

Figure S5. Multiple sequence alignment of StIA protein sequence with HMP and MetaHit homologues.

```

                10         20         30         40         50         60         70         80         90
(A) -----MDTVRQRQLGIKRSMLTALLMGFCFLFYAPEAFSQEESALPKQSGLDEDMAGVVALVKQRYGEQLRNQYMGGATTAKAYAIIGALRMENDIAKABA 97
(B) -----MNTIRHVQGAWFKKRILTVLLLVGVCCLFYPHNALSQDEPSTPPQKTGMQLDMEKGMQARRRSGDQLYNQYMGGATTAGQYAIIGAAQOMENATAEABR 97
(C) -----MNTVRQRQLGIKRSILTALLMGFCFLFYTPEAFSQDEPGSPPKKSGLAEDMEKGIQAQANQRYGQQLYQYMGGATTAKETAIIGVLRMENATAEABG 97
(D) MQAKSYPDGINTNTNITMNTVRQRQLGIKRSILTALLMGFCFLFYAPEAFSQDEPGAPSQRTAMDDDFDKSMAQARQYGNRLYQYMGGATTAKERAIIGVLRMENATAEABG 114
(E) -----MNTVRQRQLGIKRSILTALLMGFCFLFYTPEAFSQDEPGSPPKKSGLAEDMEKGIQAQADQRYGQQLYQYMGGATTAKETAIIGVLRMENATAEABG 97
(F) -----MNTVRQRQLGIKRSILTALLMGFCFLFYTPEAFSQDEPGSPPKKSGLAEDMEKGIQAQANQRYGQQLYQYMGGATTAKETAIIGVLRMENATAEABG 97
(G) -----MNTIRHVQGAWFKKRILTVLLLVGVCCLFYPHNALSQDEPSTPPQKTGMQLDMEKGMQARRRSGDQLYNQYMGGATTAGQYAIIGAAQOMENATAEABR 97
(H) -----MGFCFLFYTPEAFSQDEPGSPPKKSGLDEDMAGVVALVKQRYGNRLYQYMGGATTAKERAIIGVLRMENATAEABG 76
(I) -----MNTVRQRQLGIKRSILTALLMGFCFLFYTPEAFSQDEPGSPPKKSGLAEDMEKGIQAQANQRYGQQLYQYMGGATTAKETAIIGVLRMENATAEABG 97
(J) MQAKSCREAINNTNITMNTVRQRQLRIKRSILTALLMGFCFLFYTPEAFSQEESALPKQSGLDEDMAGVVALVKQRYGNRLYQYMGGATTAKAYAIIGALRMENDIAKABG 114
(K) -----MQAEFDKGTQAQRNRKYGDQVYDQYMGGTTTAGQRAVGAQOMENATAVABG 50
(L) -----MGFCFLFYTPEAFSQDEPGSPPKKSGLDEDMAGVVALVKQRYGNRLYQYMGGATTAKERAIIGALRMENATAEABG 76

```

```

* * ** *** * * *** ** ***** ** ** * ** *** * ** ** * ***** ***** ***** * **
(A) EGNTSLANALRQQLAMAQRTINEWYNGELQOKADAGDTRAQRELD EYYAFMKTVEGSSPAGSEFFLISIIISGLLLYGYIVLLSPKDATINRKTLLIPWCVGLAVFDMLGNWVNC 211
(B) NGNTGRANMLREQLARANSNNQWYLNEELKQKADAGDACAQRELD EYYELIKTSTPSLP--AVPFVVSIIIFGLLFYWCMLVFSRDTIILNRKTLLIPWCVGLTVEFDMLGNLTNS 209
(C) KGDTAQANALRQQLARAQSNNEWSIGHLQOKANVGDTRAQRELD EYYAFMKTVEGSSPGEAFPFLLISIIISGLLLYGYIVLLSPKDATINRKTLLIPWCVGLAIFDMLGNWVNC 211
(D) KGDTAQANALRQQLAMAQSNNEWSSTGHLLQOKANA GDTRAQRELD EYYAFMKTVEGSSPGEAFPFLLISIIISGLLLYGYIVLLSPKDATINRKTLLIPWCVGLAIFDMLGNWVNC 228
(E) KGDTAQANALRQQLAMAQSNNEWSIGHLQOKANVGDTRAQRELD EYYAFMKTVEGSSPGEAFPFLLISIIISGLLLYGYIVLLSPKDATINRKTLLIPWCVGLAIFDMLGNWVNC 211
(F) KGDTAQANALRQQLARAQSNNEWSIGHLQOKANVGDTRAQRELD EYYAFMKTVEGSSPGEAFPFLLISIIISGLLLYGYIVLLSPKDATINRKTLLIPWCVGLAIFDMLGNWVNC 211
(G) NGNTGRANMLREQLARANSNNQWYLNEELKQKADAGDACAQRELD EYYELIKTSTPSLP--AVPFVVSIIIFGLLFYWCMLVFSRDTIILNRKTLLIPWCVGLTVEFDMLGNLTNS 209
(H) KGDTAQANALRQQLAMAQSNNEWSSTGHLLQOKANA GDTRAQRELD EYYAFMKTVEGSSPGEAFPFLLISIIISGLLLYGYIVLLSPKDATINRKTLLIPWCVGLAIFDMLGNWVNC 211
(I) KGDTAQANALRQQLARAQSNNEWSIGHLQOKANVGDTRAQRELD EYYAFMKTVEGSSPGEAFPFLLISIIISGLLLYGYIVLLSPKDATINRKTLLIPWCVGLAIFDMLGNWVNC 211
(J) EGNTLANMLRQQLAAVQANQISNTELLQOKANVGDTRAQRELD EYYAFMKTVEGSSPGEAFPFLLISIIISGLLLYGYIVLLSPKDATINRKTLLIPWCVGLAIFDMLGNWVNC 228
(K) KGDTAQADMLRQQLARANSNNQWYLNEELKQKADAGDACAQRELD EYYELIKTSTPSLP--AVPFVVSIIIFGLLFYWCMLVFSRDTIILNRKTLLIPWCVGLTVEFDMLGNLTNS 162
(L) KGDTQANALRQQLAMAQSNNEWANGKLLQOKANA GDTRAQRELD EYYAFMKTVEGSSADLFPFLISIIISGLLLYGYIVLLSPKDATINRKTLLIPWCVGLAIFDMLGNWVNC 190

```

```

**** * * ** ** * **** ** ***
(A) LFLFVEILVLLVVARNFQCSWKRSFMAILGLMLVSIILIGLLPLFF 257
(B) LFLFGEVLAILLIARSEFKYSWKRTFSFLGIVLVSVLILGLLTTF 255
(C) LFLFVEILVLLVVARNFQCSWKRSFMAILGLMLVSIILIGLLPLFF 257
(D) LFLFVEILVLLVVARNFQCSWKRSFMAILGLMLVSIILIGLLPLFF 274
(E) LFLFVEILVLLVVARNFQCSWKRSFMAILGLMLVSIILIGLLPLFF 257
(F) LFLFVEILVLLVVARNFQCSWKRSFMAILGLMLVSIILIGLLPLFF 257
(G) LFLFGEVLAILLIARSEFKYSWKRTFSFLGIVLVSVLILGLLTTF 255
(H) LFLFVEILVLLVVARNFQCSWKRSFMAILGLMLVSIILIGLLPLFF 236
(I) LFLFVEILVLLVVARNFQCSWKRSFMAILGLMLVSIILIGLLPLFF 257
(J) LFLFVEILVLLVVARNFQCSWKRSFMAILGLMLVSIILIGLLPLFF 274
(K) LFLFGEVLAILLIARSEFKYSWKRTFSFLGIVLVSVLILGLLTTF 208
(L) LFLFVEILVLLVVARNFQCSWKRSFMAILGLMLVSIILIGLLPLFF 224

```

**Figure S5. Multiple protein sequence alignment of SttA and homologous sequences from Human Stool Microbiome from Human Microbiome Project (HMP) and MetaHit.** Black shading indicates regions of 100% amino acid identity. Putative transmembrane regions for SttA, predicted by TMHMM, are indicated with red boxes. Truncated or partial sequence fragments from HMP were not included (n=4). Information on the protein sequences (A) – (L) is indicated in the legend below.

**Legend:**

(A) = SttA protein sequence (B) SRS053214\_LANL\_scaffold\_17021\_\_gene\_42707 (C) SRS024549\_LANL\_scaffold\_1815\_\_gene\_4559 (D) C3406971\_\_gene\_199744 (E) SRS018656\_WUGC\_scaffold\_544\_\_gene\_591 (F) SRS015217\_WUGC\_scaffold\_30292\_\_gene\_65222 (G) SRS077730\_LANL\_scaffold\_24345\_\_gene\_72567 (H) SRS024075\_LANL\_scaffold\_21370\_\_gene\_63545 (I) Baylor\_scaffold\_22757\_\_gene\_50812 (J) C2121591\_\_gene\_151559 (K) MetaHit\_MH0011\_GL0108025 [Complete]locus=scaffold6530\_52:7938:8564 (L) MetaHit\_V1\_GL0100177 [Complete] locus=scaffold36986\_1:2178:2888