Proteogenomic re-annotation of *Shigella flexneri* 2a 2457T by GenoSuite

Data with PRIDE accessions 18992-18999 were downloaded from PRIDE database. In total 1,531,533 tandem mass spectra were searched against a 6 frame translated database of *S. flexneri* 2a 2457T genome (NC_004741.1). Search was performed in GenoSuite with OMSSA and X!tandem algorithms with following search parameters. Trypsin as protease, 1 missed cleavage, carbamidomethylation of cysteine as fixed modification, methionine oxidation as variable modification, 1 Da precursor ion tolerance and 0.8 da product ion tolerance. By applying GenoSuite, We discovered 28 novel protein coding regions and 22 gene model changes demonstrating rapid and automated proteogenomic analyses of prokaryotic genomes.

Table 1: Novel protein coding regions identified by GenoSuite

Database ID	**Locus ID**	**#Unique Peptide count**	**Significant PSMs (Unique Peptides)**	**Peptides**
chrgi:30061571:ref:NC004741.1:_1043868_1044107_3	NOVEL Protein	4	105	QAETEIADFIAQK MTGLESYDVK NTGKEVSEIR IAELNKNTGK
chrgi:30061571:ref:NC004741.1:_1320035_1320313_2	NOVEL Protein	2	2	SASNRPMNAPSWPHLR GSNWVVNSGCPQALK
chrgi:30061571:ref:NC004741.1:_1351722_1351561_-1	NOVEL Protein	2	12	LMNNEEISEEAQHEMAAEAGINPVR IDEIAEFLNQWGNE
chrgi:30061571:ref:NC004741.1:_1389606_1388686_-1	NOVEL Protein	7	16	AIGTVSAVNIGSMFR TEDVSAVIDYLTTLSYVDNTR AQYPTAPGYATLR TSDISSEEYAIMPLAPMK SIDALPHIEAGSNAR QYPAIVVSHPGGGVK TSDISSEEYAIMPLAPMKEADAPNEELR
chrgi:30061571:ref:NC004741.1:_148655_148843_2	NOVEL Protein	1	6	SVFITSLPNSMISR

chrgi:30061571:ref:NC004741.1:_1960728_1960447_-1	NOVEL Protein	2	23	NIDFTHHPAAADPVTMR QILDVLDVIAPVEVR
chrgi:30061571:ref:NC004741.1:_1972745_1972386_-2	NOVEL Protein	10	90	DGTLQALSEK IDAILVDR TNDTLAVTGEAFSR AVNDAIAEMQK LAALDLVK WFGADVTK KTNDTLAVTGEAFSR LAALDLVKK VGRIDAILVDR TYDDDPTKYQDLR
chrgi:30061571:ref:NC004741.1:_1973274_1972657_-1	NOVEL Protein	9	321	KVGVGLGTNYEEWLR HLGVEASLKPTK VGVGLGTNYEEWLR RIDVVINQVTISDER WDGMLASLDSK KYDFSTPYTISGIQALVK ESGTLLVGLEGTYPPFSFQGDDGK YDFSTPYTISGIQALVK LTGFEVEFAQQLAK
chrgi:30061571:ref:NC004741.1:_2027758_2027252_-3	NOVEL Protein	3	68	ISCPSPSASGAIILR TSADNDVDLDTDGFR AGVNVNGESGVTYENCK
chrgi:30061571:ref:NC004741.1:_204580_204356_-3	NOVEL Protein	2	68	VLNEMAADDALSEAVR YAEIASGDLGYVPDALGCVLK
chrgi:30061571:ref:NC004741.1:_2173966_2174352_1	NOVEL Protein	3	14	NLPAPDAGTHWTYMGGAYVLISDTDGK YELSSFIADFK HFKPGDTVPEMYR
chrgi:30061571:ref:NC004741.1:_2188194_2190122_3	NOVEL Protein	2	3	RATVGLDILAAIGSDIALMQLNGIAQK FPQEEALYPGLLQVK
chrgi:30061571:ref:NC004741.1:_2411849_2412037_2	NOVEL Protein	1	12	VLFPDPDDPMMLITSPAFACSETPFSTSCSP
chrgi:30061571:ref:NC004741.1:	NOVEL Protein	2	2	DGSPEAALSEFIK

_2527364_2528194_2				SFEAMGVSLVVHPHNPYVPTSHANVR
chrgi:30061571:ref:NC004741.1:_2834428_2833628_-3	NOVEL Protein	4	151	ALVEEEPSDNDLAEEELLSQGATQR LDHEPSAEEIAEQLDKPVDDVSR VHDLNEDAEFDENGVEVFDEK GLALLDLIEEGNLGLIR
chrgi:30061571:ref:NC004741.1:_2983916_2984230_2	NOVEL Protein	1	5	TLSNSHSGEVNGSLLVAGSVGNTSIAAPAR
chrgi:30061571:ref:NC004741.1:_3259632_3259811_3	NOVEL Protein	1	101	SIASTPRNSISK
chrgi:30061571:ref:NC004741.1:_3354779_3355015_2	NOVEL Protein	1	5	VAIDQAELIAQEVR
chrgi:30061571:ref:NC004741.1:_3358644_3359855_3	NOVEL Protein	3	7	TLAQQLIDFGEIKR VGDFAVAVGNPFGLGQTATSGIVSALGR SGLNLEGLENFIQTDASINR
chrgi:30061571:ref:NC004741.1:_3743922_3744953_3	NOVEL Protein	3	87	VPDSQVLADLDHVASWASR EGDPNDFADIPTLLSGLVAK QSDGPLPVVIVVQEIFGVHEHIR
chrgi:30061571:ref:NC004741.1:_4273690_4274031_1	NOVEL Protein	1	5	CSVIAFTITVRSSFPWR
chrgi:30061571:ref:NC004741.1:_4451380_4450616_-3	NOVEL Protein	11	465	WELTIPQELAYGER YLEENAK LIDGTVFDSSVAR VINQGEGAIPAR YLEENAKK GEPAEFPVNGVIPGWIEALTLMPVGSK GAGASIPPFSTLVFEVELLEIL HPAVPVDVVHR EGVNSTESGLQFR FQAMAAEGVK VRVHYTGK
chrgi:30061571:ref:NC004741.1:_4471470_4472066_3	NOVEL Protein	2	2	DHNRQTAVNPQTFDMR QTAVNPQTFDMRDGFQLCEQPVQLVLFR
chrgi:30061571:ref:NC004741.1:	NOVEL Protein	7	30	TNLLFFTK

_4529989_4528454_-3				EIWFYEHPYPAGVK LPNGVFNPYTGIK QLPHLLATTNMLLHGIEVPVQIR IFDAQEEALELEQDNYQYPIPQR HLFGDIYEQILK AAVVLPDGTLFGEGVK
chrgi:30061571:ref:NC004741.1:_512407_512015_-3	NOVEL Protein	1	10	MIIGSRIPASSTMR
chrgi:30061571:ref:NC004741.1:_745834_744317_-3	NOVEL Protein	3	23	ICLAFQEPALNLSLLR LSVGNNQLTALPELPCELQELSAFDNR LPALPPHLVALDVSLNR
chrgi:30061571:ref:NC004741.1:_821019_822005_3	NOVEL Protein	4	29	LGLTNTTFQTVHGLDAPGQFSTAR IMTSYVVGQALK TGTTAGAGYNLVASATQGDMR LISVVLGAK
chrgi:30061571:ref:NC004741.1:_932538_932275_-1	NOVEL Protein	1	6	FAAGITKNNTLMIK

Table 2: N-terminal extension of annotated gene models

Database ID	**Locus ID**	**#Unique Peptide count (All)**	**Significant PSMs (Unique Peptides)**	**Novel Peptides**
chrgi:30061571:ref:NC004741.1:_2978323_2977118_-3	S3091	4	7	TMQSVDVAIVGGGMVGLAVACGLQGSGLRVAVLEQR
chrgi:30061571:ref:NC004741.1:_2617180_2615645_-3	S2726	22	732	MQSVTLCIMPR
chrgi:30061571:ref:NC004741.1:_949746_951428_3	S0971	47	13761	LNMTESFAQLFEESLKEIETRPGSIVR
chrgi:30061571:ref:NC004741.1:_2182046_2180871_-2	S2302	11	123	GVGMNEQSQAKSPEALR

chrgi:30061571:ref:NC004741.1:_2083678_2084376_1	S2206	4	51	HPALVERSAVMNVVILDTGCANLNSVK
chrgi:30061571:ref:NC004741.1:_132026_132487_2	S0118	2	2	AAIITSMARCDFGALPGAEEHTMDYEFLR
chrgi:30061571:ref:NC004741.1:_974914_973712_-3	S0989	26	1031	WNLVMFENITAAPADPILGLADLFR
chrgi:30061571:ref:NC004741.1:_4329880_4332027_1	S4404	49	11422	NIGISAHIDAGK ILFYTGVNHK IGEVHDGAATMDWMEQEQER ILFYTGVNHKIGEVHDGAATMDWMEQEQER
chrgi:30061571:ref:NC004741.1:_2850631_2848826_-3	S2973	9	51	RMTTQVPPSALLPLNPEQLAR
chrgi:30061571:ref:NC004741.1:_677698_678033_1	S0666	4	39	ALQDFVIDKIDDLK GQDIIALDVQGK ALQDFVIDK
chrgi:30061571:ref:NC004741.1:_3408165_3409496_3	S3545	7	28	STLLSPVLPSIFILMKK
chrgi:30061571:ref:NC004741.1:_3139402_3140622_1	S3258	20	753	EQVMNNFNLHTPTR
chrgi:30061571:ref:NC004741.1:_2767041_2766622_-1	S2883	5	92	VFFIMSVMLQSLNNIR
chrgi:30061571:ref:NC004741.1:_2746739_2747119_2	S2854	2	143	VVYRPDINQGNYLTANDVSK
chrgi:30061571:ref:NC004741.1:_3215089_3213641_-3	S3331	2	2	LIFQILSILPESFIGGVVVCSR
chrgi:30061571:ref:NC004741.1:_3650780_3649722_-2	S3753	20	307	KIGVLTSGGDAPGMNAAIR IGVLTSGGDAPGMNAAIR
chrgi:30061571:ref:NC004741.1:_3594764_3593790_-2	S3701	8	12	SFIAILPIAMNYRGDGGWIMNIR
chrgi:30061571:ref:NC004741.1:_1929435_1928680_-1	S2005	5	71	GGDFCYSDGEFAAILEDVR ELGFPGLVTGVLDVDGNVDMPR
chrgi:30061571:ref:NC004741.1:_1383586_1383065_-3	S1413	12	2074	IIMSQTVHFQGNPVTVANSIPQAGSK
chrgi:30061571:ref:NC004741.1:	S4337	28	275	LLDEGNTVPFIAR

_4266014_4263684_-2				IIAGEIQARPEQVDAAVR
chrgi:30061571:ref:NC004741.1: _2969644_2971104_1	S3086	5	39	ELTDLEPMIVKK
chrgi:30061571:ref:NC004741.1: _2286634_2285597_-3	S2402	13	124	GEFIMSLDINQIALHQLIKR