Annotation of the *Bacillus thuringiensis str. Al Hakam* Genome from Locus Tags BALH_0278 to BALH_0279

Jude Kukla *, Matthias Rollins*, Sam Li , Amelia Sidonio, Shea Willis, and Peter Hentschke
The Harley School, Rochester, NY and The Western New York Genetics in Research and Health Care Partnership

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

University at Buffalo

Abstract

Using GENI-ACT, two genes of *Bacillus thuringiensis str. Al Hakam* (Locus tags BALH_0278 and BALH_0279) were annotated. In annotating these genes, we reviewed and analyzed the gene product name proposed by GenBank. We assessed the product name match in terms of 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, and the possibility of horizontal gene transfer. We did not find any results that contradicted the called gene product proposed by Genbank.

Introduction

Bacillus thuringiensis, commonly referred to as Bt, is a soil-dwelling bacterium that is often used as a pesticide. Bt operates by releasing toxins in the gut of the insect that consumes it. It punctures holes in the gut and germinates, causing death in a few days. Because of this property, Bt has been sprayed onto crops, such as corn. Recently, a specific Bt gene has also been integrated into genetically modified crops.

When the toxin producing Bt gene is integrated into crops, they express the cry gene, which then releases the pest-killing toxins. The resulting crop produces those toxins. When Bt was introduced into corn, pest species such as the European corn borer were immediately killed and prevented from doing damage to the corn. Despite their success, GMO crops such as Bt crops are still controversial. According to Hardy Hall, this is mainly because their long-term effects on humans are still unknown. As a result, some people do not want to take the risk (Hall, 2006).

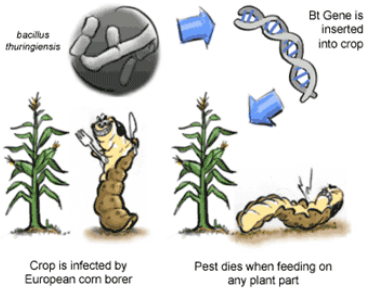

Figure 1. Application of Bt gene in agriculture. (Hall, 2005)

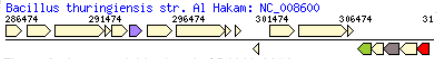

Figure 2 - Locus neighborhood of BALH_0278

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Bacillus thuringiensis str. Al Hakam* 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 been 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?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?

Results

BALH_0278 The original product proposed by Geni-Act was response regulator [*Streptococcus pneumoniae*]. PSORT-b predicted the protein would be found in the cytoplasm since the cytoplasm score was the highest. TMHMM did not find evidence of any transmembrane helices. Finally, in Phobius, there were no signal peptides on the protein. This evidence supports that this protein is located in the cytoplasm because the results from TMHMM predict that the protein is not in the cytoplasmic membrane, and the results from Phobius predict that it is not secreted to the outside of the bacterium.

In analyzing the reading frame of the gene, a Shine-Dalgarno sequence was 7 base pairs upstream of the called start codon. (Fig. 3). This suggests that the database predicted the start codon correctly and the reading frame is correct. The WebLogo of BALH_0278 ran against its top hit on blast, *Streptococcus pneumoniae*, gives a well-conserved sequence (Fig 4). This association between the response regulator [*Streptococcus pneumoniae*] and BALH_0278 is significant and their sequences are very similar and the data between the two genomes is conserved. Also, the PDB (Protein Data Bank) search resulted in the top hit being the same protein proved to be correct by previous methods. The protein is from the organism *Streptococcus pneumoniae* and is a response regulator (Fig 5).

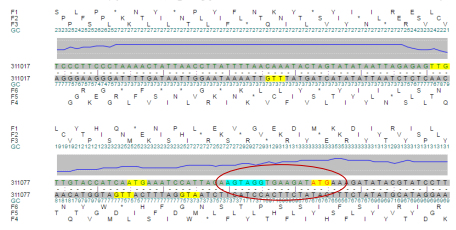

Figure 3 – BALH_0278 was analyzed and a start codon was called. The start codon is in red text highlighted in yellow, and the Shine-Dalgarno sequence is 7 base pairs upstream, in blue.

Figure 4 – BALH_0278 was analyzed and ran against the *Streptococcus pneumoniae* and base pairs from 193 to 224 are shown. They are extremely well conserved and makes this match significant.

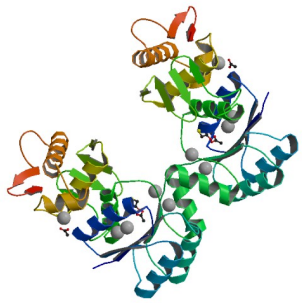

Figure 5 – The protein structure of *Streptococcus pneumoniae* response regulator that was called by the PDB.

BALH_0279: BALH_0279 has the DNA coordinates 317361..318293 as given by Geni-Act. The DNA sequence is 933 base pairs long and 310 amino acids long. The gene product of BALH_0279 [*Streptococcus pneumoniae*] as proposed by Geni-Act was permease, drug/metabolite transporter superfamily. This was supported by the top two BLASTs hits of membrane protein [*Streptococcus pneumoniae*] and EamA family transporter [*Bacillus sp.* GZT].

Both of these hits are proteins involved in transport through the membrane. The CDD search supported this with its top hit of an uncharacterized membrane protein [Function unknown]. This top hit had a COG number of COG2510 and an e-value of 1.22e-04. The WebLogo for this gene had large gaps of un-conserved areas at the c and n terminal regions. The predictions from the TMHMM and SignalP differed from the predictions of Phobius. TMHMM predicted 10 helixes and SiganlP predicted no signal peptide present, however Phobius predicted 9 transmembrane helices and the presence of a signal peptide (Fig 6).

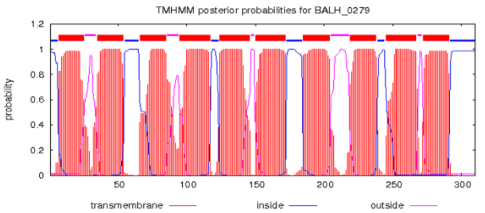

Figure 6 – The TMHMM results from the database. 10 transmembrane helices were found in this locus of the gene.

Conclusion

The GENI-ACT proposed gene product did not differ 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.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
BALH_0278	DNA-binding response regulator	DNA-binding response regulator
BALH_0279	Permease, Drug/Metabolite transporter superfamily	Permease, Drug/Metabolite transporter superfamily

References

Hall et al., 2005. BT CORN: IS IT WORTH THE RISK?. The Science Creative Quarterly Issue Two. Part III of IV.

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA)
Award - R25ODO10536
Dr. Stephen Koury

www.buffalo.edu