Partial Annotation of Four Genes within the *Roseburia intestinalis* Genome with DNA Coordinates 58672 to 66136 (or Locus Tags RO1_00690 to RO1_00770)

Kristina Carpenter, Dante Clark, Haley Costlow, Hayden Wisher, and Jared Greer
Global Concepts Charter High School, Lackawanna, NY

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

University at Buffalo

Abstract

A group of 4 (2 separate, but consecutive pairs of) genes from the microorganism *Roseburia intestinalis* (RO1_00690 – RO1_00770) were partially 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, and cellular localization data. 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 in the database. However, to better support our conclusions, further manual annotation of these genes is suggested since several modules were not completed for each of the genes.

Introduction

Roseburia intestinalis is a gram-positive bacteria isolated from human faecal material. It is anaerobic, slightly curved-rod-shaped and motile by means of multiple subterminal flagella. Phenotypic and phylogenetic characteristics reveal a low-G+C-content, butyrate-producing bacterium that show net acetate utilization during growth on media containing carbohydrates and short-chain fatty acids.

Butyrate produced by fermentation in the human colon is considered to have health-promoting properties (von Engelhardt et al ., 1998 ; Scheppach et al ., 1995). There is, however, only limited information concerning bacterial strains capable of butyrate production in the human gut. Butyrate is an important nutrient for colonocytes, as well as a signaling molecule with a central role in cell differentiation and apoptosis (von Engelhardt et al., 1998; Scheppach et al., 1995).

All strains were net acetate utilizers, removing between 9 and 14 µ mol acetate ml − "from the growth medium. Aesculin was hydrolyzed and weak fermentation of melibiose was detected. The substrates arabinose, cellobiose, fructose, maltose, raffinose, sucrose, xylose and starch were all fermented. Rhamnose, melezitose, mannitol, ribose, inulin and trehalose were not fermented, and arylamidase activity was not detected using the Rapid ID-32A system.

Phylogenetic analysis indicated that the most closely related species are Eubacterium rectale, Eubacterium oxidoreducens and Roseburia cecicola." (Duncan et al., 2002)

Figure 1. Scanning electron micrograph of Roseburia intestinalis sp. nov. L1-82 T showing a flagella bundle. Bar, 1 µ m. (Duncan et al., 2002)

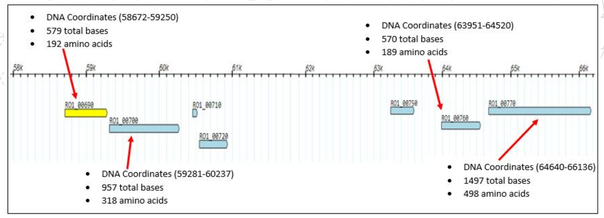

Figure 2. The locus tags and relative position of the genes under investigation in this research

Methods

Modules of the GENI-ACT (https://geni-act.org/) were used to complete *Roseburia intestinalis* genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Module 1 – Basic Information	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 Similarity	BLAST, CDD, T-Coffee, WebLogo	Is my sequence similar to other sequences in Genbank?
Module 3 – Protein Structure	TIGRFAM, PFAM, PDB	Are there functional domains in my protein?
Module 4 – Protein Localization	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?

Results

RO1_00690:
The initial proposed product of this gene by GENI-ACT was an unspecified membrane protein. Cellular localization data supported the notion that this protein is likely found embedded in the cell membrane, as TMHMM search results predicted 6 transmembrane helices. Additionally, PSORT-B tests yielded a cytoplasmic membrane score of 10.00.

The top 2 BLAST hits within the SwissProt database show a gene product name of "putative manganese efflux pump MntP" in the organisms *Desulfitobacterium hafniense Y51* and *Clostridium botulinum F str. Langeland*.

Structure based evidence yielded results of related families of proteins which served in the process of transport across the membrane. Pfam analysis produced a LysE transporter superfamily clan name with high score and significant E-value. As such, the proposed annotation is a transporter membrane protein.

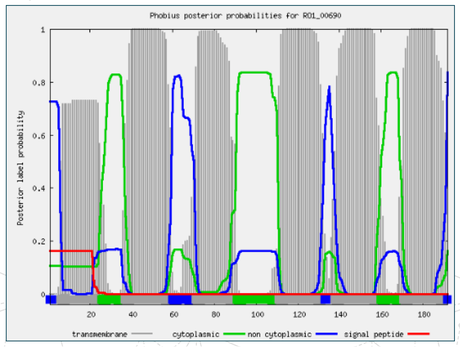

Figure 3. Phobius results for RO1_00690 indicate the presence of multiple transmembrane helices.

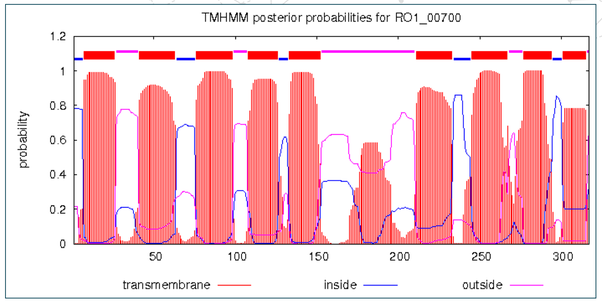

Figure 4. TMHMM prediction for RO1_00700.

RO1_00700:
The initial proposed product of this gene by GENI-ACT was a K+ dependent Na+/Ca+ exchanger related protein. The top BLAST hits showed an uncharacterized membrane protein along with a Ca2+/Na+ antiporter COG name.

Protein structure evidence showed a K+ dependent Na+/Ca+ exchanger TIGRFAM name. In addition, cellular localization data for the amino acid sequence predicted 9 transmembrane helices (TMHMM results) and a PSORT-B cytoplasmic membrane score of 10.00. See Figure 4 (above)

As such, the proposed annotation is a membrane protein that plays a role in sodium and calcium ion exchange and is K+ dependent.

RO1_00760: The initial proposed product of this gene by GENI-ACT was a glycerol-3-phosphate responsive antiterminator (mRNA binding). One of the top BLAST hits within the SwissProt database returned a glycerol uptake operon antiterminator regulatory protein with high score and low E-value in the bacteria *Bacillus subtilis*. There was one COG result with similar characteristics and noted to play a role in transcription.

Cellular localization data yielded a PSORT-B cytoplasmic score of only 1.78 and cytoplasmic membrane score of 8.16. However, TMHMM results predicted 0 transmembrane helices. These results indicate further annotation is necessary to provide clearer localization information.

Results from the structure-based evidence module further supported the proposed annotation of a responsive antiterminator. The Protein Data Bank returned a known operon antiterminator regulatory protein from *Listeria Monocytogenes Str. 4b F2365* (Figure 5 – image model).

Despite the lack of clarity with respect to localization, the proposed annotation is a glycerol-3-phosphate responsive antiterminator..

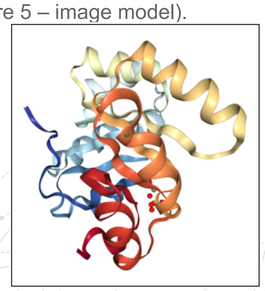

Figure 5. Model Image of protein found in *Monocytogenes Str. 4b F2365* with similar sequence as *RO1_00760*

RO1_00770:
The initial proposed product of this gene was a glycerol kinase. This was supported by the top BLAST hits for the amino acid sequence as the microorganism *Eubacterium rectale ATCC 33656* contained a similar glycerol kinase protein. The top BLAST hit and COG further detailed these types of proteins as playing a role in the production and conversion of ATP. WebLogo showed high amounts of conservation throughout the entire sequence, based on multiple sequence alignment with 10 proteins from other microorganisms mostly characterized as glycerol kinases.

All cellular localization data supported that this protein is likely found within the cytoplasm, as there were no predicted transmembrane helices, nor signal peptide markers.

Additionally, both TIGRFAM and Pfam results returned similar families of proteins that play a role in ATP production.

As such, the proposed annotation remains a glycerol kinase.

Figure 6. WebLogo for *RO1_00770* based on multiple sequence alignment with other proteins.

Conclusion

The GENI-ACT proposed gene products did not differ significantly from the proposed gene annotations 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
00690	Predicted Membrane Protein	Transporter Membrane Protein
00700	K+ dependent Na+/Ca+ exchanger related protein	K+ dependent Na+/Ca+ exchanger protein
00760	Glycerol-3-phosphate responsive antiterminator (mRNA binding)	Glycerol-3-phosphate responsive antiterminator
00770	Glycerol Kinase	Glycerol Kinase

References

Duncan et al. (2002). Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. International Journal of Systematic and Evolutionary Microbiology. 2002 Sep; 52(Pt5): 1615-20.

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

Special thanks to Dr. Stephen T. Koury and Dr. Rama Dey-Rao for assistance with training throughout the project.

www.buffalo.edu