

TransAAP: A Web-based Transporter Annotation Tool

User Guide

Qinghu Ren

April 27, 2007

qren@jcvl.org

(301) 795-7521

J. Craig Venter

I N S T I T U T E

TransAAP is a web-based transporter annotation tool. It allows users to submit their favorite genomes for transporter annotation utilizing our high-throughput transporter analysis pipeline, view automatic annotation and supporting evidence, as well as curate the transporter annotation, all of which are through a user-friendly web interface. This manual summarizes the process and illustrates the procedures step by step.

To maximize the possibility of finding out all the potential candidates of transporters, we employed four different kinds of searches with a query genome (Figure 1): blastp search against our internal curated transporter database and a general non-redundant protein database; HMM search against Pfam and TIGRfam; rpsblast search against NCBI COG; and transmembrane segment predictions by tmhmm. These searches are mutual inclusive, which means any positive hits above the cutoff from any of the four searches will be collected as potential transporters. In the meantime, there are a set of negative rules enforced to remove the false positives. Each predicted transporter will be shown on individual web pages where users can view annotation, supporting evidence, and curate the annotation.

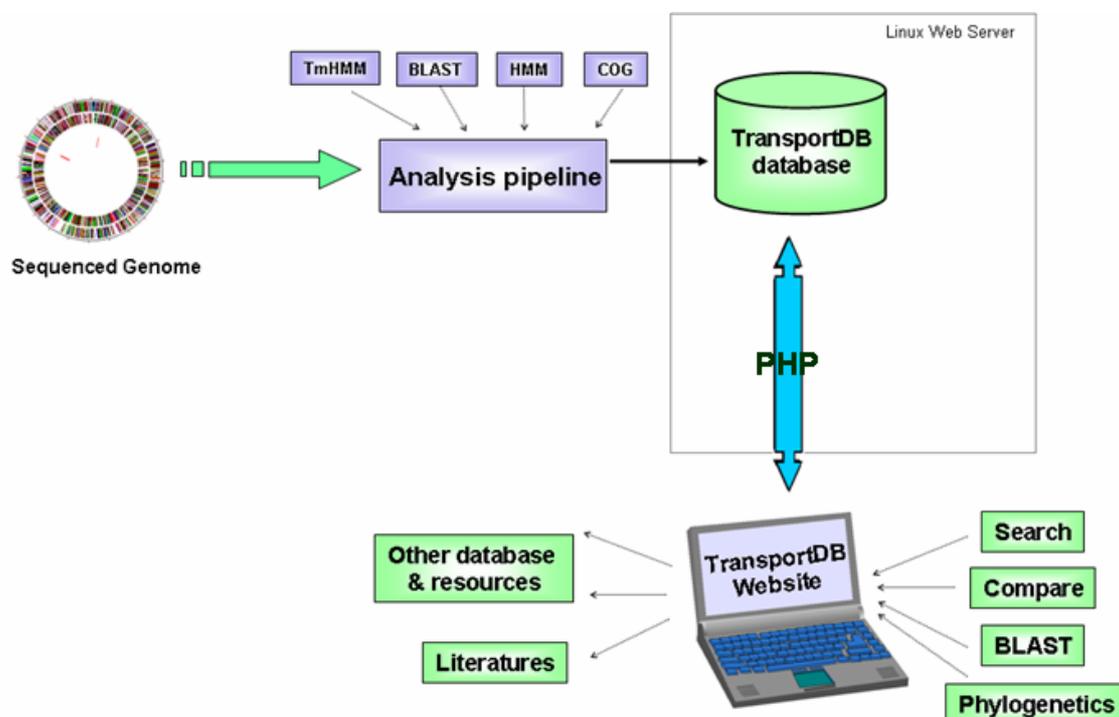


Figure 1: Illustration of TransAAP pipeline and TransportDB database

Users can go to the TransportDB web site, click on the link to TransAAP (http://www.membranetransport.org/transaap/TransAAP_login.html), and login as a registered user. To submit a request for transporter annotation, users only need to fill in simple information like Taxonomy ID for genomes deposited in the Genebank, TIGR database name for a TIGR/JCVI sequenced genome, or the FASTA format protein sequences/links to FTP protein sequence download site for unfinished genomes (Figure 2).

TransAAP: Transporter Automatic Annotation Pipeline

Submit Your Genome

User Name: test

Genus Name:

Species Name:

Strain Name:

Genome Source:

Complete Genome deposited to GeneBank
(No need for FASTA sequences)
NCBI Taxonomy ID:

TIGR Genome
(No need for FASTA sequences)
Database Name:

Unfinished Genome (Please paste peptide FASTA file below)

Figure 2: TransAAP: submit a query genome

Once we received the request, we will run all the required searches, predict the complete transporter contents, and load all the results to a database. Users can check the status of the annotation on their submitted genomes (Figure 3)

TransAAP: Transporter Automatic Annotation Pipeline

List of genomes you submitted for annotation
(Click on [GenomeID](#) or [Status](#) to see the full list of transporters)

GenomeID	Genus	Species	Strain	TaxonomyID	TIGR_db	Date of Submission	Status
N/A	Nautilia	sp.	Am-H		gamh	2007-03-13	Submitted
ngon1	Neisseria	gonorrhoeae	FA 1090	242231		2007-03-09	Automatic Annotation Finished

Figure 3: TransAAP: check the status of submitted genomes

After all the searches and automatic annotation are finished, users can access a list of predicted transporters (Figure 4), or individual transporter annotation pages (Figure 5). Users can view annotation, supporting evidence, and curate the annotation on each individual transporter annotation page.

TransAAP: Transporter Automatic Annotation Pipeline

User Name: **test**
Genome ID: **ngon1**

Complete List of Membrane Transporter in the Genome

Total transporter genes predicted: **141**

ORF Name order	GenomeID	FID order	Transporter Family	Subfamily	Substrate/Function order	Auto- annotation?	Curation? order
NGO0023	ngon1	ABC	The ATP-binding Cassette (ABC) Superfamily		iron compounds	✓	✓
NGO0088	ngon1	MOP	The Multidrug/Oligosaccharidyl-lipid/Polysaccharide (MOP) Flippase Superfamily	PST	polysaccharide export	✓	✗
NGO0110	ngon1	CitMHS	The Citrate-Mg ²⁺ :H ⁺ (CitM) Citrate-Ca ²⁺ :H ⁺ (CitH) Symporter (CitMHS) Family		proton:citrate symporter	✓	✗
NGO0136	ngon1	NhaC	The NhaC Na ⁺ :H ⁺ Antiporter (NhaC) Family		sodium ion;proton antiporter	✓	✓
NGO0142	ngon1	MFS	The Major Facilitator Superfamily (MFS)		glucose/galactose	✓	✓
NGO0143	ngon1	NhaC	The NhaC Na ⁺ :H ⁺ Antiporter (NhaC) Family		sodium ion;proton antiporter	✓	✗
NGO0168	ngon1	ABC	The ATP-binding Cassette (ABC) Superfamily	binding protein	manganese/zinc ion	✓	✗
NGO0169	ngon1	ABC	The ATP-binding Cassette (ABC) Superfamily	membrane	manganese/zinc ion	✓	✗
NGO0170	ngon1	ABC	The ATP-binding Cassette (ABC) Superfamily	ABC	manganese/zinc ion	✓	✗
NGO0181	ngon1	Tat	The Twin Arginine Targeting (Tat) Family		protein export	✓	✗
NGO0182	ngon1	Tat	The Twin Arginine Targeting (Tat) Family		protein export	✓	✗
NGO0183	ngon1	Tat	The Twin Arginine Targeting (Tat) Family		protein export	✓	✗
NGO0189	ngon1	RND	The Resistance-Nodulation-Cell Division (RND) Superfamily	SecDF	protein export (SecDF)	✓	✗
NGO0190	ngon1	RND	The Resistance-Nodulation-Cell Division (RND) Superfamily	SecDF	protein export (SecDF)	✓	✗

Figure 4: TransAAP: list of predicted transporters

TransAAP: Transporter Automatic Annotation Pipeline

Username: **test** GenomeID: **ngon1** [Return to the Transporter List in ngon1](#)

NGO0136

The NhaC Na⁺/H⁺ Antiporter (NhaC) Family
sodium ion:proton antiporter

Genome:

ORF Name:

Transporter List Index

ORF	NGO0088	NGO0110	NGO0136	NGO0142	NGO0143
Family	MOP	CRMH5	NhaC	MFS	NhaC
Subfamily	PST				
Substrate Function	polysaccharide export	proton: citrate symporter	sodium ion:proton antiporter	glucose/galactose	sodium ion:proton antiporter

Transport Annotation:

Transporter Prediction #1:

FID	TC#	Transporter Family	Subfamily	Substrate/Function
NhaC	2.A.35	The NhaC Na ⁺ /H ⁺ Antiporter (NhaC) Family		sodium ion:proton antiporter

Comments:

annotated

Delete this annotation?

Transmembrane Segment Prediction (TmHMM):

<p>Protein Length: 506</p> <p>Predicted TMS: Total: 13</p> <ul style="list-style-type: none"> TMS 1: 4-26 TMS 2: 28-50 TMS 3: 77-96 TMS 4: 117-139 TMS 5: 167-189 TMS 6: 201-223 TMS 7: 256-278 TMS 8: 299-318 TMS 9: 333-352 TMS 10: 365-387 TMS 11: 392-414 TMS 12: 453-475 TMS 13: 479-498 	<p>Topology:</p> <pre>>NGO0136 hypothetical protein [Neisseria gonorrhoeae FA 1090] MQLIDYSHSFFSVVPPFLALALAVITRIVLDSLSGIGSILVSVAFIVGQWVVDGLTHLKDMVVGGLAMADGDW SLGKPKLEWVLELLGIFTSLELRYSGSQCAPADMAKRHKKRFCKGAKMFAECLVETVEIDYFHSILAVGALIA RPVTDKFKVYSRAKILAVILLDSTASPMCLINPVSSWGASIIAFLASGLVIVYKITEYTPMCTFYAMSIMMYEA LVALINVEVYVAFPSFDIGSMARFEOAALNEAHDETAASDATKGRVYKALIIIVYALIASVTSVAMITSAQA SETFSILGAFENTDVNTSLVFGGTCGLVAVLCTFGTIKTADYFKAVWOGAKSMFGALIALLLAMLISTV VGEHMTGDYLSLTVAGNIHGGELVYILFLLASVMAFATGTSWGTFGIMLPIAAMAVKVEPALIIPCMSA VMAGAVCGDHCSPISTDTLLSSTGARCNHIDHVTQSLPYALTVAASAAASGLALGLTKSALLGPGSTGIV LAVLIFLLKDKKRAVD</pre>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

TransportDB & TCDB Blast Search

TDB NMB0531 15676438	NhaC 2.A.35 sodium ion/proton conserved hypothetical protein [Neisseria meningitidis MC58]	2514	5.7e-262
TDB HD1227 33152323	NhaC 2.A.35 sodium ion/proton conserved putative integral membrane protein [Haemophilus ducreyi 35000HP]	1813	1.1e-187
TDB PM0871 15602736	NhaC 2.A.35 sodium/proton unknown [Pasteurella multocida Fm70]	1543	4.5e-159
TDB HI1586 16273480	NhaC 2.A.35 sodium ion/proton conserved hypothetical integral membrane protein [Haemophilus influenzae KW20]	1540	9.4e-159
TDB MS1321 52425376	NhaC 2.A.35 sodium ion/proton NhaC protein [Mannheimia succiniciproducens MBEL55E]	1514	5.3e-156

[Draw a phylogenetic tree with Transporter Hits](#)

COG Search

COG1757	NhaC, Na ⁺ /H ⁺ antiporter [Energy production and conversion].	292	3e-80
---------	--------------------------------------------------------------------------------------	-----	-------

Pfam Search

2.A.35	NhaC	The NhaC Na ⁺ /H ⁺ Antiporter (NhaC) Family
PF03553	3.8e-102	Na ⁺ /H ⁺ antiporter family

NRAA Blast Search

RF YP_207306.1	hypothetical protein (Neisseria gonorrhoeae FA 1090); (exp=0; wgp=1; cg=1; closed=1; pub=0; rf_status=provisional); ^ ^GB AAW88894.1 59717489 AE004969 conserved hypothetical protein (Neisseria gonorrhoeae FA -TRUNCA	2555	1.2e-264
RF NP_273576.1	hypothetical protein (Neisseria meningitidis MC58); (exp=0; wgp=1; cg=1; closed=1; pub=1; rf_status=provisional); ^ ^GB AAF40961.1 7225756 AE002098 conserved hypothetical protein (Neisseria meningitidis MC58)-TRUNCA	2514	2.6e-260
GB CAM09770.1	putative integral membrane protein (Neisseria meningitidis FAM18); (exp=0; wgp=0; cg=1; closed=0; pub=0; rf_status=)	2500	8.0e-259
RF NP_283509.1	integral membrane protein (Neisseria meningitidis Z2491); (exp=0; wgp=1; cg=1; closed=1; pub=1; rf_status=provisional); ^ ^GB CAM07964.1 121051661 AL157959 putative integral membrane protein (Neisseria meningitidis TRUNCA	2490	9.1e-258
RF NP_873676.1	conserved putative integral membrane protein (Haemophilus ducreyi 35000HP); (exp=0; wgp=1; cg=1; closed=1; pub=0; rf_status=provisional); ^ ^GB AAP96065.1 33148546 AE017143 conserved putative integral membran-TRUNCA	1813	5.0e-186
RF NP_245808.1	hypothetical protein (Pasteurella multocida subsp. multocida str. Fm70); (exp=0; wgp=1; cg=1; closed=1; pub=1; rf_status=provisional); ^ ^GB AAK02955.1 12721182 AE004439 unknown (Pasteurella multocida subsp. -TRUNCA	1543	2.0e-157
SP P44263	Hypothetical protein HI1586. (Haemophilus influenzae); (exp=0; wgp=-1; cg=-1; closed=-1; pub=0; rf_status=); ^ ^RF NP_439731.1 16273480 NC_000907 isoleucyl-tRNA synthetas (Haemophilus influenzae Rd KW20); (e-TRUNCA	1540	4.3e-157
OMNI HI_1586	Na ⁺ /H ⁺ antiporter family (Haemophilus influenzae Rd KW20); (exp=-1; wgp=-1; cg=1; closed=-1; pub=-1; rf_status=)	1540	4.3e-157
RF YP_248941.1	predicted Na ⁺ /H ⁺ antiporter (Haemophilus influenzae 86-028NP); (exp=0; wgp=1; cg=1; closed=1; pub=1; rf_status=provisional); ^ ^GB AAW88281.1 68058028 CP000057 predicted Na ⁺ /H ⁺ antiporter (Haemophilus influen-TRUNCA	1539	5.4e-157
RF YP_088513.1	NhaC protein (Mannheimia succiniciproducens MBEL55E); (exp=0; wgp=1; cg=1; closed=1; pub=0; rf_status=provisional); ^ ^GB AAU37928.1 52307428 AE016827 NhaC protein (Mannheimia succiniciproducens MBEL55E); (ex-TRUNCA	1514	2.4e-154

Figure 5: TransAAP: individual transporter annotation page