Gene 549 (2014) 161–170

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

# The evolutionary history of the genes involved in the biosynthesis of the antioxidant ergothioneine

Gary W. Jones, Sean Doyle, David A. Fitzpatrick *

*Department of Biology, The National University of Ireland Maynooth, Maynooth, County Kildare, Ireland*

ARTICLE INFO

*Article history:*
Received 24 June 2014
Received in revised form 22 July 2014
Accepted 25 July 2014
Available online 26 July 2014

*Keywords:*
Ergothioneine
Antioxidant
Fungal
Prokaryotes
Phylogenetics
Fusion gene

ABSTRACT

Ergothioneine (EGT) is a histidine betaine derivative that exhibits antioxidant action in humans. EGT is primarily synthesized by fungal species and a number of bacterial species. A five-gene cluster (*egt*A, *egt*B, *egt*C, *egt*D & *egt*E) responsible for EGT production in *Mycobacteria smegmatis* has recently been identified. The first fungal biosynthetic EGT gene (*NcEgt-1*) has also been identified in *Neurospora crassa*. *NcEgt-1* contains domains similar to those found in *M. smegmatis egt*B and *egt*D. EGT is biomembrane impermeable. Here we inferred the evolutionary history of the EGT cluster in prokaryotes as well as examining the phyletic distribution of Egt-1 in the fungal kingdom. A genomic survey of 2509 prokaryotes showed that the five-gene EGT cluster is only found in the Actinobacteria. Our survey identified more than 400 diverse prokaryotes that contain genetically linked orthologs of *egt*B and *egt*D. Phylogenetic analyses of Egt proteins show a complex evolutionary history and multiple incidences of horizontal gene transfer. Our analysis also identified two independent incidences of a fusion event of *egt*B and *egt*D in bacterial species. A genomic survey of over 100 fungal genomes shows that Egt-1 is found in all fungal phyla, except species that belong to the Saccharomycotina subphylum. This analysis provides a comprehensive analysis of the distribution of the key genes involved in the synthesis of EGT in prokaryotes and fungi. Our phylogenetic inferences illuminate the complex evolutionary history of the genes involved in EGT synthesis in prokaryotes. The potential to synthesize EGT is a fungal trait except for species belonging to the Saccharomycotina subphylum.



## 1. Introduction

Ergothioneine (EGT) is a histidine betaine derivative with a thiol group located on the $C_2$ atom of the imidazole ring (Genghof et al., 1956; Hartman, 1990; Melville et al., 1957). EGT was first isolated from the ergot fungus *Claviceps purpurea* (Tanret, 1909) and subsequent observational studies suggested that it is primarily synthesized by fungal species and a number of bacterial species particularly those belonging to the Actinobacterial and Cyanobacterial phyla (Genghof, 1970; Genghof and Vandamme, 1964; Genghof et al., 1956; Pfeiffer et al., 2011).

Although EGT is present in plants and animals they do not synthesize it but rather obtain it from nutrients. For example, plants acquire EGT through their roots and sometimes via actinomycete symbionts (Park et al., 2010). Animals acquire EGT from food including mushrooms, garlic, wheat, oats and beans which have been shown to have concentrations of EGT ranging from 210 ng $mg^{-1}$ to 2600 ng $mg^{-1}$ (Dubost et al., 2007). After ingestion EGT is retained with minimal metabolism (Pfeiffer et al., 2011). In humans, EGT is concentrated in erythrocyte progenitor cells, monocytes, the intestine and kidneys. EGT is cell membrane impermeable and recent human studies have identified a specific plasma membrane bound organic cation transporter known as OCTN1, which is encoded by the gene SLC22A4 (Grundemann et al., 2005). Polymorphisms in SLC22A4 have been associated with diabetes (Santiago et al., 2006) and susceptibility to chronic inflammatory diseases such as Crohn's disease (Fisher et al., 2006; Leung et al., 2006; Peltekova et al., 2004). A specific EGT transporter suggests a beneficial role for EGT and multiple studies support an antioxidant function for EGT but its specific mode of action remains unclear (Cheah and Halliwell, 2012; Hartman, 1990).

Mycobacteria do not synthesize the thiol glutathione (GSH), which is known for its efficient detoxification of free radicals as well as reactive oxygen and nitrogen species. Instead it produces two low-molecular-weight thiols, mycothiol (MSH) and EGT (Genghof and Vandamme, 1964; Newton et al., 1995, 1996). In *Mycobacterium tuberculosis* evidence suggests that MSH is involved in detoxifying reactive oxygen species (Vilcheze et al., 2008). In *Mycobacteria smegmatis* MSH deficient mutants, the levels of the organic hydroperoxide resistance protein and ERG are elevated, suggesting that ERG may partly compensate for the loss of MSH and thus have a role as an antioxidant (Ta et al.,

*Abbreviations:* EGT, ergothioneine; HSP, highest scoring pair; GSH, glutathione; MSH, mycothiol.

* Corresponding author.
*E-mail addresses:* Gary.Jones@nuim.ie (G.W. Jones), Sean.Doyle@nuim.ie (S. Doyle), david.fitzpatrick@nuim.ie (D.A. Fitzpatrick).

http://dx.doi.org/10.1016/j.gene.2014.07.065
0378-1119/

2011). ERG has also been implicated in modulating the immune response (Rahman et al., 2003) and in the inhibition of metalloenzymes, preventing oxidation of DNA and protein due to its metal-chelating properties (Zhu et al., 2010), this implies that it may also act as a virulence factor. Initial investigations into the function of EGT in fungi showed that in *Neurospora crassa* it helps protect conidia during the quiescent period between conidiogenesis and germination, and protects conidia during the germination process from the toxicity of peroxide (Bello et al., 2012).

The genes (*egt*A, *egt*B, *egt*C, *egt*D & *egt*E) responsible for EGT production in *M. smegmatis* have recently been identified (Seebeck, 2010). These genes are found adjacent to one another in a five-gene cluster but are predicted not to be essential for growth of *M. tuberculosis* laboratory strain H37Rv (Griffin et al., 2011; Sassetti et al., 2003). Seebeck (2010) cloned EgtD and has shown it to be a histidine methyltransferase that converts histidine to hercynine in an *S*-adenosyl methionine (SAM) dependent manner (Fig. 1-A). Sequence similarity of EgtA to γ-glutamylcysteine ligase suggests that γ-Glu-Cys rather than Cys is a sulphur donor (Seebeck, 2010). This observation was confirmed when cloned EgtB (contains an Fe(II) binding site) was assayed with hercynine and γ-Glu-Cys in the presence of $FeSO_4$ and shown to produce *S*-hercynyl-γ-glutamylcysteine (Fig. 1-A) (Seebeck, 2010). Addition of cloned EgtC to the reaction generated hercynylcysteine sulfoxide and cloned EgtE (β-lyase) in the presence of pyridoxal-5′-phosphate produced EGT (Fig. 1-A) (Seebeck, 2010).

Recently the first fungal biosynthetic EGT gene (*NcEgt-1*) was identified in *N. crassa* (Bello et al., 2012). *NcEgt-1* catalyzes the first two steps of EGT biosynthesis from histidine to hercynine to hercynylcysteine sulfoxide (Fig. 1-B). Comparisons between wild type and *NcEgt-1* indicate that EGT plays an important protective role against the toxicity of peroxide in conidia during germination (Bello et al., 2012). Interestingly *NcEgt-1* contains domains similar to those found in *M. smegmatis egt*B and *egt*D and is most likely the result of a fusion between ancestral fungal EgtB and EgtD genes (Bello et al., 2012).

No analysis to date has attempted to fully uncover the evolutionary history of the 5-gene *egt* cluster in prokaryotes or indeed that of Egt-1 in the fungal kingdom. As providing a comprehensive analysis of the phyletic distribution of these genes in the tree of life, we have performed in-depth phylogenetic analyses of these genes. Our results give a detailed overview of the distribution of the key genes involved in the synthesis of EGT in prokaryotes and fungi. Our phylogenetic inferences illuminate the complex evolutionary history of the genes involved in EGT synthesis in prokaryotes.

## 2. Methods

### 2.1. Sequence data and database searches

Amino acid sequences from all completely sequenced prokaryotic genomes were obtained from the NCBI ftp site. The list of the genomes utilized and their taxonomic affiliations are listed in Additional file 1-A. Complete bacterial genomes were utilized to ensure that potential EGT clusters could be identified. In total our dataset contained 7,850,632 amino acid sequences from 2509 genomes (Additional file 1-A). The Proteobacterial phylum is the most densely sampled accounting for ~42% of all genomes, followed by the Firmicutes phylum ~21% and the Actinobacteria phylum ~10.5% (Additional file 1-A).

The five genes (*egt*ABCDE) from the *Mycobacterium smegmatis* JS623 EGT cluster were used as database query sequences (GenBank GI numbers 433650597, 433650596, 433650595, 433650594 and 433650593 respectively). Taking one EGT protein at a time, putative orthologs were identified using a reciprocal BlastP (Altschul et al., 1997) search with a cutoff expectation (E) value of $10^{-5}$. Each EGT gene was searched against an individual bacterial genome. The top significant hit was recorded and searched back against the *M. smegmatis* genome to ensure a reciprocal top hit. Putative orthologs are listed in Additional file 1-B.

Our fungal protein dataset consisted of 103 genomes and 1001217 individual genes (Additional file 1-C). Where available, data was obtained

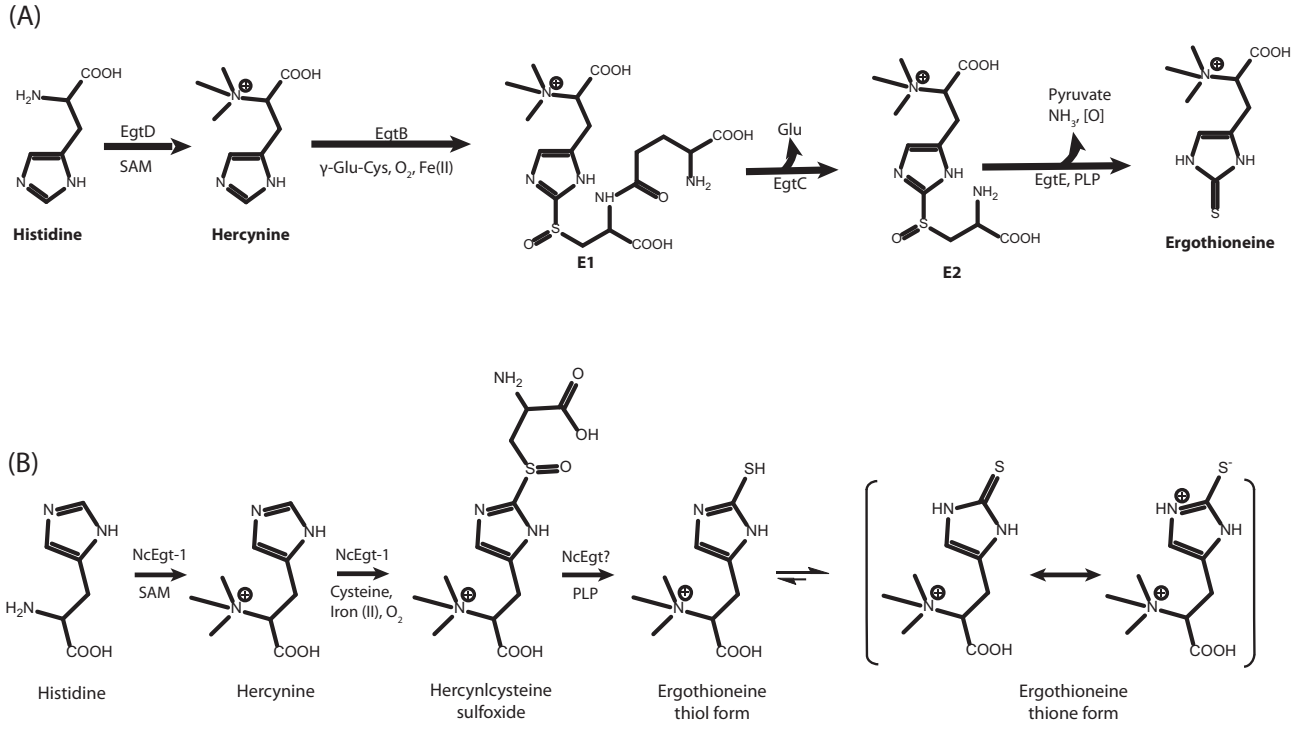


**Fig. 1.** A) Reaction sequence of ergothioneine biosynthesis in *Mycobacterium smegmatis*, redrawn from (Seebeck, 2010). B) The proposed ergothioneine biosynthetic pathway in *Neurospora crassa*.
Redrawn from Bello et al. (2012).