

Annotation of 4 *Kytococcus sedentarius* Genes: *Ksed_05080*, *05090*, *05100*, and *05120*

Kalia McCray, CJ Duggan, Mikayla Ladd, Daeja Sanger, and Margaret Diamond

The Park School of Buffalo and the WNY Genetics in Research Partnership

Abstracts

Four consecutive genes from the microorganism *Kytococcus sedentarius* (*Ksed_05080* – *05120*; DNA bases 504521 - 507698) were annotated using the collaborative genome annotation website GENI-ACT. The Geni-Act proposed gene product identification for each gene was examined in terms of the general genomic information, amino acid sequence-based similarity data (BLAST, CDD, T-Coffee, and WebLogo), structure-based evidence from the amino acid sequence (TIGRFam and Pfam), and cellular localization data (TMHMM, SignalP, and Phobius). The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for each of the three genes studied. Therefore, the genes appear to be have been correctly identified by the computer program.

Introduction

Kytococcus sedentarius is a gram positive bacterium that produces certain antibiotics. It is aerobic and requires some amino acids for growth. This bacterium was originally found in a marine environment. Even though it is often overlooked it is known to have caused serious illnesses such as valve endocarditis and pneumonia. It also causes foot odor and pitting of foot calluses (pitted keratolysis). (Sims 2009; James 2012)

The initial proposed product for *Ksed_05080* by GENI-ACT was a threonine efflux protein. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of mostly preserved domains within the amino acid sequence, and the cellular location of the amino acid sequence. The function of this protein is to move threonine from the inside of the cell to the outside of the cell.

The initial proposed product for *Ksed_05090* was as a ferredoxin subunit of nitrite reductase and nitrite-hydroxylating dioxygenase. BLAST identified it as a member of the Rieske superfamily which includes the proposed GENI-ACT identification, although the top BLAST hit identified it as cytochrome B6. Subsequent BLAST hits supported the original proposal; the Rieske superfamily includes cytochromes.

The *Ksed_05100* GENI-ACT identification was a pterin-4a-carbinolamine dehydratase (for pterin synthesis). Although the top BLAST hits were for polyketide cyclase (which is for polyketide synthesis), it is in the same superfamily as the initial proposed gene identification. Both domains were identified by Pfam within the protein.

Ksed_05120 encodes an enzyme in the pathway for making glutathione. The proposed GENI-ACT identification was supported by the top BLAST hits for the amino acid sequence, the presence of two well-described protein functional domains within the amino acid sequence, and the cellular location of the amino acid sequence. This protein is highly conserved with very high E-values when aligned with other similar bacterial sequences; it catalyzes the second step in the glutathione synthesis pathway.

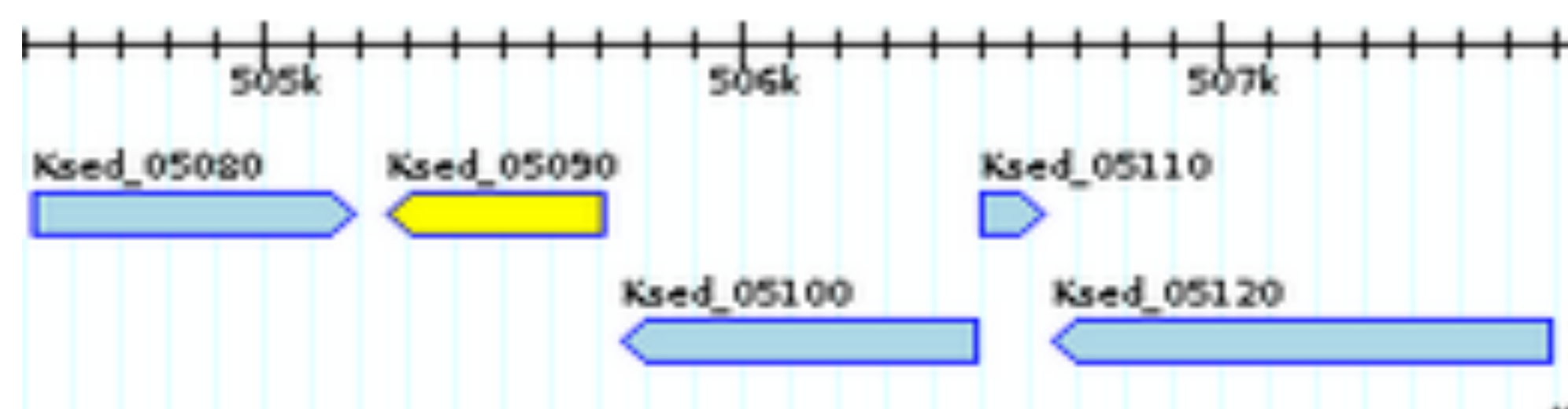


Fig. I – Gene neighborhood of the 4 *Kytococcus sedentarius* genes annotated.

Methods and Materials

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Kytococcus sedentarius* 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?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

Kytococcus sedentarius 05080:

The initial proposed product of this gene by GENI-ACT was a threonine efflux protein. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-preserved domains within the amino acid sequence, and cellular localization data that identified five transmembrane domains. The proposed annotation as a threonine efflux protein appears accurate.

Kytococcus sedentarius 05090:

The proposed GENI-ACT identification was a ferredoxin subunit of nitrite reductase and nitrite-hydroxylating dioxygenase, although the top BLAST hit identified it as cytochrome B6. BLAST identified it as a member of the Rieske superfamily, iron-containing proteins responsible for electron transfer in oxidation-reduction reactions. These proteins have two Fe atoms coordinated by two cysteine and two histidine residues accounting for the high degree of conservation of these amino acids seen in the HMM logo.

Kytococcus sedentarius 05100:

The proposed GENI-ACT identification was pterin-4a-carbinolamine dehydratase although it was noted that a polyketide cyclase domain was also identified by Pfam. The top BLAST hit for the amino acid sequence identified it as a polyketide cyclase, a member of the START superfamily as shown by BLAST CDD. Other sequence-based similarity data, structure-based evidence, and the predicted cellular location of the protein in the cytoplasm supported this identification.

Kytococcus sedentarius 05120:

The predicted product for this gene by GENI-ACT was a glutathione synthetase. This prediction was supported by the top BLAST hits for the amino acid sequence and the BLAST CDD identification, structure-based evidence, and the predicted cellular location of the protein in the cytoplasm. Two separate well-characterized domains, both contributing to the protein's function as a glutathione synthetase, were identified.

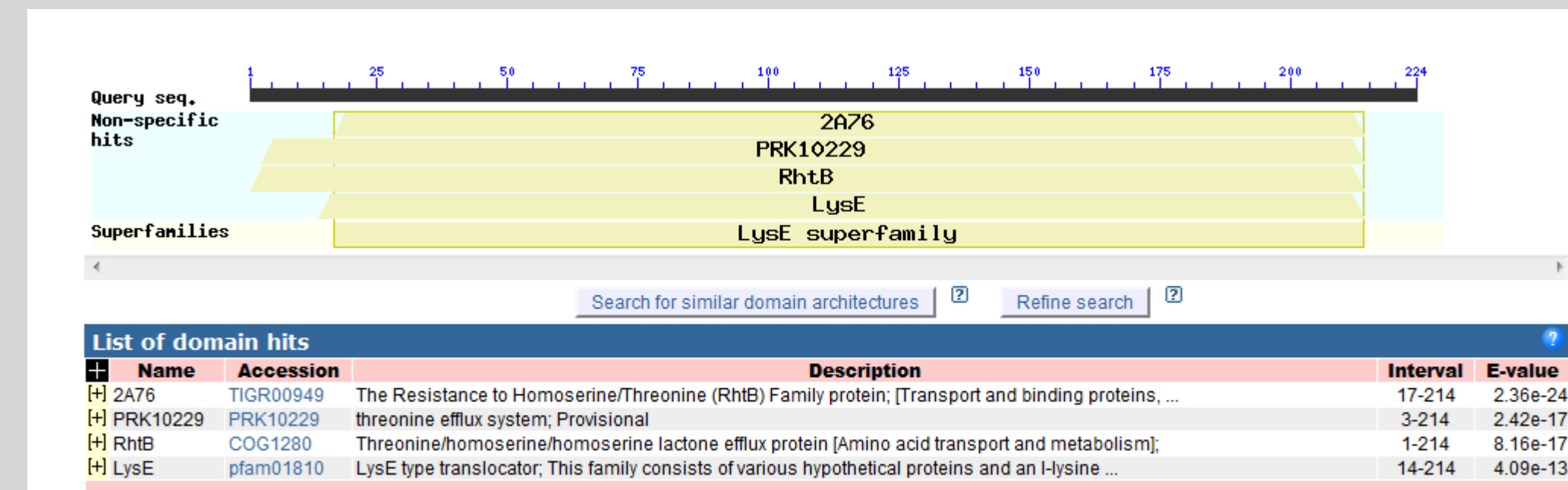


Figure I. *Kytococcus sedentarius* 05080 has a single well-characterized functional domain identified as a threonine efflux protein.

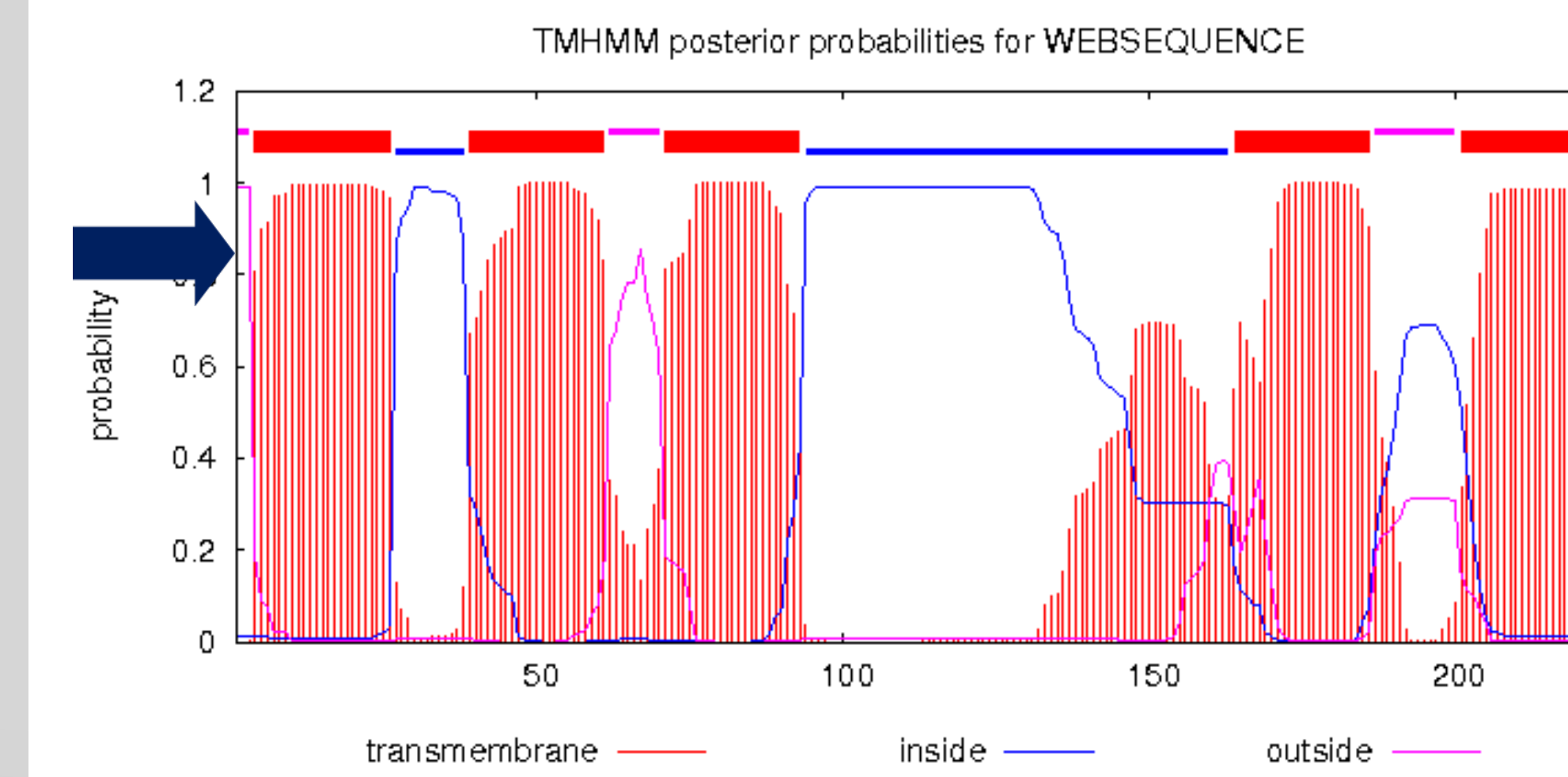


Fig. II. This graph shows the TMHMM results for *Ksed_05080*. The red shows the portions of the protein embedded in the membrane; the black arrow indicates a transmembrane domain. Five transmembrane domains are shown in this graph. The blue and pink lines indicate the portions of the protein that loop inside (blue) and outside (pink) of the membrane.

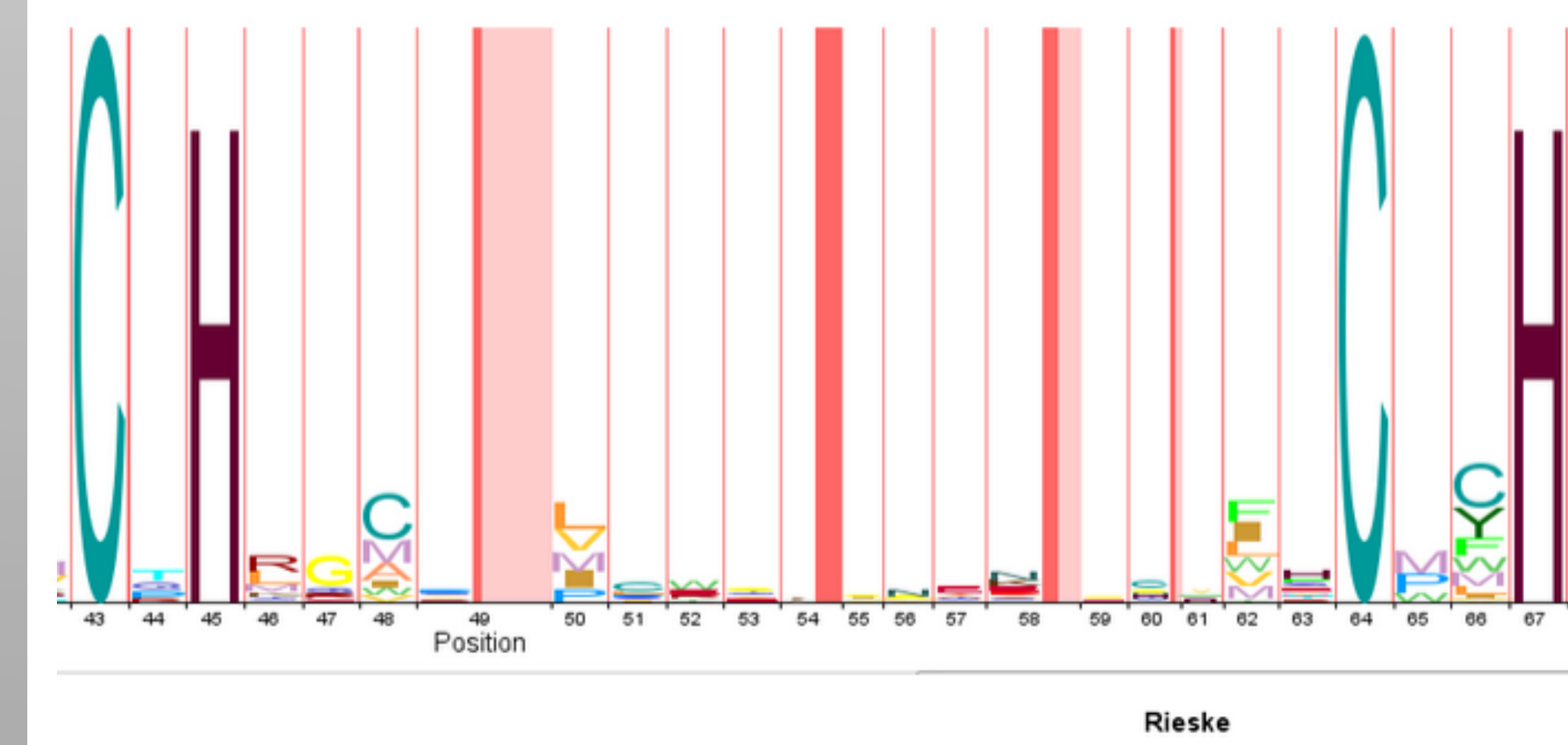


Fig. III Portion of *Ksed_05090* HMM logo showing the high conservation of two C and of two FH amino acids. They are probably responsible for the coordination of Fe atoms in this protein.

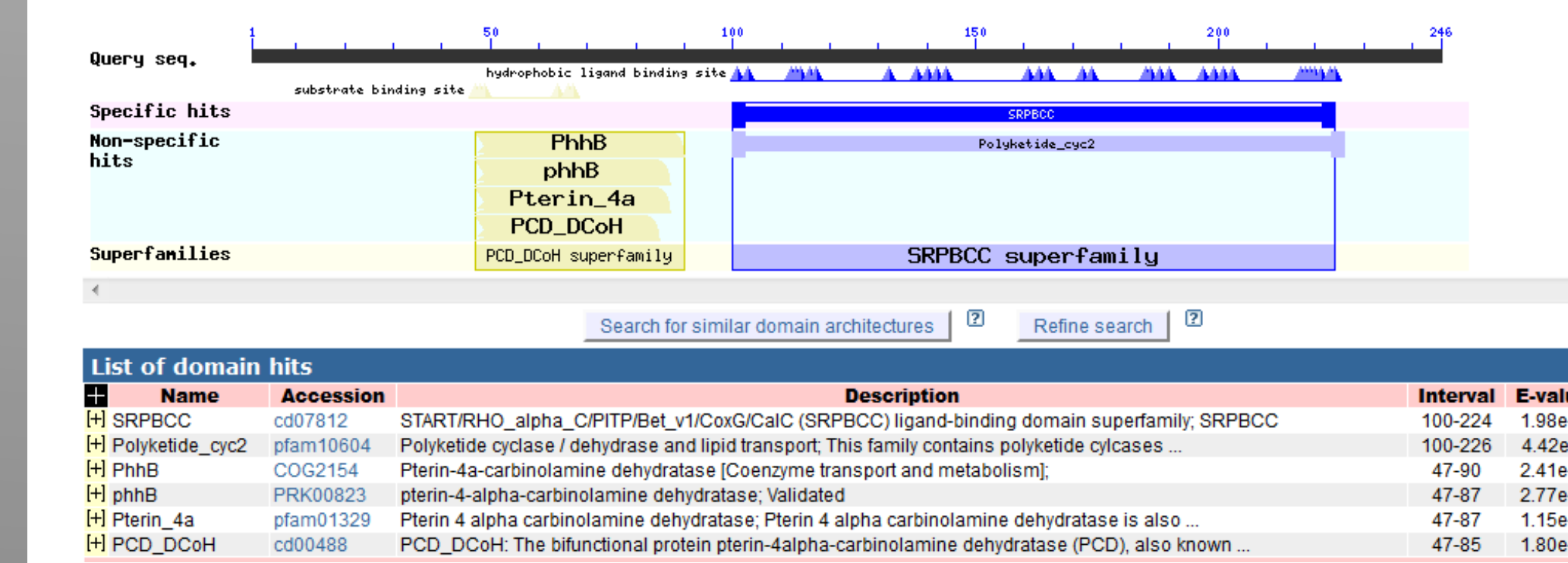


Fig. IV *Ksed_05100*, a member of the SRPBCC (START) superfamily, includes both pterin-4a-carbinolamine dehydratase and polyketide cyclase.

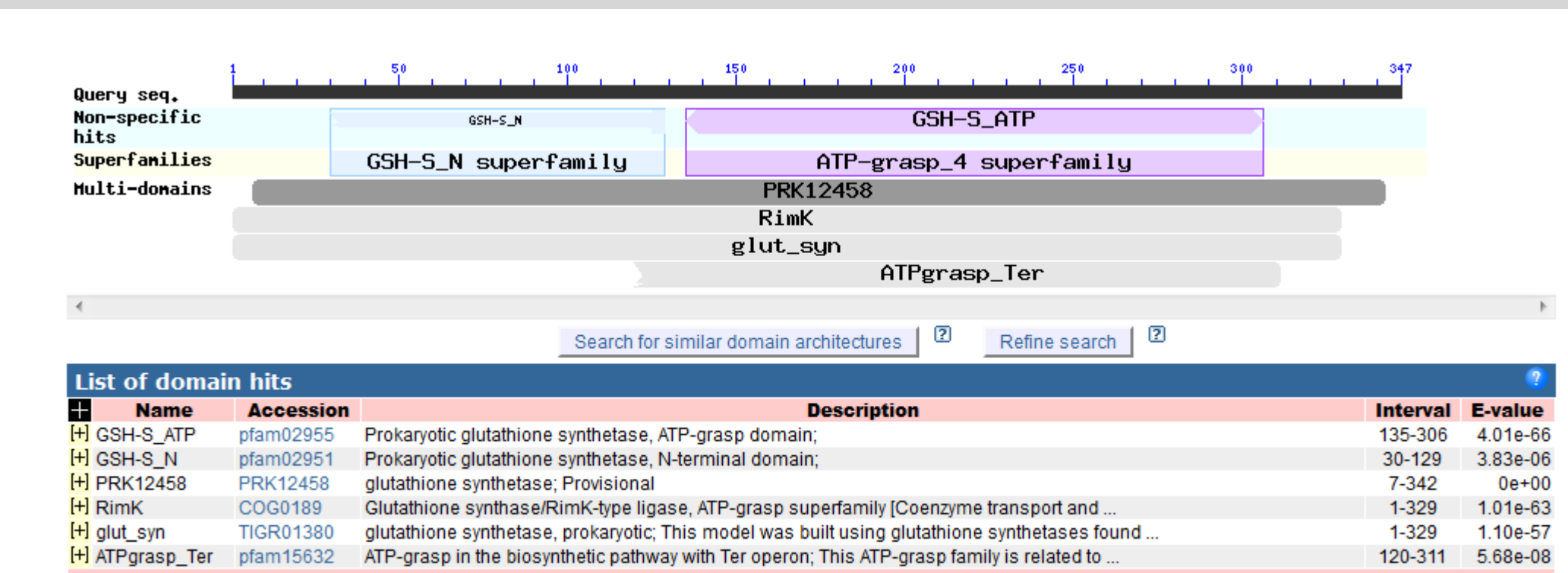


Fig. V. BLAST CDD showing two domains for *Ksed_05120*, both contributing to its function as a glutathione synthetase

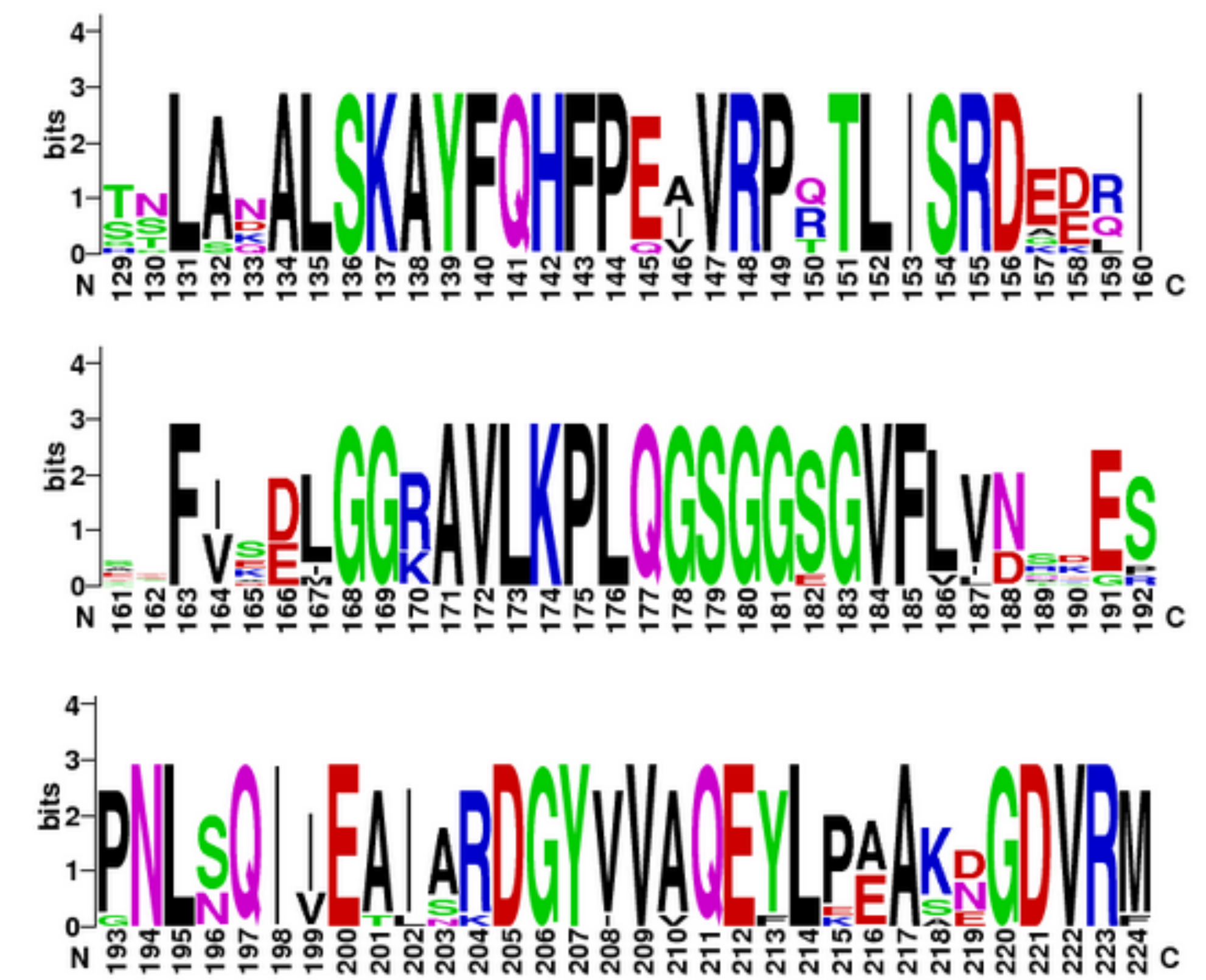


Fig. VI. A portion of the WebLogo for *Ksed_05120* showing the high degree of amino acid conservation throughout the sequence.

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 Product	Proposed Annotation
05080	Threonine efflux protein	Threonine efflux protein
05090	Cytochrome B6	Cytochrome B6
05100	pterin-4a-carbinolamine dehydratase and polyketide cyclase	pterin-4a-carbinolamine dehydratase and polyketide cyclase
05120	Glutathione synthetase	Glutathione synthetase

References

James et al. (2013). Microbiological and Biochemical origins of foot malodour. *Flavour and Fragrance Journal*, 28: 231-237
 Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). *Stand Genomic Sci*, 12 - 20.

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902
 Special thanks to Dr. Stephen Koury and Dr. Rama Dey-Rao for their help in completing the work presented here.