

Partial Annotation of Four Genes within the *Roseburia Intestinalis* Genome From Locus Tags RO1_00610 to RO1_00660

Obad Alhajjaji, Muna Alsadam, Sana Fletcher, Hyfa Alhajjaji and Jared Greer

Global Concepts Charter High School, Lackawanna, NY and The Western New York Genetics in Research Partnership

Abstract

A group of 4 (2 consecutive and 2 non-consecutive) genes from the microorganism *Roseburia intestinalis* (RO1_00610 – RO1_00660) 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⁻¹ 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, melizitose, 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)

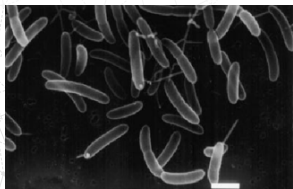


Figure 1. Scanning electron micrograph of *Roseburia intestinalis* sp. nov. L1-82 T showing a flagella bundle. Bar, 1 μ m. (Duncan et al., 2002)

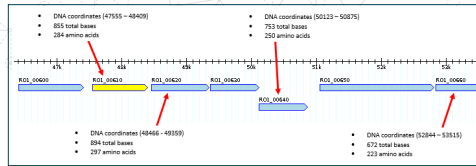


Figure 2. The locus tags and relative position of the genes under investigation in this research

Methods

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Roseburia intestinalis* 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 5- Structure-Based Evidence	TIGRFam, Pfam, PDB	Are there functional domains in my protein?

Results

RO1_00610:

The initial proposed product of this gene by GENI-ACT was a phosphate ABC transporter membrane protein (PhoT family). This gene product proposal was supported by the top BLAST hits within the SwissProt database. Figure 3 (below) shows one of many BLAST results with a high score, low E-value and high query coverage in the microorganism, *Mesorhizobium loti*. As shown, the identity of this protein is a phosphate transport system permease protein PatC.

Additionally, cellular localization data supported the notion that this protein is likely found embedded in the cell membrane. Structure based evidence yielded results of other related families of proteins which served in the process of transport across the membrane. As such, the proposed annotation is a phosphate ABC transporter membrane protein.

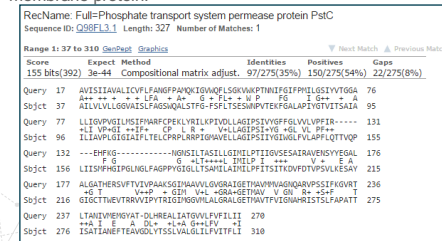


Figure 3. Top BLAST results within the SwissProt database for RO1_00610

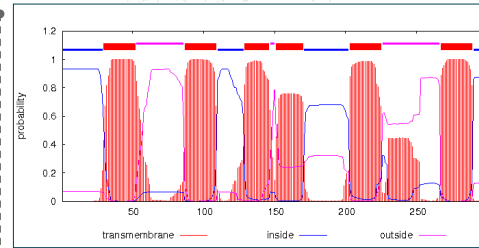


Figure 4. TMHMM prediction for RO1_00620.

RO1_00620:

The initial proposed product of this gene by GENI-ACT was another phosphate ABC transporter membrane protein (PhoT family). This gene product proposal was well supported by the top BLAST hits within the SwissProt database for the amino acid sequence and the presence of well-curated functional domains.

Cellular localization data for the amino acid sequence predicted 6 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 phosphate ABC transporter membrane protein.

RO1_00640:

The initial proposed product of this gene by GENI-ACT was a phosphate uptake regulator (PhoU). Top BLAST hits within the SwissProt database returned many similar transport and/or uptake regulator proteins with high scores and low E-values in various other microorganisms.

Cellular localization data yielded a PSORT-B cytoplasmic score of 9.97, zero predicted transmembrane helices, and low probability values for the likelihood of any signal peptides.

Results from the structure-based evidence module further supported the proposed annotation of a phosphate uptake regulator and the Protein Data Bank returned a known phosphate transport system protein from *Aquifex aeolicus* with high score and low E-value (Figure 5 – image model).

As such, the proposed annotation is a phosphate uptake regulator.

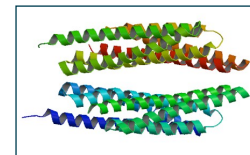


Figure 5. Model image of protein found in *Aquifex aeolicus* with similar sequence as RO1_00640

RO1_00660:

The initial proposed product of this gene was a response regulator consisting of a CheY-like receiver and winged-helix DNA-binding domain. This was supported by the top BLAST hits for the amino acid sequence as other microorganisms contained similar proteins which served as DNA-binding response regulators. In addition, WebLogo showed high amounts of conservation from amino acid 88 to 122, based on multiple sequence alignment with 10 other proteins of known response regulator function (see Figure 6 below).

All cellular localization data supported that this protein is likely found within the cytoplasm and both TIGRFAM and Pfam results returned similar families of proteins that play a regulator role, specifically with respect to transcription. More specifically, one of the Pfam top hits (PF00486) refers to a family of transcriptional regulatory proteins with a total of 7 key functional/structural residues as follows: L14, L17, L30, L53, R54, L57 and G74.

As such, the proposed annotation is a DNA binding response regulator.

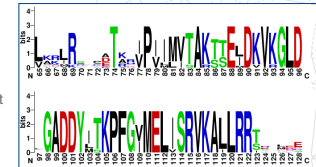


Figure 6. WebLogo for RO1_00640 based on multiple sequence alignment with other proteins.

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.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
00610	Phosphate ABC transporter membrane protein	Phosphate ABC transporter membrane protein
00620	Phosphate ABC transporter membrane protein	Phosphate ABC transporter membrane protein
00640	Phosphate uptake regulator	Phosphate uptake regulator
00660	Response regulator consisting of a CheY-like receiver domain & a winged-helix DNA-binding domain	DNA binding response regulator

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 a NIH Science Education Partnership (SEPA) Award - R25OD010536

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

www.buffalo.edu