

SUPPLEMENTAL DATA:

Characterization of a Metal-Independent CAZy Family 6 Glycosyltransferase from *Bacteroides ovatus*

Percy Tumbale and Keith Brew

Supplemental figure legends:

Supplemental Figure 1. An alignment of bacterial GT6 glycosyltransferases. The alignment was generated using MUSCLE and shaded using BOXSHADE. Positions with shaded backgrounds are identical in 70% of the sequences. The boxes denote regions of the active site that discussed in the text. The sequences are Wcmb, *Escherichia coli* Wcmb (AAV85960); Psych, hypothetical protein PscPRwf_0985 from *Psychrobacter sp. PRwf-1* (YP_001279885.1); Hs, putative glycosyltransferase from *Haemophilus somnus* 2336 (YP_001783501); Bs, *Bacteroides stercoris* (B0NSM3); Bo1, *Bacteroides ovatus* GT6a (A7LVT2); Bo2, *Bacteroides ovatus* GT6b (A7M0P3); Bc, *Bacteroides caccae* (A5ZC71); Fp, *Francisella philomiragia subsp. philomiragia ATCC 25015* (cont1.16, whole genome shotgun sequence); Hm, *Helicobacter mustelae* (from the Wellcome Trust Sanger Institute Microbial Pathogens database (www.sanger.ac.uk); Sv1 and Sv2, GT6 sequences from *Subdoligranulum variabile* (NCBI Whole-Genome Shotgun Reads database).

Supplemental Figure 2. An alignment of mammalian, cyanophage PSSM-2 and selected putative phage GT6s from the Environmental Samples database. The sequences are : bova, bovine α -1,3 galactosyltransferase (GGTA1_BOVIN); dogFS, canine Forssman glycolipid synthase (GBGT1_CANFA); huA, human histo-blood group A synthase (BGAT_HUMAN); PSSM2, cyanophage PSSM-2 (Q58M87). The sequences designated GOS are from the Marine Metagenome: GOS_A, ctg_1101668274631; GOS_B, 1096626241734; GOS_C, 1096626359338; GOS_D, 1096626608311; GOS_E, 1096626802277; GOS_F, ctg_1101667133928; GOS_G, ctg_1101668196376; GOS_H, ctg_1096626132935; GOS_I, ctg_1096626203656.

Supplemental Figure 3. A homology model of the 3D structure of BoGT6a based on the structure of bovine α 3GT showing the residues subjected to mutagenesis. The coordinates were produced through the Swiss-Model website (swissmodel.expasy.org) and the structure is displayed in ribbon format using Chimera (Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004) *J. Comput. Chem.* **25** 1605-1612). The chain is colored blue to red from N-terminus to C-terminus; the side chains of residues that have major effects on activity are purple and others are dark grey.

Sv2	1	-----MSESRI R V A V L Y L C T G A Y I V F W H D F Y P N F R Q H F L P D C D -- R T F F V F T D A A S I D Y ---E
WcmB	1	-----MVINI F Y I C T G E Y K R F F D K F Y L S C E D K F I P E F E -- K K Y F V F T D S D R I Y F ---S
Bs	1	-----M K I G I L Y V G I G R Y I R L W D K F Y S S C E S M F L P Q Y E -- K Y F I F T D Y P L ---K
Psych	1	-----M K I T E L N M S L S T S A L S N N S K P S V A I L Y I A T G R Y I V F W D Y F Y K S A E K Y L L P D C N -- K H Y I L F T D S F A L I D S F R T
Hs	1	-----M V K N I G I L Y I A T G R Y I V L W E D F Y K S A E K H F L P E L P -- K K Y F L F S D Y E G S F L G --E
Bo2	1	-----M T I A I L Y I C T G Y N Q F F K G F Y E S C E K N F L I Q T D -- K T Y F V W T D D D H L A ---D
Bcac	1	-----M K V A I L Y I C T G Y N Q F F K G F F E S C E K Y L L K D I A - Q L E Y Y V F T D D M S L ---S
Fp	1	-----M S N K K L I V L Y I A T G E Y I K F W K D F Y L S A E K F F L K N S N I S L E Y I V Y T D S T C I F A ---E
Hm	1	M Q S T A Q N T Q Q N T H F A G S S Q T T P Q A A Q S V Q Q A S L A L P K S S P T C Y K I A I L Y I C T G A Y S I F W Q D F Y D S A K V H L L P A H R -- L T Y F V F T D A D S L Y A ---E
Sv1	1	-----M T K V A A L Y I C T G Y I A F W P E F Y D S A E Q N L L P G C E -- V H Y F V F T D A P V L Y G E ---E
Bo1	1	-----M R I G I L Y I C T G Y I P F W K D F Y L S A E R Y F M Q D Q S F I I E Y Y V F T D S P K L Y D E ---E
A		
Sv2	54	D Q P D V R R F Q Q E A L P W P Y S T M Q R F D A F L S Q A E A I - A D Y D Y L F F F A N A N L H C L R D V T A G E L L P D A A K G Q E L T V V C H L P Y Y - G -- R N P I F H P Y E R R R K C
WcmB	49	K Y L N V E V I N V E K N C W P L N T L L R F S Y F L K V I D K L - Q T N S Y T F F F N A N A V I V K E I P F S T F M E S D - - - - L I G V I H P G Y K - N -- R I S I L Y P W E R R K N A
Bs	46	S S T N V Q V S F Q E D L G W P Q V L F R Y Q M F L R H K E E L - K H F D Y L F F F N G N I T E F L Q I I T P E E F L P T E K E A - - L T G L E W H T Y R - N -- K P L Q K Y H Y E R R K N S
Psych	72	K S D Q V T A L K K E A M E W P F Q T L M R F R F F L D A E N I I - K Q H D F V F F F N A N T E F L S T I T Q Y D L L P L G S H E N - L T L C L Q P H M F - H -- R N R E K Y T Y D R N P K S
Hs	52	E S G N V H R I Y Q K K L G W P Y I T L M R F D M F L K E K E R F - K E L D Y L F F F V N A N M I F I K D I H Q N E V L A S D K E - - - L I M I L Q P W F I - G -- L D I D K T T Y D R N P L S
Bo2	48	G R S N V R I Y H K E C A G F P A D S L F R F E M F L Q A E Q E I - M K Y D Y V Y F F N A N A M C I E P I G N - E I L P D E S G - - - L S M G I W G G K R L H - O H P M F Y P Y E R N R K S
Bcac	48	D E A N V H L I K K E C A G F P A D S L F R F D M F L Q V R Q E L - E K T D Y I Y F F N S N A E F K A P V G K - E L L P L N G E K - - L V G A E W P G K R - K P F K H P A F Y P Y E R N K H S
Fp	54	E K E N V L R I Y Q E N M E W P Y I T L F R F K I F K K A L P Y L - Y N S D Y I F F F N S N M K F I S D V N E - K I L P Q G S K Q - - L T F V M H P G Y Y - N -- K P V K K F L Y E K N N N S
Hm	91	E A S D V R K I Y Q E N L G W P F N T L K R F E M F L G Q E E A L - R E F D F V F F F N A N C L F F Q H I G D - E F L P I E E D - - - I L V T Q H Y G F R - D -- A S P E C F T Y E R N P K S
Sv1	51	A N P R I H R C P Q E A Y S W P F A T L R R F E I F L S R E E E L - K A F D Y I F F F N A N A Q I M T T I T P E M F L P R A D R G E H L L V V Q H P S F Y - T -- K P N Y E F T Y D R N P R C
Bo1	52	N N K H I H R I K Q K N L G W P D N T L K R F H I F L R I K E Q L E R E T D Y L F F F N A N L L F T S P I G K E I L P P S D S N G - - L L G T M H P G F Y - N -- K P N S E F T Y E R D A S
B C D		
Sv2	145	R A G I - P Y N C G T Y - Y V A G G I N G G A S A F L E M C R E L K A R T D E D L Q R G I A R C H D E S Q L N R L V A E C P E R F R I L P P E Y C T P E E T - - - P T G K E A I R V L Q K
WcmB	135	T C Y L - G Y L K K G I - Y Y Q G C F N G G K T A S F K R L I Q C N M M T M A D L K K N L I A K V H D E S Y L N Y Y Y Y N K P L - - L L S E L Y S W P E K Y G E N K - - D A K I I M R D K
Bs	135	Q A Y I - P Y N S G E H - Y Y Q G G L I G G E S K A Y I E L L E Q C S L M T E T D L K R N I T A R W H D E S Y L N K Y L L D K Q I K - - I L S T E Y G R E Q E W T V P P - - T P K I I F R D K
Psych	162	T A Y I - A Y G E G K Y - Y F T G A L N G G K S A A F L D L C H T L Y N N T Q S D L K Q D I A L W H D E S H L N K F A L G R E D I - - K I L P P Y - F T R G E R Y W K K T S K L M F S D K
Hs	140	T S Y I - P Y G E G E H - Y V M G S F N G G K T S A F L E M C Q Q L N D N I Y L D L K K D I A L W H D E S H L N H Y C W R N K E K I F I L P D S Y L L D E G Q Y F P - - - - - S K K
Bo2	136	W A Y V A P Y G K D Y T - Y F M G G L N G G R P K E Y L E M V R T L S V N I R D D Y D R G I I A L V H D E S H I N A Y M R S H P C K - - I L P V E L N R P E E W - - - A D E H T K L I F R E K
Bcac	138	S A Y I P P R E D K P Y I Y M G G I N G G E A A D Y L E M A E T L S E N I H A D Y E K G I V A L V H D E S H I N K Y A I T H E K L I H V L S P O Y G Y P E G L - S N P D F E I K L V I K N K
Fp	142	L A F V - P K S L G K Y - Y F M G G V N G G T T K A Y I K L I N D L E Y A I D Q D L K K N V I A I W H D E S H L N K Y A I T H E K L I H V L S P O Y G Y P E G L - S N P D F E I K L V I K N K
Hm	178	L A Y V - P F G K G K A - Y Y Y G S T N G G K A G A F L A L A R T L Q E R I Q E D L S R G I I A I W H D E S H L N A Y I I D H P N Y - K M L D Y G Y G P E G Y G R V P G G G V Y I F L R D K
Sv1	142	R A F I - P M G L G R Y - Y V C G G I N G G E A A A F L K L C H T L D K R I R K D L A H N V I A Q W H D E S H I N R Y I L W R R D V - R V L S P S Y C W P E G W N L P L - - P C R I L I R S K
Bo1	142	T A Y I - P E G E G R Y - Y Y A G G L S G G C T K A Y L K L C T T I C S W R R D A T N H I I P I W H D E S H I N K Y F L D N P P A - I T L S P A Y L Y P E G W - L L P - F E P I I L I R D K
E F		
Sv2	235	S H Y I - - D M S A V R Q Q G R R Q N Y L - - - - Q R K W E A F C L N W L P Y L W W A R D T L L R R R V D P P R T R
WcmB	224	E R - - - - - E S W Y G N I K K - - - - -
Bs	224	N T I L - - G A S Y I C S - - - - - L K K R N R L K L I S K A I R N I L N K L L K R - - - - -
Psych	252	S H Y R F G G H A Y L R S E T D E K I T Q A E W N K K N A K R R R K L K F R A K Q Y I S S L L F R Q - - - - -
Hs	225	S H Y I - - N P L Y L K N - - - - - H - - - - D V R L V M R N K N H Y R F G G E L H E R D Y - - - - -
Bo2	225	T H - I - - D P Y F N K G - - R K T S L - - C A R F K K G T T I L F N A I R W Y V K L - - - - -
Bcac	230	V K - L - - D P Y F N K G - - - R D Y S L - - - K G K L K K T V S V I Y R A I R W Y L - - - - -
Fp	234	E N Y I - - D I H T I K G - - - - - S K R R N V F L N I Y S Y I K G L L S E N N - - - - -
Hm	270	S R V I - - D V N A I K G - - - - - M - - - - G S P A N R R L K N A L R K L K H F S K R L L G R - - - - -
Sv1	232	A R Y F - - D V Q Q L R K - D A P A T E L - - - - - P R Y V V R C N D F M K R A A R W L Q R R L P P K K E D I N D E -
Bo1	232	N K - - - - - P Q Y G G H - - - - - E L L R R K N S L W E R I K L I C Q K F K S A D - - - - -

Figure 1

```

bova      1 ---ESKLLSDWFNPFKRPEVVTMTKWKAPVVWEGTYNRAVLNDNYAKQKITVGLTFAVGRYIEHYLEEFLLSANKHFVGHVPVIFYIMVDDVSRMPLIELGP
dogFS    1 QVTRSPHPQPKLLEQ-RPELLELTLPWLAPIVSEGTFNPELLOHIYQPLNLTGLTVFAVGKYT-RFVQHFLAESAEQPFMOGYQVYYIFTDNPAGIPRVPLGP
huA      1 --LPRMVYQPQKVLTP-CRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTGLTVFAIKKYV-APLKLFLFETAEKHFVGHVHRVHYVFTDPAAVPRVTLGT
GOS_E    1 -----MKVAITFIIGTNKYL-NFLPKYVENIHKYFIPGVVEKTFLLVFTDGEDGDFP-----
GOS_F    1 -----MKVAITFIIGTNKYL-DPLPRYVENIKEYFIPNTKKIFLVFTDGDGDFP-----
GOS_G    1 -----MKVAITFIIGTNKYL-DPLPRYVENIKEYFIPDTEKIFLVFTDGDGDFP-----
PSSM2    1 -----MKICILTIATNKYI-QFVEKLYDNIHDDHFLNGHEIEGIFFTDQEVES-----
GOS_I    1 -----MKICILTIATNKYI-QFVERLLDNIENFLNGHEIECLLFTDHEVEAS-----
GOS_A    1 -----MKIDILTIATNKYI-QFVERLYDNIADNFLNGHEIQGLLFTDHEVDVETS-----
GOS_B    1 -----MKIDILTIATNKYI-QFVERLYDNIEEKFLNGHEIQGLLFTDHEVDVETS-----
GOS_H    1 -----MKIDILTIATNKYI-QFVERLYDNIHEDKFLNGHEIQGLLFTDHEVDVETS-----
GOS_C    1 -----MKICILTIATNKYL-QFVEKLYDDIAEKFLPDAEVCNLLFTDHEIEET-----
GOS_D    1 -----MKICILTIATNKYI-QFVEKLYEDIAEKFVCPGAEINCLLFTDHEIEES-----

```

A

```

bova      102 LRSFKVFKIKPEKRWQDISMRRMKTIGEHIVAHIQEVDLFLCNDDVDQVFDKFG-VETLGE--SVAQIQAWWKA--DPNDFT--YERRKESAAYIPFEGE--D
dogFS    104 GRLLSIPIQRHSRWEESTRRMETISRHIAQRAHREVDYLFCDVDVFRNPNWG-PETLGD--LVAALHPGYAV--PROQFP--YERRHISTAEVAENEG--D
huA      102 GRQLSVLEVRAYKRQDVSMMRMEMISDFCERRFLSEVDYLVCDVDVDFRDHVG-VEILTP--LFGTHHPGFYGS--SREAF--YERRPQSQAYIPKDEG--D
GOS_E    48 -DDVKVYPQE-HLSWPFITLKRFEILNN--AREEILKNDWLVFIDADALVVDEISAEEFFTDKPLFGVHPCHFLGMPHDQPSGAFETNKSSBAYVDVYKEAPL
GOS_F    48 -DDVKVFKQE-HLEWPIITLKRFEIINK--AREEIKCDYLVEFDADALVVDKITEEEFFTDKPLFGVHPCHFLGMNPHNKLPGAFETNQSSBATLDLEKYQPQ
GOS_G    48 -DDVKVFEQE-HLEWPIITLKRFEIINK--ARKEIKDCDYLVFIDADALVVDKITEEEFFTDKPLFGVHPCHFLKMOPHNQLPGAFETNKNNSAALDLEKYKPD
PSSM2    48 -DNIKISQIE-HEPWPFITLKRNYFMK--EAESHKYDYCFYFDVDMGIVDKVG-DEVLGD--LVATMHP--YQSFAPKIORS--YDRNPKSLAYVPLYDEGE-E
GOS_I    48 -DNVRVCOID-HEPWPFITLKRNYFVK--EKEFISQFDYCFYFDVDMGLVDKVG-DEVLSD--LVATMHP--YQSFYPKQORS--YDRNPKSLAYVPGQEG-E
GOS_A    48 -DNIKVSQIE-HEDWPFITLKRNYFIK--EKEYISQFDYCYFDDVDMGIVGKVG-DEVLGD--LVATMHP--YQSFYPKQORT--YDRNPKSLAYVPAGQEG-E
GOS_B    48 -DNIKVSQIE-HEPWPFITLKRNYFIK--EKEFISQFDYCYFDDVDMGIVSNVG-DEVLSD--LVATMHP--YQSFYPKQORT--YDRNPKSLAYVPAGDEGE-E
GOS_H    48 -DNIKVSQIE-HEPWPFITLKRNYFIK--EKEFISQFDYCYFDDVDMGIVSNVG-DEVLSD--LVATMHP--YQSFYPKQORT--YDRNPKSLAYVPSGDEGE-E
GOS_C    48 SENVRVHYID-HEPWPFITLKRNYFVK--EKDFLEHDYCFYFDADMRIADAPV-EEVLS- --LVATMHP--YQSFYPKQORT--YDRNPKSLAYVPSGDEGE-E
GOS_D    48 GDNVRVHYID-HEPWPFITLKRNYFMK--EKEFIMEHDYCFYFDADMRIADGIVTPDEVCGD--LVATMHP--YQSLASPAQOS--FDRNPKSLAYVAPNEST-V

```

B C D

```

bova      198 FVYHAAIFGCTPTQVLNITQECFKGILKDKKNDIEAOWHDESHLNKRYFLNKPTK-IILSPYCWYDHYHIGLPADIK--LVKMSWQIREYNVVRINV
dogFS    200 FVYGGAVFGGRVAKVYEFTTGCHMAILADKANGIMAAWQESHNRFRSHKPSK-VLSPENLWDDRKPQPPSLK--LIRFSTLDKATSWLRS--
huA      198 FVYLCGFVCGSVQVQRLTRACHQAMVDAQANGIEAVWHDESHLNKRYFLNRKPTK-VLSPENLWDDQQLLWPAVLK-KLRFSTLAKNQAVRNP
GOS_E    149 VYVYOGCLWGGKVPEVCAMIDTLMDRTNRDLKDVVALWHDESHLNKRYFLNHQEDVHTHGPEFAYPEVVFKEYCEF---KPRIVHLLAKDNSEYQC---
GOS_F    149 VYVYOGCFWGGKVPEVCAMIDELEDRVNDLKRRIAVWHDESHLNKRYFLNHQDKVHTFGPEFAYPEVVFKEYHCTF---KPRIVHLLAKDNS---
GOS_G    149 VYVYOGCLWGGKVPEVCAMIDELMNRNDLKKDIAVWHDESHLNKRYFLNHQDKVHTFGPEFAYPEVVFDDYCDF---KPRIVHLLAKDNSKYV---
PSSM2    141 HYVAGGFNGGSKTRFLEMAEVIADRVNKDLENDVIALWHDESHLNRYLIDNPPTL-SLTPSYCFABEQMSNLEXPY-KPKRIALAKDNHNLRS--
GOS_I    141 NYVAGGFNGGSKTRFLEMAEVLADRVTKDLENGVIALWHDESOMNRYLIDNPPTL-SLTPSYCFABEQMGNPQYPY-EPKRIALAKDNHNLRS--
GOS_A    141 LYVAGGFNGGSKTRFLEMAEVLADRVNKDLENDVIALWHDESOMNRYLIDNPPTL-SLTPSYCFABEQMNSDYPY-EPKRIALAKDNHNLRS--
GOS_B    141 LYVAGGFNGGSKTRFLEMAEVLADRVTKDLENGVIALWHDESOMNRYLIDNPPTK-SLTPSYCFABEQMNSSEYPY-DAKRIALAKDNHNLRS--
GOS_H    141 LYVAGGFNGGSKTRFLEMAEVLADRVSKDLENGVIALWHDESOMNRYLIDNPPTK-SLTPSYCFABEQMNSSEYPY-DAKRIALAKDNHNLRS--
GOS_C    142 TYVAGGFNGGKTIKFMEMSEVIADRVNKDLEKDI VALWHDESHLNRYLIDNPPTL-ALNPEYCYABEFIG-TDYFFQEPKRIALAKDNHSELRS--
GOS_D    143 TYVAGGFNGGRTENFMAMSEKIANNVNTDLENGVIALWHDESHLNRYLIDNPPTL-DLDPYCYABEFIG-TDYFFSPGRIALAKDNHAAALRS--

```

E F G

Figure 2

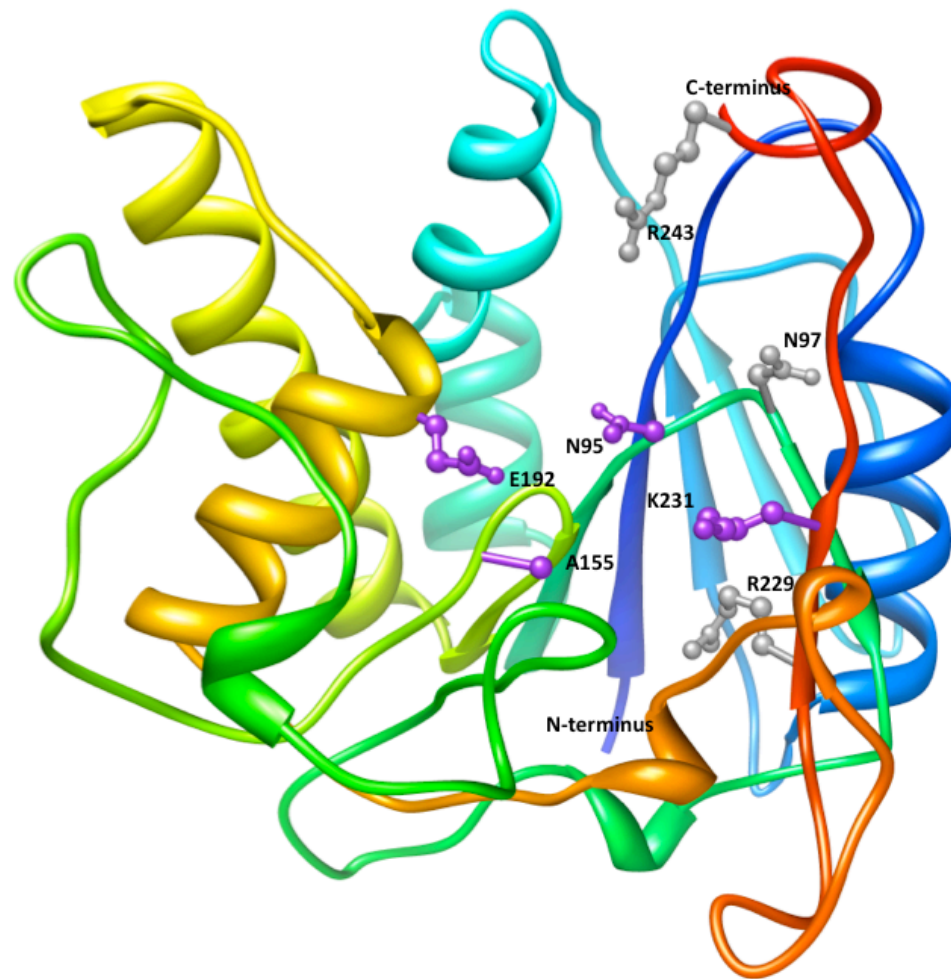


Figure 3