

Salt Lake Community College

Salt Lake Community College Library Digital Archives

Science, Mathematics and Engineering
Symposium

Student Experience

2017

2017 - Annotation Project for Locus Hbor_07980 of the Halogeometricum borinquense genome ATPase AAA - Poster Presentation

Jacob Gustavus

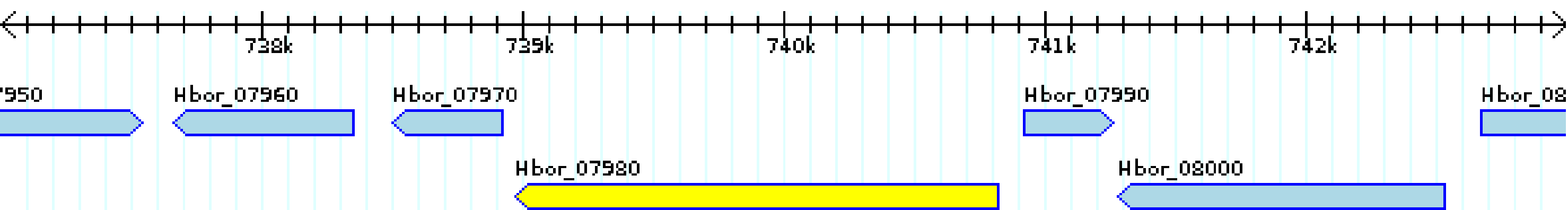
Follow this and additional works at: <https://libarchive.slcc.edu/sme>



Part of the [Biology Commons](#), and the [Chemistry Commons](#)

Abstract:

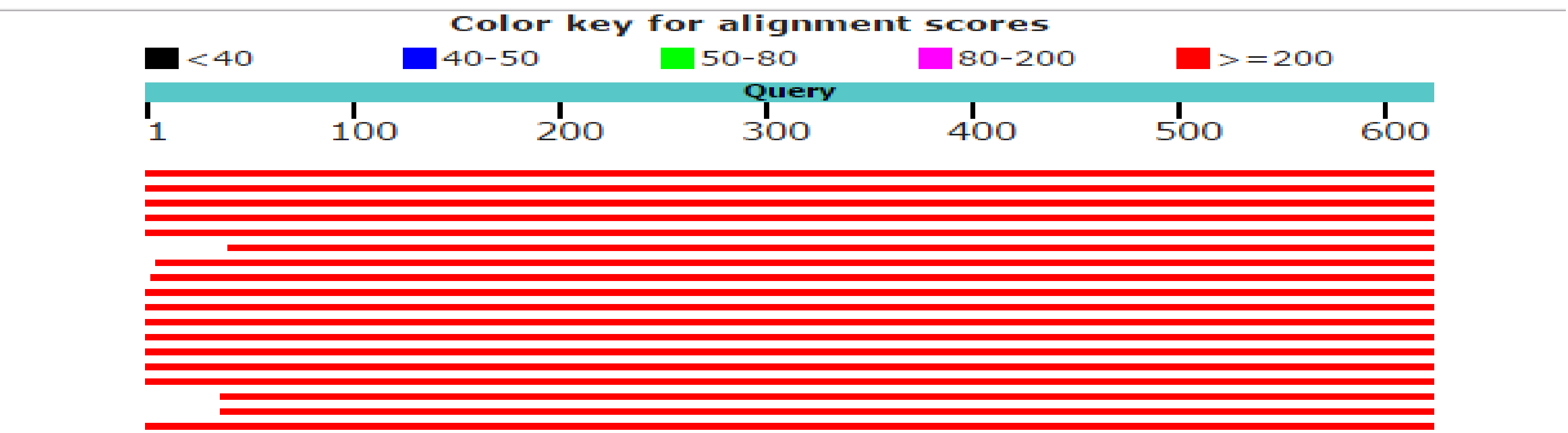
Each semester students here at SLCC examine a 50,000-nucleotide segment of the Halogeometricum borinquense genome to determine likely protein coding genes and to develop a scientific hypothesis concerning the location of the protein coding genes in the segment. By using open reading frames (ORFs) and bioinformatic programs to identify the portion of locus Hbor_07980 putative conserved domains have been detected to the P-loop_NTPase Superfamily. According to NBCI “The P-loop NTPases are involved in diverse cellular functions, and they can be divided into two major structural classes: the KG (kinase-GTPase) class which includes Ras-like GTPases and its circularly permuted YlqF-like; and the ASCE (additional strand catalytic E) class which includes ATPase Binding Cassette (ABC), DEXD/H-like helicases, 4Fe-4S iron sulfur cluster binding proteins of NifH family, RecA-like F1-ATPases, and ATPases Associated with a wide variety of Activities (AAA).” Locus Hbor_07980 also had a specific CDD hit to YjgR a “Archaeal DNA helicase HerA or a related bacterial ATPase that contains HAS-barrel and ATPase domains often used in Replication, recombination and repair.”



Locus: **Hbor_07980**
DNA Location: **738973..740829**
Nucleotide/Peptide Sequence length: **1857/618**

Methods Used

BLAST – Alignment Distribution and Sequences



ATPase [Halogeometricum pallidum]
Sequence ID: **WP_008384613.1** Length: 638 Number of Matches: 1
▶ See 1 more results

Range 1: 1 to 638	GenPept	Graphics	Expect	Method	Identities	Positives	Gaps
1122 bits (2903)	0.0		Compositional matrix adjust.		557/638 (87%)	583/638 (91%)	20/638 (3%)
Query 1	MSDLGDFDTDFEDDPDTSSSHITDASATEEAQ-----TREQSPQSDSEPSDTDQGVSFET	55					
Subjet 1	MSDLGDFDTDFESGSDRDSSTQADASGAERGEASDGAAADAAAEASDAADAEVAFER	60					
Query 56	YDVPAGEDRGIGAVSVSGLRVAEDGDDTTLRAFYVTGNRRDVRRLGKYLVPVPDDLL	115					
Subjet 61	FDVAPEGEDRGIGAVSVSGLRVAEDGDDTTLRAFYVTGNRRDVRRLGKYLVPVPDDLL	120					
Query 116	FCRITALEYAQEFQTDDATEIHARRAMRQOGEERDYKFVASLDPVAVLFEQGD-----	169					
Subjet 121	FCRITALEYAQEFQTDDATEIHARRAMRQGEERDYKFVASLDPVAVLFEQNETRDASE	180					
Query 170	-----ELKRRMVDVRVPGKAIVAEATDPEQIKTGLKIPAEGVFLGHLVSGGEKVRT	220					
Subjet 181	GRTESSGDDGELKRRMVDVRVPGKAIVAEATDPEQIKTGLKIPAEGVFLGHLVSGGEKVRT	240					
Query 221	AAEPPTIDYRLKDDYTDDPLVFRHRTLVAAGGTGSGKTHASKNLLRQLLSDRSTYEMDDGR	280					
Subjet 241	AAEPPTIDYRLKDDYTDDPLVFRHRTLVAAGGTGSGKTHASKNLLRQLLSDRSTYEMDDGR	300					
Query 281	DARMVAVQDPDQEAQIHDDNPENTSVARRYEREGVAHGGDDTLCLVPKEGDPVYGG	340					
Subjet 301	DARMVAVQDPDQEAQIHDDNPENTSVARRYEREGVAHGGDDTLCLVPKEGDPVYGG	360					
Query 341	DNHRAEQLEFTTFPHSHARDPVLVAGSSLNENOPPAKELLRRFFROYGDEGTYDEF	400					
Subjet 361	DNHRAEQLEFTTFPHSHARDPVLVAGSSLNENOPPAKELLRRFFROYGDEGTYDEF	420					
Query 401	LDDPALKEELHESGRVHEATDAVKRRVRGVPNGVFDQSGARPITELDHVLVRPGLTVIP	460					
Subjet 421	LDDPALKEELHESGRVHEATDAVKRRVRGVPNGVFDQSGARPITELDHVLVRPGLTVIP	480					
Query 461	TYHLSTSRRAKEMFVLAVSSMLIDDKLSNTPDQSRITKQPLVLGDEAHNFLADADNRQAR	520					
Subjet 481	TYHLSTSRRAKEMFVLAVSSMLIDDKLSNAPRSDRITKEPTVLGDEAHNFLADADNRQAR	540					
Query 521	KVYSKFTEAAKQGRKERLGLFLITQDPQGVAEVFKQVNTKLVNLGDEDAIKSVNTPPN	580					
Subjet 541	KVYSKFTEAAKQGRKERLGLFLITQDPQGVAEVFKQVNTKLVNLGDEDAIKSVNTPPN	600					
Query 581	LEDKVPYMEKGQHVVPSPDNEPVELVGLSTCVTRHGE	618					
Subjet 601	LEDKVPYMEKGQHVVPSPDNEPVELVGLSTCVTRHGE	638					

Proteins used for T – Coffee: Multiple

Sequence Alignment

SFR60101.1 hypothetical protein SAMN04487947_2706 [Halogeometricum rufum]

SFR52883.1 hypothetical protein SAMN04488124_2144 [Halogeometricum limi]

WP_008384613.1 ATPase [Halogeometricum pallidum]

SFG52608.1 hypothetical protein SAMN04488063_2309 [Haloplagius inordinatus]

SDQ12322.1 hypothetical protein SAMN05216278_0522 [Haloplagius longus]

SDY02958.1 hypothetical protein SAMN04487946_105183 [Halobellus clavatus]

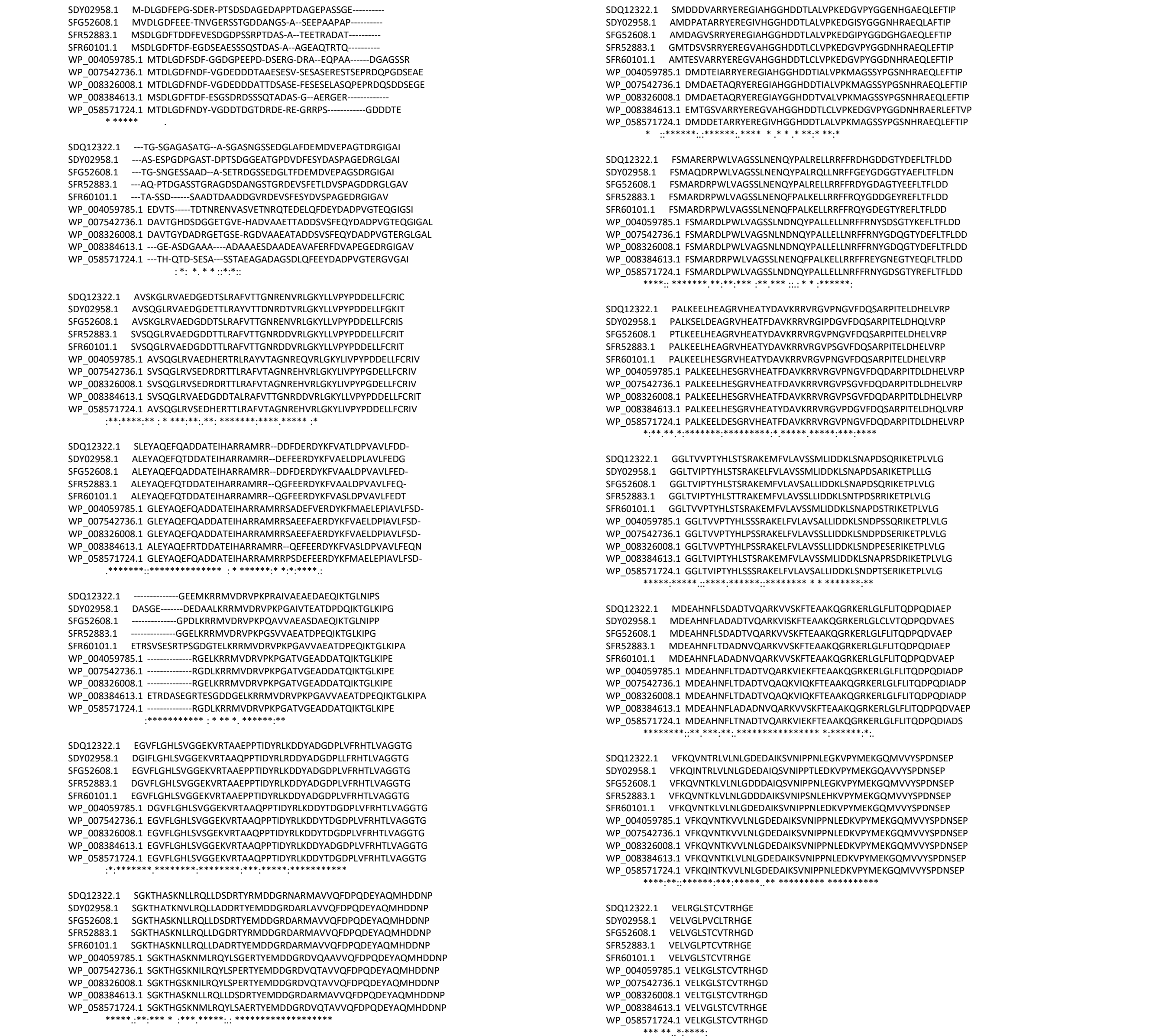
WP_004059785.1 ATPase AAA [Haloferax mediterranei]

WP_008326008.1 ATPase [Haloferax elongans]

WP_007542736.1 ATPase [Haloferax larsenii]

WP_058571724.1 AAA family ATPase [Haloferax sp. SB29]

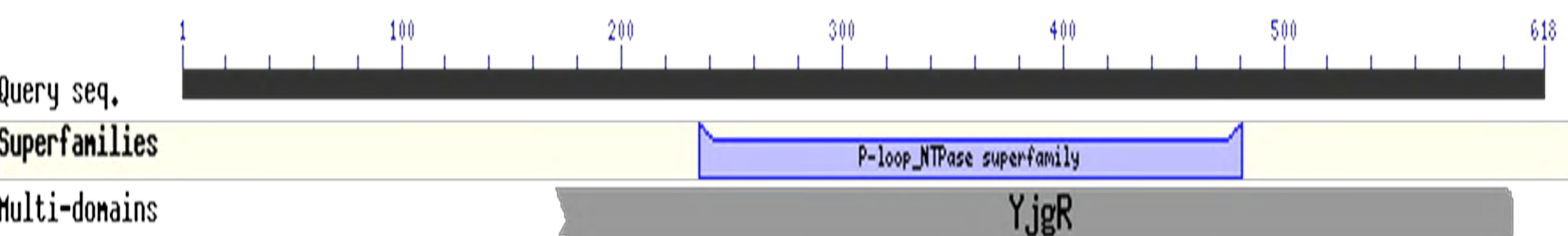
T – COFFEE Multiple Sequence Conservation



WebLogo – Multiple Sequence comparison

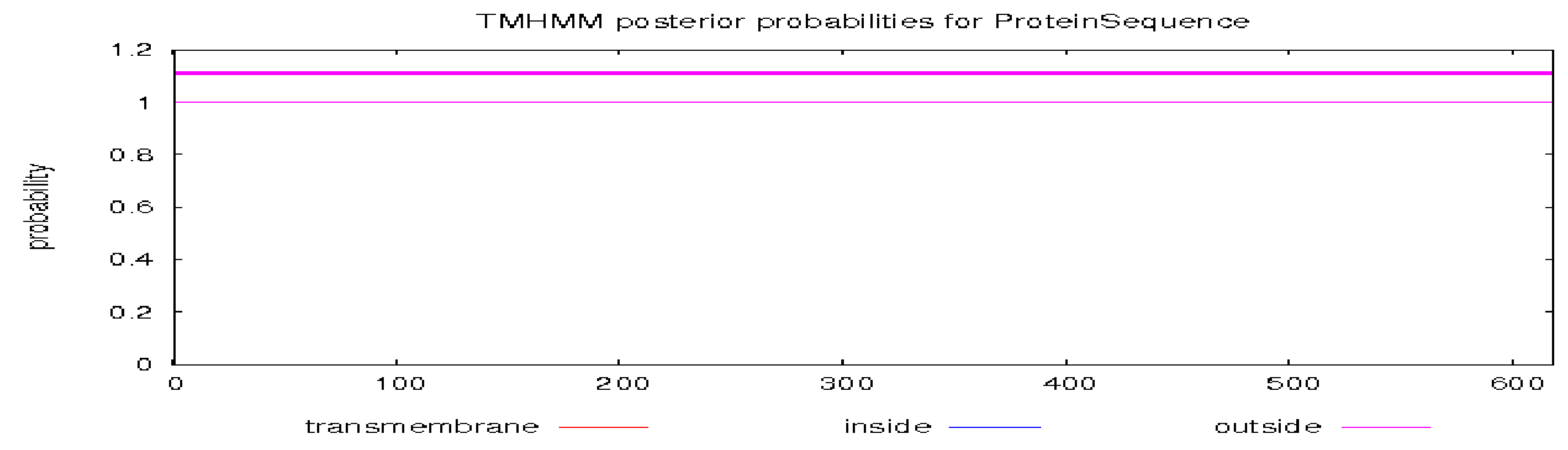


CDD – Conserved Domain



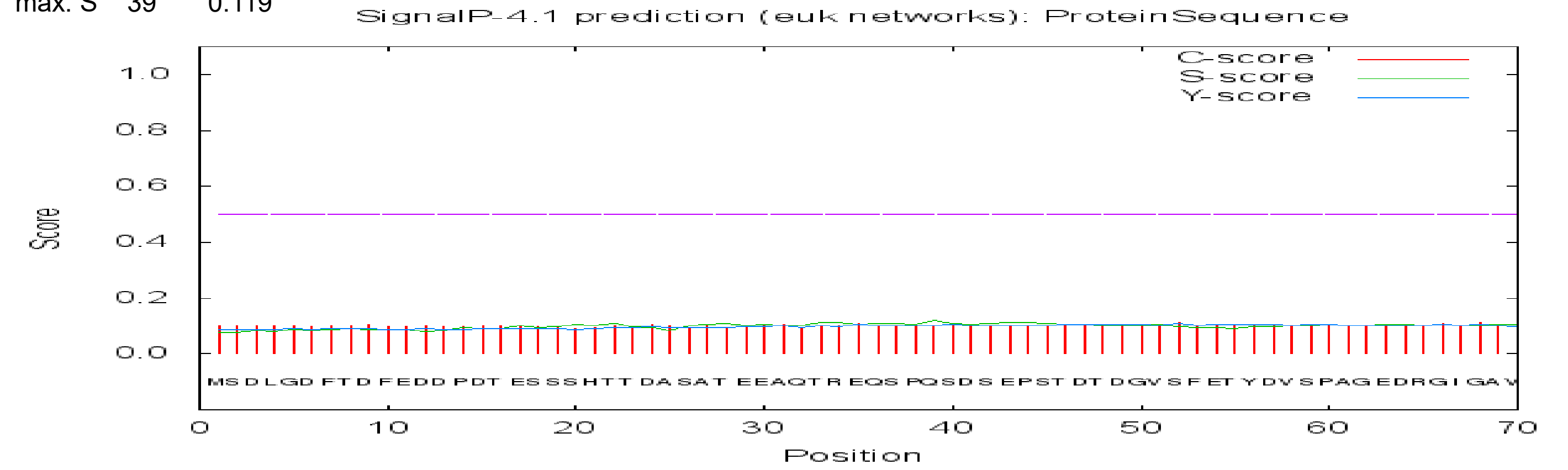
TMHMM – Transmembrane Helices

Total prob of N-in: 0.00011; Outside: 1.0 (618 Amino Acids)



SignalIP

Measure Position Value Cutoff signal peptide?
max. C 68 0.112
max. Y 52 0.109
max. S 39 0.119



PSORTb – Cellular Localization Data

Cytoplasmic 7.50
Cytoplasmic Membrane 1.15
Cell wall 0.62
Extracellular 0.73

Final Prediction:

Cytoplasmic 7.50

IMG/EDU – Alternate ORF

