomerase encoded by *trpF*, these mutations may have provided subHisA with an adaptive mechanism to avoid productive binding of PRA within a promiscuous-prone enzyme

active site. More generally, our model emphasizes the need for an integrated view on the evolution of enzyme substrate specificity, which shall accommodate prokaryotic genetics and physiology, as well as HGT.

In conclusion, we anticipate that the discovery of a novel ( $\beta\alpha$ )<sub>8</sub>-isomerase sub-family, i.e., subHisA, will open the door to explore the mutational effects, within a naturally occurring evolutionary trajectory, upon the fitness landscape both in

terms of the biochemical and biophysical features of enzymes, something that has remained elusive.<sup>25</sup> Our proposed evolutionary model of enzyme sub-functionalization after HGT or lineage-specific gene loss emphasizes the need to further investigate the neglected role of prokaryotes in evolutionary theory. It remains to be seen whether our model can be formalized within bacterial population genetics theory, once more examples similar to the

evolution of subHisA from PriA become available.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Acknowledgments

This work and related studies were supported by grants from The Royal Society of the UK, The UCMEXUS and Conacyt, Mexico.

#### References

- Hughes AL. The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci* 1994; 256:119-24; PMID:8029240; <http://dx.doi.org/10.1098/rspb.1994.0058>
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999; 151:1531-45; PMID:10101175
- Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* 2008; 9:938-50; PMID:19015656; <http://dx.doi.org/10.1038/nrg2482>
- Khersonsky O, Tawfik DS. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu Rev Biochem* 2010; 79:471-505; PMID:20235827; <http://dx.doi.org/10.1146/annurev-biochem-030409-143718>
- Dittmar K, Liberles D. (2010) Evolution after Gene Duplication. Hoboken, New Jersey: Wiley-Blackwell
- Jensen RA. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* 1976; 30:409-25; PMID:791073; <http://dx.doi.org/10.1146/annurev.mi.30.100176.002205>
- Piatigorsky J. (2007) Gene sharing and evolution: the diversity of protein function. Cambridge, Massachusetts: Harvard University Press. 320 p.
- Depristo MA. The subtle benefits of being promiscuous: adaptive evolution potentiated by enzyme promiscuity. *HFSP J* 2007; 1:94-8; PMID:19404414; <http://dx.doi.org/10.2976/1.2754665>
- Des Marais DL, Rausher MD. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 2008; 454:762-5; PMID:18594508
- Treangen TJ, Rocha EP. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* 2011; 7:e1001284; PMID:21298028; <http://dx.doi.org/10.1371/journal.pgen.1001284>
- Näsval J, Sun L, Roth JR, Andersson DI. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* 2012; 338:384-7; PMID:23087246; <http://dx.doi.org/10.1126/science.1226521>
- Barona-Gómez F, Hodgson DA. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep* 2003; 4:296-300; PMID:12634849; <http://dx.doi.org/10.1038/sj.embor.embor771>
- Glasner ME, Gerlt JA, Babbitt PC. Mechanisms of protein evolution and their application to protein engineering. [xii-xiii]. *Adv Enzymol Relat Areas Mol Biol* 2007; 75:193-239, xii-xiii; PMID:17124868
- List F, Sterner R, Wilmanns M. Related ( )8-barrel proteins in histidine and tryptophan biosynthesis: a paradigm to study enzyme evolution. *ChemBiochem* 2011; 12:1487-94; PMID:21656890; <http://dx.doi.org/10.1002/cbic.201100082>
- Due AV, Kuper J, Geerlof A, von Kries JP, Wilmanns M. Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *Proc Natl Acad Sci U S A* 2011; 108:3554-9; PMID:21321225; <http://dx.doi.org/10.1073/pnas.1015996108>
- Xie G, Bonner CA, Song J, Keyhani NO, Jensen RA. Inter-genomic displacement via lateral gene transfer of bacterial trp operons in an overall context of vertical genealogy. *BMC Biol* 2004; 2:15; PMID:15214963; <http://dx.doi.org/10.1186/1741-7007-2-15>
- Xie G, Keyhani NO, Bonner CA, Jensen RA. Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiol Mol Biol Rev* 2003; 67:303-42; PMID:12966138; <http://dx.doi.org/10.1128/MMBR.67.3.303-342.2003>
- Noda-García L, Camacho-Zarco AR, Medina-Ruiz S, Gaytán P, Carrillo-Tripp M, Fülöp V, Barona-Gómez F. Evolution of Substrate Specificity in a Recipient's Enzyme Following Horizontal Gene Transfer. *Mol Biol Evol* 2013; 30:2024-34; PMID:23800623; <http://dx.doi.org/10.1093/molbev/mst115>
- Noda-García L. (2012) Estudio de la evolución molecular de la función enzimática usando como modelo una enzima con características ancestrales. PhD Thesis, Irapuato: Cinvestav-IPN. 268 p
- DePristo MA, Weinreich DM, Hartl DL. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet* 2005; 6:678-87; PMID:16074985; <http://dx.doi.org/10.1038/nrg1672>
- Barkman T, Zhang J. Evidence for escape from adaptive conflict? *Nature* 2009; 462:E1, discussion E2-3; PMID:20010636; <http://dx.doi.org/10.1038/nature08663>
- Ikeda M. Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. *Appl Microbiol Biotechnol* 2006; 69:615-26; PMID:16374633; <http://dx.doi.org/10.1007/s00253-005-0252-y>
- Brune I, Jochmann N, Brinkrolf K, Hüser AT, Gerstmeir R, Eikmanns BJ, Kalinowski J, Pühler A, Tauch A. The lclR-type transcriptional repressor LtbR regulates the expression of leucine and tryptophan biosynthesis genes in the amino acid producer *Corynebacterium glutamicum*. *J Bacteriol* 2007; 189:2720-33; PMID:17259312; <http://dx.doi.org/10.1128/JB.01876-06>
- Jung S, Chun JY, Yim SH, Lee SS, Cheon CI, Song E, Lee MS. Transcriptional regulation of histidine biosynthesis genes in *Corynebacterium glutamicum*. *Can J Microbiol* 2010; 56:178-87; PMID:20237580; <http://dx.doi.org/10.1139/W09-115>
- Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 2007; 445:383-6; PMID:17251971; <http://dx.doi.org/10.1038/nature05451>