Annotation of the *Cupriavidus metallidurans* Genome from Locus Tags RMET_RS03075 to RMET_RS03095

Sabrina & Savannah Battleson, Rosalinda Mormina, Rachel Wach, Jillian Weber and Keith Kwas
Eden Jr/Sr High School Eden, N.Y. and The Western New York Genetics in Research and Health Care Partnership

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health

Abstract

A group of five consecutive genes from the microorganism *Cupriavidus metallidurans*, RMET_RS03075 – RMET_RS03095 were annotated using the collaborative genome annotation website GENI-ACT. The GenBank proposed gene product name for each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, potential alternative open reading frames, enzymatic function, presence or absence of gene duplication and degradation, the possibility of horizontal gene transfer, and the production of an RNA product. The GenBank proposed gene product name did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes appear to be correctly annotated the computer database.

Introduction

Cupriavidus metallidurans (previously *Ralstonia metallidurans*) type strain CH34 (hereafter *C. metallidurans*) is a Gram-negative , motile, non-spore forming, rod-shaped bacterium known for its ability to resist toxic heavy metals; metallidurans literally translates to "metal-enduring". When cultured on trypticase soy agar (TSA) at 30°C for 24h, *C. metallidurans* forms round, smooth, flat and convex colonies that are transparent and about 0.5mm in diameter. *C. metallidurans* is a mesophile; it has not been shown to grow under conditions outside of a range of 20-37°C.

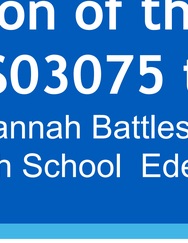


Figure 1. *Cupriavidus metallidurans* ultra-thin section containing a gold nanoparticle.

This bacterium's high number of heavy metal resistance genes and ability to thrive in toxic environments make it an important model organism in studying the mechanisms by which diverse microbes deal with heavy metal stress. *C. metallidurans* is also able to precipitate solid gold from aqueous gold (III) tetrachloride, which is a potent heavy metal toxin to most organisms [1].

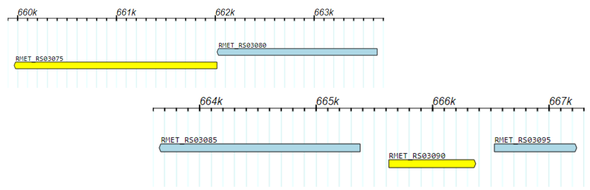


Figure 1. The locus tags and relative position of the genes under investigation in this research

Methods

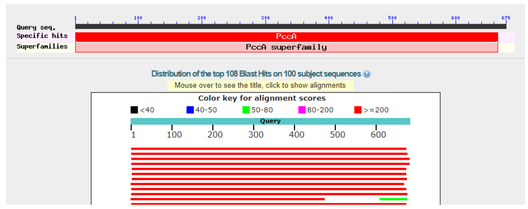
Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Cupriavidus metallidurans* genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of the gene and protein? Where is it located in the genome?
Sequence-Based Similarity	Blast, CDD, T-Coffee, WebLogo	How similar is the protein under investigation to other proteins in GenBank?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Cellular Localization	Gram Stain, TMHMM, SignalP, LipoP, Psortb, Phobius	Is the protein under investigation located in the cytoplasm, secreted, in the periplasm or embedded in the cell membrane or cell wall?
Enzymatic Function	KEGG, MetaCyc, E.C. Number	In what process or structure is the protein under investigation involved?
Duplication and Degradation	Paralog, Pseudogene	Are there other forms of the protein under investigation in the same genome? Is it functional?
Horizontal Gene Transfer	Phylogenetic Tree, Gene Neighborhood, GC Content	Has the protein under investigation co-evolved with the rest of the genome or has it been obtained in a different way?
RNA family	Rfam	Does the gene under investigation encode a functional RNA?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

RMET_RS03075:

The initial proposed product of this gene from GENI-ACT was a Methylcrotonyl-CoA carboxylase alpha. This was supported by the fact that we could use Swiss-Prot. The top six results of the Blast are shown below.



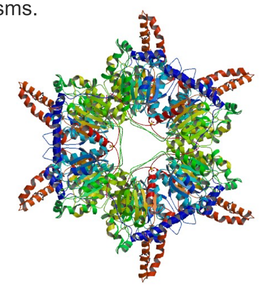
	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
†; AltName: Full=3-methylcrotonyl-CoA ca	678	678	98%	0.0	51.81%	Q54KE6.1
†; AltName: Full=3-methylcrotonyl-CoA ca	627	627	98%	0.0	47.15%	Q5I0C3.1
†; AltName: Full=3-methylcrotonyl-CoA ca	622	622	99%	0.0	47.51%	Q96RQ3.3
†; AltName: Full=3-methylcrotonyl-CoA ca	615	615	99%	0.0	47.54%	Q99MR8.2
Flags: Precursor [Oryza sativa Japonica	609	609	98%	0.0	47.60%	Q2QMG2.2
Flags: Precursor [Arabidopsis thaliana]	606	606	98%	0.0	46.33%	Q42523.2

Organisms from Blast Results:

1. Dictyostelium discoideum (Soil-dwelling amoeba, slime mold)
2. Rattus norvegicus (Norway rat)
3. Homo sapiens (human)
4. Mus musculus (house mouse)
5. Oryza sativa Japonica Group (Japanese rice)
6. Arabidopsis thaliana (thale cress)

RMET_RS03080:

The initial proposed product of this gene from GENI-ACT was a Methylcrotonyl-CoA carboxylase beta. Again Swiss-Prot was used for the Blast results showing a great similarity amongst a variety of organisms.



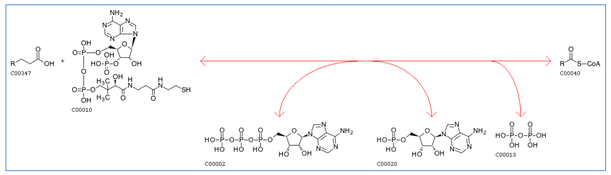
Crystal structure of P. aeruginosa 3-methylcrotonyl-CoA carboxylase (MCC), beta subunit

Both RMET_RS03075 and RMET_RS03080 are biotin dependent carboxylases that work in the mitochondria on the catabolism of leucine. Leucine catabolism is widely spread among bacteria and has been thoroughly studied. Its pathway is comprised by multiple reactions and converges with other catabolic routes, generating acetoacetate and acetyl-CoA as its final products. The initial three steps are conserved in most bacteria, constituting the first steps of the branched-chain amino acids catabolic pathway. The main product of these sequential reactions is the 3-methylcrotonyl-CoA metabolite, which undergoes further enzymatic steps towards the production of acetoacetate and acetyl-CoA.[2]

RMET_RS03085:

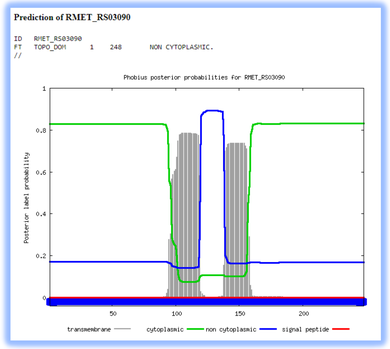
The initial proposed product of this gene from GENI-ACT was an Acyl-CoA synthetase. This gene was also blasted with Swiss-Prot returning many similar results. The oxidative degradation of fatty acids is a two-step process, catalyzed by acyl-CoA synthetase. First, the fatty acid reacts with ATP to form an acyl phosphate. This intermediate reacts subsequently to give acyl-CoA:

Fatty acid + CoA + ATP ⇌ Acyl-CoA + AMP + PPi



RMET_RS03090:

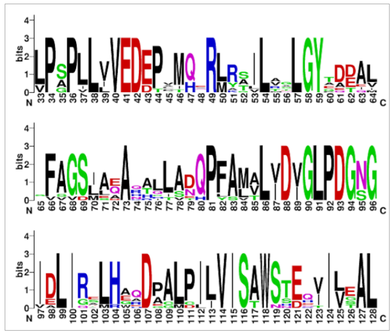
The initial proposed product of GENI-ACT was a TetR/AcrR family transcription regulator. Structural analyses have revealed that the helix-turn-helix (HTH) signature is the most recurrent DNA-binding motif in prokaryotic transcriptional factors, since almost 95% of all transcriptional factors described in prokaryotes use the HTH motif to bind their target DNA sequences.



Phobius results of RMET_RS03090 possibly showing the HTH structure of this gene.

RMET_RS03095:

The initial GENI-ACT results showed this to be a response regulator transcription factor of the LuxR family. Most LuxR-type regulators act as transcription activators, but some can be repressors. This gene will work in conjunction with RMET_RS03090.



Partial WebLogo of RMET_RS03095 showing a good degree of conservation between sequences. The majority of these amino acids are black which represents a hydrophobic region.

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes appear to be correctly annotated by the computer database.

References

1. Janssen PJ, Van Houdt R, Moors H, Monsieurs P, Morin N, Michaux A, et al. (2010) The Complete Genome Sequence of Cupriavidus metallidurans Strain CH34, a Master Survivalist in Harsh and Anthropogenic Environments. PLoS ONE 5(5): e10433. https://doi.org/10.1371/journal.pone.0010433

2. L. Díaz-Pérez, Alma & Díaz Pérez, César & Jesús, Campos-García. (2015). Bacterial l-leucine catabolism as a source of secondary metabolites. Reviews in Environmental Science and Bio/Technology. 15. 10.1007/s11157-015-9385-3.

University at Buffalo The State University of New York
Jacobs School of Medicine and Biomedical Sciences
buffalo.edu

Acknowledgments -
Special thanks to Dr. Stephen Koury and Dr. Rama Dey-Rao.
Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536