Annotation of the Enterobacter aerogenes KCTC 219 Genome from Locus Tags EAE_15860 to EAE_15870)

Andrew Bialy, Demetra Jacobs, Angeleah Lograsso and Mr. Philip Archabald Eden Central School, Eden, NY and The Western New York Genetics in Research Partnership



SEPA SCIENCE EDUCATION PARTNERSHIP AWARD

Abstract

A group of 3 consecutive genes from the microorganism Enterobacter aerogenes (EAE_15860 – EAE_15870) 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. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for each of the genes in the group, except for EAE_15870 which was proposed as a Transporter protein but BLAST found it as a binding protein, as such, the other genes appear to be correctly annotated.

Introduction

Enterobacter aerogenes is a rod-shaped bacterium with a length of approximately 1-3 microns forming gram negative coccoid bacterium, found predominantly in tetrad formation. This organism is a nosocomial and pathogenic bacterium that causes opportunistic infections including most types of infections. Some of the infections caused by E. aerogenes result from specific antibiotic treatments, venous catheter insertions, and/or surgical procedure. It has been found to live in various wastes, hygienic chemicals, and soil. The commercial significance of this bacterium is that it produces hydrogen gas during fermentation.





Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete Enterobacter aerogenes KCTC 219 genome annotation . The modules are described below:

Modules	Activities	Questions Investigated
Module 1- Basic Information Module	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Module 2- Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to other sequences in Genbank?
Module 3- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 4- Alternative Open Reading Frame	IMG Sequence Viewer For Alternate ORF Search	Has the amino acid sequence of my protein bee called correctly by the computer?
Module 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?

Results EAE_15860:

The initial proposed product of this gene by GENI-ACT was a extracellular solute-binding protein family 1. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated protein functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, and the cellular location of the amino acid sequence. As such, the proposed annotation is a Iron(III)-binding periplasmic protein.

EAE_15865:

The initial proposed product of this gene by GENI-ACT was a binding-protein-dependent transporters inner membrane component. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. As such, the proposed annotation is a Iron(III)transport system permease protein.

EAE_15870:

The initial proposed product of these genes by GENI-ACT was an ABC Transporter related protein. This gene product proposal was not supported by the top BLAST hits for the amino acid sequences, the presence of well-curated functional domains within the amino acid sequences, the cellular location of the amino acid sequences, and the enzymatic function of the amino acid sequences. But the TIGERFAM BLAST supported the gene product proposal. EAE 15870 was proposed to be a transporter protein while the top blast supported the bacteria as being a binding protein. It contained only a single well-curated functional domain, in other microorganisms possessing the enzyme formate dehydrogenase, there are two well-curated functional domains. The domain present in this gene is an FbpC domain that is found within the alpha subunit of the enzyme

Range 1: 1 to 338 GenPept Graphics								Next Match	A Previous	Match	
Score		Expect	Method			Iden	tities	Posi	tives	Gaps	
574 bit	s(1479	0.0 (Composition	nal matr	ix adjus	it. 278	338(829	b) 302/	338(89%)	0/338(09	5)
oery	1	MNFRLSAN	CAAMLAASAL L AS++	FTAPOVI	AADNG	GIVer	AQUENLA AQUENLA	KSNVDG	FTKETGIK	T 60	
bjet	1	MKLRISSI	GPVALLASSN	MLAFGA	AASADO	GIVIY	AQHENLY	KSWVDG	FTKDTGIK	T 60	
bery	61	LRNGDDSI	LGNQLVQEGS	ASPADVI	LTENSI	PSMVLVI P+MVLVI	NAKLFAF	LDAATL	NOVAPOYRS	E 120	
bjet	61	LENGEDSI	LCNQLVQECS	ASPADVI	LTENSI	PAMVLVI	NAKLFAF	LDAATL	AQVEPQYRI	\$ 120	
overy	121	HGRWIGI	ARSTVEVYNE	AKISES	LPHSLA	DLAKPI	WKGRNAJ	SPSGAD	FOAIVSAM FOAIVSAH	A 180	
bjot	121	EGRWIGI	ARSTVFVYNF	AKLSDA	LPKSLI	LDLAKPI	WKGRWAN	SPSGAD	FOAIVSALL	E 180	
bery	181	LKGEKATI	DWLKAMKTNF WLKAMKTNF	VAYKGN:	TVMKA	VNACQ11	GVIYHS	YRFVDQ Y FVD	AKTGENSKI AKTGENS I	T 240	
bjet	181	LXGENATI	AMLKAMKTNF	TAYKGN	STVMKA	INACOVI	DEGVIYEN	YPEVDG	AKTGENSN:	I 240	
bery	241	RLYYFKN +LYYFK+	DPGAFVSISG DPGAFVSISG	CGVLAS	SKEKEQJ	MOAFIK MOAFIK	TTGKQGC	DELRTN	NAFEYAVG	N 300	
bjet	241	KLYYFEB	OPGAFVS180	CCVLAS	SKEQQQJ	MAPIK	TTCKQCC	EILRIN	NAFEYAVC	G 300	
bery	301	AASNPKL: AASNPKL	PLKELDAPKV PLK+LDAPKV	EPSTLN: + + LN:	SKKVIDI	LNTQAGI	LL 338				
bjct	301	AASNPKLI	PLKDLDAPKV	DAAQLNS	XXVVE	LNTEAG	L 338				

Figure 2 - The pairwise alignment of EAA_15860 with its top nr top blast



Figure 3 - EAE_15870 well-curated functional domains which have many similarities among hits.



Figure 4 - EAE_15865 cell localization diagram.



Figure 5 - RCSB Protein data bank 3D image for nucleotide binding domain 3FVQ, the PDB name for EAE_15870. Crystal structure of the nucleotide binding domain FbpC complexed with ATP

Conclusions

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group except for EAE_15870 which was proposed as a Transporter protein but BLAST found it as a binding protein. As such, the other genes appear to be correctly annotated by the database.

References

Sims et al. (2009). Complete genome sequence of *Enterobacte*. Aerogenes KCTC 219. Standards Genomic Sciences, 12 - 20.

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R250D010536

www.buffalo.edu