

Supplemental TABLE I. Primers for PCR amplification of TcOGNT RNA, DNA and mutagenesis

<u>Purpose</u>	<u>Primer</u>	<u>Species</u>	<u>Sequence</u>
Expression			
full-length TcOGNT2 ¹	E5S1	<i>L. tarentolae</i>	5' - <u>acagatct</u> ccaccatggataaaaagaagc
TcOGNT2cat ²	E5S2	<i>L. tarentolae</i>	5' -gaggcgccgggtcccaccatggaattg
both	E5AS1	<i>L. tarentolae</i>	5' - <u>gactcgagtt</u> atccaagtgttgcccttttc
full-length TcOGNT1	LiFullS	<i>L. tarentolae</i>	5' - <u>atccatgg</u> agcttgaggaaaccaatttgta
full-length TcOGNT1	LiCATAS	<i>L. tarentolae</i>	5' - <u>atctcgagtc</u> atccgcaccatcgaccg
full-length TcOGNT2 ²	E5SATG2	<i>D. discoideum</i>	5' - <u>ctggtagc</u> ctgatcaatggataaaaaggaagcttgac
full-length TcOGNT2 ²	E5AS2	<i>D. discoideum</i>	5' - <u>gtagatctt</u> ccaagtgttgccctttca
TcOGNT2cat ¹	E5U2	<i>D. discoideum</i>	5' - <u>gaagatct</u> tagatttgacaggcgtgggtccc
TcOGNT2cat ¹	E5L1	<i>D. discoideum</i>	5' - <u>gtgagctc</u> ttatccaactgttgcccttttcgc
TcOGNT2cat ¹	E5S3	<i>E. coli</i>	5' - <u>gacatatg</u> agatttgacaggcgtgggtccc
TcOGNT2cat ¹	E5AS3	<i>E. coli</i>	5' - <u>gtggatcct</u> ccaactgttgcccttttcgc
Mutagenesis and expression			
TcOGNT2cat ²	TcE5D234Afor	<i>L. tarentolae</i>	5' -gtgattgctagccattcaagatttgtgcctgagtgg
TcOGNT2cat ²	TcE5D234Arev	<i>L. tarentolae</i>	5' -cttgaatggctagcaatcaccatatagtaatcctgg
TcOGNT2cat ²	TcE5D234Nfor	<i>L. tarentolae</i>	5' -gtgattaatagccattcaagatttgtgcctgagtgg
TcOGNT2cat ²	TcE5D234Nrev	<i>L. tarentolae</i>	5' -cttgaatggctattaatcaccatatagtaatcctgg
5'-splice site mapping			
<i>TcOGNT2</i> ¹	TcE5AS415	<i>T. cruzi</i>	5' -gagagcgagtcaacacaatgttct
<i>TcOGNT1</i>	TcE2NAS	<i>T. cruzi</i>	5' -tgaaaacccaaccttcggac
miniexon primer	TcME33	<i>T. cruzi</i>	5' -cgctattattgatacagtttctgtactatattg

¹ Haplotype Tc00.1047053511309.70 was cloned.

² Haplotype Tc00.1047053511759.30 was cloned.

Restriction enzyme sites for cloning are underlined.

Supplemental Table II. *TcOGNT2cat* identification in gel bands

Gel bands in Figure 7E were excised, trypsinized and identified by MALDI-TOF-TOF-MS as described in Experimental Procedures. MS/MS data consist of at least 5 expected b or y ions detected after fragmentation. Peptides detected are mapped in Suppl. Figure 2.

TcOGNT2 position	MH ⁺ observed	MH ⁺ expected	Additional evidence	Bands
66-87	2314.0	2314.3	MS/MS	Top, middle
88-101	1698.6	1698.8		Top, middle
117-131	1712.6	1712.8		Top, middle
217-238	2645.3	2644.9	MS/MS	Top, middle
279-300	2416.8	2417.2		Top, middle
301-327	3062.3	3062.6		Top, middle
306-327	2378.9	2379.2		Top, middle
350-371	2708.9	2709.2		Top, middle
432-442	1319.6	1319.4	MS/MS	Top, bottom
450-468	2412.9	2413.1		Top, bottom

SUPPLEMENTAL FIG. 1. Alignment of sequences corresponding to the NRD2- and Gal/GalNAc-like domains of select family GT27 and GT60 sequences. Human (Hs) pp- α GalNAcT2 and murine (Mm) pp- α GalNAcT1 initiate *O*- α GalNAc mucin type glycosylation and belong to CAZy family GT27 (Canteral et al. 2009). Dd- α GlcNAcT2 (DdGnt2) initiates the related *O*- α GlcNAc mucin type glycosylation in *Dictyostelium*, and belongs to family GT60. The three related sequences from *T. cruzi* most closely resemble the GT60 sequence class. Amino acids are colored according to group properties, and shaded (using a distinct color scheme) at positions where this property is conserved in either the pp- α GalNAcT or the pp- α GlcNAcT sequences (West et al. 2004), not including TcOGNTL which is predicted to be catalytically inactive. Positions of sequence identity are bolded, and positions in Mm pp- α GalNAcT1 where amino acid substitutions inactivate the enzyme (Hagen et al. 1999) are marked with an asterisk. Amino acid positions relative to the full-length protein are indicated in parentheses.

		NRD2 Motif-1	
Hs-ppGalNAcT2	(139)	S VVIT---FHNE----- A RSALLRTVVS V LK K SP P HLI--- K EIILV- D DYS(178)	
Mm-ppGalNAcT1	(119)	S VVIV---FHNE----- A WSTLLRTV H SVINRS P RHMI--- E EIVLV- D DAS(158)	
DdGnt2	(63)	T IFV S LAA Y RD V F----- C SDTINYIF N H- A NR P E K IF I G I V D Q G S E F L E E D F (109)	
TcOGNT2	(78)	T IFV S IA A F R D K E----- C V T L L E S L F T R - A KN F RRV L G I S E ER F EM- D S S C(123)	
TcOGNT1	(357)	S L F L N IA S F R D K E----- C W P S I D H M I Q R - S T N M F RV V W G V A Q Q H Y YS- D L P C(402)	
TcOGNTL	(182)	S IFV G IT-- H S O E K V T S S E M E T A C A A T V R N I Y D A - A H W P L G V F T G I V E IT L S K T S P T T(236)	
		NRD2 Motif-2	
Hs-ppGalNAcT2	(199)	D RR E GL M RS R VR G A- D AA O A K V L T F L D SH C E C NE H W L E P LL E R V A E D R ----- T R V V S P I I D V I (256)	
Mm-ppGalNAcT1	(184)	E OR S GL I R A R L K G A- A V S R G Q V I T F L D A H C E C T A G W L E P L L A R I K H D R ----- R T V V C P I I D V I (241)	
* *			
DdGnt2	(143)	A Q S K G P T L A R Y Y A T T L Y N N E T Y F M Q V D S H L R F I K G W- D S L I I N D L W L T K S Y A P I G E N G I P R T V L T H Y P M A Y(212)	
TcOGNT2	(208)	D D A R G P T Y G R Y L T S L F Y N N Q D Y M V I D S H S R F V P E W - D I K M I E R A R L M P T ----- W G V M S Y P N G Y(267)	
TcOGNT1	(448)	F T A L G P A F G R Y M S M L L Y C G E D Y M L V L D S H N R F V Y A W- D A R I I A M Q V A L E H P----- K A V L S H Y P E S Y (508)	
TcOGNTL	(318)	T L L S S A A A Q R Y A T L A L Y R G E T Y V M F V R A G L Q L V Y K W- D V L T R L L W L Q L P S ----- R S A V L S Q P V Q I (378)	
		Gal/GalNAc Motif	
Hs-ppGalNAcT2	(290)	P I K T P M I A G G- L F V M D K F Y F E E L G K Y D M M D - V W G G E N L E I - S F R V W Q C G G S L E I I P C S R- V G- H V F - R K Q H P Y T F F (369)	
Mm-ppGalNAcT1	(285)	P V R T F T M A G G- L F S I D R D Y F Q E I G T Y D A G M D- I W G G E N L E I - S F R I W Q C G G T L E I V T C S H- V G- H V F - R K A T Y T F F(354)	
*			
DdGnt2	(254)	P A E C F Y I A A G F F T - S G E A I K L V P - F D P H L S N L F E G E E I L Y- S V R M Y - S A G - F R F F A P T L N V C F H Y S R P K S P K F F W(324)	
TcOGNT2	(308)	P V L Q G F V A A G Y I F G- D A Q F V K D V P - F D P Y L P F L D G E E I L Y - T V R L W - T S G - W D S Y C P G T A F L F H N Y M R T D A P R F F W (378)	
TcOGNT1	(575)	P L A Q F W A A G G F L F A- N A S V M R E V P - F D P H L P H L D G E E V M Y - S V R L W - T H G - Y D I F S P K R G I C Y H F Y D R P N E P K L L W (645)	
TcOGNTL	(627)	Y L Q S S W V T P D F L F T- R A E -21- V P- L D P F L S F L G A D E E A V L L S A R L W- T H G - W D F F S S T E P I A F I V T - K P P A E G D A S (716)	

SUPPLEMENTAL FIG. 2. Annotated nucleotide and amino acid sequences of the two TcOGNT2 haplotypes. Nucleotide and amino acid sequences are numbered beginning with the start codon. 5'-splice acceptor sites for TcOGNT2-1 (violet), primer hybridization sites, N-glycosylation sequons, site of mutagenesis, peptide sequences (green or blue) identified by MS (Suppl. Table II), and key motifs are denoted.

>Tc00.1047053511759.30 (*TcOGNT2-2*)

>Tc00.1047053511309.70 (*TcOGNT2-1*)

-240 aacatgtacatttacctgttggttaaactttgaggatgtggtgaattgactcgtatttcat
-180 tcatctctttcctatgcttatatatatatatattttgatttctattattgatgaacttgg
-120 gatccggcgagcggcccagctgaatatttggagaggttttacggaaaagggtgtagttc *5'-splice donor nt*
-60 gacacacaaaaagcaatcaagtaaagggtgaagtaacaactcatatttacatcagcccatg

1 atggataaaaaggaagccttggacggggccttgaggattcatcgtggggggaggggtacaaaag **E5SATG2**
atggataaaaagaagccttggacggagccttgagggttcatcgtggggggaggggtggaaaag **E5S1**
1 M D K R K L G R G L R I H R G G R G T K
K S V G

61 agattttttcgagtgtttttggttttcgtggcatttttttactgctgttttttagattt
agatgttttcgagtgtttttggtttgcgtggcatttttttactgctgttttttagattt **E5U2, E5S3**
21 R F F R V F L V F V A F F L L L F F R F
C C

121 gagaggcgt**ggtcccaccatggaattg**gaagatgcagacggcattcttcagacgataccg **E5S2**
gacagggcgt**ggtcccaccatggaattg**gaagatgcagacggcattcttcagacggtacct
41 E R R G P T M E L E D A D G I L Q T I P
D V

181 gtattttcttcgaaggatcttccggctgcaaataataaccagtggacattgctactatttt
gtattttcttcgaaggatcttccggctgcaaataataaccagtggacattgctactatttt
61 V F S S K D L P A A N I P V D I A T I F *tryptic peptides*
(MS of 40KD band)

241 gtatccattgcggccttttcgagacaaggaatgtgtcactacactagagagtctctttacc
gtatccatcgcggccttttcgagacaaggaatgtgtcactacactagagagtctctttacc
81 V S I A A F R D K E C V T T L E S L F T *tryptic peptides*
(MS-MS of 40KD band)

301 cgcgcgaagaatcctcggagagtgtatttgggtatcagtgaggagcgatttgaaatggat
cgcgcgaagaatcctcggagagtgtatttgggtatcagtgaggagcgatttgaaatggat
101 R A K N P R R V Y L G I S E E R F E M D

361 agcagttgcattgcttcaccggaggtgcttaagaacattggtgtgactcgcctctcgtaaa **E5AS415**
agcagttgcattgcttcaccggaggtgcttaagaacattggtgtgactcgcctctcgtaaa
121 S S C I A S P E V L K N I G V T R S R K

421 cttcactggaaggatgttgttccctttgcctacaatcccacagacaagcagaaatatccc
cttcactggaaggatgttgttccctttgcctacgatcccacagacaagcagaaatatccc
141 L H W K D V V P F A Y N P T D K Q K Y P
D

481 aagcacacacctgtgctgcatgcaaaacgcgatgaggatgtcatcacgtgtctactcagt
aagcacacacctgtgctgcatgcaaaacgcgatgaggatgtcatcacgtgtctactcagt
161 K H T P V L H A K R D E D V I T C L L S

541 gaaggttcttccgacgcaatgcccgccctgggtgaatctgttttgaacggatgtcaggtc
gaaggttcttccgacgcaatgcccgccctgggtgaatctgttttgaacggatgtcaggtc
181 E G S S D A M P G P G E S V L N G C Q V

601 ttgacacgcattggacacccggatgatgcacgaggcccgacttacggtcggttacctaaca
ttgacacgcattggacacccggatgatgcacgaggcccgacttacggtcggttacctaaca
201 L T R I G H P D D A R G P T Y G R Y L T

661 agtctcttctacaataaccaggattactatatggtgattgacagccattcaagatttgtg
agtctcttctacaataaccaggattactatatggtgattgacagccattcaagatttgtg

221 S L F Y N N Q D Y Y M V I D S H S R F V DxD-like motif
N/A mutagenesis

721 cctgagtgggatataaagatgattgaacgtgacggttattgccacttggggcgtaatg
cctgaatgggatataaagatgattgaacgtgacggttattgccacttggggcgtaatg

241 P E W D I K M I E R A R L M P T W G V M

781 tcgtactacccgaatgggtacacaccggaggatccagatgaggagatcaacaagacggat
tcgtactacccgaatgggtacacaccggaggaccagatgaggagatcaacaagacggat

261 S Y Y P N G Y T P E D P D A E I **N K T D** **N-glycosylation sequon**

841 gttatggccatgtgtaaggggttatcttagataatgacattccaaagtggggcgcg
gttatggccatgtgtaaggggttatcttagacaatgacattccaaagtggggcgcg

281 V M A M C K G V I L D N D I P K L G A R

901 tggattgatataaggaccatcctgttctccaaggctttgtagctgctggatataat
tggattgatataaggaccatcctgttctccaaggctttgtagctgctggatataat

301 W I D I R D H P V L Q G F V A A G Y I F

961 ggggatgagcagtttgtgaaggatgtcccttttgatccgtatcttccatttttattgat
gggatgagcagtttgtgaaggatgtcccttttgatccgtatcttccatttttattgat

321 G D A Q F V K D V P F D P Y L P F L F D

1021 ggagaggagattctgtacactgtacgctgtggacaagcgggtgggacagttactgtcct
ggagaggagattctgtacactgttcgcttgggacgagcgggtgggacagttactgtcct

341 G E E I L Y T V R L W T S G W D S Y C P

1081 ggcacggcatttctttttcataactacatgcaaccgatgagcggaggttttggtccgtc
ggcacggcatttctttttcataactacatgcaaccgatgagcggaggttttggtccgtc

361 G T A F L F H N Y M R T D A P R F W S V

1141 ttgacaagagatgaggtgaaggttcatcggaatttgaacggcagacctccataaaacgt
ttgacaagagatgaggtgaaggttaccgggaatttgaacggcagacctccataaaacgt

381 L T R D E V K V H R E F E R Q T S I K R

1201 gcgttgatttaaatgagcgggaatgtcttgaactccaccaaccggttgtagcgaagat
gcatttgatttaaatgagcgggaatgtcttgaactccaccaaccggttgtagcgaagat

401 A L Y L M R R N V L **N S T** Q P V V T E D **N-glycosylation sequon**

1261 gaggcgatcgattgaaccggcgattggacgtgagatggaacactttggattggggatg
gaggcgatcgattgaaccggcgattggacgtgagatggaacactttggattggggatg

421 E A H R L N P A I G R E M E H F G L G M **tryptic peptide**
S **(MS of 15KD band)**

1321 ttgacgacgatttccgattactggcgttttggtggaacttagtgatgagtttgtaagaa
ttgacgacgatttccgattactggcgttttggtggaacttagtgatgagtttgtaagaa

441 L R R I S D Y W R F V E L S D E F V K E

1381 aaagatgatgagaatcgctggaagggtggtgaaggctctgtgaaaaggcaacacttgga
aaggatgatgagaatcgctggaagggtggtgaaggctctgtgaaaaggcaacacttgga
gcaaaaaggcaacagttgga **E5AS1, E5AS2**
E5AS3, E5L1

461 K D D E N R W K G G E G L C E K A T L G
V

1441 taa
taa
-

SUPPLEMENTAL FIG. 3. Synteny of *OGNT* genes in *T. brucei*, *T. cruzi* and *L. major*. *OGNT1* (green boxes) and *OGNTL* (blue boxes) (**A**) and *OGNT2* (pink boxes) (**B**) genes are shown relative to neighboring predicted coding regions. Shaded gray boxes represent conserved hypothetical protein orthologs; white boxes are hypothetical proteins that are not conserved among these chromosomes; red boxes represent casein kinase 2 (CK2); and yellow boxes are phosphoglycan β 1-3-galactosyltransferase-like protein (PG β 1-3GalT) and putative phosphoglycan β 1-2-arabinosyltransferase (PG β 1-2AraT). Numbers above each box represent the identifier protein numbers for the corresponding *T. brucei* and *L. major* chromosomes (indicated at the right), and for *T. cruzi* haplotypes Tc00.1047053508741.340/Tc00.1047053508741.360 for *OGNT1/OGNTL* and Tc00.1047053511759.70 for *OGNT2*, respectively.

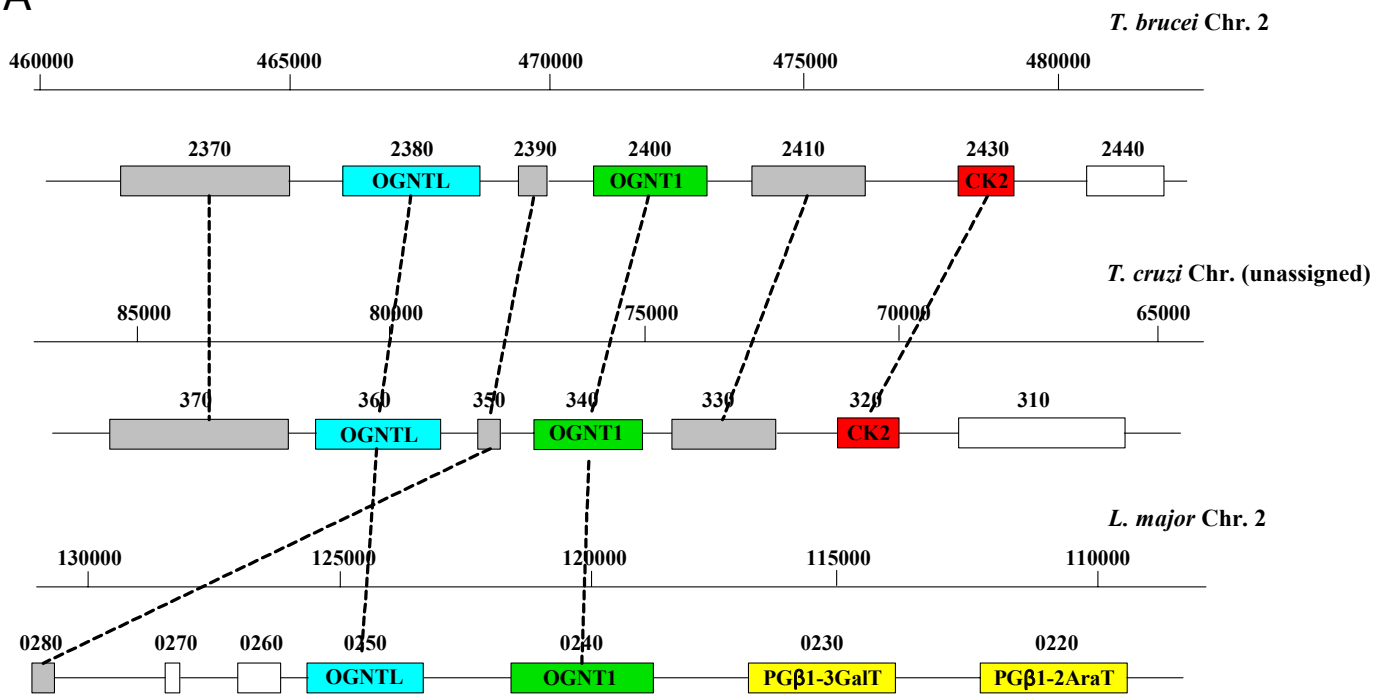
SUPPLEMENTAL FIG. 4. Mapping the 5'-region of the *TcOGNT2* transcript. (**A**) Ethidium bromide-stained agarose gel of the reverse transcription and PCR amplification of total epimastigote RNA using the *TcOGNT2*-specific reverse primer TcE5AS415 and the miniexon primer TcME33 (lane 2). Lane 1: same without reverse transcriptase (from a separate gel). Arrow marks position of the ~500 bp product. (**B**) The amplification product was vector-ligated and transfected into *E. coli*. Ten clones were sequenced, and digested with EcoRI and BamHI to release the insert and analyzed by agarose gel electrophoresis. Ethidium bromide staining revealed minor size heterogeneity of the ~500 bp fragments (arrow), suggesting multiple splice sites. Alignment of the sequences with the 5'-UTR of CL-Brener *TcOGNT2-1* gene revealed splice donor sites at positions -70, -101, -109 relative to the translation start site (see Supplemental Figure 2).

SUPPLEMENTAL FIG. 5. Expression of *TcOGNT2* in *Dictyostelium*. A soluble protein fraction (300 μ g) prepared from sonicated microsomes of *D. discoideum* was assayed for 2 h in the absence or presence of peptide T29. *modB*⁻, DdGnt2-null strain DL118; *modB*⁻ cat, transfected with *TcOGNT2*cat-myc(C-terminal). Similar absence of detectable activity (data not shown) occurred with the other constructs tested (Figure 1B), and with tagged *TcOGNT1*, and tagged *TcOGNT1* or *TcOGNTL* coexpressed with *TcOGNT2* (unpublished data). Protein expression was verified in all cases (not shown). Results are representative of five independent experiments.

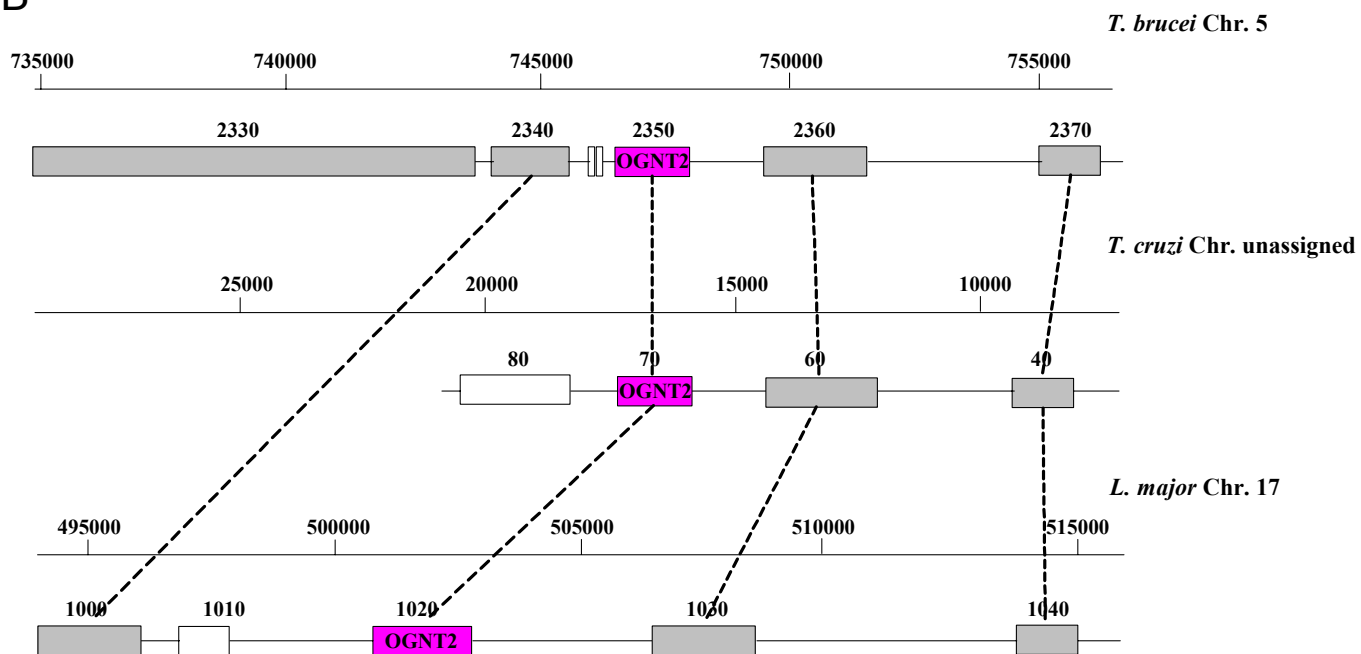
SUPPLEMENTAL FIG. 6. Temperature dependence of *TcOGNT2*-dependent and endogenous enzyme activities. (**A**) Transferase activity was assayed in the presence or absence of peptide acceptor as described in Figure 4A. (**B**) UDP-[³H]GlcNAc hydrolysis activity was assayed as described in Fig. 4C. Data are plotted relative to values at 22°C, which are shown in corresponding vertical order in the 22° column. Similar results were obtained in a second trial using *TcOGNT2* from hygromycin-resistant cells.

Supplemental Figure 3

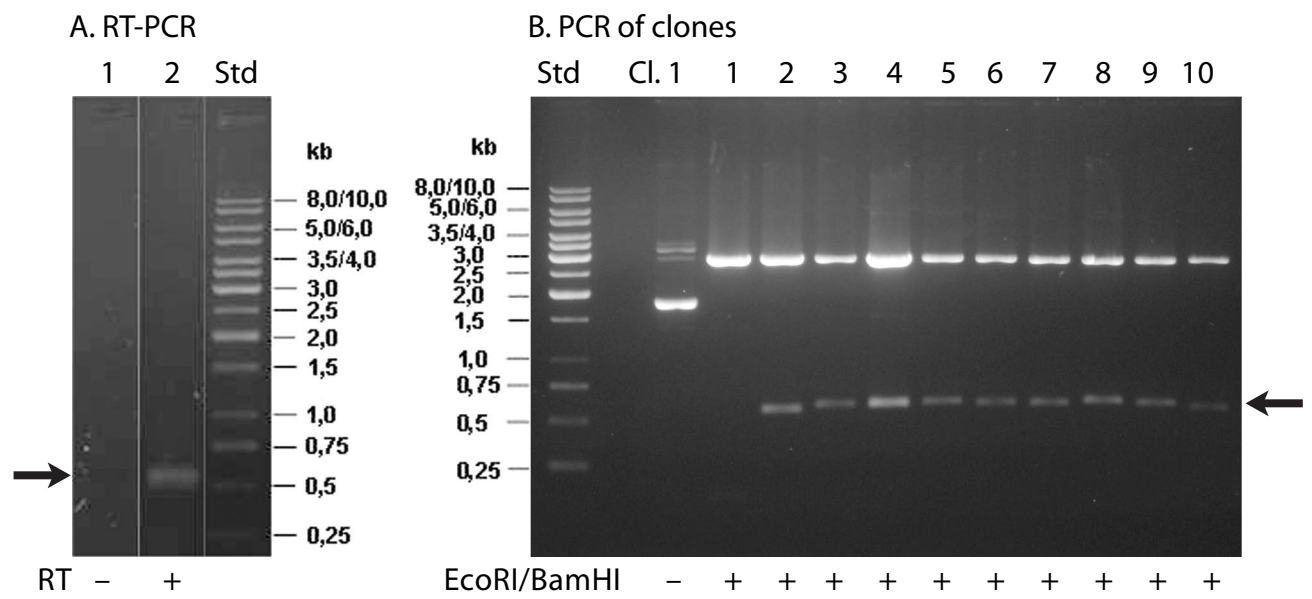
A



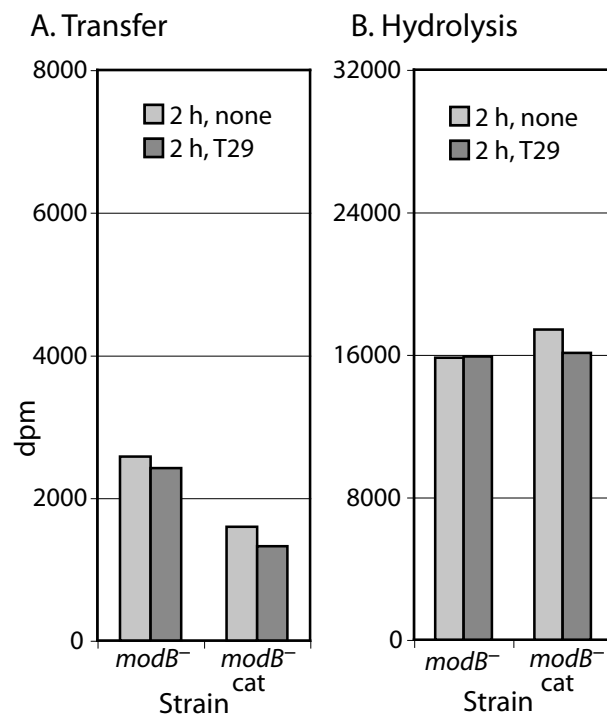
B



Suppl. Figure 4



Suppl. Figure 5



Suppl. Figure 6

