

Needleman & Wunsch (J. Mol. Biol. 48:443 (1970)); by the search for similarity method of Pearson (Proc. Natl. Acad. Sci. USA 85: 2444 (1988)); by computerized implementations of these algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.); ClustalW (CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, Calif., described by, e.g., Higgins, Gene 73: 237-244 (1988); Corpet, Nucleic Acids Res. 16:10881-10890 (1988); Huang, Computer Applications in the Biosciences 8:155-165 (1992); and Pearson, Methods in Mol. Biol. 24:307-331 (1994); Pfam (Sonnhammer, Nucleic Acids Res. 26:322-325 (1998); TreeAlign (Hein, Methods Mol. Biol. 25:349-364 (1994); MEG-ALIGN, and SAM sequence alignment computer programs; or, by manual visual inspection.

[0047] Another example of algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul et al, J. Mol. Biol. 215: 403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>; see also Zhang, Genome Res. 7:649-656 (1997) for the "Power-BLAST" variation. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al, J. Mol. Biol. 215: 403-410 (1990)). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff, Proc. Natl. Acad. Sci. USA 89:10915-10919 (1992)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands. The term BLAST refers to the BLAST algorithm which performs a statistical analysis of the similarity between two sequences; see, e.g., Karlin, Proc. Natl. Acad. Sci. USA 90:5873-5787 (1993). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[0048] In a preferred embodiment of the present invention, a nucleic acid molecule of the present invention encodes the homologue of a known protein. Table 1 sets forth a list of nucleic acid molecules that encode *D. v. virgifera* proteins or fragments thereof which are homologues of known proteins

[0049] In a preferred embodiment of the present invention, a *D. v. virgifera* protein or fragment thereof of the present

invention is a homologue of another insect protein. In another preferred embodiment of the present invention, a *D. v. virgifera* protein or fragment thereof of the present invention is a homologue of a fungal protein. In another preferred embodiment of the present invention, a *D. v. virgifera* protein or fragment thereof of the present invention is a homologue of a mammalian protein. In another preferred embodiment of the present invention, a *D. v. virgifera* protein or fragment thereof of the present invention is a homologue of a bacterial protein. In another preferred embodiment of the present invention, a *D. v. virgifera* protein or fragment thereof of the present invention is a homologue of an algal protein. In another preferred embodiment of the present invention, a *D. v. virgifera* protein or fragment thereof of the present invention is a homologue of a plant protein.

[0050] In a preferred embodiment of the present invention, the nucleic molecule of the present invention encodes a *D. v. virgifera* protein or fragment thereof where a *D. v. virgifera* protein or fragment thereof exhibits a BLAST probability score of greater than 1E-12, preferably a BLAST probability score of between about 1E-30 and about 1E-12, even more preferably a BLAST probability score of greater than 1E-30 with its homologue.

[0051] In another preferred embodiment of the present invention, the nucleic acid molecule encoding a *D. v. virgifera* protein or fragment thereof exhibits a percent identity with its homologue of between about 25% and about 40%, more preferably of between about 40% and about 70%, even more preferably of between about 70% and about 90% and even more preferably between about 90% and 99%. In another preferred embodiment, of the present invention, a *D. v. virgifera* protein or fragment thereof exhibits a percent identity with its homologue of 100%.

[0052] In a preferred embodiment of the present invention, the nucleic molecule of the present invention encodes a *D. v. virgifera* protein or fragment thereof where the *D. v. virgifera* protein exhibits a BLAST score of greater than 120, preferably a BLAST score of between about 1450 and about 120, even more preferably a BLAST score of greater than 1450 with its homologue.

[0053] The degeneracy of the genetic code, which allows different nucleotide sequences to code for the same protein or peptide, is known in the literature. (U.S. Pat. No. 4,757,006).

[0054] In an aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleotide sequence from those encoding a *D. v. virgifera* protein or fragment thereof in SEQ ID NO: 1 through SEQ ID NO: 9112 due to the degeneracy in the genetic code in that they encode the same protein but differ in nucleotide sequence.

[0055] In another further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleotide sequence from those encoding a *D. v. virgifera* protein or fragment thereof in SEQ ID NO: 1 through SEQ ID NO: 9112 due to fact that the different nucleotide sequence encodes a protein having one or more conservative amino acid changes. It is understood that codons capable of coding for such conservative amino acid substitutions are known in the art.

[0056] It is well known in the art that one or more amino acids in a native sequence can be substituted with another amino acid(s), the charge and polarity of which are similar to that of the native amino acid, i.e., a conservative amino acid substitution, resulting in a silent change. Biologically func-