

TABLE 5-continued

| Toxins and Toxin Homologs |          |          |               |              |             |         |        |   |
|---------------------------|----------|----------|---------------|--------------|-------------|---------|--------|---|
| SEQ ID NO                 | Gene Id  | NCBI gi  | aat nap Score | BlastP Score | BlastP-Prob | % Ident | % Cvrg | NCBI gi description   |
| 694                       | Bt1G917  | g1665720 | 1623          | 1644         | 4.70E-169   | 95      | 93     | (D17312) diarrheal toxin [ <i>Bacillus cereus</i> ]   |
| 1552                      | Bt1G3948 | g2507017 | 210           | 200          | 1.50E-15    | 51      | 25     | HEMOLYSIN BL BINDING COMPONENT PRECURSOR (ENTEROTOXIN 40 KD SUBUNIT) [ <i>Bacillus cereus</i> ] |
| 2056                      | Bt1G4495 | g97193   | 96            | 105          | 9.60E-05    | 51      | 6      | leukotoxin B - <i>Pasteurella haemolytica</i> [ ]   |

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.

% Ident: The entries in the "% Ident" 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 "NCBIgi desc" column provides a description of the NCBIgi referenced in the "NCBIgi" column.

#### LENGTHY TABLES

The patent application contains a lengthy table section. A copy of the table is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US20100261614A1>). An electronic copy of the table will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

#### SEQUENCE LISTING

The patent application contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US20100261614A1>). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

1-40. (canceled)

41. A method for identifying plasmid DNA sequences of a *Bacillus* species, the method comprising the steps of:

- identifying a first *Bacillus* species strain which does not contain plasmid DNA;
- generating a library of chromosomal genomic DNA from said first *Bacillus* species strain;
- obtaining the nucleotide sequences of said chromosomal genomic DNA from step b);
- identifying a second *Bacillus* species strain which contains plasmid DNA;
- generating a library of genomic DNA from said second *Bacillus* species strain, wherein said genomic DNA contains chromosomal and plasmid DNA;
- obtaining the nucleotide sequences of said genomic DNA from step e);
- subtracting any common sequences identified in the genomic DNA from step f) which are also identified in

the chromosomal genomic DNA from step c) from the nucleotide sequences of step f) to obtain a subtracted set of plasmid DNA sequences; and

- constructing contigs and sequences of said subtracted set of plasmid DNA sequences; wherein said contigs and sequences comprise the plasmid DNA sequences of said second *Bacillus* species.

42. The method according to claim 41, wherein said first and second *Bacillus* species is selected from the group consisting of *Bacillus thuringiensis*, *Bacillus subtilis*, *Bacillus cereus*, and *Bacillus anthracis*.

43. The method according to claim 42, wherein said *Bacillus* species is *Bacillus thuringiensis*.

44. The method of claim 41, wherein the nucleotide sequences of said chromosomal genomic DNA obtained in step c) comprise the sequences selected from the group consisting of SEQ ID NO:1 through SEQ ID NO:8283.