TABLE 4-continued

Antibiotic, Chemical, and Heavy Metal Resistance

SEQ ID NO	Gene Id	NCBI gi	aat nap Score	BlastP Score	BlastP-Prob	% Ident	% Cvrg	NCBI gi description
1854	Bt1G4261	g1945096	446	406	7.20E−38	59	35	(D88802) *S. lividans* chloramphenicol resistance protein; P31141 (492) transmembrane [*Bacillus subtilis*]
1914	Bt1G4327	g2668553	130	68	0.85	37	6	(U62929) multidrug resistance protein 1 [*Filobasidiella neoformans*]
1923	Bt1G4339	g115950	133	153	1.80E−10	55	18	GLUCOSE-RESISTANCE AMYLASE REGULATOR (CATABOLITE CONTROL PROTEIN) [*Bacillus subtilis*]
2151	Bt1G4629	g3925779	238	181	5.00E−13	29	31	(AL034353) putative major facilitator family multi-drug resistance protein [*Schizosaccharomyces pombe*]
2179	Bt1G4659	g2632985	98	107	9.60E−05	30	11	(Z99107) similar to acriflavin resistance protein [*Bacillus subtilis*]
2211	Bt1G4695	g4416482	101	115	2.40E−06	38	24	(AF125999) daunorubicin resistance protein A [*Mycobacterium avium*]
2304	Bt1G4796	g115950	204	269	2.40E−23	51	32	GLUCOSE-RESISTANCE AMYLASE REGULATOR (CATABOLITE CONTROL PROTEIN) [*Bacillus subtilis*]

SEQ ID NO: A sequential SEQ ID NO: is assigned to each contig or singleton and the SEQ ID NO: corresponds to that set forth in the sequence listing.

Gene ID: Refers to an arbitrarily assigned Gene ID number.

NCBI gi: Each sequence in the GenBank public database is arbitrarily assigned a unique NCBI gi (National Center for Biotechnology Information GenBank Identifier) number. In this table, the NCBI gi number which is associated (in the same row) with a given contig or singleton refers to the particular GenBank sequence which is the best match for that sequence.

aat_nap score: The aat_nap score is reported by the nap program in the aat package. It is an alignment score in which each match and mismatch is scored based on the BLOSUM62 scoring matrix.

Blastp-Prob: The entries in the "Blastp-Prob" column refer to the probability that such matches occur by chance.

BlastP Score: Each entry in the "BlastP Score" column of the table refers to the BLASTP score that is generated by sequence comparison of the designated clone with the designated GenBank sequence.

% Iden: The entries in the "% Iden" column of the table refer to the percentage of identically matched nucleotides (or residues) that exist along the length of that portion of the sequences which is aligned by the BLAST comparison to generate the statistical scores presented

% cvrg: The % coverage is the percent of hit sequence length that matches to the query sequence (% coverage = (match length/hit total length) × 100).

NCBI gi description: The "NCBIgidesc" column provides a description of the NCBIgi referenced in the "NCBIgi" column.

TABLE 5

Toxins and Toxin Homologs

SEQ ID NO	Gene Id	NCBI gi	aat nap Score	BlastP Score	BlastP-Prob	% Ident	% Cvrg	NCBI gi description
73	Bt1G100	g419952	125	135	1.30E−09	29	13	alpha-latroinsectotoxin precursor - black widow spider (fragment) [*Latrodectus tredecimguttatus*]
212	Bt1G260	g2507017	330	263	1.00E−22	68	29	HEMOLYSIN BL BINDING COMPONENT PRECURSOR (ENTEROTOXIN 40 KD SUBUNIT) [*Bacillus cereus*]
316	Bt1G386	g2507017	990	1027	1.10E−103	68	78	HEMOLYSIN BL BINDING COMPONENT PRECURSOR (ENTEROTOXIN 40 KD SUBUNIT) [*Bacillus cereus*]
410	Bt1G503	g2507017	126	156	1.10E−10	60	13	HEMOLYSIN BL BINDING COMPONENT PRECURSOR (ENTEROTOXIN 40 KD SUBUNIT) [*Bacillus cereus*]
479	Bt1G597	g2507017	1829	1732	2.20E−178	98	98	HEMOLYSIN BL BINDING COMPONENT PRECURSOR (ENTEROTOXIN 40 KD SUBUNIT) [*Bacillus cereus*]