

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
9 December 2010 (09.12.2010)

PCT

(10) International Publication Number
WO 2010/141294 A1

(51) International Patent Classification:
C12N 15/82 (2006.01) A01H 5/00 (2006.01)

(21) International Application Number:
PCT/US2010/036275

(22) International Filing Date:
27 May 2010 (27.05.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/182,872 1 June 2009 (01.06.2009) US

(71) Applicant (for all designated States except US): **NORTH CAROLINA STATE UNIVERSITY** [US/US]; 920 Main Campus Drive, CB 8210, Venture Building II, Suite 400, Raleigh, North Carolina 27606 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ROE, R., Michael** [US/US]; 1695 Woodards Dairy Road, Middlesex, North Carolina 27557 (US). **DONOHUE, Kevin, V.** [US/US]; 100 Village Circle Way, Apt. 723, Durham, North Carolina 27713 (US). **VAN KRETSCHMAR, Jacob, B.** [US/US]; 2735 Rosedale Avenue, Apt. A, Raleigh, North Carolina 27607 (US). **MAGALHAES, Leonardo, C.** [BR/US]; 505 Tartan Circle, Apt. 22, Raleigh, North Carolina 27606 (US).

(74) Agent: **SIBLEY, Kenneth, D.**; Myers Bigel Sibley & Sajovec, P.A., P.O. Box 37428, Raleigh, North Carolina 27627 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: INSECTICIDAL TARGET GENES IDENTIFIED FROM BUG WHOLE BODY AND DIGESTIVE SYSTEM TRANSCRIPTOMES

(57) Abstract: A transgenic plant such as a cotton plant is described. The plant contains a polynucleotide construct that comprises a heterologous nucleotide sequence operably linked to a promoter that drives expression in the plant. The heterologous nucleotide sequence encodes an active agent toxic to an insect pest such as a plant bug, leaf bug, grass bug, shield bug or stink bug. The active agent may be: (i) an RNA (e.g., a dsRNA) that inhibits expression of a target gene in said insect pest, and reduces or inhibits the expression of the expression product (e.g., protein, ribosomal RNA) (ii) an antibody that binds to the expression product of said target gene and inhibits the activity thereof, or (iii) an aptamer that binds to the expression product of said target gene and inhibits the activity thereof. Insecticidal compositions and bait compositions containing such active ingredients are also described.



WO 2010/141294 A1

INSECTICIDAL TARGET GENES IDENTIFIED FROM BUG WHOLE BODY AND DIGESTIVE SYSTEM TRANSCRIPTOMES

R. Michael Roe, Kevin V. Donohue, Jacob B. van Kretschmar,
and Leonardo C. Magalhaes

Field of the Invention

The present invention concerns transgenic plants such as transgenic cotton, and methods of controlling insect pests such as plant bugs and stink bugs.

Background of the Invention

Since the mid 1990s, transgenic *Bt* cotton has been widely expanded and highly efficient in controlling lepidopteran pests. For this reason—and the success of the boll weevil (*Anthonomus grandis* (Boheman)) eradication program—chemical insecticide use is reduced considerably. As a result, a new group of pests has recently emerged in cotton. Piercing-sucking insects, such as plant bugs and stink bugs, which were indirectly being controlled have now become important cotton pests in the US.

Among other injuries, *Lygus lineolaris* impact cotton by releasing salivary toxins that causes abortion and malformation of the attacked part, usually flowers. Losses can exceed \$40 million depending on the region and time of infestation. Stink bugs damage cotton mainly later in the season by feeding on developing bolls, which reduces yield and fiber quality. Total cotton losses in 2004 related to stink bug damage were estimated around \$10 million. Insecticide applications and yield losses in cotton production due to plant and stink bugs have been increasing sharply, and tarnished plant bug resistance to both pyrethroid and organophosphate insecticides is now widespread.

The absence of a transgenic approach using *Bt* toxins to control piercing-sucking insects is problematic for the sustainability of crop production. Also, even if a protein toxin was available to control piercing-sucking insects, there are important advantages to a nucleic acid approach for insect control including the significant ease with this approach of

preventing target site insect resistance not possible with protein insecticides. Hence, there is a need to identify novel insecticide modes of action for use in developing nucleic acid-based, transgenic plant technology for insect control.

Summary of the Invention

A first aspect of the invention is a transgenic plant that contains a polynucleotide construct that comprises a heterologous nucleotide sequence operably linked to a promoter that drives expression in the plant, the heterologous nucleotide sequence encoding an active agent toxic to an insect pest; the active agent selected from the group consisting of: (i) an RNA that inhibits expression of a target gene in the insect pest, and reduces or inhibits the expression of the expression product (*e.g.*, protein, ribosomal RNA) (ii) an antibody that binds to the expression product of the target gene and inhibits the activity thereof, and (iii) an aptamer that binds to the expression product of the target gene and inhibits the activity thereof.

In some embodiments of the foregoing, the insect pest is a member of the Heteroptera suborder of the Hemiptera order.

In some embodiments of the foregoing, the insect pest is a member of the Miridae family.

In some embodiments of the foregoing, the insect pest is a plant bug, leaf bug, or grass bug.

In some embodiments of the foregoing, the insect pest is a member of the Pentatomoidea superfamily.

In some embodiments of the foregoing, the insect pest is a shield bug or stink bug.

In some embodiments of the foregoing, the plant is a monocot; In some embodiments of the foregoing, wherein the plant is a dicot.

In some embodiments of the foregoing, the plant is cotton.

In some embodiments of the foregoing, the dsRNA is a double-stranded molecule with a first strand consisting essentially of a ribonucleotide sequence which corresponds to a nucleotide sequence of the target gene and a second strand consisting essentially of a ribonucleotide sequence which is complementary to the nucleotide sequence of the target gene, wherein the first and the second ribonucleotide strands are separate complementary strands (*e.g.*, on the same molecule or on separate molecules) that hybridize to each other to form the double-stranded molecule, and the double-stranded molecule inhibits expression of the target gene.

In some embodiments of the foregoing, the plant further comprises:

(a) plant cells containing nucleotide sequences encoding immunoglobulin heavy- and light-chain polypeptides, wherein each polypeptide contains a leader sequence forming a secretion signal; and

(b) immunoglobulin molecules encoded by the nucleotide sequences, wherein the leader sequence is cleaved from the immunoglobulin molecules following proteolytic processing.

In some embodiments of the foregoing, the plant further comprises:

(a) plant cells containing nucleotide sequences encoding a biologically functional multimeric protein comprising at least two different polypeptides not naturally produced by plants, wherein each nucleotide sequence encoding a polypeptide of the multimeric protein encodes a leader sequence forming a secretion signal that is cleaved from the polypeptide following proteolytic processing; and

(b) biologically functional multimeric protein encoded by the nucleotide sequences and resulting from assembly of the at least two different polypeptides wherein the multimeric protein is an abzyme or immunoglobulin.

A further aspect of the invention is a transgenic seed of the plant of described above, wherein the seed comprises the heterologous polynucleotide.

A further aspect of the invention is a method for protecting a plant from an insect pest, the method comprising: introducing into the plant at least one polynucleotide construct that comprises a heterologous nucleotide sequence operably linked to a promoter that drives expression in the plant, wherein the nucleotide sequence encodes an active agent toxic to an insect pest; and wherein the plant is a plant as described above.

A further aspect of the invention is a method for protecting a plant from an insect pest, the method comprising administering to the insect pest an active agent selected from the group consisting of: (i) an RNA that inhibits expression of a target gene in the insect pest, and reduces or inhibits the expression of the expression product (e.g., protein, ribosomal RNA), (ii) an antibody that binds to the expression product of the target gene and inhibits the activity thereof, and (iii) an aptamer that binds to the expression product of the target gene and inhibits the activity thereof.

In some embodiments of the foregoing, the administering step is carried out by applying the active agent to the plant.

In some embodiments of the foregoing, the applying step is carried out by applying an insecticidal bait composition to the plant, the insecticidal bait composition comprising the active agent.

A further aspect of the invention is an insecticidal composition comprising a carrier media and an active agent, wherein the active agent comprises a compound selected from the group consisting of: (i) an RNA that inhibits expression of a target gene in the insect pest, and reduces or inhibits the expression of the expression product (*e.g.*, protein, ribosomal RNA), (ii) an antibody that binds to the expression product of the target gene and inhibits the activity thereof, and (iii) an aptamer that binds to the expression product of the target gene and inhibits the activity thereof.

A further aspect of the invention is, in an insecticidal bait composition comprising an insecticide and one or more of insect food, insect attractant, carrier, and an insecticidal active agent, the improvement comprising employing as the active agent a compound selected from the group consisting of: (i) an RNA that inhibits expression of a target gene in the insect pest, and reduces or inhibits the expression of the expression product (*e.g.*, protein, ribosomal RNA), (ii) an antibody that binds to the expression product of the target gene and inhibits the activity thereof, and (iii) an aptamer that binds to the expression product of the target gene and inhibits the activity thereof.

In some embodiments of the foregoing, the target gene encodes a protein of the insect pest selected from the group consisting of: transcription factors, translation factors, microvilli proteins (*e.g.*, mucin, pseudomucin), electron transport proteins, *Bacillus thuringiensis* toxin (Cry toxin) receptors, insecticide target proteins (*e.g.*, acetylcholine esterase, GABA-gated chloride channels, sodium channels, nicotinic acetylcholine receptors, chloride channels, midgut membrane proteins, mitochondrial ATP synthase, citin biosynthesis enzymes, ecdysone receptors, mitochondrial complex I, III, and IV electron transporters, Mitochondrial complex I electron transporters, voltage-dependent sodium channels, acetyl CoA carboxylase, *etc.*), and ribosomal RNA.

In some embodiments of the foregoing, the target gene is given in **TABLE 1** herein.

The present invention is explained in great detail in the specification set forth below. The disclosures of all United States Patent references cited herein are to be incorporated by reference herein in their entirety.

Brief Description of the Drawings

Fig. 1. Percentage of ESTs matching putative functions listed.

Detailed Description of the Preferred Embodiments

A. DEFINITIONS.

"Plant" as used herein includes both gymnosperms and angiosperms, and includes both monocots and dicots.

Dicotyledon (dicot): A flowering plant whose embryos have two seed halves or cotyledons. Examples of dicots are: tobacco; tomato; the legumes including alfalfa; oaks; maples; roses; mints; squashes; daisies; walnuts; cacti; violets; and buttercups.

Monocotyledon (monocot): A flowering plant whose embryos have one cotyledon or seed leaf. Examples of monocots are: lilies; grasses; corn; grains, including oats, wheat and barley; orchids; irises; onions and palms.

Eukaryotic hybrid vector: A DNA by means of which a DNA coding for a polypeptide (insert) can be introduced into a eukaryotic cell.

Extrachromosomal ribosomal DNA (rDNA): A DNA found in unicellular eukaryotes outside the chromosomes, carrying one or more genes coding for ribosomal RNA and replicating autonomously (independent of the replication of the chromosomes).

Palindromic DNA: A DNA sequence with one or more centers of symmetry.

DNA: Deoxyribonucleic acid.

T-DNA: A segment of transferred DNA.

rDNA: Ribosomal DNA.

RNA: Ribonucleic acid.

rRNA: Ribosomal RNA.

Ti-plasmid: Tumor-inducing plasmid.

Ti-DNA: A segment of DNA from Ti-plasmid.

Insert: A DNA sequence foreign to the rDNA, consisting of a structural gene and optionally additional DNA sequences.

Structural gene: A gene coding for a polypeptide and being equipped with a suitable promoter, termination sequence and optionally other regulatory DNA sequences, and having a correct reading frame.

Signal Sequence: A DNA sequence coding for an amino acid sequence attached to the polypeptide which binds the polypeptide to the endoplasmic reticulum and is essential for

protein secretion. This sequence may also be referred to herein as a secretion signal or secretion signal sequence. The term "signal sequence" may also be used to refer to the sequence of amino acids that determines whether a protein will be formed on the rough endoplasmic reticulum or on free ribosomes. And while a "leader sequence" generally means a sequence near the 5' end of a nucleic acid strand or the amino terminus of a protein that functions in targeting or regulation, the term is sometimes used herein to include a "secretion signal" or a "signal sequence".

(Selective) Genetic marker. A DNA sequence coding for a phenotypic trait by means of which transformed cells can be selected from untransformed cells.

Promoter: A recognition site on a DNA sequence or group of DNA sequences that provide an expression control element for a gene and to which RNA polymerase specifically binds and initiates RNA synthesis (transcription) of that gene.

Inducible promoter: A promoter where the rate of RNA polymerase binding and initiation is modulated by external stimuli. Such stimuli include light, heat, anaerobic stress, alteration in nutrient conditions, presence or absence of a metabolite, presence of a ligand, microbial attack, wounding and the like.

Viral promoter: A promoter with a DNA sequence substantially similar to the promoter found at the 5' end of a viral gene. A typical viral promoter is found at the 5' end of the gene coding for the p21 protein of MMTV described by Huang et al., Cell 27: 245 (1981).

Synthetic promoter: A promoter that was chemically synthesized rather than biologically derived. Usually synthetic promoters incorporate sequence changes that optimize the efficiency of RNA polymerase initiation.

Constitutive promoter: A promoter where the rate of RNA polymerase binding and initiation is approximately constant and relatively independent of external stimuli. Examples of constitutive promoters include the cauliflower mosaic virus 35S and 19S promoters described by Poszkowski et al., EMBO J. 3: 2719 (1989) and Odell et al., Nature 313: 810 (1985).

Temporally regulated promoter: A promoter where the rate of RNA polymerase binding and initiation is modulated at a specific time during development. Examples of temporally regulated promoters are given in Chua et al., Science 244: 174 181 (1989).

Spatially regulated promoter: A promoter where the rate of RNA polymerase binding and initiation is modulated in a specific structure of the organism such as the leaf, stem or

root Examples of spatially regulated promoters are given in Chua et al., Science 244: 174 181 (1989).

Spatiotemporally regulated promoter: A promoter where the rate of RNA polymerase binding and initiation is modulated in a specific structure of the organism at a specific time during development. A typical spatiotemporally regulated promoter is the EPSP synthase-35S promoter described by Chua et al., Id. (1989).

Single-chain antigen-binding protein: A polypeptide composed of an immunoglobulin light-chain variable region amino acid sequence (V_L) tethered to an immunoglobulin heavy-chain variable region amino acid sequence (V_H) by a peptide that links the carboxyl terminus of the V_L sequence to the amino terminus of the V_H sequence.

Single-chain antigen-binding protein-coding gene: A recombinant gene coding for a single-chain antigen-binding protein.

Multimeric protein: A globular protein containing more than one separate polypeptide or protein chain associated with each other to form a single globular protein. Both heterodimeric and homodimeric proteins are multimeric proteins.

Polypeptide and peptide: A linear series of amino acid residues connected one to the other by peptide bonds between the alpha-amino and carboxy groups of adjacent residues.

Protein: A linear series of greater than about 50 amino acid residues connected one to the other as in a polypeptide.

Immunoglobulin product: A polypeptide, protein or multimeric protein containing at least the immunologically active portion of an immunoglobulin heavy chain and is thus capable of specifically combining with an antigen. Exemplary immunoglobulin products are an immunoglobulin heavy chain, immunoglobulin molecules, substantially intact immunoglobulin molecules, any portion of an immunoglobulin that contains the paratope, including those portions known in the art as Fab fragments, Fab' fragment, F(ab')₂ fragment and Fv fragment.

Immunoglobulin molecule: A multimeric protein containing the immunologically active portions of an immunoglobulin heavy chain and immunoglobulin light chain covalently coupled together and capable of specifically combining with antigen.

Fab fragment: A multimeric protein consisting of the portion of an immunoglobulin molecule containing the immunologically active portions of an immunoglobulin heavy chain and an immunoglobulin light chain covalently coupled together and capable of specifically combining with antigen. Fab fragments are typically prepared by proteolytic digestion of

substantially intact immunoglobulin molecules with papain using methods that are well known in the art. However, a Fab fragment may also be prepared by expressing in a suitable host cell the desired portions of immunoglobulin heavy chain and immunoglobulin light chain using methods well known in the art.

F_v fragment: A multimeric protein consisting of the immunologically active portions of an immunoglobulin heavy chain variable region and an immunoglobulin light chain variable region covalently coupled together and capable of specifically combining with antigen. Fv fragments are typically prepared by expressing in suitable host cell the desired portions of immunoglobulin heavy chain variable region and immunoglobulin light chain variable region using methods well known in the art.

Asexual propagation: Producing progeny by regenerating an entire plant from leaf cuttings, stem cuttings, root cuttings, single plant cells (protoplasts) and callus.

Catalytic site: The portion of a molecule that is capable of binding a reactant and improving the rate of a reaction. Catalytic sites may be present on polypeptides or proteins, enzymes, organics, organo-metal compounds, metals and the like. A catalytic site may be made up of separate portions present on one or more polypeptide chains or compounds. These separate catalytic portions associate together to form a larger portion of a catalytic site. A catalytic site may be formed by a polypeptide or protein that is bonded to a metal.

Enzymatic site: The portion of a protein molecule that contains a catalytic site. Most enzymatic sites exhibit a very high selective substrate specificity. An enzymatic site may be comprised of two or more enzymatic site portions present on different segments of the same polypeptide chain. These enzymatic site portions are associated together to form a greater portion of an enzymatic site. A portion of an enzymatic site may also be a metal.

Self-pollination: The transfer of pollen from male flower parts to female flower parts on the same plant. This process typically produces seed. **Cross-pollination:** The transfer of pollen from the male flower parts of one plant to the female flower parts of another plant. This process typically produces seed from which viable progeny can be grown.

Epitope: A portion of a molecule that is specifically recognized by an immunoglobulin product. It is also referred to as the determinant or antigenic determinant.

Abzyme: An immunoglobulin molecule capable of acting as an enzyme or a catalyst.

Enzyme: A protein, polypeptide, peptide RNA molecule, or multimeric protein capable of accelerating or producing by catalytic action some change in a substrate for which it is often specific.

"Aptamer" as used herein refers to nucleic acid molecules having specific binding affinity to molecules through interactions other than classic Watson-Crick base pairing. Aptamers, like peptides generated by phage display or monoclonal antibodies ("MAbs"), are capable of specifically binding to selected targets and modulating the target's activity, e.g., through binding aptamers may block their target's ability to function. Created by an in vitro selection process from pools of random sequence oligonucleotides, aptamers have been generated for over 100 proteins including growth factors, transcription factors, enzymes, immunoglobulins, and receptors. A typical aptamer is 10-15 kDa in size (30-45 nucleotides), binds its target with sub-nanomolar affinity, and discriminates against closely related targets (e.g., aptamers will typically not bind other proteins from the same gene family). A series of structural studies have shown that aptamers are capable of using the same types of binding interactions (e.g., hydrogen bonding, electrostatic complementarity, hydrophobic contacts, steric exclusion) that drive affinity and specificity in antibody-antigen complexes. *See, e.g.,* US Patent No. 7,538,211.

"Administering" as used herein with respect to active agents can be carried out directly or indirectly by any suitable means or route, including but not limited to orally or topically administering the active agent, by orally or topically administering a microbial vector such as a virus that produces the active agent (*e.g.,* a dsRNA), etc.

B. ANTIBODY ACTIVE AGENTS.

Plants containing antibody active agents of the present invention can be produced in accordance with known techniques, or variations thereof that will be apparent to those skilled in the art. *See, e.g.,* US Patent No.s 7,037,722 and 6,417,429 to Hein and Hiatt; *see also* US Patent Nos. 7,101, 688; 7,005,560; and 5,959,177.

In general, such methods can be carried out by: 1. Inserting into the genome of a first member of a plant species a gene coding for a first polypeptide to produce a first transformant 2. Inserting into the genome of a second member of a plant species a gene coding for a second polypeptide to produce a second transformant 3. Producing a population of progeny from the first and second transformants. 4. Isolating from the population, a progeny having the multimeric protein. The multimeric protein is typically an immunoglobulin molecule comprised of an immunoglobulin heavy chain and an immunoglobulin light chain. The immunoglobulin heavy and light chains are associated with each other and assume a conformation having an antigen binding site specific for, as evidenced by its ability to be

competitively inhibited, a preselected or predetermined antigen (that is, the product of the insect gene). When the multimeric protein is an antigen binding protein its affinity or avidity is generally greater than 10^5 M^{-1} or usually greater than 10^6 M^{-1} , and preferably greater than 10^8 M^{-1} .

In a further embodiment, the multimeric protein is a Fab fragment consisting of a portion of an immunoglobulin heavy chain and a portion of an immunoglobulin light chain. The immunoglobulin heavy and light chains are associated with each other and assume a conformation having an antigen binding site specific for a preselected or predetermined antigen. The antigen binding site on a Fab fragment has a binding affinity or avidity similar to the antigen binding site on an immunoglobulin molecule.

In yet another embodiment, the present transgenic plant contains a multimeric protein that is a F_v fragment comprised of at least a portion of an immunoglobulin heavy chain variable region and at least a portion of an immunoglobulin light chain variable region. The immunoglobulin heavy and light chain variable regions autogenously associate with each other within the plant cell to assume a biologically active conformation having a binding site specific for a preselected or predetermined antigen. The antigen binding site on the F_v fragment has an affinity or avidity for its antigen similar to the affinity displayed by the antigen binding site present on an immunoglobulin molecule.

In still another embodiment, the multimeric protein is an enzyme that binds a substrate and catalyzes the formation of a product from the substrate. While the topology of the substrate binding site (ligand binding site) of the catalytic multimeric protein is probably more important for its activity than affinity (association constant or pK_a) for the substrate, the subject multimeric protein has an association constant for its preselected substrate greater than 10^3 M^{-1} , more usually greater than 10^5 M^{-1} or 10^6 M^{-1} and preferably greater than 10^7 M^{-1} .

When the multimeric protein produced in accordance with the present invention is an abzyme comprised of at least a portion of the immunoglobulin heavy chain variable region in association with another polypeptide chain, this other polypeptide chain includes at least the biologically active portion of an immunoglobulin light chain variable region. Together, these two polypeptides assume a conformation having a binding affinity or association constant for a preselected ligand that is different, preferably higher, than the affinity or association constant of either of the polypeptides alone, i.e., as monomers. Useful multimeric proteins contain one or both polypeptide chains derived from the variable region of the light and

heavy chains of an immunoglobulin. Typically, polypeptides comprising the light (VL) and heavy (VH) variable regions are employed together for binding the preselected antigen.

1. Inserting Genes Coding for a First Polypeptide into a First Member of a Plant Species. Methods for isolating a gene coding for a desired first polypeptide are well known in the art. See, for example, Guide To Molecular Cloning Techniques in Methods In Enzymology. Volume 152, Berger and Kimmel, eds. (1987); and Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, New York (1987) whose disclosures are herein incorporated by reference.

Genes useful in practicing this invention include genes coding for a polypeptide contained in immunoglobulin products, immunoglobulin molecules, Fab fragments, F_v fragments, enzymes, receptors and abzymes. Particularly preferred are genes coding for immunoglobulin heavy and light chain variable regions. Typically, the genes coding for the immunoglobulin heavy chain variable region and immunoglobulin light chain variable region of an immunoglobulin capable of binding a preselected antigen are used. These genes are isolated from cells obtained from a vertebrate, preferably a mammal which has been immunized with an antigenic ligand (antigen) against which activity is sought, i.e., a preselected antigen. The immunization can be carried out conventionally and antibody titer in the animal can be monitored to determine the stage of immunization desired, which corresponds to the affinity or avidity desired. Partially immunized animals typically receive only one immunization and cells are collected therefrom shortly after a response is detected. Fully immunized animals display a peak titer which is achieved with one or more repeated injections of the antigen into the host mammal, normally at two to three week intervals.

Usually three to five days after the last challenge, the spleen is removed and the genes coding for immunoglobulin heavy and immunoglobulin light chain are isolated from the rearranged B cells present in the spleen using standard procedures. See Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, New York (1987) and Antibodies: A Laboratory Manual, Harlowe and Lane, eds., Cold Spring Harbor, N.Y. (1983).

Genes coding for VH and VL polypeptides can be derived from cells producing IgA, IgD, IgE, IgG or IgM, most preferably from IgM and IgG, producing cells. Methods for preparing fragments of genomic DNA from which immunoglobulin variable region genes can be cloned are well known in the art. See for example, Herrmann et al., Methods in Enzymol. 152: 180 183 (1987); Frischauf, Methods in Enzymol. 152: 183 190 (1987); Frischauf,

Methods in Enzymol. 152: 199 212 (1987). (The teachings of the references cited herein are hereby incorporated by reference).

Probes useful for isolating the genes coding for immunoglobulin products include the sequences coding for the constant portion of the V_H and V_L sequences coding for the framework regions of V_H and V_L and probes for the constant region of the entire rearranged immunoglobulin gene, these sequences being obtainable from available sources. See for example, Early and Hood, Genetic Engineering, Setlow and Hollaender eds., Vol. 3: 157 138, Plenum Publishing Corporation, New York (1981); and Kabat et al., Sequences of Immunological Interests, National Institutes of Health, Bethesda, Md. (1987).

Genes coding for a polypeptide subunit of a multimeric protein can be isolated from either the genomic DNA containing the gene expressing the polypeptide or the messenger RNA (mRNA) which codes for the polypeptide. The difficulty in using genomic DNA is in juxtaposing the sequences coding for the polypeptide where the sequences are separated by introns. The DNA fragment(s) containing the proper exons must be isolated, the introns excised, and the exons spliced together in the proper order and orientation. For the most part, this will be difficult so the alternative technique employing mRNA will be the method of choice because the sequence is contiguous (free of introns) for the entire polypeptide. Methods for isolating mRNA coding for peptides or proteins are well known in the art See, for example, Current Protocols in Molecular Biology, Ausubel et al., John Wiley and Sons, New York (1987), Guide to Molecular Cloning Techniques, in Methods In Enzymology, Volume 152, Berger and Kimmel, eds. (1987), and Molecular Cloning: A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1982).

The polypeptide coding genes isolated above are typically operatively linked to an expression vector. Expression vectors compatible with the host cells, preferably those E 15 compatible with plant cells are used to express the genes of the present invention. Typical expression vectors useful for expression of genes in plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers et al., Meth. in Enzymol. 153: 253 277 (1987).

However, several other expression vector systems are known to function in plants. See for example, Verma et al., PCT Publication No. WO87/00551; and Cocking and Davey, Science, 236: 1259 1262 (1987).

The expression vectors described above contain expression control elements including the promoter. The polypeptide coding genes are operatively linked to the expression vector to allow the promoter sequence to direct RNA polymerase binding and synthesis of the desired polypeptide coding gene. Useful in expressing the polypeptide coding gene are promoters which are inducible, viral, synthetic, constitutive, temporally regulated, spatially regulated, and spatiotemporally regulated. The choice of which expression vector and ultimately to which promoter a polypeptide coding gene is operatively linked depends directly, as is well known in the art, on the functional properties desired, e.g. the location and timing of protein expression, and the host cell to be transformed, these being limitations inherent in the art of constructing recombinant DNA molecules. However, an expression vector useful in practicing the present invention is at least capable of directing the replication, and preferably also the expression of the polypeptide coding gene included in the DNA segment to which it is operatively linked.

In preferred embodiments, the expression vector used to express the polypeptide coding gene includes a selection marker that is effective in a plant cell preferably a drug resistance selection marker. A preferred drug resistance marker is the gene whose expression results in kanamycin resistance, i.e., the chimeric gene containing the nopaline synthase promoter, Tn5 neomycin phosphotransferase II and nopaline synthase 3' nontranslated region described by Rogers et al., in *Methods For Plant Molecular Biology*, a Weissbach and H. Weissbach, eds., Academic Press Inc., San Diego, Calif. (1988). A useful plant expression vector is commercially available from Pharmacia, Piscataway, N.J.

A variety of methods have been developed to operatively link DNA to vectors via complementary cohesive termini. For instance, complementary homopolymer tracks can be added to the DNA segment to be inserted and to the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules.

Alternatively, synthetic linkers containing one or more restriction endonuclease sites can be used to join the DNA segment to the expression vector. The synthetic linkers are attached to blunt-ended DNA segments by incubating the blunt-ended DNA segments with a large excess of synthetic linker molecules in the presence of an enzyme that is able to catalyze the ligation of blunt-ended DNA molecules, such as bacteria phage T4 DNA ligase. Thus, the products of the reaction are DNA segments carrying synthetic linker sequences at their ends. These DNA segments are then cleaved with the appropriate restriction

endonuclease and ligated into an expression vector that has been cleaved with an enzyme that produces termini compatible with those of the synthetic linker. Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including New England BioLabs, Beverly, Mass.

2. *Inserting a Gene Coding for a Second Polypeptide into a Second Member of a Plant Species.* Useful genes include those genes coding for a second polypeptide that can autogenously associate with the first polypeptide in such a way as to form a biologically functional multimeric protein. The methods used to introduce a gene coding for this second polypeptide into a second member of a plant species are the same as the methods used to introduce a gene into the first member of the same plant species and have been described above.

3. *Producing a Population of Progeny from the First and Second Transformants.* A population of progeny can be produced from the first and second transformants of a plant species by methods well known in the art including those methods known as cross fertilization described by Mendel in 1865 (an English translation of Mendel's original paper together with comments and a bibliography of Mendel by others can be found in Experiments In Plant Hybridization. Edinburgh, Scotland, Oliver Boyd, eds., 1965).

4. *Isolating Progeny Containing the Multimeric Protein.* Progeny containing the desired multimeric protein can be identified by assaying for the presence of the biologically multimeric protein using assay methods well known in the art. Such methods include Western blotting, immunoassays, binding assays, and any assay designed to detect a biologically functional multimeric protein. See, for example, the assays described in Immunology: The Science of Self-Nonself Discrimination, Klein, John Wiley and Sons, New York, N.Y. (1982).

Preferred screening assays are those where the biologically active site on the multimeric protein is detected in such a way as to produce a detectible signal. This signal may be produced directly or indirectly and such signals include, for example, the production of a complex, formation of a catalytic reaction product, the release or uptake of energy, and the like. For example, a progeny containing an antibody molecule produced by this method may be processed in such a way to allow that antibody to bind its antigen in a standard immunoassay such as an ELISA or a radio-immunoassay similar to the immunoassays described in Antibodies: A Laboratory Manual, Harlow and Lane, eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1988).

C. RNA ACTIVE AGENTS.

RNA active agents for carrying out the present invention are generally RNAi agents or dsRNAs (including but not limited to siRNAs, shRNAs and miRNAs, *see, e.g.*, US Patent Application 20090118214). Production of such RNA active agents in plants is carried out in accordance with known techniques, or variations thereof that will be apparent to those skilled in the art. *See, e.g.*, US Patent No. 6,506,559 to Fire et al., *see also* US Patents Nos. 7,456,335 to Kogel et al., 7,122,719 to Hakimi et al., and 6,846,482 to Liu et al.; US Patent Application 2006/0095987 to Niblett, and K. Gordon and P. Waterhouse, *RNAi for insect-proof plants*, Nature Biotechnology **23**, 1231-1232 (November 2007).

Inhibition of gene expression refers to the absence (or observable decrease) in the level of protein and/or mRNA product from a target gene. Specificity refers to the ability to inhibit the target gene without manifest effects on other genes of the cell. The consequences of inhibition can be confirmed by examination of the outward properties of the cell or organism (as presented below in the examples) or by biochemical techniques such as RNA solution hybridization, nuclease protection, Northern hybridization, reverse transcription, gene expression monitoring with a microarray, antibody binding, enzyme linked immunosorbent assay (ELISA), Western blotting, radioimmunoassay (RIA), other immunoassays, and fluorescence activated cell analysis (FACS). For RNA-mediated inhibition in a cell line or whole organism, gene expression is conveniently assayed by use of a reporter or drug resistance gene whose protein product is easily assayed. Such reporter genes include acetohydroxyacid synthase (AHAS), alkaline phosphatase (AP), beta galactosidase (LacZ), beta glucuronidase (GUS), chloramphenicol acetyltransferase (CAT), green fluorescent protein (GFP), horseradish peroxidase (HRP), luciferase (Luc), nopaline synthase (NOS), octopine synthase (OCS), and derivatives thereof. Multiple selectable markers are available that confer resistance to ampicillin, bleomycin, chloramphenicol, gentamycin, hygromycin, kanamycin, lincomycin, methotrexate, phosphinothricin, puromycin, and tetracyclin.

Depending on the assay, quantitation of the amount of gene expression allows one to determine a degree of inhibition which is greater than 10%, 33%, 50%, 90%, 95% or 99% as compared to a cell not treated according to the present invention. Lower doses of injected material and longer times after administration of dsRNA may result in inhibition in a smaller fraction of cells (*e.g.*, at least 10%, 20%, 50%, 75%, 90%, or 95% of targeted cells).

Quantitation of gene expression in a cell may show similar amounts of inhibition at the level of accumulation of target mRNA or translation of target protein. As an example, the efficiency of inhibition may be determined by assessing the amount of gene product in the cell: mRNA may be detected with a hybridization probe having a nucleotide sequence outside the region used for the inhibitory double-stranded RNA, or translated polypeptide may be detected with an antibody raised against the polypeptide sequence of that region.

The RNA may comprise one or more strands of polymerized ribonucleotide. It may include modifications to either the phosphate-sugar backbone or the nucleoside. For example, the phosphodiester linkages of natural RNA may be modified to include at least one of a nitrogen or sulfur heteroatom. Modifications in RNA structure may be tailored to allow specific genetic inhibition while avoiding a general panic response in some organisms which is generated by dsRNA. Likewise, bases may be modified to block the activity of adenosine deaminase. RNA may be produced enzymatically or by partial/total organic synthesis, any modified ribonucleotide can be introduced by in vitro enzymatic or organic synthesis.

The double-stranded structure may be formed by a single self-complementary RNA strand or two complementary RNA strands. RNA duplex formation may be initiated either inside or outside the cell. The RNA may be introduced in an amount which allows delivery of at least one copy per cell. Higher doses (e.g., at least 5, 10, 100, 500 or 1000 copies per cell) of double-stranded material may yield more effective inhibition; lower doses may also be useful for specific applications. Inhibition is sequence-specific in that nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition.

RNA containing a nucleotide sequences identical to a portion of the target gene are preferred for inhibition. RNA sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found to be effective for inhibition. Thus, sequence identity may be optimized by sequence comparison and alignment algorithms known in the art (see Gribskov and Devereux, *Sequence Analysis Primer*, Stockton Press, 1991, and references cited therein) and calculating the percent difference between the nucleotide sequences by, for example, the Smith-Waterman algorithm as implemented in the BESTFIT software program using default parameters (e.g., University of Wisconsin Genetic Computing Group). Greater than 90% sequence identity, or even 100% sequence identity, between the inhibitory RNA and the portion of the target gene is preferred. Alternatively, the duplex region of the RNA may be defined functionally as a nucleotide sequence that is capable of hybridizing with a portion of the target gene transcript (e.g., 400 mM NaCl, 40 mM PIPES

pH 6.4, 1 mM EDTA, 50 degree C. or 70 degree C. hybridization for 12-16 hours; followed by washing). The length of the identical nucleotide sequences may be at least 25, 50, 100, 200, 300 or 400 bases.

As disclosed herein, 100% sequence identity between the RNA and the target gene is not required to practice the present invention. Thus the invention has the advantage of being able to tolerate sequence variations that might be expected due to genetic mutation, strain polymorphism, or evolutionary divergence.

RNA may be synthesized either in vivo or in vitro. Endogenous RNA polymerase of the cell may mediate transcription in vivo, or cloned RNA polymerase can be used for transcription in vivo or in vitro. For transcription from a transgene in vivo or an expression construct, a regulatory region (e.g., promoter, enhancer, silencer, splice donor and acceptor, polyadenylation) may be used to transcribe the RNA strand (or strands). Inhibition may be targeted by specific transcription in an organ, tissue, or cell type; stimulation of an environmental condition (e.g., infection, stress, temperature, chemical inducers); and/or engineering transcription at a developmental stage or age. The RNA strands may or may not be polyadenylated; the RNA strands may or may not be capable of being translated into a polypeptide by a cell's translational apparatus. RNA may be chemically or enzymatically synthesized by manual or automated reactions. The RNA may be synthesized by a cellular RNA polymerase or a bacteriophage RNA polymerase (e.g., T3, T7, SP6). The use and production of an expression construct are known in the art (see also WO 97/32016; U.S. Pat. Nos. 5,593,874, 5,698,425, 5,712,135, 5,789,214, and 5,804,693; and the references cited therein). If synthesized chemically or by in vitro enzymatic synthesis, the RNA may be purified prior to introduction into the cell. For example, RNA can be purified from a mixture by extraction with a solvent or resin, precipitation, electrophoresis, chromatography, or a combination thereof. Alternatively, the RNA may be used with no or a minimum of purification to avoid losses due to sample processing. The RNA may be dried for storage or dissolved in an aqueous solution. The solution may contain buffers or salts to promote annealing, and/or stabilization of the duplex strands.

RNA may be directly introduced into the cell (i.e., intracellularly); or introduced extracellularly into a cavity, interstitial space, into the circulation of an organism, introduced orally, or may be introduced by bathing an organism in a solution containing the RNA. Methods for oral introduction include direct mixing of the RNA with food of the organism, as well as engineered approaches in which a species that is used as food is engineered to express

the RNA, then fed to the organism to be affected. For example, the RNA may be sprayed onto a plant or a plant may be genetically engineered to express the RNA in an amount sufficient to kill some or all of a pathogen known to infect the plant. Physical methods of introducing nucleic acids, for example, injection directly into the cell or extracellular injection into the organism, may also be used.

Physical methods of introducing nucleic acids into plant cells, and the production of transgenic plants, are discussed in greater detail below, but generally include injection of a solution containing the RNA, bombardment by particles covered by the RNA, soaking the cell or organism in a solution of the RNA, or electroporation of cell membranes in the presence of the RNA. A viral construct packaged into a viral particle would accomplish both efficient introduction of an expression construct into the cell and transcription of RNA encoded by the expression construct. Other methods known in the art for introducing nucleic acids to cells may be used, such as lipid-mediated carrier transport, chemical-mediated transport, such as calcium phosphate, and the like. Thus the RNA may be introduced along with components that perform one or more of the following activities: enhance RNA uptake by the cell, promote annealing of the duplex strands, stabilize the annealed strands, or otherwise increase inhibition of the target gene.

D. APTAMER ACTIVE AGENTS.

Aptamers can be produced in accordance with known techniques or variations thereof that will be apparent to those skilled in the art. *See, e.g.*, US Patent No. 7,538,211; *see also* US Patent No. 6,949,379; 5,840,867; and 5,582,981.

One method for generating an aptamer is with the process entitled "Systematic Evolution of Ligands by Exponential Enrichment". The process is a method for the in vitro evolution of nucleic acid molecules with highly specific binding to target molecules and is described in, e.g., U.S. Pat. No. 5,475,096 entitled "Nucleic Acid Ligands", and U.S. Pat. No. 5,270,163 (see also WO 91/19813) entitled "Nucleic Acid Ligands".

The process relies as a starting point upon a large library or pool of single stranded oligonucleotides comprising randomized sequences. The oligonucleotides can be modified or unmodified DNA, RNA, or DNA/RNA hybrids. In some examples, the pool comprises 100% random or partially random oligonucleotides. In other examples, the pool comprises random or partially random oligonucleotides containing at least one fixed and/or conserved sequence incorporated within randomized sequence. In other examples, the pool comprises random or

partially random oligonucleotides containing at least one fixed and/or conserved sequence at its 5' and/or 3' end which may comprise a sequence shared by all the molecules of the oligonucleotide pool. Fixed sequences are sequences such as hybridization sites for PCR primers, promoter sequences for RNA polymerases (e.g., T3, T4, T7, and SP6), restriction sites, or homopolymeric sequences, such as poly A or poly T tracts, catalytic cores, sites for selective binding to affinity columns, and other sequences to facilitate cloning and/or sequencing of an oligonucleotide of interest. Conserved sequences are sequences, other than the previously described fixed sequences, shared by a number of aptamers that bind to the same target.

The oligonucleotides of the pool preferably include a randomized sequence portion as well as fixed sequences necessary for efficient amplification. Typically the oligonucleotides of the starting pool contain fixed 5' and 3' terminal sequences which flank an internal region of 30-50 random nucleotides. The randomized nucleotides can be produced in a number of ways including chemical synthesis and size selection from randomly cleaved cellular nucleic acids. Sequence variation in test nucleic acids can also be introduced or increased by mutagenesis before or during the selection/amplification iterations.

The random sequence portion of the oligonucleotide can be of any length and can comprise ribonucleotides and/or deoxyribonucleotides and can include modified or non-natural nucleotides or nucleotide analogs. See, e.g., U.S. Pat. Nos. 5,958,691; 5,660,985; 5,958,691; 5,698,687; 5,817,635; 5,672,695, and PCT Publication WO 92/07065. Random oligonucleotides can be synthesized from phosphodiester-linked nucleotides using solid phase oligonucleotide synthesis techniques well known in the art. See, e.g., Froehler et al., Nucl. Acid Res. 14:5399-5467 (1986) and Froehler et al., Tet. Lett. 27:5575-5578 (1986). Random oligonucleotides can also be synthesized using solution phase methods such as triester synthesis methods. See, e.g., Sood et al., Nucl. Acid Res. 4:2557 (1977) and Hirose et al., Tet. Lett., 28:2449 (1978). Typical syntheses carried out on automated DNA synthesis equipment yield 10^{14} - 10^{16} individual molecules, a number sufficient for most processes. Sufficiently large regions of random sequence in the sequence design increases the likelihood that each synthesized molecule is likely to represent a unique sequence.

The starting library of oligonucleotides may be generated by automated chemical synthesis on a DNA synthesizer. To synthesize randomized sequences, mixtures of all four nucleotides are added at each nucleotide addition step during the synthesis process, allowing for random incorporation of nucleotides. As stated above, in one embodiment, random

oligonucleotides comprise entirely random sequences; however, in other embodiments, random oligonucleotides can comprise stretches of nonrandom or partially random sequences. Partially random sequences can be created by adding the four nucleotides in different molar ratios at each addition step.

The starting library of oligonucleotides may be either RNA or DNA. In those instances where an RNA library is to be used as the starting library it is typically generated by transcribing a DNA library in vitro using T7 RNA polymerase or modified T7 RNA polymerases and purified. The RNA or DNA library is then mixed with the target under conditions favorable for binding and subjected to step-wise iterations of binding, partitioning and amplification, using the same general selection scheme, to achieve virtually any desired criterion of binding affinity and selectivity. More specifically, starting with a mixture containing the starting pool of nucleic acids, the method includes steps of: (a) contacting the mixture with the target under conditions favorable for binding; (b) partitioning unbound nucleic acids from those nucleic acids which have bound specifically to target molecules; (c) dissociating the nucleic acid-target complexes; (d) amplifying the nucleic acids dissociated from the nucleic acid-target complexes to yield a ligand-enriched mixture of nucleic acids; and (e) reiterating the steps of binding, partitioning, dissociating and amplifying through as many cycles as desired to yield highly specific, high affinity nucleic acid ligands to the target molecule. In those instances where RNA aptamers are being selected, the method further comprises the steps of: (i) reverse transcribing the nucleic acids dissociated from the nucleic acid-target complexes before amplification in step (d); and (ii) transcribing the amplified nucleic acids from step (d) before restarting the process.

Within a nucleic acid mixture containing a large number of possible sequences and structures, there is a wide range of binding affinities for a given target. A nucleic acid mixture comprising, for example, a 20 nucleotide randomized segment can have 4^{20} candidate possibilities. Those which have the higher affinity constants for the target are most likely to bind to the target. After partitioning, dissociation and amplification, a second nucleic acid mixture is generated, enriched for the higher binding affinity candidates. Additional rounds of selection progressively favor the best ligands until the resulting nucleic acid mixture is predominantly composed of only one or a few sequences. These can then be cloned, sequenced and individually tested for binding affinity as pure ligands or aptamers.

Cycles of selection and amplification are repeated until a desired goal is achieved. In the most general case, selection/amplification is continued until no significant improvement

in binding strength is achieved on repetition of the cycle. The method is typically used to sample approximately 10^{14} different nucleic acid species but may be used to sample as many as about 10^{18} different nucleic acid species. Generally, nucleic acid aptamer molecules are selected in a 5 to 20 cycle procedure. In one embodiment, heterogeneity is introduced only in the initial selection stages and does not occur throughout the replicating process.

A variety of nucleic acid primary, secondary and tertiary structures are known to exist. The structures or motifs that have been shown most commonly to be involved in non-Watson-Crick type interactions are referred to as hairpin loops, symmetric and asymmetric bulges, pseudoknots and myriad combinations of the same. Almost all known cases of such motifs suggest that they can be formed in a nucleic acid sequence of no more than 30 nucleotides. For this reason, it is often preferred that procedures with contiguous randomized segments be initiated with nucleic acid sequences containing a randomized segment of between about 20 to about 50 nucleotides and in some embodiments of about 30 to about 40 nucleotides. In one example, the 5'-fixed:random:3'-fixed sequence comprises a random sequence of about 30 to about 50 nucleotides.

Variations on the foregoing aptamer selection procedures include, but are not limited to, those described in U.S. Pat. Nos. 5,707,796; 5,763,177; 5,567,588; 5,861,254; 5,496,938; 5,705,337; and 5,580,737.

The aptamers may incorporate containing modified nucleotides conferring improved characteristics on the ligand, such as improved in vivo stability or improved delivery characteristics. Examples of such modifications include chemical substitutions at the ribose and/or phosphate and/or base positions. Aptamers containing modified nucleotides are described, e.g., in U.S. Pat. No. 5,660,985, which describes oligonucleotides containing nucleotide derivatives chemically modified at the 2' position of ribose, 5 position of pyrimidines, and 8 position of purines, U.S. Pat. No. 5,756,703 which describes oligonucleotides containing various 2'-modified pyrimidines, and U.S. Pat. No. 5,580,737 which describes highly specific nucleic acid ligands containing one or more nucleotides modified with 2'-amino (2'-NH₂), 2'-fluoro (2'-F), and/or 2'-OMe (2'-OMe) substituents. Other modifications include, but are not limited to, 2'-position sugar modifications, 5-position pyrimidine modifications, 8-position purine modifications, modifications at exocyclic amines, substitution of 4-thiouridine, substitution of 5-bromo or 5-iodo-uracil; backbone modifications, phosphorothioate or alkyl phosphate modifications, methylations, and unusual

base-pairing combinations such as the isobases isocytidine and isoguanidine. Modifications can also include 3' and 5' modifications such as capping.

In one embodiment, oligonucleotides are provided in which the P(O)O group is replaced by P(O)S ("thioate"), P(S)S ("dithioate"), P(O)NR₂ ("amidate"), P(O)R, P(O)OR', CO or CH₂ ("formacetal") or 3'-amine (-NH-CH₂-CH₂-), wherein each R or R' is independently H or substituted or unsubstituted alkyl. Linkage groups can be attached to adjacent nucleotides through an --O--, --N--, or --S-- linkage. Not all linkages in the oligonucleotide are required to be identical. As used herein, the term phosphorothioate encompasses one or more non-bridging oxygen atoms in a phosphodiester bond replaced by one or more sulfur atoms.

In further embodiments, the oligonucleotides comprise modified sugar groups, for example, one or more of the hydroxyl groups is replaced with halogen, aliphatic groups, or functionalized as ethers or amines. In one embodiment, the 2'-position of the furanose residue is substituted by any of an OMe, O-alkyl, O-allyl, S-alkyl, S-allyl, or halo group. Methods of synthesis of 2'-modified sugars are described, e.g., in Sproat, et al., Nucl. Acid Res. 19:733-738 (1991); Cotten, et al., Nucl. Acid Res. 19:2629-2635 (1991); and Hobbs, et al., Biochemistry 12:5138-5145 (1973). Other modifications are known to one of ordinary skill in the art.

E. INSECTICIDAL COMPOSITIONS AND BAIT COMPOSITIONS.

Active agents as described above can be orally administered to an insect pest by any suitable means, such as by expressing the active agents in the plant through transgenic techniques (discussed below), or by applying the active agent to a plant (e.g., on said plant or immediately adjacent said plant, such as by applying the active agent to the soil or media in which the plant is grown).

The active agent can be applied to a plant as an insecticidal composition (e.g., a composition comprising the active agent plus insecticidal carriers and/or adjuvants), or as an insecticidal bait composition.

Insecticidal compositions including bait compositions are known and can be produced in accordance with known techniques. In general, an insecticidal composition comprises the active agent, a carrier (which may be simply water), and optionally a sticker. An insecticidal bait composition generally comprises the active agent and one or more of insect food and/or insect attractant (e.g., sugar or carbohydrate), carrier, etc.

Suitable carriers and adjuvants can be solid or liquid and are substances useful in formulation technology, e.g. natural or regenerated mineral substances, solvents, dispersants, wetting agents, tackifiers, thickeners, binders or fertilizers.

A preferred method of applying an active agent of the invention, or an agrochemical composition which contains at least one of said compounds, is foliar application. The frequency of application and the rate of application will depend on the risk of infestation by the corresponding pathogen. However, the active agents can also penetrate the plant through the roots via the soil (systemic action) by drenching the locus of the plant with a liquid formulation, or by applying the compounds in solid form to the soil, e.g. in granular form (soil application). In crops of water such as rice, such granulates can be applied to the flooded rice field. The active agents may also be applied to seeds (coating) by impregnating the seeds or tubers either with a liquid formulation of the fungicide or coating them with a solid formulation.

The term locus as used herein is intended to embrace the fields on which the treated crop plants are growing, or where the seeds of cultivated plants are sown, or the place where the seed will be placed into the soil. The term seed is intended to embrace plant propagating material such as cuttings, seedlings, seeds, and germinated or soaked seeds.

The active agents are used in unmodified form or, preferably, together with the adjuvants conventionally employed in the art of formulation. To this end they are conveniently formulated in known manner to emulsifiable concentrates, coatable pastes, directly sprayable or dilutable solutions, dilute emulsions, wettable powders, soluble powders, dusts, granulates, and also encapsulations e.g. in polymeric substances. As with the type of the compositions, the methods of application, such as spraying, atomizing, dusting, scattering, coating or pouring, are chosen in accordance with the intended objectives and the prevailing circumstances.

Typical rates of application are normally from 5 g to 2 kg of active ingredient (a.i.) per hectare (ha), preferably from 10 g to 1 kg a.i./ha, most preferably from 20 g to 600 g a.i./ha. When used as seed drenching agent, convenient dosages are from 10 mg to 1 g of active substance per kg of seeds.

The formulation, i.e. the compositions containing the active agent and, if desired, a solid or liquid adjuvant, are prepared in known manner, typically by intimately mixing and/or grinding the compound with extenders, e.g. solvents, solid carriers and, optionally, surface active agents (surfactants).

Suitable carriers and adjuvants may be solid or liquid and correspond to the substances ordinarily employed in formulation technology, such as, e.g. natural or regenerated mineral substances, solvents, dispersants, wetting agents, tackifiers, thickeners binding agents or fertilizers. Such carriers are for example described in WO 97/33890.

Further surfactants customarily employed in the art of formulation are known to the expert or can be found in the relevant literature.

The agrochemical formulations will usually contain from 0.1 to 99% by weight, preferably from 0.1 to 95% by weight, of the active agent, 99.9 to 1% by weight, preferably 99.8 to 5% by weight, of a solid or liquid adjuvant, and from 0 to 25% by weight, preferably from 0.1 to 25% by weight, of a surfactant.

Whereas it is preferred to formulate commercial products as concentrates, the end user will normally use dilute formulations.

The compositions may also contain further adjuvants such as stabilizers, antifoams, viscosity regulators, binders or tackifiers as well as fertilizers, micronutrient donors or other formulations for obtaining special effects.

F. TRANSGENIC PLANTS.

Transgenic plants are produced in accordance with known techniques, or variations thereof that will be apparent to those skilled in the art. *See, e.g.*, US Patent No. 5,750,386 to Conkling et al.; *see also* US Patent No. 7,476,781 to Abad et al.

Vectors which may be used to transform plant tissue with DNA constructs of the present invention include, but are not limited to, both *Agrobacterium* vectors and ballistic vectors, as well as vectors suitable for DNA-mediated transformation.

Agrobacterium tumefaciens cells containing a DNA construct of the present invention, wherein the DNA construct comprises a Ti plasmid, are useful in methods of making transformed plants. Plant cells are infected with an *Agrobacterium tumefaciens* as described above to produce a transformed plant cell, and then a plant is regenerated from the transformed plant cell. Numerous *Agrobacterium* vector systems useful in carrying out the present invention are known. For example, U.S. Pat. No. 4,459,355 discloses a method for transforming susceptible plants, including dicots, with an *Agrobacterium* strain containing the Ti plasmid. The transformation of woody plants with an *Agrobacterium* vector is disclosed in U.S. Pat. No. 4,795,855. Further, U.S. Pat. No. 4,940,838 to Schilperoort et al. discloses a binary *Agrobacterium* vector (i.e., one in which the *Agrobacterium* contains one plasmid

having the vir region of a Ti plasmid but no T region, and a second plasmid having a T region but no vir region) useful in carrying out the present invention.

Microparticles carrying a DNA construct of the present invention, which microparticle is suitable for the ballistic transformation of a plant cell, are also useful for making transformed plants of the present invention. The microparticle is propelled into a plant cell to produce a transformed plant cell, and a plant is regenerated from the transformed plant cell. Any suitable ballistic cell transformation methodology and apparatus can be used in practicing the present invention. Exemplary apparatus and procedures are disclosed in Sanford and Wolf, U.S. Pat. No. 4,945,050, and in Agracetus European Patent Application Publication No. 0,270,356, titled Pollen-mediated Plant Transformation. When using ballistic transformation procedures, the expression cassette may be incorporated into a plasmid capable of replicating in the cell to be transformed. Examples of microparticles suitable for use in such systems include 1 to 5 μm gold spheres. The DNA construct may be deposited on the microparticle by any suitable technique, such as by precipitation.

Plant species may be transformed with the DNA construct of the present invention by the DNA-mediated transformation of plant cell protoplasts and subsequent regeneration of the plant from the transformed protoplasts in accordance with procedures well known in the art.

Any plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a vector of the present invention. The term "organogenesis," as used herein, means a process by which shoots and roots are developed sequentially from meristematic centers; the term "embryogenesis," as used herein, means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristems, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem).

Plants of the present invention may take a variety of forms. The plants may be chimeras of transformed cells and non-transformed cells; the plants may be clonal transformants (e.g., all cells transformed to contain the expression cassette); the plants may comprise grafts of transformed and untransformed tissues (e.g., a transformed root stock

grafted to an untransformed scion in citrus species). The transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, first generation (or T1) transformed plants may be selfed to give homozygous second generation (or T2) transformed plants, and the T2 plants further propagated through classical breeding techniques. A dominant selectable marker (such as npt II) can be associated with the expression cassette to assist in breeding.

Plants which may be employed in practicing the present invention include (but are not limited to) cotton (*Gossypium hirsutum*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), soybean (*Glycine max*), peanuts (*Arachis hypogaea*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), Avocado (*Persea americana*), Fig (*Ficus casica*), Guava (*Psidium guajava*), Mango (*Mangifera indica*), Olive (*Olea europaea*), papaya (*Carica papaya*), Cashew (*Anacardium occidentale*), Macadamia (*Macadamia integrifolia*), Almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), corn (*Zea mays*), wheat, oats, rye, barley, rice, vegetables, ornamentals, and conifers. Vegetables include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuea sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.) and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherima*), and chrysanthemum. Conifers which may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*).

The present invention is explained in greater detail in the following non-limiting Examples.

EXAMPLES

Objective 1: Sequence and characterize the expressed genes from the head, salivary gland and the digestive system of the tarnished plant bug and the digestive system of the green sink bug from different developmental stages. Transcriptomes are built in accordance with known techniques. We published the ontogeny for an EST library for the digestive system of the cabbage looper in 2003, where plasmid isolation was performed using a Qiagen BioRobot 9600 followed by sequencing on an ABI PRISM 370 DNA analyzer in our Genome Research Laboratory (GRL). In this project, 957 ESTs were sequenced where 12% were digestive proteases, 9% ribosomal proteins, 7% matched previously sequenced genes of unknown function and 29% were novel proteins (**Figure 1**).

Over 29% of the ESTs matched lepidopteran sequences and 21% noctuid genera. More recently, using Roche/454 GS FLX sequencing, we were able to sequence in a week or less starting with mRNA, 114 million bp from the tick brain and associated endocrine tissues which represented 530,000 reads assembled into over 21,000 contigs (**Table A**).

Table A. Roche/454 GS FLX sequencing metrics from <i>Dermacentor variabilis</i> synganglion cDNA library.	
Total reads	532,136
Bp sequenced	113,636,431
Bp after primer trimming	98,567,133
Assembled contigs	21,112
Putatively identified transcripts	10,674
Novel transcripts	10,445
Mean contig length (bp)	347
Mean number reads/contig	11.8
Large contigs (>500 bp)	3216
Mean large contig length (bp)	989
Largest Contig (bp)	7731

The goal of Objective 1 is to apply the same effort to build transcriptomes for the plant bug head, salivary gland and digestive system, with a comparative analysis of the digestive system of the stink bug.

Methods. Objective 1 is achieved by separate 454 sequencing of at least 500,000 cDNAs synthesized from expressed genes in the head, salivary gland and digestive system of the tarnished plant bug, *Lygus lineolaris*, and the digestive system of the green stink bug, *Acrosternum hilare*, where the cDNA will be synthesized from feeding male and female last instars and adults. All insects for these studies are collected directly from the field in North

Carolina, so gene expression is most relevant to natural conditions. Cotton is scouted in June for plant bugs. Once they appear in good numbers, the grassy areas closest to the field edge are swept for feeding stage, last instars and adults of both sexes. For green stink bugs, adults are collected in July by sweeping cotton or by placing white cloths on the ground and shaking the plants. Feeding will be determined for both species by the presence of food in the digestive system. The insect collections are made in the early morning, maintained with water available *ad libitum* for transport to the laboratory, and dissected in the afternoon. Any residual food in the digestive tract is washed from the opened gut with physiological saline to minimize plant contamination. The cDNA library is constructed from the entire digestive system including the malpighian tubules and rectum but not that part of the digestive tract found in the head. The head does not include the salivary glands but contains part of the salivary ducts and the most anterior portion of the digestive tract found in the head. cDNA library construction is conducted from equal amounts of tissue from both feeding stage immature and adult insects of both sexes.

Ultra high throughput pyrosequencing. As a sequencing platform, the Roche/454 GS FLX has no requirement for reference genomes for transcriptomics and is therefore well suited for Objective 1. A specific protocol for the preparation of cDNA suitable for sequencing on the 454 GS FLX sequencer (posted by the NCSU Genomic Sciences Laboratory, <http://gsl.cals.ncsu.edu/media/cDNA.pdf>) has recently been developed, has been successfully used to construct two transcriptomes, and is used in the carrying out Objective 1.

In the current project total RNA is purified separately from each tissue using TriReagent (Sigma), and messenger RNA purified using an Oligotex mRNA isolation kit (Qiagen). Modified primers (Beldade et al. 2006) are used for first-strand synthesis and second strand synthesis/PCR amplification to limit homopolymer stretches from the polyA tail and to produce sufficient quantities of cDNA for sequencing (3-5 µg). Aliquots of the purified mRNA and resulting cDNA are stored at -80°C for further use in Objective 3. An alternative to this PCR based amplification is the Eberwine method of linear mRNA amplification followed by RNA amplification to produce cDNA. However, in our previous 454 sequencing efforts, the PCR approach produced excellent results where rare hormonal and other transcripts were identified and large full length contigs (>5 kb) could be constructed as determined by comparison to previously sequenced tick genes.

The cDNA synthesized by PCR is purified and adjusted to 50 ng/µl which is a suitable starting concentration for library preparation using the Roche/454 GS FLX library

preparation protocols. Sequencing is conducted in the NCSU GRL. Contiguous sequences are assembled using GS Assembler ver. 1.1.03 (Roche) initially with default parameters. Custom PHP and PERL scripts are written to manipulate the resulting sequences, which are compared to the non-redundant and EST GenBank databases using DeCypher accelerated blast algorithms (Time Logic). To further aid in the characterization of the resulting genes, a PHP script is written that translates the resulting contigs into all six reading frames which can be searched for conserved domain architectures and signal sequences.

Data analysis and broad impact of proposed work. Objective 1 provides a comparative transcriptome for the digestive system of two agriculturally-important, emerging pest species and that for the head and salivary gland of the plant bug. Data is given in **TABLE A** herein. These data provide a number of new opportunities for future studies of digestion, the regulation of salivation and digestion, novel insecticide targets, and neuroendocrinology for an insect group where minimal sequence data are available.

As noted above, an aspect of the invention is a transgenic plant that contains a polynucleotide construct that comprises a heterologous nucleotide sequence operably linked to a promoter that drives expression in the plant, the heterologous nucleotide sequence encoding an active agent toxic to an insect pest. In some embodiments, the encoded active agent is an RNA (*e.g.*, a dsRNA) that inhibits expression of a target gene in said insect pest, and reduces or inhibits the expression of the expression product (*e.g.*, protein, ribosomal RNA) of that gene. Such plants are produced by providing a construct in accordance with known techniques, and producing transgenic plants with such a construct in accordance with known techniques. The promoter can be any suitable promoter as described above (including inducible promoters, viral promoters, synthetic promoters, constitutive promoters, temporally regulated promoters, spatially regulated promoters, and/or spatiotemporally regulated promoter, etc.). Since, for insects, a dsRNA is in some embodiments from 200 to 300 nucleotides in length (though shorter RNAs of at least 21 or 24 nucleotides in length; or longer RNAs up to 500 or 1000 nucleotides in length, or more, may also be used), the heterologous nucleotide sequence in the construct may be any of the nucleotide sequences present in **Table 1 (SEQ ID NO: 1 through SEQ ID NO: 133)** herein, where sequences are given 5' to 3'), or an active fragment thereof. Examples of active fragments thereof include continuous fragments of the disclosed sequences of at least 21 or 24 contiguous nucleotides in length. Such fragments can be selected in accordance with design guidelines known in the art. For example: In some embodiments the fragments are those that begin with two

contiguous "A" residues; in some embodiments the fragments are those that do not contain 4 or more contiguous "T" or "A" residues; in some embodiments the fragments are those that contain 30 to 50 percent GC content (as opposed to higher G/C content); and/or in some embodiments the fragments are those that do not contain more than 16-17 contiguous base pairs of homology to other coding sequences, particularly other coding sequences in the host plant, as determined by homology searching such as by BLAST searching). *See, e.g.,* Elbashir et al., *EMBO J.* **20**, 6877-6888; *see also* Applied Biosystems Technical Bulletin # 506, siRNA Design Guidelines. Minor changes can be included in the sequences or fragments thereof (*e.g.*, 1, 2 or 3 nucleotide substitutions) in accordance with known techniques. Where a dsRNA is desired, the complementary strand can be provided in accordance with known techniques, such as by an additional self-hybridizing segment on the same nucleic acid, by transformation of the plant with an additional construct containing a nucleic acid encoding the the complementary strand driven by a promoter, etc.

A further aspect of the invention is a dsRNA compound or product encoded by a nucleotide sequence of **Table 1 (SEQ ID NO: 1 through SEQ ID NO: 133)** as described above, or by an active fragment thereof as described above. Such products can be produced in accordance with known techniques, including but not limited to those described above, and are useful as an active ingredient in an insecticidal bait composition as described above.

TABLE 1: NUCLEIC ACIDS ENCODING TARGET INSECT PROTEINS

Transcription factor:

contig00163 length=1989 numreads=341

CGCACAAACCGTGTGCACAGAGTTCTTTATAAAAAATGGGTGCCATCATGGCGTATT
TACTTACTTCAGTCGCACAGTCGACACAAAATATAATAAAGAAGAATGGATATG
GAGTCGTCGGTCAGCCTCGATTTGAAGGTGGAGCCGCGAGCCCCCTTATGGGAG
CTTGAGCTGCTTGATGAGGTCGGGGAGATCTGGGAAGGCGATGTCACGGTCATC
GAGGGGGAACTGGATGTGGACCTGGTCGAGGATGTGGACACCGACGCCTTCTGG
GGTCAGCAGGACATCTCCGAGCTCATTGGCAAGGAAGAGCAGATACTTGGACTG
GAACTTGACGAAATCACCTTGGAGGATCAGGACGTGAAGGAAGAACCTTTCACT
AGCTGGTTGGACAATGGCGCAATCTACGATGGACCGAATTCCACCGTTTCAGGTC
TTCCAACCTTCAACACGCCATCTTTGTGGTCAGCACCAGAGTCGCCCTCCGAAGA
AAAGGCCGCGAGCGTTACCGGTGCCTGCGGGTGGGCCTGCAGCGCCGGACCCGCT
GCAAGGAGGGACTGCGGACCTGTTCACGGAGTTCGAGGACGTGCTGAGCCAAGC
AGCGGGTTCCTCACGCCGCCCCGACTCTCCCCGCTCTGCATCAGCATTCGGGGCG
CAGGCGCTGCTTCAGCTCGCCCCCACTTGCAGCCGGTGCAGCTCTCGCCCCGCTT
CTGGCCAGAGATTCATCGTGTTCAGGAAATGGCGAACTCGCCTGGCGTCAACTA
TGAGTTGGATGAATTGATGGCGGAGACCTACCAGAAGAACGGAGGTATCTCCGT
CAGCAGTAGCGGCAACAGCAGTCCCATCAGCAGCAGCAGCAGCGGTACCAAAAC
CAAAGGACGTCCTTGGGACTCACCATCTGCCTCAAGTTCTGGCTTCAGTGATAGC
GGCTACGACGATCCTGAGTGGAGCTCCACAAAATCTACGGAATCCCGTGACTCTT

CTGATCCGTCGCCAGCGAAAAAAGGTGTTTGAACCTCGACTGAAGAGCGTCGAT
TGAGGAAGAAGGAGCAGAACAAGAACGCTGCTACCCGCTACCGCATGAAGAAG
AAGATGGAAATGGAAGAAATCATCGGCGAAGAGAAGGAGCTGATGGACCGCCA
CGACACCCTCAAGTCCGAAGTCGTCGAGATCACCCGAGAAATCAGATACCTCAA
GGGCCTCATGGGAGACTTATTCAGAGCCAAGGGCCTCATTGCGGCCAAGTAATTT
CCAAATCAACTCGCTCTCCCCAACCAAAAAATCAACCTGTTCTTCTAGTAGTG
AAAATTCAATTAGTGGATGTACTAGTAGACCATAATTTTCATGATATCAATAACCT
AGTTATTATAACCTAGATTTCGTGTATTTTTTCGGGGTTGTTTTTCTTCAATGAACT
CCTCATTTCTCGTTAAGTATCGTGAAATGTTTTCAATCGCTTGAAAAGTGAATTT
TTGCAGTGACATCTTCCTTGTGTGTTTTGAGGCCCTTGATCTCTTCTTCTCAAGTT
AAATCCCTTTTTTGTACTCGTCCAAGCTTGAATTTTCATTCCCCTCGCCTTTCCATG
GCCTCCCCTACTTCTCGCCATCTTGTGTACGGGTTTTTCAACCCTCTATGACGCAT
TGTATTCCAAATATGCTTTTTTGGGTTCGGTTCGCCCCCGAAAAAGTTGGAATCAG
CCGATGTTTTTGGTTGATAATCCTTACATTTAGAAGGCTCAAGCATTTCAGTTTC
CATGATCTATTTTATTTTTTAATATAACCGGTCAATCCCAATCAAATTTTTTATTG
ATCAACTTTTCTTTTATGAACGAGGGTCATCCTCGTTTTCCATCTGTGTCAAATT
TGCCTAAATTTTCACGTTGCACTTTGGTAATCAGGATTGTACCGGAACCTCACTTT
TATTGGTTATACAATATTTAGTTATGTAATAATGTAACTTTTTTTAAAATAAATG
TGGATCTAACTCCC (SEQ ID NO: 1)

contig06596 length=697 numreads=52

AGAGAGAGCTCGGCTTCTTCAAATTCGTCTGACTCTTCAGAACTTCAAGAAGCA
GTTCCGGTCTCCAGCCCAGGATCAAGATCTAGAGCTGAGCAACCTCCTCGTAAAAA
TGGAGGTCCTGCGAAAAAAGAGAAAAAGTCAGCCCCCTCCTCAACTTGCCTCGGA
CAGAGACTCGGATATTGCTCCTACCAGGAAGAGCACAAGGTCCTCGAGTACTAG
ACGCAGCAAACACGTTGTAGCCAGGCTTGCAAGCGACTCAGACTCTGATACAGG
TAGTAAATTGAGCTCTACAAGCAAGTCCCCTGCCAAAAAAGTTACTCCCAAAGG
ACCCAACGTGAGAAAACCAAAGAAGTCTCACATCTCAGAGACAAGCCATGTAAT
TGAGAATGAAGAAAGAACATGTCCGTTTGACGGTTGTGACTCTTCTGGTCATTTA
GGTGGTAGGCTTGAAAAGCATTTCACTTTGGAGGCCTGCCCTGCTTACCACAATA
CAACACCTCAACAATGCAAGGAAGCATAATTGAACATTTAAAGAGAGATGAGG
AACGGAAAAAAGCCATTGGGAACCTCAAACCGGGACAGGCAACTACACCAACTC
CTGAACAAAAGGAGTATCGGGATAAAGTTAAGGAAATCAGGATCAAATTTGTGC
CTAAGGATGATGCTCAATCGGAACCAGACAATGGAAAAGAACGAC (SEQ ID NO:
2)

contig05742 length=783 numreads=38

CGGCCGGGAGGAGTTTCGAAAAGGCTGCCTGTGGGCGATGAACCCTACGAAAATA
TCCAAGATGGACGACGAAGTCGCCAAATGGTCGAGGAAAGACCCTTTGGCTATT
CGGAGAGCAATGAGATATCCAGATGATCTCGAAAAGCTGGAAAGGGGAGAAAT
GAAATTCGGCTCCATATCGCACGCTGTTGAAGAGGAGAGCGGAGATGATGAAGT
TCCCGAAGAAGAAGAAAGCGACGAGAACGGTGAAGCTGCTAAACGGATTAACA
TGTTTCGAACTTGCTCACATACTTCGGGACGAAGAGAGTACCAGAGCGTTCCTGTA
CGACTATAAACTGCTGCCGCAGCAGAAGATATGCCCCAGGTGCAATTCGAAAT
GAAGCTGACTTTCGTGAATAATTTTTACGGATGGTTCAAGTGCAATAATGGCCAC
AGGTATAACGCTGTCAAAGGAACGTGGTTCGAAGGTACGCATTTGGGTGTGCGA
GAGATCTTATTGTCGACGTACTGCTTCCTTCAGAAAGTTCTCGATGGATCAAACGA
TGGCAGAAGTGACGGCGGGGAAGGACGTGTTGGAGAGGGAGGTTATAGTGAATT
GGTGGTATAACACGAGAGAGGTCTGCATAGACGCCATCGAGCGGGAGTACGACG
AACTAGCCAGCCTAGGGCCGATAGGAGGCGTCGGTTGTACGGTCATCGTCGACA

AAATGAAGATCGGCGAGTTGAACAATTCCACAGGAGGATCAAAGACAACGTCT
TCATTTTGGTGATGCTCGAAAAAGGG (SEQ ID NO: 3)

contig01506 length=333 numreads=4

GGAGACGACATCGAGCGGAAACAGCAGGAAAACAGGAGGAGCTCGCACCAGGC
GTTGATGAGGAGGGAGGTTTGGTCGTCGATCGAAATGTTTTGAAGAGTGGGAG
CGACTTACACCACTTGACTATCTTGTAAGTCTATGGTCGGCAATGTTACATAAA
TTGGCCATGAGATCCGTAGGCTGGGGTGCAGAGGGTCTTGGTATCCTGCCTACGT
ACTTGGGGTCGGCTTCGTTGTACTGCCAGAGGTGTTCAACCTCATAATTTCTCTGC
AGAAGTTGAGGGACGAGGTTGGTTTGTGGAGCGGATTGCTGCTGCTGTTGAAGCT
CCGCG (SEQ ID NO: 4)

contig01035 length=736 numreads=15

ACTGCCGAACATGTTGGTGCCACCACCAGGCGACCCCATGGACGACGCTGGGGG
TGGGTTGATGCGCTTTTCCTTCTGCCTCCTATTACAGAACCAGACTCGAACGACTT
CCTTCTCCATCGTCAGACTGTCCGCTAAAATTGAGATCTCTTCAGACGTGGGTTTC
GGGTTCTGCAGGAACGCCTTCTCCAGGCTATCCGGACGCTCGTTTCTATGGACGT
CCGCTTCTTCTCCGCTCCCGATGGCCTCCGGGGTCGTCATCGGGTTCGTCAGA
GAGTTCGGGTTGGTCAGCGAGTTGTCCGCGTCCAACAGCCATTTTTGGAGGAGCG
GCTTCAGCTTGCACATGTTCTTGAAAGAGAGGTTTCAGCGCTTCGAACCTCGAGAT
GGTGGTTTGAGAGAAAGTCGTTTCCGTACAGTTTGCCCATCGCTAGGCCGACGTCA
CCCTGCGTGAAACCTAATTTGATTTCGTCGCTGCTTGAAAGTTTTGGCGAACTGTT
AAGCTCTTCGAGATCTGTCGTTTCATCGGGCTCCATAGAGTGAGGTTGGGGCATT
GGGGTTCTATGTATTTTGGCAACGGGGCTACCCTTGTTATCGCCTAAAATACCTC
CCTGTACTGAATGAGAAAGTCTAGGTGGGGGAACGTTCTGTCGTAGGCTAGGTGT
GAGGAGACCCTGATTGTGGAGGAAATACTGTTGAGGGTGATGGTGTTGGGGTGT
GGATGTGGGAGGTGGAGGAG (SEQ ID NO: 5)

contig05994 length=1316 numreads=44

GTGTAAGAAATGCAGAATGTATACACAGAATGGGGAAAAGGGAGGAAATTATG
GTAAGTTTCGTTTCATGTTACCACCACCTTCAAGTGCTAAAGTTATTGCGGTCGGAA
TGTTACGAATATTATTCAACTGAAGTAGATTGCTACAAAATGGGGGACGAATTG
ATGATTTTTACTTTCAACACATTCATAAGAAGAATAATGACCAGATGAGTACCT
ATAATAAAATGAAAATAGATAGGTAATAAGTATGTGCGGTTTAAACAATAGCAC
ATACCTACCTACTACCTAAGTATTATCCATAAGTACTCCCATCAGAAATTATCAA
TGATACGATGATTTTTACAAAATATTGCAACCATGTATCCAATTACCAATCCAC
TTTCAATTCCTTACCATGTCAATACAGCATTGTGGGAAACTTGGAATCTGACTCC
ATATTATGCTCTATTTAAGAGCATTCCTTGTTTACACCTCCCAAGCGAGCTATG
ACGTTTTGTAATCTATTCACTTGGGATGTTAGACTACAATTAGCATCTTCAAGTGA
GCAAATTTGCTGTTTGAAATCTGCATTCTCGGTAGAAAGATTTGACTTTTCTCTC
AAGCGCATCCATGTATTCCTTCTTTTTCTCCGTGATTCTTGAGCAGAGATCTTAT
TCTTGATTTTTCTGCGAATTTTCTTCAGACATTTTCTTCGGCTTTTGTGAGAGGTA
GTCTCACTGGTATAGGGTACCCTTCTGCCACCAACGTTCTCTTTCTCTTCGGTG
AGCAATAACATCCCCGTGGATCCTTTTGGCTGAGAACTAATGAGAGGAGTTTGTA
TGGGCTGCCGACTTGTGGTGGAACAGAGAGGAACAGACTCCGATCCGATGATG
ATGGATGAAGTGGTGACACATTTCTTCATTATCAGAAGATGCGCAAGAAGGAG
GTGTTGGTGGTAACGGAAAGCGTGTGCTGGAAGTTGATGATGGTTCAATTTTTTA
TAATATTAGCTGAGATCATCTGGTGCTTCCTCTTCATGCCAAGACTTTTATTGCTC
AACATAGTGGTAGTTGGGAGTTGCTTGACCGGAGTTTGTGTCATAGGCTTGGAAT
TTAGAAACATATCGGTGTAGATTATCGAGGCTTTGACAGATCTAGGAGGCGGTG

GGGATGGAGAAGATGGACATGATGATGCTGGAGAGCTGGCGCCGCTCATTGGCT
CGGACTTGATGACCACGAATGGGATAGGAGAAGACACACGACCACTGGTGGTTT
TAGGAGGTACTGCTGGAAGTAGACAATCGTCTTCTACGTCATCTTTGATGA (SEQ
ID NO: 6)

contig05201 length=274 numreads=2

TGGTGTCGTTTGTAAACGACACCATTCTAGCGACTGGAGTATTTTCCTCTAAGGAG
CTCTGGAGACTGTCTAAAGCATCTATTGTACAAAACCTGCAACGGTATGCAGCCCC
CTATAAATGTAATTAAAGTAGAAAACCCCGGCTACTATTGCAGCAGCTATACACC
CCCAGGTCAAGACCACACCACCATGATTCCTTCGGCTCAATATAGCCCTTTAGGA
CTCCCCACGAAGTGGTCAAAGCCCTGGCGAACC GCGCTCCAAGAGGGTCCGTCG
(SEQ ID NO: 7)

contig03255 length=595 numreads=12

CCCGGTTACGCACTATAGTGGGCCAAAAGCCAGGGGCGGTGGCACAGGTTGTAA
CCTCTTCTGGAAGTGTGATTCAAGGGACCTCGATGTTGCGAACTGTTTCTCTGAC
AGGAAAACCAAGCACGCCAGTACGTCACCATTCCCTCCGAATCAGATGCCCAAGAT
TCAAGAAAGGTCCGCTACTGTTATTGCATCTCCATCTCCTAGGAATGTGGCTACT
AGCTCCCGGCCGATGACATCAGCGGGTGACAGATCACTGAAAAGAGAGATGAGC
ATTGATTTTCGTGCCCGATGGGCCTAGCTCAAACAAGCGGAGGAAAACCTGAAAAA
GGAGGAAAAGGCTTGAGGCACTTCTCCATGAAAGTTTGTGAAAAGGTCAAGAAG
AAGGGACTCACCTCTTACAATGAAGTGGCGGATGAACTCGTGGCGGGAGTTCAC
TGATCCCAAACACATTCTTAACAGCGTTGATCAACAATATGACCAGAAGAACAT
CAGACGTCGAGTCTACGATGCGCTGAACGTGCTCATGGCGATGAATATTATATCT
AAAGAGAAGAAAGAAATCCGATGGTTGGGGCTTCCCATGACCACTGCTCAGG
(SEQ ID NO: 8)

contig02985 length=278 numreads=8

CGCTTGGGTGGGCACATCCGAGGATTCGAAAATAAGCGAGGTGTTTCATGTTGCTG
CCATTACAAAGCAAGGACTTTTCCATCAGGAGTCTTGCTGATCGGATATCTGACT
TACCCACCTCCAGTACCTCTACCCCGACATCTGTAAGACTCAAGCTTTCAGCAA
GTACTACACCCCCTTTAGTGATAATCAGACGACCACGAATAACGGTTACGTGAAA
CCAATGCTGGTGACTCAGATACCACGATTAGGAGGAGGCAACAGCCTCCCTCCA
ACGC (SEQ ID NO: 9)

contig06813 length=444 numreads=268

GTATCATACCTGGTGAAGTGACAGACGCCGGCAGTCATGATGGCAGGATTATAC
TCGGCCTACGACCACGACGCCAACTACGACGCGGCTGGCTCCTTCCTCAAGACT
ATCACACGGACGAGGGATATGGGTTGTTGGACTCGATCAAGATGCCACCTCAGG
GAAGGTCCGGGTTCACATCGTCGACATACTTGATCTCAACGACGCCGCCAAAAC
ACAACACACGGATGGCGCCGCTCTCAACCCCAACCAAGCTCACGACGATATTTCA
GGAGGTGAAATGACAAACCTCCAGCAGCTTGGGAACCACCCCGGCTGCTGTTC
GCCGGTGGTCAAGTCACCACTCCCATGGAAGCCCTTCACCATCATTGGGGCAACT
CCATCACCGAACTTAGAGGTCACCAGCAAGCCAGCCCCGACAGCACATCACCCG
GACCGTAA (SEQ ID NO: 10)

contig05957 length=899 numreads=24

GCATTTCTGCTTGCACTTCTGACCGCAGCAATGAGGTGTCACTCCAGGAGTGGGT
TTAGGTTTTTTGTAGACGTCCGCAGAGACACTTATATCGCAGTTTATTGGGAGTGC
AGAATTTTGGTCTTCTGCAGACTGGACAAATCCACGCTCCTCCCATTTCAGTACTT

CGATTTGTTTTAGCCCATCGCCTGATGCAGTGCAGATGGAAAATGTTGAAGCAGT
TGTTGCAAGGCCAACTCTCTCTTCGGCGAGGATCAGGTCAGTGCAGATGACGCA
CTCGGCAGTGTTTGAGTTGAGTTCATCGATTAAGTTAGCTCTTTGACAATCGACTG
ATTCAGGAAGTTGGCGGCGAGCTGGTTCATATTGAATCTCGGTGGAGTAAATCC
TCTGAAACCATCGCCTTGCCCTGCAGTTGGATTTGCCCGGAGTCTGGTCCGGAAG
TTTTGGTGCATGTTGGAGAACGCATCGGCCAGCTCTGCCATGCTGCTGAGCAGCT
GCACAGAGATGTTGTTGTGGAGGAATATGATGGGGTTGCCGACGTTTCTCATACT
CGTGATGATGGCTCTCTCTAAAGTAGGTTGGTCGGAATCGTCGTCATCATCAGCC
GCTTCTTCTTCTTCTTCTCGTGGGCATTTCGGTTTCTGCGGATGACAGAGATGGTG
TTGGTGAAGGATGGAGGGCATTAAGTGATAACGACACAAGCGAATGAGGTGAAG
CAGCTGTGGAGGCAGATTTCGATGCTCTCCCCTGAGTCAGACTCAACGCGGTCAGT
GTCCGAAAAGAACACGTCAGACATTTACGCGTCAGAAATGCTGCTGACGAACGT
CGTGGCGTCGCTGTCACCTGAAAGAGAGAACGTGGGGTCGTTGGATTCACTGTCT
TCATCTTCGTCCGTTCCA (SEQ ID NO: 11)

contig02526 length=277 numreads=2

ACAACAAGTGGGCCACCAAGGTCCCATCGCCGTTGGGCTCCAAGGCCTTCCCTTG
GACCGTCAACCTCTGACGACGTCGCTCAACCACCACCAGGGCGGATGCTTCAAC
CCGACGGCCGGCCAGACGATGCCGGTGTGCTCTCCGGTGACGTCCCCCGTCACCT
CGCCCTGCTACGGTGGGGCCCCGGTACATGTACCACCCCCGCCCTCCCGACGCCG
TCCTGCATGAGCTCATCCCTCGCGTCTTTGAGGCTCAAGGCGAAGCAGCACTGTC
CG (SEQ ID NO: 12)

contig02094 length=249 numreads=5

CCGTGTCCACGTCACAGTCATTTGACGAGGCGGCCGGCTTTGTTTTTGCCATTTTT
TCAATATATTTTTCCATACATTTTGGGACCAGCTGTTTCACTTTCTCCAGATTTGC
TTTTGTGATACTGCCAGTGTCAGACCGTGAGTCATGCCGATTTGCTTGAGTACTT
CAAAATCGTCTTTGATCTCAAATTTATCATTGAATACAAACAGGTACTCTGTCTTG
TCATTGGTGATAGAGCAATCAATCA (SEQ ID NO: 13)

contig00531 length=903 numreads=54

ACAGCAATCTGGTTTTCCAAAGAAGCAATCTGAATTTGTTCCCTGGTCGATTGTT
CCACAAACTGTTTTGTGATGCGTCTGTGTTTGTTCGATTTTCGATAAATAGTTT
TCAACGCTTAATGCTGGCAGTGGGGTGCTGGAGACAACTGCCTTATGGTGGCCA
TTGAAGGGGTTTGGTTGATGGATAGATCAGGAGTCACTGTTTCAGTAGGAGGACT
TGGTACTTCCCTCTTCAATTCTATGTGTTCTTCCAAGATCGTGCCCATTTGTCTGGG
ATGGTCGCCTCACAGGAGAGAGTACAGCTGGTGGTGGAGGCATCAACATTACAG
GGTGCACATTGGAGTCATCAGTACATTTGCTGTTGTCTGGCTGAATGTATGACGC
AGTCGGAAGGGAAGAGTGACTCCCCAGCTGGATGAGAGGTATACCAAGAAGAA
GATTCCTCGGTGGGAAGGATACAGGGGGTTGGTTCACGCATTGAGCAAGAGCTG
CTCTGGCTGCTACTAAGGGTTCTCGTGGGAGAGCTTGGAGAGGTGGTGTACCCAG
AGGGGTATAGAGAAAATTGTGAGCTTCACTACTTCGCCCACTAGACATTACATTC
AACTTGACTACTACCGGCTCTTTGATATCCTCAACCTTGTAAGGAGGAATCGTAC
ACACTAAGTGTGACTGTTGTAAAACTCCTTATCAGGCTGTACAACCTTCTTCCCA
ATAATTCTCTGCATTCATATCCCCTTCAGAGAATATCACTCTGGTATCTTTCAAAA
AGTTTTTACCGAGTATGAACAGTTCAAGGCCTCCAGCGGAGTGACATGAACTAC
AGACTTCTTACATATTTACAGGAAGTCCAGGAGGTTGTGTGCAAATTATAGGTTGA
GAAGTCACTTGTAATGTTTCCAGG (SEQ ID NO: 14)

contig04146 length=312 numreads=4

GCGTTCTCATCTTCAGCTTGGCTTCTTGGAGTGCGAAGTCTCGAACTTGCTCCAGA
TGGTTACGCAAATTGAGAACATACTCGGTTCGATGGTCTTAGGCGTTGTTGAGGCG
GTCTTACTAGGGCTTGAGGCAGTCTCATATCCCGGCCATAGATTAGCTGTGCTGG
GGATAATCCTGTTGTGGCGTTTGGAGAAGATCTATAGGCCAACAGGAAAAGAGG
AATTTTCTCGTCCCAATCAGTTTGGCTCTTGTGACAAACATTGATATGTGTTTAG
TCAAGGTTTCGGTTCAGCCTTTCAACAAGCCCAGCAG (SEQ ID NO: 15)

contig03476 length=652 numreads=8

AGTGGCGACTAGGTTCTTCCTGGCATCCTCATGAATAATGACACTGTGCTGGCCA
CAGTGGGCAGTAAACAGCAGGCAATCATGAACACAAAAAGGGATACCGAGCCA
GCTGTTCGAACTGGAGTGCCAGTTTATTCTTCGTCTACCAGAAGAACCTGGCAAAG
CCCTCCGAGACATGTTAAGAAGTGGGATGTCCAACTTGAAAGACCGTTTCCTTGT
GAGATTGGAGAATGATGTTTCGATATGGCGAAGTCCGTGTCGACAACTGGCTCCTA
CATGCCAAAGTTGTTGATCTCCTTACAATTGTTGAATCTCTGAAGACCATTGATA
ATAAAAGCTTTTACAAAACCTGCTGACATCTGCCAGATGATGATATGCAAAGTGG
AAGATGATGCCCCCACAGACGAAGAATCCCCCTCCAAACATAAGAAGAAGGACC
CGAATAAAGTCGACAAAAAGTTTCTCCATCCACACGGCATCACGCCTCCTTTGAA
AAACGTCAGGAAAAGACGCTTTCGGAAGACTTTGAAAAAGAAATATGTCGAGGC
CCCAGAAATTGAAAAAGAGGTTAAACGACTGTTACGTGGCGACAATGAAGACCG
TTTACTGTCAGGTGGGAGGTGATCAACGAAGAGGACGACAAAAAAGGGAGTAA
(SEQ ID NO: 16)

contig02585 length=824 numreads=46

AAAAATAATTTTTATTTCATCGCTTACAGATACCTAATAGTATTATTTACAGATAAT
GTAAATAATTATGTACAGTCTTAATGTAAATAGGAGAGGTTATGATTTACAAAAA
TAATATTTTCTTTGGTGTTTTCAATTATACAGTTTTTCAGCGTGA CTCTGGTTTTGTC
AATTTTTTAGTGAGAAGTCGAAAGGGTAGTGGTGAGGAGGGAAGGAGGGGTGGT
GGGAGGGGATTGAAGGGAGGGAAGGGAGGTGGAAGGGTGGAGGGTAGTAAGCG
AGGGAGGGGTGGTAGAAGGCAAAGGGCTGGTAAGGCCAAAGGGCGGTAGGGTA
GTAGGGCGGAGGCGGCAACTCCCCAGGACGGGATAAGGGCCGGGGGCGGCGA
AGGGGTGAGGAAACATCGGAGGCGGCATCGGCGACGTCGAAGGCGAAGTCAGC
GGGGCGGGAGAGGCTGAGCCGACGGACACGCCGAGTCATCCGTCGGAGGCGG
CGGCGGGCGGGGCTTTGACAGTCGGAAGCGGGGATCCCCTTTCGCGGCGGGCCG
GCTCCTCGCGACACACGGGAGATTTTCTCTCATCGGCTGTTTCGTCGTCGCTCGCT
TCCTTCTTGGGCGTCTCATCGTCCGCTGTGCGACAGCTTCGACTTCTTCTTGGTATG
CGCCATCCCCGTCTCCCGAAATCCTTTGGCGAAAGGGTTGTTATCGATCTTCAAC
TTCGTAATTTTCTCATTTTGGTAAGCTGTAAACGGCGATGAATTCAGTTTCTGGGAA
GGTAACTGTGAGTGTGGGGCTCCAATGAAGGGCGAGGATGTCTGAGGCTTTGAT
GACGTG (SEQ ID NO: 17)

contig02033 length=254 numreads=2

GAGGAGGAGAGGAGCAAGAGAGGAGAGAGAACAGGAGGAAAGAAGAGGAGGAGA
GGAGTAAAGAGAGAGGAGAGGAGAGGAGAGGAGAGAGTGTGAGATAGAGGAG
GAAGAGGAGAGAGGAGTAGAGAGGGGGAGGAGTGTGTTAGGGAGGAGAAAGTA
GAAAACAGGAGAGTGAAGAGAGATCGGAGGAGTAGAGAGGAGGGGAGACGAA
GAGGGTGACGCGAGAGAGGAGGATGAGATGCGAGGGAGGAGAGGA (SEQ ID
NO: 18)

contig01290 length=942 numreads=36

TCAGCAAGGAGGTTGTCCAATTCTTCGGCGGTCTTCATTTCTCTCGAGTATTCGTA
AGCTTCAACAGGAACAGTTTGAGCGGGTTTGGGCAGAGGTGTTTCGCCGTCAGC
GAATCTGATGGCCATCGCTCTCTTCCAAAGTTCCAGAGCTCCGATCATGTCCCTCT
TCTTGTC AACCAACGTCGCACCGAGCAGCTCCAAGGCGTCGATACTCTCCTGCCT
GCTCACGATCTCAGGGAGGCCGATGAGGTACTCGACGATGTGGGAGTGTCCGGT
GACGCTGGCCGCCAGCAGAGGTGTCATACCGTAGGAATCGACGTCCATTTTAGC
ACCGTGGTCGACTAGCATCTTGAGAATGTTGAGAGATCCAGATTCGGCGCAGTCG
TGGAGAGCCGTGTTGCCTCTGACGGACTTTCTGTAAACGTTGGCGCCAAGTCCCA
GAAGGTAGTGGGCGATCTGGTGGTGTCTTTGTAGCAGGCGATCATCAAACACGT
GTGACCGTGCCTGTTTGAGACCTCAATGTCAGCTCCATGTTCAACGAGGAATTTG
ACGATTTTCGTAATGTCCATCAAAGCAGGCAGCTCTGAGCGGCGTGGAATTTGTTT
TGGTGGTGGAGTTGACGCAGGCGCCGCGCTTCACCAAGAATTTGACAACGGCGA
AGTGACCGGCAGCCGCTGCGCACACAGAGGAGGCGCTCCTTCGATCGTCTCTCC
ATCAAACACAACCTGAACCAGCTTGTTTCGAGATCGGCGTTGCACCTTTTCGACGAGG
TATTCGACAACATCATAGTGACCGTTCCGACAAGCCATGACCAGAGGTGTGCGAC
CGTTCGTTTTTGGTGGCGATCAGGTTCACTACTTCTTCGGGCGTCTATGATCGAG
GAAAACCTGAGAGTTTGATGATTCCAAGGAGTTGGAGCTCCGGTTCCACCATCCA
AGCCTCCAC (SEQ ID NO: 19)

contig00731 length=1376 numreads=65

ACTGGGCCTGGGGCCAACCGTATAGTGCCGAACCTCAGTTTGTGTTGTTGCCTAAC
TCAGTTTGTCTGTTGTTGTCGTCTTCAATTGGTAGTGCCAACCGAATTATTTTGT
GATCTGTGTTTACTGTCCGTCGAGAGCAAGTACCTGCCATACTTCACACAGTC
CTTCAAGCATTTTCATCAAAAATGCATACCTAGAACTGCACGAGATGATGCCTCGG
AATCATCTCATTCCACAACTATCCGACATACTGACAACCAGGATTTGGAACGTC
CTGAAGCAAAGTCGCCGATTGAATACTGGACATCTAACGATATTGCCAGCTGGCT
GATTCGGGTCTCGGATTGGTTGAAGGTGGATTTACCCGACCTACATGGAGAAAA
ATTCTGTACATTTGTCCGGCACGACGCTGCTCAATATGACGGAGTCAGAGTTCACG
GATGTAACGCCGAGTTTGGGAAAAGTCTATACGAGCACATCAAGATGGAATCAA
GTCTTACCCCAAGTTACATCAGGTGATTACCTCGTTCAATCGGAGACCATCAAACC
TGTTGAGCAGTGGAACGATTCGGACATCGCAGATTGGCTTGTTAGAACGTCCATT
GACATTGGAGTTGACTACACAGAACTGTGCGGTGAGAAATTCGTCCATCTAACTG
GCCCTGAGATTCTGGACATGACGGAACGAGATTTCTTGACGATCTGCCAGGAAC
ATGGAAAGCAGCTCTTCAACAGCATCCAACGTGGATTATCCTCGTGCCAAAATG
ACGTCCACGTGCCACCAATGCAGACATATGAAGATGAAAATGGCTCTGACAGGG
AAAACACTCCGTCTTCAGGTTTGTCAAAACCTGTGGTCCCCGAACATTCAAACCC
ACCAGGAATCGTGCAGAAACGAAGACGGGGAAGACCGTCCAACCCTAATAAGA
AAGGGAAGAAAGAAACCAGACAATTCTTACGGACGTCTGTGGGAATTTTGTAGAG
ACCTCCTCAAGAACCCGGACATGTGCCCCAGAATTATTAAATGGGAAAACACG
ATGATGGCTTGTTTCGATTTGTACAGAGCAATGAAGTTGCCAGAATTTGGGGTGA
GAGGAAAGGAAACAAAGATATGACCTACGAGAACTCAGCCGTGCCATGAGGTT
TACTACGTGAAAAAACCTATCCTATCTACCTCAGAGGGCACTCCACGCGTCTA
GTGTACCAGTTCGGACCAGGCGTACCAAATTGGAGAGTTGAGGACCCAACTTC
GAAAAGAGGAATAAGGACCTCGAGCAGACGTTAGACTCTATCCAGGGACTCCTG
GACAGTCTCGCTTTTGAATCCACCAACATTGGAAAACTAATTTATAAATACATG
ATGTACGATAGT (SEQ ID NO: 20)

contig01861 length=237 numreads=9

ATATTCAGGTGTCGGCCCCTGTTTCACCGGTTTCAGGCCCCATTTCAGCTCTGGCAA
TTCCTGCTGGAGCTTTTGACTGACAAATCTTGCCAAGGATTTCATTTCTTGGAAGTGG
AGACGGGTGGGAGTTCAAATTAACCGACCCTGATGAAGTTGCCAGGAGGTGGGG
AGTCCGAAAAACAAGCCTAAGATGAACTATGAAAACTGAGCCGAGGGCTTAG
GTACTACTACGACAAGAA (SEQ ID NO: 21)

contig01108 length=1392 numreads=57

GTGCTTCACATTCTTCTCCACATTCCATACACTCCTTTCCTGCGCTCTCTTTGTCCG
GAGCTTCAGAGACTAGATGGCAAGGTAAAGCACACTCGTGGTTAGAGCAGGCAA
GGATTTCGGCCACAGTTTCTGCCACACTGCTGCGAAACAGCCCGATGACAGGGTA
AGGAGATCTCTTCGTGGCCACCGGGACACCATACTCCGACAGGGACCACACATG
GAGGACAAGGGAGTTCTTTGATTTCAACTGGAGCAGTCACTCTTTCCCAGGGTGT
CGCTGGCTTGTAGTTTGGAACTGAATTGACCAGGACAGAAGAGTGACAGGGCAG
AGGGCAACGGTGACCGCACTCCTGAGGTTTGTACAGATTTGTGCGGCACGAAGG
GCAGGGGGCCCATATGACATTTATGCTCTTCCCTTTTAGGATGGTGACATTCAGGA
GGGATCATGCACAACCTTGTTACATTTTCAGGTTCTTCTTTTTTCGATGAGTTCCACA
GGGGACTGTGACGACAGTAGCGCCACACCGACAGGAAATTTGCTTGGTCTGTTCC
AGGGGTAGCAGGGTCCACGATGGCAGACTGAGGCACACTTGTGTTGACCACACC
GGAGTGTTTTGCCGCAAAGTTTCTCACATGGAGGACACCGGCCATCGCAACATTT
TCTATTGCAAGGATGGATGCCACAATCTCGAGTTCCTTTACACTTGGTGTCGCAC
AGGTATTCTTTGATGCAGGCAATTTCTTTCTTGTGGAGACCGCAACGGCACGACT
TCTCCAGGAACCTCCATGCAGGCGCCGACGAGACCCTCGGTGGCACCGCTGGTTGCA
CGAATGAAGGCCACAATCCAACAATTTGCCGCAAGTGTTGCCACACGTGGCAAC
TTCCTGAGTACAGGGGATGACCGTTTCCTCAGACCCGACAGGGACAAGTCTTGGGC
AGGGTGCGTGGACAAGGTCCACACTCTTCTTTGTGGCAGACTAGATCACAACCTGT
GTTTGTACAGTCCAATAATTTATTGCAAACCTTGTGCGCATTTCCATGAAAGCTCA
GCGCAGGGAGTGAGGCGAGATTGACTTTTACAATCGCAGGTCATCATTGATGTTT
TTGGACAAGCTGGACACTCTCCAGGATGGCAAACCGTTTCGCACTTGTGCTTTTC
ACAACCTGAGGAGCTTACTGCATTTCTGACCGCACGACCATTCTTTGTAGCAGCAC
CTCACCATTTTGTCTCCTCCATTAGAGCAGTAGCATTTGGCTCTGATCATTTTGGG
GCACGGCGGACAAGGACCTGGGTGGCACAGCAACAGACACGTATGCCACAGCT
GGGTACCAAGTCCTTGCCACATACTTCACCGCAGGAATGTGGAACGAGCCAAGG
CTGATAGGGAGGATCCAAGA (SEQ ID NO: 22)

contig02251 length=256 numreads=4

GCCTGCGGAGGAAGAATGCGTTCACTTTGGCTTCGGACATCCTCTGTCGAGGG
TAGATTGGAATGTACACGTCGTTTTCTACAGCAACCTTCATCTTGTGCGACGACGG
ATTTCAACAGTGCGTTGAGAAATTTGAAGAGTGCGTCACTGTAGGGTAATCCTG
GATGAATTTGGTGACCAGACCAACGAGGGAGACCGTCTGAGTGCTGGACAATGG
GTCCCAACACTGCGTCACTAGTTGAGTTAATTTTCGGT (SEQ ID NO: 23)

contig04665 length=272 numreads=3

GGTGGCAGGTGCCGACGGCAGGGTCCGCCGACGCCCCGGGCGTCACCGCCGTCGA
GTATTTGTACTCTTGAGCTGACGACGCGGGGTTCCAGAACAGCTTTCGGTTCGCCT
CTCGAGTCCGGGAGGTTGTAGTACGTGGGTGCGAGTACTGGTCGAACGGTGCTG
TCGGGTGGTGACCGTCAACAGTACTCGAATAGTTCCAAGGCTGAGACCCCGGAT
GGTGCCGGATGACACCAGAAGATGGTTCCATGCACTCATACCCGAGCCGTTTCG
(SEQ ID NO: 24)

contig00974 length=302 numreads=5

CATCACCCGGCGTATCGGAGCTGACGGACCCGGGTCCGATGATCCACGGGATCC
CGTCTCCAGTCTCAAAGACGATCGGATAGAGGACGACGAAGAGGACCAATCCA
GGGAGCAGATGGCGTCCGGACCCGGACACAAGAAACGGAAGAGGAGGGTCCTC
TTCAGCAAACAGCAGACCTACGAGTTGGAGAGGCGGTTTCGACAGCAGAGGTAC
CTTTCGGCCCCCGAACGGGAACACCTCGCGTTCGATTATCCACCTTACGCCGACCC
AAGTGAATAATCTGGTTCCAAAATCACCGATAC (SEQ ID NO: 25)

contig02631 length=259 numreads=2

TCAAGGCGTCGTTCGAAATCAGCCCGAAGCCTTGTCGGAAAGTTAGGGAAGGTC
TGGCTAAAGATACAGGGCTCTCCGTGAGAATTGTTTCAGGTTTGGTTCCAAAACCA
AAGAGCTAAAATGAAAAAAATACAGAAGAAAGCTAGGATGGATGTAGAAAAAA
CAAACAAACTACTAGGAAGCAACGGAGGTAGGACTGATGGCGACACTACAGAC
GTCCACAAGAGTGAAAGGATAAAGGACGAAGAACACAGCGGTGG (SEQ ID NO:
26)

contig02098 length=428 numreads=17

GTAAATACCATGAAACTGGGTACCTATATAAATCATTATGTCATTTTTTTCATGTTT
CGAATTTTTATACAAAAAGGAAGAAAAAGAAATACCATCAGATTGTCCTGGTG
TAGGTTCCATACTCCACTGAATACGAAAAGGTTTTTTATTTGTTTTTATATTTTCGA
TAACAAAGCGGATGGTAGGGGTTTCACAACCTAAGTTACGCAGAGACGTACAACC
GCATCGATTGAGAAGGTGGCAAAAAAACTGTGGTATTTCGGTCAACACTGCTGA
TAACGGTTACTTGCTGAAGGAGAGAGGTGTTGTGGAGTGGGCCATTGACTCTGAC
TACAGGTGGTGCCCATAGTATTGGTGAGGGTAGTGATACTCGGAGTGGAAGGCG
GGGTCTCCGCTCCTGTGCATCATCTGCTGGCCGTCCATCATGTAG (SEQ ID NO:
27)

contig04401 length=249 numreads=3

TGACTCTGCTATATTCGTTTCAGAGCCCAAACCTGCAACCAGAGATACGGCTGGCAT
CCTGCAACCGTTTGCAAGATCCCTCCTGGTTGTAATTTGAAAATTTTCAACAACC
AAGAGTTTGCAGCATTGTTATCACAATCCGTTGGACAAGGCTTCGAGGCTGTTTA
CCAATTAACCAGAATGTGCACCATACGAATGTCTTTTGTAAGGATGGGGTGCT
GAATACAGGCGCCAACTGTGACATCCAC (SEQ ID NO: 28)

contig02932 length=398 numreads=7

TTGTTGTGGGCGACGGCAGGAACGAGAAGATGGTCACTTGAGAAGGAGAAGACG
TAAAATGTGTCGTCACGACGTGGAAGCGCCCCCTGGGAAAGGCGGTCTGTAGAAG
TCGAGGGCGCTGACAGTCGCCCCGAATTTTTGAGCCTTGATTGGTTTCTTTTTGGG
AATGTTGTTGTGGAGTATGAGCTGCCGCAAGGAAGGTTTAGGATGGTAGGCAGC
AGACATGTTGTGGTACCTCACCTCTTCTCGTCCACTGCTATCCACCTTCGGAGTT
CTTCTCTAATCTGAGAGATTTCGTTTTGATTTATCCGCAAGGGACAAGTCGGGCC
TTGCTGCGAAGAGCTGGTGATGTTGAGTGGATCTTCGGCCACCCACAGTAATGTC
CGGCCTATCCGCGT (SEQ ID NO: 29)

contig06264 length=792 numreads=99

CAACGCAGAGTGAAATCAGATCACACCCTTGCTCAAAGTGCGATTTTGTAGCTCG
ATGTCCTGCTAGCTTGAGAAATCATCTGACGTTGCACACGAACGAGAGGCGCTAC
TCGTGCAAGAAATGCAATTACAAAGCTAAACAATATTCTCATCTGAGGAACCAC
CTGTTGATACACAATTCCATCAAACCACACGCTTGTCGCACTGTGATTACCGAG

GAAAGCGCATAGATGCCCTCAGGAATCATCTTTTGAGTCATTCCAACAATGCTGC
CGCGGAAACCTACTTATGTGAAGACTGCGATTATAAGACATCTAAAATGGCCAA
TTTAAGAAGCCACAGGTTGACCCATTCCAGAATCAAGTCATTCTCATGTTCCATT
TGTGACTATAAATCTTCTTATCAGAGCGACATTTCAAGGCATACGAACAACGTTC
ATAACAAACCTGATAAACGATTTTTTTGACGGTTTGCAACTACAAAACGTCTAG
GGCGGATCACTTGAAAGCTCATGTCAAAACTCATTCTAGTCTTAAGAAAAAGTGT
CCTCATTGTGAATACAAAGCAGTAAAAAGTATATCTGTTAGAACTCATATCAAGA
AAATCCATCCGGACAAATCTTGAGATTGTTATTGCTTCTGAAACCAAGTTTACA
TGTTACCATAGAAAAGCTTCGAGCTTCGACTGATTTTGAGTATTGGAAAGACTAC
CTACGCTTTCCTTTAGCTTTACAAAATCTAAAAATTGTATAATTCTTTGAAAAAGA
TAGCAATAAAATGATATTTTTTT (SEQ ID NO: 30)

contig02728 length=287 numreads=9

TTCTCGAAAGTGAAAGAACACTAGCAACAAATGAGTGGCGGATAGACCAAGATC
CCTGGACAAATAATTTTGGAGTGAGCGAAACCGCATACAAATTTGGAGGAGATA
CATACCAGCTCTGTTGGATTGGTCTATCATAGGTTGTACGATTCTGCGTCTCGCGT
TGATGAACCAATTGTTTACTTGGAGTATGGTCAGCCCAGTGTCTGCGCTAATTG
CTTTTTCTGATCTTCAGAAGGGTAAGGGTGAGTGAGATGCTGAAAGAGCCAAGCT
CGGAGAATGTTGG (SEQ ID NO: 31)

contig06226 length=511 numreads=55

GAATGGAGGTCGTCAGACTACAGGGAACCGCACGACTCGTGGAGCGGTCCTTCC
ACTTCCAGGGCGGAGAAGCAACAGCTCCGTCAGAAGACGCAGGAAAGAAGGCGC
TGGAGATGCTGATCGCGATGAATGGTGCAGCCAGAAGATGACGCTGACGAACGA
GCTCTGCGCTGGTGTAGCGGAATGCAGCATTGCGTGGAACCTATAGATGCGCAG
GAGCAGGTGTACCCGTGTAAAAAATGTTTCAACATTTTCCATCTGCCCTGCATTC
GAAAATGGGCAGATCAATGCATATCAAACAATGGCACGCATTGGACATGTCCAA
ATTGCAAATATCAAACCACGCAGACCCCAAGACCATCCTGCTTGTGTGGTCAAAC
CGCCAAACCCATATCTAGTCCCGGTCGTACGCCCCATTGTTGCTCAAAGAAGTGC
AGTAGATGCAACAAGCCCTGCCATCCTGGAGCGTGTGATTGATGAGCAGGATGA
CCTACTACGTAATATACTTGCT (SEQ ID NO: 32)

contig04962 length=247 numreads=2

AAGAACTCAAGCAGCTCCCTAACATGCGTATGCACAAAACATATAGAAATGGGA
AAGTTTGAAATGGAGGTCTGGTACCAGAGCCCCTATCCAGATGACTACGCGCGC
CTACCAAAGCTGTACTTATGTGAATATTGCCTTCGATACATGAAATCGAGAAGTA
CTTTAGACCGACACTGTGTGAAGTGTGTTTGGCGCCACCCGCCTGGTGAAGAAGT
TTACAGGTGCATATTGGAATATTACGGTC (SEQ ID NO: 33)

contig04941 length=264 numreads=2

ATCGGGTTCCCTTTGATGCTCCATGAAAGCAGAGAATTCAACTAGCCGCAGCTTG
TGAGTTGCTATCGCTCGTCCTTCCCATGGTGGTGGGGCTGGTACTCCAAGCTCAG
CAGAACTCCCAGATACTGAACCTGTGGGTTTCCCCTTGGGTAGGCAGGTGCACC
AGAAAAGGGCTTGACGTCTTGAGAGGTTTCTGGCTGAAGACCTGGTTGCCAAAA
ACACTTTTAATTTGGCCTGAATCTCCCGGAGTTTCTCCGCGCAA (SEQ ID NO:
34)

contig04132 length=246 numreads=2

GCAAGGATGCTTCCCGCAGGAGAGGTTGTTGGGGCAGCGGGAGCCGCAGGAGTA
CGACGTGGGACCTCCCATCTCCGGTGGGACGGTGTGTCCGTCGCAAGTCAGATCT

CGGTCGGTTCGACCGCAGTGGCACGTTTGGTGAATGGATTTGTTCGCAGGGAGGA
CAAGGGCCTTGATGGCACACGGAGGTACAGGGATGCAGCCCGCAGTTGAGCGCT
CGCCACAGACGTCTCGCAGTCGAACGT (SEQ ID NO: 35)

contig00668 length=668 numreads=78

CGGACGGTCTGAACGAAAAGCAGTTGCACACGTTGCGGACATGCTACTCGGCC
AACCCGCGACCGGACGCCCTCATGAAAGAGCAACTCGTCGAGATGACCGGCCTC
AGTCCGCGAGTCATCAGGGTCTGGTTCCAGAACAAGCGGTGTAAAGACAAGAAG
AGAGCCATCGCCATGAAACAGCAGATGCAGCAGGAAAAGGACGGGAGGAAGCT
GGGCTACGGGAGCATGCAGGGTATCCCCATGGTGGCGTCTAGTCCTGTTTCGGCAC
GAGAGCCCACTGGGGATGAACCCGATCGAAGTGACGTCCTACCAGCCGCCTTGG
AAAGCCCTCAGCGACTTCGCCCTGCACACCGACCTCGACCGCCTCGACCCCTCCG
CACCGCCCTTTCAACACCTTGTTAACCAACTCCATGGCCACGGATACGACATCCA
CGGACCCGCGGGACCCGCGCCACACGAAGTCCTGGGCCCCAACACCGACATGGT
CCACCCGGACTCAACGGATAGCTACGTTACCTACCTCGAAAGCGACGACAGTTTG
CAGCATGATACATCCAGCCCATAACGGCCACAAGATTTGTTCGCAGAAGTGTA
TAATACGATTTCACTAGAAATTTCTCACGTTTCGTCAAGGAACTCTTCCTGTTTT
CAACAAGAACAAGCA (SEQ ID NO: 36)

contig02090 length=290 numreads=5

TCCGCACCGAGGGAGATGCCAGAGTCACGCTCGTCCGAGGAGCCTTCGCCGTCC
CAGGGAGGGCTCGCCCGAGCGCCGGGTTCTTGCTTGATGTTGTGGAGGCTGAGG
AGCGCCGTCGCTGCCACGTCCTTGTCGCCGAGGTGGCTCTTGTCGAGCTTGT
GAGGGAGGCAGGAGGCAGCTGGCGGCTCGAGCGGCAGAGGGAGCTCGTAGGGG
CTCGGCGACGACCGGCTTCGACTCAGGTTACGACTGAGTTGCCGGCGGCTTCGA
GGACCGGAGGGTTGAGCGG (SEQ ID NO: 37)

contig05463 length=611 numreads=118

TAATTTAGCTTTGCAAAACACTTTTTAAGACCGTTGCCGTTTATATCGGAAGGCA
AACGTTTTTCTAGTCAAAACGTGTGCGGGTAGCAAAGACCTTCACACAAAAATTGT
TCTAGCGCACACTTTTTTTAAACATTTTTTTAAATTAATTTTTTTATGTTTTTTGCTT
TTTATGAACATTTAAACCGCCTGGAGGGATTTGAAGTCCTATGTTTTGTGCTGA
AAATTGTGAGTTCAATTTTTTAAAAATTAAAAACAATTTTTTTTCTTCCTAAAA
CTTTTGTACTCCTCTCATTTGGGACCCTGTGGGCTCTCGGGACCGCTGACTTTGT
CCGAACCGTCCTTCATTTGTTCGCAATAGTCAGCGAATATGTCGCGGAAGCGTC
CCCATGACTTCTCAGGCACTGTTATCGCTGTTCTAAAATTGGTTTTAACCTCTGAA
ATGCGCATGTAGATACCGCGATTATTTTGGCCGATGTCGAAATAAAAGTTTTTAT
TATCTACTCGCATGTGCTTCCCCTCAGGTAGCTCCCCCTTGAACCCTCCGTCATCC
GTTCCAAATTCGACGAGAAGATCGGTGAGAGCGTCTCTAAATTCTATCATCCCC
(SEQ ID NO: 38)

Translation factor:

contig00310 length=243 numreads=4

AGGAATGGTCGTAAAACGCTAACCACTGTGCAGGGGCTCTCATCGGAGTATGAC
TTGAAGAAAATCGTCCGTGCGTGTA AAAAGGAGTTTGCATGCAATGGAACAGTG
ATTGATCATCAAGAGTACGGTGAAGTTCTGCAGCTGCAAGGTGACCAGCGGGAG
AACATATGCCAGTGGCTCACCAAAACAGGCCTCGCCAAGCCCGACCAAGCTCAAA
GTTACGGTTTACTAAAAAAATACCT (SEQ ID NO: 39)

contig05381 length=705 numreads=34

CCGGGTATTACAGAAAATTACGGCTTGTGTGATACTTAGCGTATCGTACAGATCG
CAAAGTGTTCCTAATTTCCAATCTTCTCTCTCCACAGAGACGTAAAACGTGTTGAT
ACCTTCAAGGGTCAACTCTTCCTTCTTGACCAAGATACGGACGGGCTGCCTCATG
AAGCACGAAGTAACTTCGAGCACGTCGCTCGGCATAGTAGCAGAAAGCAGGATA
ACCTGCACTTCCTGCGGCAGATTTTTGAAGACATCGTGGATCTGCTCCTTGAAAC
CTCGAGACAACATTTTCATCGGCTTCATCCAACACGAAAATCCTGATAGTGTTGGT
TTTCAGAGCTCGTCTGGTTATCATGTCCAAAACCTCTTCCAGGAGTTCCGACCAGA
ACATGGACACCCAATTCCAGCTTGCGCATGTCCTCCCTGACATTGGTGCCTCCAA
TGCAAGCGTGACACATAGCCCCCATGAAATCTCCGAGGGTCAACACAACCTTTCTG
GATCTGCTGGGCCAACTCTCGAGTCGGAGCGAGAATGAGTGCCTGACATTCTCGC
AAGTTGATATCAATTTGCTGGAGAATAGAAATTGAGAACGTAGCTGTCTTTCCAG
TACCAGACTGAGCCTGGGCAATGACGTCATGACCCTTGATGCAGGGGATGATAG
CGCGCTGCTGGATAGCGGAAGGCCTTTCAAACCATAGGCGTAGAT (SEQ ID
NO: 40)

contig02633 length=462 numreads=8

AGTATCGAGGAATGTAACTTATCCTTCCCAATCACAACCTGAAAATGCACCAATG
TGCTGAGTTATACCGCCGAATTCCTGTTCAACTATAGAAGACTGTCTCAGAGTAT
CCAGAAGAGTGGTTTTCCCGTGGTCCACATGACCCATGATAGTCACCACAGGAG
GACGAGGCCTGGGATCAACGTAATCCGAAGGATTTCTTCGTCCAACATCTTTAAA
ACTTTTTTCTTCAGTTTCTGTAGGCTTACTAGCCATCCTACATCGATGACCCGACT
TTTTAACTATTTCTTCAATAACTTTGAAGTTGTCGATGACATCACGGTCTTTTCTG
TACTGTTCCGTGTTGTCAATGAACGACATAAATTCTAAAACGTGTCCTACATCTTT
ACCCAGTGTATTTGCTACTTCCGCCACTGTCATACGCTTCCACACGTCAACGGGC
TCGCCTTGTTTCCATTTTGA (SEQ ID NO: 41)

contig02589 length=242 numreads=7

CTCCTGAAGGCCCTCACTAATTTACCCCATACCGACTTCATTTTGTGCAAATGTCT
TCTGAATCATGAGAAGTGCAGCGAATCTCCAATCAAGGATATCATCTATTTGGCG
GACTTGCTGGAATCCTGTGAATTTCAAATATTCTGGCAGCAACTGCAGGACCGCA
AAATGCAGGAATTGTGTTCTCGAATCTTAGGATTCCAGGATTCAATAAGGAAGTT
TGTCTGTCACGTTGTGGGTAT (SEQ ID NO: 42)

contig02032 length=252 numreads=3

CGGCGGCGGCGACCGGTTTGTGTTGGTGGGTTTGTGTTCCCTTGAGTGGAGATGCTTT
GGCTCTGTGTCATATTCAATCTCCAAGTCATCATCATCGCTTTCGTCACTGGACC
CTTCTTCCTCTTCAGCCTCTTTCAGCCATGTGAGAAACGGTTCAGCCTTGTTGTGT
ATTTCTGACTAAGCTCTTTAGACACATATTTCTTGCTCACTTTCCTCCCCATTCC
AACAGGACTCCTTCATCCAGTATATCCT (SEQ ID NO: 43)

contig01435 length=212 numreads=9

TGATGTCTCTTACATCCTTCCTGTTTCATTGTCTGCCAATCATACAACCTGTCTATTT
TGTTCAAAGCAACGACAAAAGGAGTTTTCTGTCTTCAACAAGTTCAAGGACTC
GATGGTTTGGGGTTCAAGCCCATGCATGATGTCCACGACAAGGATGGCGATGTC
ACAAAGGGAAGACCCTCTGGATCTCAGGTTGCTGAAAGATTTCGTGA (SEQ ID
NO: 44)

contig01089 length=945 numreads=27

GGGAAATCTGGACGTCGCTCAGGGTGCCTGCTCGGTCTCGTAGTGGAACACAG
GTTGGAAATCACCAATTGCTTCCCTTTCCCAAAAACCTTACGATGAACTGTCGAC
GAAGAGGAGTACCAATTGGATATGATGAGGAAATTGAGGAGAGTCAATGTGGAT
CACTATCACGTAGGATGGTATCAAAGCTCGGACGTCGGTTCCTTCTTGAGTTTGC
CTTTGTTGGAGTCGCAGTACCATTACCAAACGTCGATAGAAGAATCCGTTGTTCT
TATTTACGACACACAGAAATCCGAAAGAGGGTTCCTTACCCTAAAAGCCTACAG
GCTCACCCCTCAAGCCATAACCATGTACAAGGAAGGAGAATTTTCGCCTGATGCG
ATGAGAAACCTCAAAGTATCTTATGAGAGCCTCTTGATCGAAGTGCCTGTGGTCA
TCAGGAATTCCAGCTTGACTAATATTTTGATGTGCGAGCTAGAAGAAATGATTCC
TTCTCGTAAAGGTTTCAACTTCCTTGATCTTCAACCTCGTCTGTCCTCGAAGGAC
AGCTGAGGAGCTTGATGGAGCACGTAGATGATTTGAACCAAGAAGCGATCAAAT
TCAACAGGTACCAACACGCTGTAGTTAGGCAAACCCAAGACAAACATCGCTACC
AACAAAAGAGGAATCAGGAAAATGCTGCTCGTGCTGCCAAGGACGAGCCTCTTT
TGCCCGAAGAGGACATCAACAAACTCTTCAGGCCGATCATGGTCCCTCCTCGCCT
GGTGCCCATGATCATGTCTGCCCAAGTAGACGCCTATTCCCAACACATTTACAG
TTCTGTTCTCAAGCTTTGGCTAAATTGTACATCATGCAGGGACTTCAGACACCCG
ACAAATAATAATCCAAGACCGAAAATTGGTATTGATTATACTTTACGTTTAATAT
TCCATTAAATCATAA (SEQ ID NO: 45)

contig06240 length=1462 numreads=126

TCCTTCCGAGAAACCAAAGATGTTTGCCAAGGATGCTGAAATAGATATGAACCT
AGTGCTTAAAAAGTTGGGAGAAATCGTTGCAGCCAGAGGTAAAAAGAGGACCGA
CCGGAGAGAGCAGATCGAATTGCTCCATGAGCTTGCAGCCGTTGCTGACGCTCAC
CATTTAGGTCTGCTATCAACATCAAATCAAACCTGGCCATAATATCTTCCATTTT
CGACTACAACCCGAAAGTTTCAGACGCTATGAAGCCAGAATATTGGGCCAAACT
TCTGGAAAGGATATCTGAAACATTGGACCTGCTAATCTCGACTGAAGATATTACG
ATTGGAGAAAATATCACTGAGGAGAACGAAGAGTTCGATAAAGCTCCATACAAA
GTGAGAGGTTGTGTATTGACTGTTGTGGAGCGCCTTGATGAAGAGTTTACGAAGC
TGCTGAAAGAATGTGACCCTCATTCCAATGAATACGTTCAACGCTTGACGGACGA
AAAGAGAGTTTGTGTCATTATCGATAAGGTTTCAGAAGTTTTTGGAAAGAACGAGT
GTTCTTCTGATCTGTGTAGGATTTACCTTCGTAAAGTAGAACACTTATACTACAA
GTTTCGATCCGAATGTTATCAGACAAAAGAAGGGAGAACTTGAGATCGGTTTCGCC
CACTAGTATCCAAATCATGGATAAGTTGTGCAAATACATCTATGACAAGGATGTG
ACTGACAGGCTGCGTACGCGTGCCATCCTCGCCCATGTTTACCACCACGCTCTTC
ATGACAACCTGGTTTCAAGCTAGGGATCTCATTCTCATGTCCCACTTACAAGAAGC
TATCCAGCATTCTGATCCATCCACACAGATTTTGTACAACCGTACTATGGCACAC
TTGGGTCTTTGTGCTTCCGGCATGCGAATATCAAGGATGCCCATAACTGCCTTGT
GGACCTAATGATGACCGGAAAAACCAAGGAGTTGCTTGCACAGGGTCTTATGCC
ACAAAGACAACATGAAAGAAGCAAGGAACAAGAAAAAGTGGAAGGCAAAGA
CAAATGCCATTCCATATGCACATCAACCTTGAACCTTATCGAATGTGCCTACTTGG
TGTCAGCTATGTTGATTGAGATCCCTTACATGGCAGCTCACGAGTTTGACGCTCG
GAGAAGAATGATCTCCAAAACATTCTACCAACAATAAGAAGTAGCGAGCGCCA
GTCCCTCGTTGGTCCCTCCAGAAAGCATGAGGGAGCACGTAGTAGCTGCTTCCAAA
GCAATGCGAAATGGAAATTGGCAAGCTTGTGCGGAGTTCATCATCAACGACAAG
ATGAATGCTAAAGTGTGGGATTTGTTCTATGAAGCGGACAAAGTCAGAGAAATG
CTCAAGAACTCATCAAAGAGGAGTCGTTGAGAACCTACCTCTTCACGTACTCGC
ACGTCTACGCCTCCATCTCAATGGAAACGTTGTCAGAAATG (SEQ ID NO: 46)

contig00950 length=726 numreads=27

TTAAGGGCCAACATTTTTATTACATCCCATAAGATCTTGCAATTCGTATTTTCAT
CATTTTTTCAACAAAATTATAAGCACTCCAAATTCTACTAAAAATTGCAATACAT
AAAATAAAAATGTATACAAGTAATATAAACAAAAATTAACCTTCCATCCCTAAT
GTATAGCCACAATACTTCAACTTAATGACTGTTGATTACCATACATGTATTATA
CTTACTATTAAAAAAGAGTACGAGAAACCACCATATATATATTTTTTAATAAATAC
GATCTAACGTCACACATTCCGGGAATAAATCAATTTTTCAAAAGTTTCCGTATGA
TGCGTTTAGAGATAAACTGAAATCGCCGATTTTAAATCATTTCAACTTGTAATTT
ATTAAATTTTTGTATATAATGTGTGTATGTGTTTCAGAAATCTTGAGGAATTCATTG
AAGTCGACCCAAATTATCCACATCAGATAGCATCTACAGCATCCCCTTCACCTTC
GTCATAACTTCGTAACCAAATTCGATATCTTCGTCGTATCCATCCTCCAAAAAC
GTGACGGTGTCTGTTGATACGAACTGTTTCAGGGAAGCTCTCCATAAGTTTTGAGAT
TACGCGCCTCATCTGGAGTGTATTTCTGGATAACATCGGCTTTAGTATCCTGGTA
GTCCCTGAGGCCGATAAGAATAATGTCGCCTTGATTATCCACACCTTTTTCTAA
GTTTCCCCC (SEQ ID NO: 47)

contig00844 length=376 numreads=6

TATGGTTTTTATTCCTTGGTACGAAGTAGGGGCTTCAAATGTTTTGGGTCTTGA
GGCTTCCATCCTGAAGGATAAACTTCTTTGGGTTCAGGAGTCGGAACACATTTGG
GTGGTAGAAGTCTCTCTACGAATTTCCATTCTCAGGATTCTTGCTGATCTCAAAT
TCAGTTTTATAGGTGGGCTCGAAAGGTTTGAGGAATTTGTAATTGGTACTGGTTC
TGCACATCATTGTACCTCTCAGCGAATTCGATGAAAGAGACTTCACAAGGGTCCT
AGCTAGTTTGGATCCCACTTCAATCAAAGCCATCGTAATTCAATAAATTAGAGAC
TTCCAGTAAAACGTGAAATCAAGCACTGGAAGCCCGTTGAGACT (SEQ ID NO:
48)

contig00169 length=373 numreads=4

CATCATAAAATGGGGTAGTTTAATACGTAATATTAACAAAAGGCGTAACAAAAG
CGTGGATGAAGGATTTCAATCTCCCTCTTCAGCGCTCTCCGCGGCGTCTCGTCAT
CGTGATCGCCGCCGCTTCTCTTCTCCTCATCATCGTCGTCGCCGGCGACTTCGG
CGTCTCCGCTTCCGCTCTCTCCATCTGTTTGGCGAGTTCCGCTTCGTCCGACGCG
TCCACCACTTTGGGTGGTTTTTCAACGCGGAACACTCCACCTAGATTGTTAATTT
GGCTTCTATTACATCTATGGCATCCTGTAACGCTTTCAGTCCGTCAGACTTCTCAG
GTGTCGCAGTGGTCATAACATACAAAGGAGGAGCTATCA (SEQ ID NO: 49)

contig05794 length=521 numreads=30

ATTTACAACTTTGTATAATGTGTACCTGAATGGATATAAAGGCTTTCCAAATGT
ATTTGCACTCCAATCTGTTCTTCATCTAATAAGTATTTAACAGTAGTAGGTAGACC
TTCAGAGTACGCCTTGAACAGTTTTGAATGTAAGAAACATTATGATTGAAGATTT
TAATAAGCCCTCTGCTGATCGCCTCTTTCTCTTCGTCCCCGTCCCCAATCTTGACC
AGGGCGGTTGAAATTCTGCCTGTAAGGTCCTTGATCCCACCACGTTGGAAGAAG
TTCCCTTGCTTCATCTCGAATATTCTTTCGTTTCGAGTCCACCAAATTGTTGACCTT
GTCAGCCAATTGGAGAGCCAAGGACTGCAGACGACTGGGCTCACTTCTGTGCAT
CACTACAGTTTGAGAAGGATCGTCCAAAGAAGCCATCAGTTCCTCATTGATTATC
ATTTTGCTGACAATAGAGTGAATAACAAATCGGTCCAAGTAAACATTTCTGACA
ACGTTTTCCATTGAGATGGAGGCG (SEQ ID NO: 50)

contig02102 length=280 numreads=2

AGGAATACGCGGACTGTGTCAAGAGGGGGAAGGAGAATATTTGGATCCTTAAGC
CTACGTCGTCCCTCCTCCTAGAAAGAGGACTTAAGGAAGTTGGTGACCCCGGAGA
ATATCTGCCAGTACGAGTCGATGATAGCCGGGGAGAGACGCCTGGAGGACTGTG
GGTATAAAAGTAGTAAGGGGGAGGAGAGTGAGGAAGAGGATGACGTCTACTCC
GTCCCATGGCAGTTGAGTAGGAATTTCTATAAATGCGGCTAATGGCAAGGGGCTTC
TTGAACTAAG (SEQ ID NO: 51)

contig00970 length=579 numreads=10

CGGCTTCCACGGTCTTCTGCAATTCAATCACAAACTCAACACTTACATCATCAAG
AACCCGCCCAACTTGAACCCAGGAAGTCAAGGGCGCTTCTTTGACTGAAGGCAAG
CGCTCAAAGCGTAAGGGCAACAAGCAGAATGGTGACGCCACCAACGGAGACCG
CTCTGGATCTCCCAAATCAGAAGAAGAACCTGAAATCATTGTGGCACTCGGCAA
GAAGTCATCTAAGAAGGACGACGACGATGATGAAGTTGACTGGGCGGTTCGATAC
ATCCGAGGAAGCCGTCCGAGCCCGACTGCAGGACCTCACTGACGGAGCCAAGCA
TCTCACTCTCACTGACGATTTGGAGAAAAGTGAAGAGGAGCGCATGGACATCTTC
TACGGTTTCGTGAAGCATCGCAAGGACACTGGAGATTTAAGCAACATTGCGACT
GCCAAGGAAATTCTGTCTAGAGGCCGAGCGTTTGGAGATCAAGAGCAAAGCTCCT
TTGGTCTTGGCCGAACCTTTTGTTCGATGACAAAATTCACGTTCAAACCAAGCGTT
TCCGAGGACTACTTCTTCGTTTACTCATGAGGATGC (SEQ ID NO: 52)

contig06228 length=1144 numreads=48

AGGAGGACTGGATCCACACCTCAAATATTAATTTTATTCAATTAGAAAATGGTTA
CATCTTCCGGATGGATTATATTTATTTACTGGGTAGGTATGATACTGCACGAAAA
TGTTTCTCCAGACCACATATTACTGGAAGAACGTTCCCTGCATCTTCTTCTTCGGCA
TCTTCGTTCTCAGATTTCGAAGGTGTTGTCTAGGAATGTCATAGAGACGAATGAGGG
GCTTGTTAGGGTCCTTCATGATGAGGTATTTCCCTCCCCTTGCTTCATGCACAAGT
CGATGATGCAACGGAGGATCCCCCAGGCGTTGTCCATGTTCAACTTGATCTGAGT
TGCGAACTCCGCTGGTTTGTACTGCTGAGTCCCGAGGATGACGTGCTTCGAGGAG
TCGCGAACGTGAGCGCGGGAGACGTAGCCAACTTGACTTGCTCAGAGCCGGCC
AGGACAGCCTGCACAGTCCACTTGCCAGCTTACAGGCGTTGTTTCGGAGCTCGT
TGGCTAGAACTGCACCTCGCTGGATATCCAGTTTTTGCCTCCATTTCGACGCCACC
AGAGAGCTTGGAGTCCCATTTCATTGAGCGCCTTAATTGTCAGGAATTGAGTGTCA
CCAGCAGGTGTAGCAAGGACAGCATCGTGCTCGCACCTCGCCACGAGGATAATA
CCATTGTCAAGATCCCATTTCTGTAAACGGTATGCGACTGACGCCTCCTCACCTTC
TTCTGGGGAAAGTAATGAAGGGATTGGGATTATCGAACTTGTAACGGGGGCTCCCC
AGTTTTCAAACTTGCTGAGAGAAATTGTGGTTGATGAAAGTAGCTTCGAGCGCC
AGGTTCCCTTGGGCTGTTGAGACTGTTGGAGTCGTCCATAGGCGGATCCACAGAGG
TCTCGTTGACAGTCAAAAGATCGAATTCAGTGTTGTACGCTTATCAAAGAACAG
CTTGCCCTCCGATTTTTTCAATGACGATGTCCCAGGAGTAATTGGACCGAGTGCAG
CACATCAGAGTAGCCAATATAGCATCAGTAGCATAAACGTTGCCTTCAGTTTTAG
AGAGCTTACGAATGATAGGGTCGTCTAGTAGTAGTGACAGTGTGGAAAATCCTAT
CGATACGCTGCAGCGGTTTCTCATTCTTGACGTTGACCCGGCCGT (SEQ ID NO:
53)

contig05347 length=377 numreads=143

CCGATTTGTATCTCCTAAAAAATTGACTTAGCAGATCAGATCTGCAACGTTTCATC
GGCATTTCCTTCTATGATAGTGTTGTAAACTGTTGAATGTCCGTGAGAGTACGCT
TGTCATCTTCGGTGACGAAATTGATAGCAACCCCTTACGACCAAATCGTCCTCC

TCGTCCGATCCTGTGAATGTAGTTTTTCGCGGTTCTGAAGGTAAGTCATAGTTTATA
ACTAAGGAACTTGCTGAACGTCAATGCCTCGAGCGAGAAGGTCAGTCGTAATC
AGGACACGGCTGGATCCTGTTCTGAAAGCCTTCATAATCAAGTCTCTTTCTTTTG
TCCCATGTCTCCGTGCATAGCAGACACTGTGAAATCCCGGCCGTAAT (SEQ ID
NO: 54)

contig01100 length=1317 numreads=45

CTTCAACTCAACCATGAGCTGCCAGTCCGTTTTCACTAAGTCGTCTCGGAAGTAG
TCTTTGCAGGCGTCGATGGATTGCCGGGAGATTTTGGAATTAGGAAGTCTTTTT
CATCAAAGTGGCGCCCGAACATTTTAGGGGATTCGCCAGGAATGGGCTCAATCTT
TATGCACACTTCCATCCCTTTCCGCGCACTCTCCACCGGCTTGTGGTTATTTTCTA
TGCTCGTAACTATTCCAAGATCCACGAATTCTTTGCTAGGGACACATATCGGAGT
CCCTTCTTTGACTATTCCAGCTTCGATAATAACACCCATGACGATCGGGTCACGT
GAGTTGAAAACGAACGTGTGGAAGGATTTTAAGTTTGCAAGGGAAGACAGCAATC
GACTTGAACCTCTCCCTCTTACGGGCCTTCAACTCTTCGCGGAAGGCCATGAACT
TGTCGAAGAGGTGGTAGATGATGTCCGCTTGGAAGATCTTCACGCCCAGCGTGTG
AGCCAGGTCTTGCGCGTCTCTCTCGATTTTTACGTCAAAAGCGAGAATCACGGCG
TATTGGCTGTCATGCTCCAACATGATGGAAGCCTTCATAACATCTTTTTTGACAAC
TGGACCTATCCTGATGTTGGAGTAAGGTATCGAAGAGGTCCTAAGGAATTCGAG
AAGAGCTTCGAGGGAGCCGAGAGTCGATGCCTGGACGTAGACACCACGTTCCGGC
CATTTTGATGGAGGCTCAGAGCGTGTTTTAGTTCCCGCGCTACCTCCTCTTTTAGA
ACCTCAACTTCATCGGGAAATTTTCGCAACCCGGAGATTCAGTCCTGCTATAACTT
TTTCTAGATCTTTAGCAGCAATTTTACACCCTGTGCAGCCCTTATTTCTTTGTATT
CGATATACGCGTTCTTAACTCTAAGTTCCTTCATGGGTTGAGGCATGAGGAGGGA
TCGGATCTGTGTGACTATGGGGGCTTCCGTGCCCCGCCAGGATGGCTGTGTCTCCT
TCTCTGAGCGTACCATTACAGAAATGACGTCGATAGTCGTTCCAGCCCAGAAA
TGGCTTTCACCTCGAGAACGGTAGCCTGCAGTTCCTTCGGAGAACATCAGACGTTT
GGCAAGCATCGTCTGACATGATTCAACGATAAGAGCCAGGAGGTTCCCCATACC
TTCCCCAGTGACGGCGGACGTGGGAACGAGGGAAACGTAACCTTACGTCTGG
ATTATTATAGAAGAGAGCAGCATTACATACCCTGTTTCGGCAAACCTGGAGGATAACT
TCGTTAGTCCTCTGTTCAAACCTCCATTTGAGTGTTGGGGGCTTGTGCTTTGA (SEQ
ID NO: 55)

Microvillar proteins:

contig02056 length=449 numreads=19

TTCAAGGTCGTGGTTGCTTCTGTGGATTTTCCTCCTTCCGTCATTTCCCTGTCTGGA
AGCGTCAGCAGATCTCCAGAGAATCTTCCTGGGCAGCCCGTTTCTGACGAGCAC
CGTCAGTATTGATGGTACCAGTCGTATTATTTTCTTCCGCGGTTGAATTTTCTTGA
ACTCCACTGCCTTCCGATGAGTCAGGTTTAGAACTGCGGCAGCATCATCTGGGA
CGCCACGAACCTGGGCGAGGTCGAATTTGCATAATGTGAGACTCCGATAATAACAG
TTTCCATTACATATGGTGAGGATTGAGCTGTGCGCAGGCGTTTCTTCATCAATTCCA
GGCTGCGTTGATGACGCTGGGTCACTGCTAGCATTGATAGAGATTGCAGCAGTAC
TGGAGGTTACTGTAGCATCGCTAGTATTGATGGACAATGTGGTAGTTGTTCACTA
GCATCG (SEQ ID NO: 56)

contig01526 length=314 numreads=12

GCACCAACGGATGTTACCACTGTAACCTGAAATAACGGAGTCTTCCTCTGTTACTG
GATCAGGAAGTGAAACTGTAGGTACCAGCTTGATGCCACCTCTACGATTCTCTCCA
ATGCTACAGAAGCAAAGAGCCATCCCCCATCAAACGGACTGTTCTCTATTTTAC
TCCTGTTCAAATAACGAACCAATTCTCATGAAGTGATGCTGCGGATGCATTTC

ACAAATTATTAAAAATCTGTGATGTTCTCAAAGAGCTAATTGTGATTTGCCATA
TTAGTATTACTTATATTAAATTATAATATGTATAATATG (SEQ ID NO: 57)

contig01939 length=333 numreads=7

CAATATGGACATCAGCCCAATGGACAGGACCAATATGGACAGGACCAATATGGA
CAGCCCAATGGACAGGACCAATATGGACATCAGCCCAATGGACAGGACCAATAT
GGACAGCCCAATGGACAGGATCAATATGGACAGCCCCGATATGGACAGCCCAAT
ATGGACAGAACCGATATGGACAGCCCAATATGGACAGAACCGATATGGACAGC
AGCCCATGGACAACAGCCCCAAAACCGACCCCGTTCCCAACCAACATCCAACCA
AACCCCAACTATTACAACAACAACAACAGTGACTACCAACCTCAATCACC
ATCCCATTC (SEQ ID NO: 58)

contig03364 length=521 numreads=10

CTCCCGACGTCGTTACGCGTTTGAAGATCTCTGATGTAGTCCCGACATATTTTTTC
CATCTTACTATCACATGGGTTTGATTTACACTTGTGACCACCAACTCTGCCTCCGC
AATTTGGATTTTGGCAATTACAGCGTTTCTTCCGCCGGAACAAGCAACACCAGCA
AGGACAGCAACCTCTCTTGTCATGGACGCTAACGGTGGAATCTTCTATAATAACC
TGACGTTCAGGACTACAAACTGTTTTATGGGCTTTAGTGTCTTCCCACAGCAACA
ACAGATACCCCTCCCCCTTACAGACCGGAGACACTTATACACAGAAGTTCCACAT
GGTTTTTTCGGTTTCTTGATTTTAATACGGATTCTGCTTTTCCCAATTAGCCGTTTT
TTCATCAGTTTCCCTCCCTTTCCTTTTATCGTCAGACTTGGTTCCCGATAGAATAA
TTCTTCTTCATCATCACTTGTCTCTGACATACCGCTGCTTGAACAGAAACGACCGA
AAGCTCTCTTGAAAATACAC (SEQ ID NO: 59)

contig04612 length=275 numreads=2

TGGGCTGGTGGCGTCGGCGCAGGAGCTGGCGGCACCAGCCCAAGTAATGTCGGT
GGAAGGTTTCGCTGACGCCAGCAGCAGCCCCTCCAGCCTTTGAATCGCCACCACCT
CCTCCCATTGCGTCGCCCCCACCAGTCAGTCCTCCTGCTGTAGTCGCGGCAC
CCCCGTCCCCTCCACCGCCGGCTCCCCCGTCGCCACCTCCGGTTCGTGGAAGAGCC
GGCACCGACGAGACCTGCCGAGGAGGCGACACCTAGCGAACCAGTTGCCGTCCC
CG (SEQ ID NO: 60)

contig05924 length=388 numreads=8

CTATCACTGCTGCCACCATCCACAATATTCTGAGCACATTGTGCCAACTTAAACT
ATTTCAGTCTGTGCGCTATACATCCCCTAAGAAGAATTTAATGCTTGGTGCCTA
GCCACCCTATTTGTAATAAGTAATGTAGGTATTGCAATCGTTATTTATCATGTATT
TTTGAATTGATTTTAGAATCAATTTTCACTTCCTATGTTTCATATCTAATAACATTC
CCTTGAGGGTTGTGTTACAAAATACTGTTTAATTTGTTAGTTCATGAATGGTGTG
TTATCGTTATTTTCAAACCTTTTATGTGTAAATTGTTATGGTTCCACGTTTGTCTTGT
TAACACCTCTCTTGGGGTACCTTTTGAACCTATATTGTGCTGGAGTTTTGTCA
(SEQ ID NO: 61)

contig06813 length=444 numreads=268

GTATCATACCTGGTGAAGTGACAGACGCCGGCAGTCATGATGGCAGGATTATAC
TCGGCCTACGACCACGACGCCAACTACGACGCGGCTGGCTCCTTCCTCAAGACT
ATCACACGGACGAGGGATATGGGTTGTTGGACTCGATCAAGATGCCACCTCAGG
GAAGGTCCGGGTCCACATCGTCGACATACTTGATCTCAACGACGCCGCCAAAAC
ACAACACACGGATGGCGCCGCTCTCAACCCCAACCAAGCTCACGACGATATTTCA
GGAGGTGAAATGACAAACCTCCAGCAGCTTGGGAACCACCCCGGCTGCTGTTC
GCCGGTGGTCAAGTCACCACTCCCATGGAAGCCCTTCACCATCATTGGGGCAACT

CCATCACCGAACTTAGAGGTCACCAGCAAGCCAGCCCCGACAGCACATCACCCG
GACCGTAA (SEQ ID NO: 62)

contig06941 length=155 numreads=40

AGAGTCGCTCGGGCAAATCATCGTGCTGCCGCAGCCGTAAGTGTACAGCAGGTT
CCTGGCTCTCTGACTGTCTGGAGCCACTGTCTCGAGCTTCATGGTGCCACAGTTAG
GCGGCAGAGGACACATCGAGGGCGAGCAAACGCAAACCTCCGGCCG (SEQ ID NO:
63)

contig06960 length=154 numreads=53

GGAGTCGCTCGGGCAAATCATCGTGCTGCCGCAGCCGTAAGTGTGCGCAGCAGGTT
CCTGGCTCTCTGACTGTCTGGAGCCACTGTCTCGAGCTTCATGGTGCCACAGTTAG
GCGGCAGAGGACACATCGAGGGCGAGCAAACGCAAACCTCCGGCC (SEQ ID NO:
64)

contig06961 length=155 numreads=31

GGAGTCGCTCGGGCAAATCATCGTGCTGCCGCAACCGTAAGTGTGCGCAGCAGGTT
CCTGGCTCTCTGACTGTCTGGAGCCACTGTCTCGAGCTTCATGGTGCCACAGTTAG
GCGGCAGAGGACACATCGAGGGCGAGCAAACGCAAACCTCCGGCCA (SEQ ID NO:
65)

Electron transport protein:

contig00337 length=632 numreads=82

TTTTTTTTTTGAAATGAAAAAATATATATTTAGTATAAAAAATATCAATGTTTTTCC
CCCTTACGATGCTTATCTTCCAAAGGAAATTTCTGCTGTGGTGAATATAAATGA
AAAGAGAGAGACTGTGCAGAGCAGACCATTTTCATTCAACCCAAGAAGGGTATT
CAATAGTAACAAAATAATAGGTAAAGTACCTATTTATGAATACTTGGTCCGATT
GATGTGTTTTAGTTTGGTGAAGGGTGATTTGTCTTTGGAACCTCAAATTTTTTCGGT
GAGTTCTGGCACTGCTTTGAAGAGGTCAGCTACAAGACCATAATCAGCTACTTGA
AATATGGGCGCCTCAGGGTCCTTGTGATAGCCACAATGGTCTTGGAATCTTTCA
TACCGGCCAAGTGCTGAATAGCACCAGATATCCCAACGGCAATATAGAGCTCAG
GCGCCACTATCTTGCCAGTTTGTCCAACCTGTAGATCATTAGCCACAAAACCTGC
ATCAACTGCTGCTCGGGAGGCACCTACAGCAGCGCCCATCTTGTGAGCAAGAGT
ATAAAGAAGTTGGAAGTTATCTCCAGATTTTCATACCTCTTCCTCCACTAATTACG
ATTTTTCGCGCAGTTAGTTCCCGGCCGT (SEQ ID NO: 66)

contig06602 length=413 numreads=384

ACGGTCTGGGCGAGAGGGCCAAAATCCCCGATTGGACCATCTACAAAGTGGAAG
ACTGCCCTGAACTGGTCAGACACCGCGATAATTTGGCTTTGAAAGGCCTGAAAG
ATCCTTGGATCAGAAATGAAGCCTGGAGATTCAACCGCAAAGAGTTCGGATCTG
CTGGCTCGAGGGCCCTGCTTGTCTTCAGAGGACTGAAATGGGGCGTGTTGGCCAC
CATCATCACCGTAGGCATTGAGAAAGGCCTCAACTATGGCAAAGACGATCACCA
CGGCCACCATTGAGTTTTTTTTTAAATCAATATCTATGTATTTCTGTTGATTGTAGA
AGAGTGTACATTTTCGTAGAAATCTATCTACCTATACTTATTATTACCTACCTATTG
TTGTTAATAAAATACATCTAAAAGAACTAT (SEQ ID NO: 67)

contig01234 length=447 numreads=8

CCCCAACGAAGAGGCTGCCAGAGCTGCCAACAATGGCGCCTACCCACCTGACCT
TAGTTACATCGTTCTGGAAGACATGGAAAGGAGGATTACATCTTCTCATTGTTG
ACCGGTTACTGTGATGCGCCTGCTGGTATGGTGATGCGTGAAGGACAATATTTCA

ACCCTTACTTCCCTGGTGGTGCTATTTCTATGGCGCAAGCCTTGTATAATGAAATT
ATCGAGTACAGCGATGGTACTCCCGCGACAGCCTCACAACCTCGCCAAAGACGTA
GCCACGTTCCCTCAAATGGACCGCAGAGCCAGAGCATGATACTCGAAAACAGATG
GCTATCAAGATCATTGCTATCTTCACGTTCTTCTGCTCCGTTGCCATCTACTGGAA
AAGACACAAGTGGAGCACAAATGAAATCCAGAAAAATCGTCTACCAGCCTCCGAA
AAATAATGT (SEQ ID NO: 68)

contig00750 length=493 numreads=24

TCTGTATCTAGAGAACGTTTATTAGTGCTTTAGGTAACTACCAAATTCATCATAAT
AATGCCATAAGCATCAATAATAATTATTATGCCAATTGTTCTTTGATGGGTAGTC
ACTTTTATCAACTCCAAACTTCTTGCTGGTACCGGTTTTCTCTGTCCATCTTCTCCA
ACCATTTGTCAGCTTCTGAATCGGACATCCTCCCAGCTTCAATGGCGGACTTCTTA
ACTGCGTTGCGCACGGCTTCTGGCATGTCTTTGGAGCTCCCTGCCACGAAGAAAT
GACCACTGTGAGGAGACAAGCTGGTTCACAATGAGCTCTTTTTGATCCAAAAGTA
CGTGTTGGACATAGATTTTGTGAGGTTGATCTCTGGAAAAAGCGGTGAAAACATT
GAGTTGGTGGACCATAGCCCAATCTTTCCTGAAGAAATAGTCGGCTTCTCGGTTC
CTGCTGCCATAGAACAAAGTTAGCTGTCGTTTCTCTGCACCAGACTCTT (SEQ ID
NO: 69)

contig02025 length=357 numreads=10

AGGGTTTACGTGTATTGTGGTAAGTTGGTGGAACCTCAGGTTCAATTTGTGGGTGAG
CACCACCTCCTTCAATTTGTGCGAAGTTGCTTCCGGAAAATGAGGTCAACTTCTTCT
TTTCTTTGATGAAAACGAAAGTCGGCATAACAGAAATCTCGTAGGCAGCAGCAA
CGTCTTCACACTCGTCGACATCGACCTCAGGAAACGATTTTCAGGGTGCGCCTGGG
CGAGTTCCTCCAGTTTAGGAGCTATCATCTTGCATGGGCCACACCAGGTTGCGTA
GAAATCGATGACAACTAGGTTGGTTCCTGCTTCGGTAAGCTTAGCTTCCAAATCA
TCCCTTCGACAGACTAAAATTCGCTTT (SEQ ID NO: 70)

contig04863 length=265 numreads=3

TCACGATTTCAAAAGCCGAGGAATCGAAGTGTCTAATGTAACCTTGAACCTTGGAG
GCCATGATGGCTGCCAAATCCGGTTCAGTTAAGGCCCTCACTGGTGGAATCGCGC
ATCTTTTCAAGGCAAACAAGGTTACCCTGGTGAAAGGACACGGAAAAATCACTA
GTACCAACGAGGTCATGATTGAAAGAGGATGGATCGAGCGAAGTCATCAAAA
CAAAGAACATCCTCATCGCTACTGGCTCAGAAGTCACACCTTTCCCG (SEQ ID
NO: 71)

contig06115 length=1689 numreads=200

GGCCGGGACGTTGAGCCTGTCGCGTCTTGCGATCGTCAAGCCAAATTTGGGGTCT
GTCGTTCCCTGTTTGACGGTGCAGAGCAGAAATCAGTCGACTGAAGCTCCAGTCA
CCAAAGCGAAATATGGACCTCTTGCCGACGAGGACAGAATTTTACCAACCTGTA
CGGACGACATGATTGGCGACTCAAAGGTGCCTTGAAAAGGGGTGATTGGTACAA
GACCAAAGAAATCATTGAGAAAGGACCTGATTTCAATTCTGAATGAGGTCAAAAC
ATCTGGCTTGAGAGGACGGGGTGGAGCAGGCTTCCCATCGGGCATGAAGTGGAG
TTTCATGAACAAGCCATCTGACGGAAGGCCCAAGTACCTCGTCGTCAATGCTGAC
GAAGGTGAGCCAGGAACCTGTAAAGATAGGGAGATCATGCGTCACGACCCGCAC
AAGCTTGTTGAGGGGTGCCTGGTTGCTGGGAGGGCGATGGGAGCTCGTGCTGCTT
ACATCTACATCCGCGGAGAATTCTACAATGAGGCCTCCAACATGCAAGTGGCCAT
TGCTGAGGCCTACCAAGCTGGTCTGATAGGCAAGAATGCTTGCGGATCAGGCTA
CGATTTTGACGTGATCATGCACAGAGGAGCCGGAGCTTACATTTGCGGCGAAGA
AACTGCTCTCATCGAGTCTCTGGAAGGTAAACAAGGAAAGCCCAGGCTCAAGCC

GCCATTCCCTGCTGACATCGGTGTCTTTGGATGCCCTACCACTGTTGCCAACGTCG
AAACCGTGTCAGTCGCACCTTCGATTTGCCGCAGAGGTGGAAAGTGTTTCGTTGG
TCTGGGCCGTCCTCCGAAATTCCGGAATAAATTGTTCAACATCTCGGGTCACGTC
AACAAATCCTTGCACAGTGGAAGAGGAGATGTCAATACCGCTGCGCGAGCTGATA
GAACGGCACGCCGGAGGTGTGCGTGGCGGATGGGAAAACCTATTGGCTGTCATC
CCTGGAGGGTTCATCTACACCTCTGATTCCACAAAGTGTCTGCAATGAGGCCATAA
TGGATTTTCGATGGGTTGGTTGCGGTCCAGACGAGCTTGGGAACTGCTGCTCTGAT
TGTGATGGACAAGTCGACGGACATTGTCAAGGCTATCACTCGACTCATCATGTTC
TACAAGCACGAGTCCTGCGGCCAGTGCACACCATGCAGGGAAGGAATAAGCTG
GATGAACAAGATCATGCACAGGTTTCGTAAGTGGCAACGCTCAGCCCTCTGAAAT
CGACATGCTGTGGGAACTGAGTAAGCAGATTGAAGGACACACCATCTGTGCTTT
GGGAGACGGTGCTGCCTGGCCTGTACAGGGTCTGATTCGCCACTTCCGACCAGAG
CTGGAAGAGAGGATGCGCGTCTTCCACCAGAAGGACAAGAAAGCAGCAGCCTCC
AATTAATCTCCTCCTCACTCCAAAATGTAGAACTCCACCTCACTATTATGGTA
TTACCTCTTGAGAGGAACTCTACTAGGATAAAGCATAATAAATAAGTACCTAGCA
TTCGAACTCTGCTGTACGCGCTTAACGTTGTAATTTATTGTTGCAAAACAAGTAG
GGGATTTAAAAACAGGAATGTTGATTCTATTCAAGTTCTGCTTTATCCTAAATATC
ACACATGAAATGTATAATGAACTCGAAAATGTTAATAAAGTTATGTTTCAA
(SEQ ID NO: 72)

contig05464 length=362 numreads=6

TTTTTTTTTTTTTTACTAACAAAGTGTTCATGTATTTGAACCGATAAAACGAGAA
ATGAGACGATCAACAATGACGAAATATAGGTAGTTACTTATTTTGGCGTTGGTTG
ATGCCAGGATGAAACGGATGTTTGATAACTGACTAACGAAGCGAGTGGTGACAC
CCCAGAAGGAACTGCCGAGACACTTCACTTCTTTTCTTTAGGGGCGTCCGCCTT
CTCGACGGGGATTTCCGAAGAGGCGTCTTTGAAGTTGGGGAAGTGTCCACGGA
GTCGAGAGATCGAATTTACGGAAGTCTTGAGCCAAGTCTAGCGGCTCGCAAACA
CCCTTTTCTTTCATCGTCGTAGCGGACCTCGA (SEQ ID NO: 73)

contig01256 length=352 numreads=28

ATACTAAAAATTATCATATCAGTTTCAATAACAATTATATTATCAATCATCATAA
TAACATTATGTATATTAATCAGAAAAAAATCAATAATAGATCGTGAAAAAATAT
CACCGTTTGAATGTGGGTTTCGACCCAAAAAGATCATCACGAGTACCTTTTTCTT
CACAATCTTCTTAATTGCAGTCCTATTCCTAATCTTTGATATCGAAATCGTAATT
ATCTTACCAATAATCATCACGATCAAAACAAGAAATATAATATATTGATTTATAA
CATCAACCATATTCTTAATCATTTTAATTATTGGATTATATCACGAATGAAAAAA
TGGGATCCTAGAATGAGCCAAC (SEQ ID NO: 74)

contig05979 length=375 numreads=12

TTCTTTTATTTTTTGTGATTAGATCTTGTGTTGGCAGAAACAAATCGGTCACCCT
TTGATTTTGCTGAAGGTGAATCTGAGTTGGTTTCGGGATTTAATGTCGAATATAG
AAGAGGGGGATTTGCTTTTATTTTTTATCAGAATATATAAATATTATTTTTATATC
TTTGTTAACTTGATTGTTTTTATAGGTTGTAATATTTATTCTATCACCTTTTTTG
AAGTGGGTTTTTTTGGTTTTTTCATTTATTTGGGTTTCGCGGGACTTTACCCCGCTTT
CGTTATGATAAGTTAATATATTTGACTTGAAAGATATTTTTGCCTTTATCTTTAAA
TTATTTGTTATTTTTTATTGGATTATATTTCTATATAA (SEQ ID NO: 75)

contig02238 length=224 numreads=7

TGAATTTTCACCCCATCATCATCTAAATCAAGCTCCCATGCTTGGAAGTGGTTTCC
CTGAGCAGCCAATAGAGCTTTAGCTTTTGTACACCAGGGGCACCAAGACTTGCTG

AAAATTGCAATCTGATTACCCTTAACGAACTCTTCGATTCTGTTGACAAAAATCT
TGCTGTCTGTGTTTGAAGAATCAGATGACCTTCCAACGTTGATCAATTCCTTAAGT
CG (SEQ ID NO: 76)

contig06162 length=307 numreads=20

CGGCAAAGGTCCAGATCAGCCATGCGGTGGCATTAGCGGAAGATACCTCGGGCG
GGGCAAATGCGGCTCCGGAATGGAATGCAGCGGAGAATACAAGGGTTACTGCAA
GCACTGCTCATCGGTAACGCTGGAATGTGCCAAATGGCAGCACGCACGCAGAGA
ATTGGATGATATTGACAGCGGGTTGATGGACTACGGGTTCTGAGCAATAACAAA
ATACACCCAAAGTCTCCACCAACTGCGACTTCAAAGAGGAAGATGAAGACTATA
TCTTGTACCATTATTTACGACATTAAACATGCTCGTA (SEQ ID NO: 77)

contig01186 length=1349 numreads=17

CTTGAGAGAACCTTCCTTCACATAGTTCTCCAGCTCTTCTTCGTATAAGAAGTCTT
CGCTCTTCTTTCTGCATCCAAAATACAGTATCGTTTCTCCAACCTGGTTTACTGTCT
TTTTTGGCGAGGTGGCGCTCCTGAATGAAACCCCGGAAGGGAGCAAAACCTGTT
CCTGGCCCCGATCATGATAATTGGCGTTTGTGGCCTTGTAGGAAGTTTGAATTGCG
ATTTTCTGACGAAGATAGGAGCAACCACAGGGGCGACGCCATTTTCTGAAGGTA
CTTTGTTAGACAGCCAGGACGTGCACACACCCTTGTTCTGCGACCAGTAGGAGT
ATCGTACTCTACTTTGACTGCCGTCACGTGCACTGCCGTGGGAGAAACCTTGGGG
GATGACGAAATTGAATAGTAACGACACTGGAGTCTGGGAAGCAACTCGCAAAGG
CGATCCAATTGAGGTTTGACGGAAGGAAGATCTTCAAGGATATGGACGATGTTA
CGGTTGTCCTGAACTATCCACTGTTGGTACAGAGCTTTGCCCTCCGGGGTTGTGCT
TGCCATAAGCTTCAATTTTTTCAGCTTTCCTTGGGATCTTTAGCGTACTCTGCCAATT
CTTTGAGGATGTGGGTTCTGGGGTTTGACGTGATGTCGATGTAGTAAGTGAGAGC
CGTCTTGTAGGAGCACGGACAAGGGAATGGGTGCTTCTTACTAGATTCCCTCATCT
GTATTCGTGAGAGTGAAAATAGTTTCCAAAGCTGTATTCGACATATCGAGACATT
TTCCAATCCTGTTTACAAGCTCCTCGGAGTTCTGTGGGTATATCGCGAGATGGTC
GCCAACATCATATCTGATTTTAGAGCCCTCAATGTGCAATTCAATGTGCATGCAG
GAACGATCGCCTCCTTTGTACAACTCGCGGTTGACTTTGATCGGTGCCAGGTAAG
GGTTCTTCAGGTCGAATGGTGGTCTTTGATTGGTGAATGAGTAAAGCCTGGCGAC
CTCTCCGGTGTACACACGTTCTGGGATCTCCTGGATGTGTTCAAGTGAAGTTTGTATT
GACGGATGGACACATCTTCTCCTGCTCCTTCGATACCGAAGTACTCACATACAGT
GGGCCAGAACTTGTCCTTCCATGTGATGAAGTCATCTTCAATGTTAGCATCATCG
TCTCCAAGTCCAAGTTCAATGACCCTGGTGGCACCCAGTTCTTCGAGGCGTTTAT
CAACGTAGATCCCCACTTCGTTGTAATGCTCGTACGTTTTGTTCCCCAACCCAAAC
ACAGCGTAGTTCAGGCCAGTGAGATCGGCCTCTCCGTTTTGTAGCCACTCGTAAA
AGTCCATGGCGTTATCAGTAGGGTCAC (SEQ ID NO: 78)

contig01654 length=356 numreads=6

TTGAGGACCGTAAAAAACTGGATGGACTGTACGAATGTATCCTATGCGCCTGCTG
CAGCACCTCCTGTCCATCATACTGGTGGAAACGGTGACAAGTACCTGGGGCCCTGCA
GTGTTGATGCAGGCTTACCGCTGGGTGTCGACTCTCGCGATGAAATGACTCAAG
AGAGATTGGAAGGTTGAAGGACCCTTTCTCAGTTTTCCGGTGTACACCATTTCT
CAACTGCACTAAAACGTGTCCTAAGGGTTTGAACCCGGCCAAAGCCATCTCCGA
GATGAAACAACATCTCTGGCACAAAGCACAAAGCCGGCCCCCTGAACTCAAAGC
TGTGTAATAATCGCCGACACCGTTACCT (SEQ ID NO: 79)

contig00571 length=255 numreads=44

CCGGGAAGAATTTCGATAAGAACAAGGCTGTGAGGGACATTCGTGTGATCGATAT
GCTGGTGATAAAGGGTCAAATGGAGCTTAAAGAAACAGTCAACATATGGAAGCA
GAAAGGTACTGTCATGGCTTATTTCAAGGAGACTATCGAACCCCAGCCCAAAGA
TTTCTTGTCCAAGTTCTTCTCGAATGTAAAGTAAAAGATTAGTTCATTCAAATTA
TGTAATAATTGTTTTAATACTTAGTTGTTGGTTCTT (SEQ ID NO: 80)

contig04258 length=340 numreads=15

CGAAGACCATCTTGGTGCCGGGGATGTATTTCTTGGGGTCTCGAGGTACTCGAA
GAGAGTGTCTTGCTCCAAGTGATGCCTTTACTGACGTTGGCGTCGGAGTAGGAG
AATCCTGGGGCTTGACCAGTCTTCCTTCCAATAAGACCACTGAGGTTGGGTCTTA
CTTTATGTTTGCCTCCCTTCTCCACCGTGTGGCACTGGGAGCATTCTGGACGAAC
ACCTTCTTTCCCTTCTCAGCATCTCTGCGGGAACACCCATGACTTCGAGTTATGG
ATTAGGTAGGTGGCTGAACGTCCGAAAAATCGAAGAGGTGCGTCGACTGCGCGA
TCACTGGCA (SEQ ID NO: 81)

contig06671 length=469 numreads=16

AAAACCATTAAGAAATAATATCAATTTTAAAAAAAAGATATTTAAAGGAAAT
CAATGTAAAAATAACAAATAATATTCAACATAATAACCATAAAAAACCACAAAAC
AAACCCCTATAAAACAACCCCATGATTAATAGAAGAATATAGATAAAATACTATAT
ATACATCCTATAAAAGAAGCAAATATCAAATAGAAAAAAGAAATATATCTCCAA
GATATAATTCCATTAATCAATAAAATTTACCTATCAAATTCATTGAAGGAGGTC
TAGCCATATTATTAATCATCAATAAAATCAAAATAAACATATTCTAGGTATAAA
AGTTAATAAACCCTTATTAATAAAAAATTCTTCGAGATAAAGTTCGTTTCATAGGCA
ATATTTGCTAAACAAAATATACCAGAAGAACATAAACCATGACCAACCATCAGA
ATCAATGAACCAATAATACCTAATATATTCACAC (SEQ ID NO: 82)

contig06264 length=792 numreads=99

CAACGCAGAGTGAAATCAGATCACACCCTTGCTCAAAGTGCGATTTTGTAGCTCG
ATGTCCTGTAGCTTGAGAAATCATCTGACGTTGCACACGAACGAGAGGCGCTAC
TCGTGCAAGAAATGCAATTACAAAGCTAAACAATATTCTCATCTGAGGAACCAC
CTGTTGATACACAATTCCATCAAACCACACGCTTGTCGCGCACTGTGATTACCGAG
GAAAGCGCATAGATGCCCTCAGGAATCATCTTTTGAGTCATTCCAACAATGCTGC
CGCGGAAACCTACTTATGTGAAGACTGCGATTATAAGACATCTAAAATGGCCAA
TTTAAGAAGCCACAGGTTGACCCATTCCAGAATCAAGTCATTCTCATGTTCCATT
TGTGACTATAAATCTTCTTATCAGAGCGACATTTCAAGGCATACGAACAACGTTT
ATAACAAACCTGATAAACGATTTTTTTTGCACGGTTTGCAACTACAAAACGTCTAG
GGCGGATCACTTGAAAGCTCATGTCAAACTCATTCTAGTCTTAAGAAAAAGTGT
CCTCATTGTGAATACAAAGCAGTAAAAAGTATATCTGTTAGAACTCATATCAAGA
AAATCCATCCGGACAAATCTTGAGATTGTTATTGCTTCTGAAACCAAGTTTTACA
TGTTACCATAGAAAAGCTTCGAGCTTCGACTGATTTTGAGTATTGGAAAGACTAC
CTACGCTTTCCTTTAGCTTTACAAAATCTAAAAATTGTATAATTCTTTGAAAAAGA
TAGCAATAAAATGATATTTTTTTT (SEQ ID NO: 83)

contig04729 length=261 numreads=3

TCCGAAGGAATTCAACTAATCCGAAGGGTGTTTTAGGGTCTAAAGCCACCGGCG
AGCCACCGCTGTGTATGACGATGAGCCTGCAGCACGCACTTAGACAAGCATTGA
GGTGTGCGAGAGTGGAATCAGGCCTAGCAGACGAATGGTTCGACTACCATCAAC
CATGCACTTTTGAGAAGACATTACTAACATCAGGAACCAGCCCGAAACAGTTTA

ATTTGTGACGGCACCACGATAAACTCAACAAAGTTGGAGGGTTCG (SEQ ID NO: 84)

contig01756 length=577 numreads=19

AAATTTATTAGTAATATAAATAATGGGAATATTATATTAATATATAATATATTAG
TCGTTGATATTTTGTGTTTATTTAAAATAACTCATTTTTGTGAGTTTCTTATTA
TTATGTTGATAGATCTAATTATTCGTATATAATATGAAAGTGTGACTAATGATGA
CAATATTATCACTGTTAATATTATTATTTTCATTTGAAATAATTATATATTCGATAG
TTATTCATTTTGGTATGAACCCAAGAAAAGGGGGCATCCCTCCTATTCTTATTATT
ATAATTATAAATGATAATTTTTCTGTTATTCTTAGGGATTAAATATTAATGATT
AATATATATAATATTATATTTACTTAAATTTATGCATATTGGAATGATTATTAGTG
AATATAATATTATATATATTATTCATAATTTTTTATTGATTATAGCAGCTGCTATC
ATTCAACCTACATGATTAATAGATGAGTAAGCTATGATTTTCCGGATTGAGGTAT
GATTTAATCCTAAAATTGCACCTATGATGGTAGATAATATAATTATCATTATTATT
AATTTGTTGTTTTCTATAA (SEQ ID NO: 85)

contig01582 length=718 numreads=22

TACATTTTTTCGGATGTTATAGTGCAATTTACTTGATTTTTTCGAATTAATTTGGTG
GTGGGAATTTTAGAAATGGCTCTTTTGAAGCCTACCTTGTTGTTTAGCATTACCCG
TCGATCCATTGGCTCCTCCCATGTCTTGGGAGCTTACTGGAACAAGGACTGGGAG
CCTGGCCCCCTATCCACGGACTGAGGAAGAGAAGGCGGCTGCAGCAAAGAAGTAT
GGTTTGTTGAAAGAAGAATATATACCCATCCCCGATGATGGGGACGGTGCTGGA
GATTACCCTGCATTCAAGCCGTATTCTGTGGAAGGTAGAGACCCCCATTACCCTT
GGGATTTCCCAGAGTACCGAAGGAACCTCAATGAAGCCATGCACGAGAACTTCA
ATATGTACACTGAAGAAAAATATAACGTCGAGAGGAAATTTTCGGTATCCGATGA
GTGCCATGTTGGCCGCCTTCCTTGGAGTCATGGGAGGCACTTTCTTCCTGGTCGA
CCAATGCAAATACACCGCGTTCCTTCCTGCTATGCCCAGACACCTTCAGAGCCT
GGAGTCTCGCACTACAAGTACTCTCGCACTTAAGTAACGTACCCCCATATTAACA
TAGTTGTACAATATTTCAACCCTTTCCTACCCTATCTACCGATTGGTCTTAGACCA
TTGTAATCATTTATGGGGGAAGTGTAAAATTTTCGTTATTAATAAATGTTGATATT
TAC (SEQ ID NO: 86)

contig02061 length=508 numreads=5

ACTCTCCAATATTTGATGGCTGGAGCATCGGCTGTTTCAGATATGTAGTTCTGTGC
AAAATCAAGATTTACCGTCGTGGATGATTACATCTCTGGCTTGAAAGCTCAAAT
TTACATTAAATCCTTGGAACACTCCGAAAAATGGTTGGGACAGTCACCACCTACC
CAAGTCCACCAAAAAGGAAAACCGACTTTGACAACAAAGAACCAGAATGGGAA
GACACTGCCTGTTTTTCGGAAAATACAAAGAAGAACGAGAAAAAATCATCGCTCT
TTCAAAGAAATCAGGACAAAAAGCGAAAAACACTCAAGCCCCGGGTGATATCTGA
GCAAGTGAGGCCAGTCATGAGGCTCGAGGAGCTGATAGGCTCCTCATTGGAGCA
CATCACGAATTACAAAAGTCTGGACAACACCGAACAAGTTGTAGCCCTAGTCAA
CGAGGATATGTGCATCAATTGTGGGAAGTGCTACATGGCATGCAATGATTCTGGT
TATCAAGCAATCACATTG (SEQ ID NO: 87)

Known chemical insecticide targets:

contig01103 length=312 numreads=15

TATCTGGCGAGCGAGCAATGTGGTGTCTCCTCTCGTTTTTCTGTGGTCTGTCGCC
CTTTCGCGCGGGGCGTCGGCGCCAGTGGTTGACACCCCGGCGGCGCCAGTCGTCTG
TCACCCCTTTGGGGAAAATCAGCGTTGGGAGAGCGATTGGGCGACGCCGAGCG
GGCGCGCGTACCAGGGTTGGACCAAAGTGCCCTACGCTCAGCCTCCAGTTGGAA

GTCTCAGATTCAAGCCACCCAAGCCTTTAGAAAAATGGGACGGCGTGTTGGAGG
CCACAGGAGAGATGCCCCATTGCATACAGCGCGGTCCTG (SEQ ID NO: 88)

contig05244 length=510 numreads=9

CGTGATGCACGGGGACGAGATCAATTACATCTTTGGCGAACCGCTGAACCCGAC
CAAAAATTACCAGCAGAACGAAGCCGACCTGAGCAAAAGGATAATGAACTATTG
GGCGAATTTGCGCAAAACGGGGAAACCCAAACCTGCGGAGTAATAATTCGTGGAC
ATCGACGTACTGGCCGCAGCACACGGCATAACGGCCGCGAGTACATCACCTCGA
TATCAACACTACCGCCTCGACGGGGAGGGGCCCGCGCGTGAAGCAGTGTGCCTT
TTGGAAAAAATATCTCCCCAGCTGATTTCTATGACAGAACAACGACAGCAAAAC
GTGCAAGAATGCCCCGCCCATAGCTTCTGGAGTTTCCCTGAGGGACTCGTACACGT
ACATCCTGGCGCCCATCTTCAGCTTCTTCACGTTCCCTCCTGGTCCTCCTGACATCT
GTGTAAACTGACAACAGCCGAGGAGACCACCCGAAATGAAATTTTAAGAGAGAG
GCACCAGAAAAACCTACAAAT (SEQ ID NO: 89)

contig00372 length=361 numreads=4

ATCCAAGGTTTGTGCGAGGGCGCATCGACAGATCTTCAGTAGCAAACGCGAGCTC
CGAGCCTTCAGGTTTCGATGACTGGCACGTAGGGGATGGTGTGAGGTGGACTACC
ATCGTTTGGGCTTGTCCAATAGCCATTAAAAGAGCCTTGGCGTCCACCTCCTGCA
AGCACTTTAGAGTTTCTTCGTTGTCTTCCGGGAGCAGCCTGTCAATTTCACTACA
TTCATCGTGGTGGATTTGACTTTTTTCGGTTCGCGAAGAGAGACCAAACGTGGTTGA
TAGTCCCTGAATGTGATACAGCCCTCGTCACAAGGTCTTCGCTTAGTGGAGATT
CAGCAAAAAGTGCGACGAGGCGCCCCCAGCAG (SEQ ID NO: 90)

contig05930 length=890 numreads=27

TGTTTCCTTGGGTGACACAGCGTACAGACTTCGATCGTTGAAATAGTATAGCTTT
TGATCGGCGTGCAGTGCAACATTTATGCTGTTTCCTTTAGGCCGTAAAGTTCTTGA
GCTCGAGCCTCGAGATTTGCGCAACATACTCAAAATGGTAGTACCACTTTGGACC
TTGGTGATACTGAAGATTGTAAAGTGCTGGGAAGAAAAACATGGTATCAGCTCC
AACTTTTGCTAGTTCGATAGAGAAATTTTGTAAGGGCTCCATAACGGGAAAGTAT
AATTTCTTCAATATCGCAGCTCCTTCTCTCATTCTCTCTGGATCCTCATAGCCCAA
ACGAGTGGCGATTACCCGTTCCAAGTAAGCGTCCAAATTTTGCAGTGCATCGGAC
ACGTCGTTAGGGGACATCATAAAAATATTTAAGAGAAATTCGCCATTAGCGTTAC
TCGTGATCCATGGTTTATTGGAAGGAAGCTCGAACAGGTCTTCTTTGATCACAGC
ATCCTCCGCGTCTCAGGTTCAAAGCAGGGTGCAAACTATAAACCCAGACAG
TCCGTCTGGCATTGCACCAATCAAGTCTTACCGCTGGCTCCTTGAAGGCACTCG
AAAATCTCCAAGTTGCTGCTTGTGATTTCGACAGCCAAGTTTCGCAGCTAACTGCA
ATGTGCCATTTCTGATGATTTCCGGTTTAATTTTCGGACCATATGTAGTTAATTGTT
CCAGAATCAGAAAGTGCCCTGTGTACCATCCCTTCAGTAGCTGGTGTCTTCAACA
GGATATGACTGCTAGCACCTCCAGCTGATTCACCAACGACTGTAACGCTATTCGG
GTCACCACCGAAAGAAGCGATATTCTGCTGCACCCATTTCAACGCCATAACTTGA
TCCTTAATC (SEQ ID NO: 91)

contig03360 length=249 numreads=6

CGTCGCAGTGATGGTTGGGCTAACTTCTCTGTGGCGCTCAGCGCCGTTTTTCGCCT
GTTTCGGTCGCTCAACATGAACAACAACCCAAGTAACGACGTCTCTCGGGACGTTG
CAAGGATCGGTTTCGGCATTCAATTTTCAGGCAGAGAATTCCACGCTTTCGAGGGCA
TACCCTTCGCAAAACCTCCAGTAGGAAAACATCGGTTCAAGCGAGCTGTTCCGAA
TGGCCCGTGGACTGGAGTCCTCCACGCAA (SEQ ID NO: 92)

contig05458 length=272 numreads=7

TCCAAAACGGCCGACATCGAGAGGATGATCAGGTCCTCCGCCTCTTCGGCGACTT
CGTCCTCGAAGGGAAGCAGTTGGTCGAGCAGCCAACGAGCCGTGCCAGGAGCGT
CTGCCGAGAGCAAACGGAGGCTAACGGAATCCGCGACCCGGGCGCTGACGGACG
CCCCGCCACGCGCCCAAGTCGGCGTCTGGAAACGCAGGGAACATCGCATCCC
AACCAAAGAGCCCCCTCTAGTCATTTACCCGAAATCAAAAAACCCATCTTTTTAT
(SEQ ID NO: 93)

contig01440 length=249 numreads=3

TAGGAGGTGAGATACTCCTCTTTCTGGTTTGCATTCTCATAGGGGACGAACCCCC
TCCACTGCTCCTCCCCTTCCTCTCAGCTTCTCGGGCACTGCTCGAACTCATCTTCC
TTGGTGGAGAATCTTTATGGCTGCTGCTTTTCATGATGGATTCTTCTTGTCTTCT
GGGGATGACGAGGACTGACGCCTCTTCTTATCGTCGGTCTTCTTGACAACGCTGG
CACTCGACCTCGACGAGGACGAGCTGG (SEQ ID NO: 94)

contig02231 length=430 numreads=6

ACCTCCGTATCAAACACTACAAGGTAGTAACAAAATTTCATCATGATGACAGGAG
TCATCGTTCTGCTGCTCACTGTATTCGCGTGCTTCGCCGCCAGCAGTGTCGTCAGC
ATTCAAGGACCACGTGTGCGCCCGATCATAACACCCAGAGGATCGCAGTTGCA
CGTGGCAAGCTTGTGTGACGGATAAATCAGTTTGTCTGTCATGCCATCCCTCCTC
GGCGTCCCTCAAATCAAGTGCAATCCCGTTAAGGATCCGAAACAAGACGTTTTCC
CGAAGTGCATAGATGTGATGTTGGGATGATCATCATCGACGGATTGATCAAGGC
GATTGATCCAGATGGAGAAACTGAGTCTTTTTGGACGTATTATATTATTGTCGCC
TAAATTTTGTATGTAATAAAACATTTTAATGCATAATAAAAATTCT (SEQ ID NO:
95)

contig05130 length=260 numreads=2

CTCGACGAAAGCCTGGTAGTAAAATTCAAGGGAGGGGCAGACCTTGAAACCCTT
GGACTACAGACGGTTTCGTTGTCTGACATCGAAGAGGATGACGAGCCCCATCCG
GGTAAAAGCTTCATGCAAACCTACCCACGTCTACACGTATACCAACAGCACAGTA
ATGCATCCGGACGACTGTACAACCTACAACCAACGTAGTTTCCAGCTACCGAGGA
AGCCGCATGAGCACAGCAGCCAGCAGCAGGATGAGTTCCGTGTG (SEQ ID NO:
96)

contig00542 length=270 numreads=3

ACCAATGATTTTGACGACTCTCCAGACGACGGGTCGCCAGTCGATCTCAAGTTTG
AAGTACTTAAACGGAAAATTGGCCAGCTTGAAGATGAAAACAAGAGCCTGAAAA
ATGAGACTTTCCGCCTTTCTGAGGAAACGGAGGCTTGCGAGGAAGCTGAAGCCA
AACTCGTGGAAGACCTTGTGGCACAACCTCGAAGCCTGCAGGTTTCGATTTGGATT
TACGGGTATCGACATCGAAAAATCCAAACAACGAGAGGCTGAGCTGAAAGAG
(SEQ ID NO: 97)

contig04701 length=259 numreads=2

GCCAATGACGACATCTACTGACATTCTCACCTCCTTCGGTATGGACACGCCTGAG
AATGACAGAAACGTCGAAAACCTACCCGATCCCGAAAGAAATTAGTTCACGTTCT
AAGATTAGAGAATTTGATGAGAGGAAAAGAAAGGTACCGACAGGAAGGAAATT
CGGTCCTCTGGCTCCCATACCAAGTTCCCAAAGAACTGGAAGATGTGGACATTACT
GAACTCAATTCAATTGAAGAGGTGGCTAAGCAGATAGCAGAG (SEQ ID NO: 98)

Glycolysis/gluconeogenesis:**contig05492 length=915 numreads=475**

CCTAGTTCCTCTTCGATACGAAGAATCTGGTTGTATTTGGCACAGCGTTCAGATCT
GCATGGTGCACCAAGTCTTGATCTGTCCAGTTCCAAGTCCAACGACCAGATCAGCG
ATGAAGGTGTCCTCAGTTTCTCCGGACCTGTGCGACACCATGGTGCCCCAGCCGT
TCTTTTTGGCGAGGTTGTGGGCTTCAATCGATTCCGTCAACGAGCCAATTTGATTG
ACTTTGAGGAGGAGACAGTTACAAGACTTTTTGCCACCGCCATTTCTATCCTCTT
GGGGTTCGTGACAGTAAGGTCGTGCGCAACAATCTGAATGCTGGTACTTCCAGTA
AGTCCAGACCAAGCTTCCCAATCGTCCTGGTCGAAAGGATCCTCGATGGAAACG
ATAGGGAATTCCTTGATGAAAGACTGGTAGAGATTAGTCAGCTCAGCAGGGCTG
ATGTAGGTGGAAGGGTCGGACTTTGGGTTCTTGAAGTCCAAGTCATACTTCCCAT
CCTTGAAGAATTCAGAAGCAGCGCAATCCATTCCAATCTCGATTTTACCGGTGTA
GCCGGCCTTCTCGATGGCAGTGGTGATCAAAACGAGACCTTCTTTGTTGTGCGAGG
ATGTTGGGTGCGAACCCACCTTCGTGCGCCACAGCTGTGGCGTCAAGGCCAAACT
TGCCTTTGATGACCGCCTTCAAGTGGTGGTACACTTCAGAACCCATCCTCATCGC
CTCGCAGAAATTGGAAGCTCCAGTTGGAAGGATCATGAATTCCTGCATGGCGAG
TTTGTGCGCGCATGGGAGCCACCGTTGATGACATTGAAGGCAGGGCAGGGAAG
GATGATGTTTTTGTGCGCAGCAAGTTCAGCAATCCATTTGTACAGAGGAAGGCCT
TTTTTAGCAGCTCCCGCTTGCATACGGCAAGTGAC (SEQ ID NO: 99)

contig01170 length=499 numreads=19

GATAGTTTTACTGTATTTTTCATTACTGACAAAATGTAGAAAATGAGCATTATGC
TGCTTTAAAATATATACTTTTTACATCAGTTTGTGTTCTGCTTATAGCATGGAAG
TTGTAACACTCCCAAGTGTTAAGTTGAGGTTCAAGTACAGATAACTAGTATGGGAG
ACAGAAAGCGACTTTATGGCAGTTTACTTGTCTTGGTCTGCATGTACTTGATGA
GGTCGACAACACGGTTGGAGTAACCAAATCGTTATCATACCAAGACACAAGCT
TGACGAATTTGTTGTTGAGAGGGATACCAGCCTGAGCGTCAAAGATGGAGCTGT
GGGTGTCACCAATGAAATCAGCGCTGACGACCTGGTCTTCAGTGTAAGCCCAAGAT
TCCCTTGAGAGGTCCTTCGGAAGCCTCCTTCACTTTCTGCTTGATTTTCGTGTAAG
TGGCTTCTTTGGTCAGCCTGACAGTGAGGTCCACGACTGAAACGTTGGGGACAGG
AACT (SEQ ID NO: 100)

contig06703 length=1080 numreads=132

AAAAAAAAAAGAACTATTGGTAAGAAAAGTTTACGGAGTAGGTACAATATGGTA
GCAAAGTACCTACTCAAATCAATAGGTACATAATATATACTCATAATATCATTT
AATTGGACATTGAATGTCATCTCAGTCGATCGAATGCAACAATATCGCAGCTTTC
TCAATTCATTGCACATCCGCTTGTAAGAACTCACAGTTTCACGGCAACTTTAAGGC
AAAATAAATTGTTGGAACGACGTTTTTTCCAGGTAAGTACGAAGTTCCGAACCT
AGGGTCCTTGTAAGCTAAGAAAAACAGTGCGATGGAACGAAAACAAATATCAT
TACAACTATATACATAATGTAATTTTACACAAGTTATTTTTTTTTGTTTTTTAAAT
TTTTTTTCTTTTTCATTTTTCATTTTTTGAATAAATTAAATAAATATCTTACTCA
TTTTAACGAACCTGATAGGGCAAATAATAAGGTGTCAGCTACAGAATGCGAAAGA
ATTGACTATCTCTTAATGCTCGTAATAAATTCTGTTTTTTAATGATAAAAATTAT
AAGGTGAACCTTTTAAATAAAAAATTCAAAAATCACTAGACGAGTTGAAGCACTA
ACGTGTTGATGTGGGATGAATAGGTTGTTTTAAAAACGTAGTCGGACATCTTGAA
GAAAATCCGATGAATAAAAAAACTCCACGGAGGAGTTATCAGCGAATAATTCT
AAACTGTTTAAACCTTCTTGCGACTGTGAAAGGTAACATTTTTAAAGGTTATATTT
CTTTGTGATGCGGGGTTCGGAGATGAGGGTGAACCTTGGGGAAAGTCTATTTAGA
GGGACATGGACTCTTCCAAGAGGGAAGCGAGTTCTTTGAGATGCATCATTTTGA
GCCGGTGGCCGGAGAGGCGGCAACGGGTCTGCCAGCATAGCTGACATCCCGACT

CCTGCTGAGGGTGGTTTGGAACTGGGGCAGACGTAAGACCATGCTCCCTGGTTC
TTGTGCTCTTCTTGGGCCCAGCACAGCTTGGCATTAGAGTACTTGGCGCATTCCTT
CTTGATGAGGTCGTAGGGGAAAGGAGAAATCTGTT (SEQ ID NO: 101)

contig06954 length=149 numreads=90

TTGATGGCTCCTGCAGTTCTAATGAAAGATGCGAAGACCATGTGCGACTCCTTGcT
GAACTCCAAACTCCAAATCCTGTTTgCCTTTTCAGACACGGCAGGAAGATCGAC
ATTGATGCCTGGAAGgTTGACACCTTTCTTGCTTCCCAA (SEQ ID NO: 102)

contig06212 length=1248 numreads=63

ATAGTATTTTATTGTAGATTTATAAGACGCATGGAAAATATGGGATAAGGGGTAA
ATTATTGATAGTTCAACGTGAGGAAATCGTAAATGATTGATAATTAGTAAGCGTG
GTTTTTGATAAACAGATCTTGGTCGGCTGCTGATCCTTTGACGGTGCCGGCGACG
TACTTCCCAAGGGCGGCCTGTCCGTTGGCACGCGCCCTGTTACGACGCTCCTGCT
GAGCAGCTTTCAAATTGCCATCTTTGCCTTCCAAGCGCGAGGACGGAGGCTTGGA
GAGCTCTACCAAACGAGAATGTGAGCGGCCAGGGTTTCTTTCCTTGGAATTTGTT
GATGGCATCCAAATTCAGGGTCGCTTCTTCCTCGGACTGACCGCCAGACAAGAAG
GTGATGCCCCGCGACGGCTGCTGGAAGTGTCTGGAGAAGGCGAGACAAGTGGCC
CTAGCTACTTCCTCTGGCCTGGCTCTTTTGGGGCATTGCATGCCTGGAGTGACCAT
GTTGGGCTTGAGAAGCGTTCCTCGAGGTAAACGTGGTGGTCGTTCAAGTGCCCTG
TACACGGCTGCCAGTACAGTCTCCGTCACTTTTTGCGCTCTGTCCAGGTCGTGCG
AGCCATCAGGCAGTACCTCTGGCTCCACGATTGGCACGATTCTGTTGGCCTGGCA
GATCGATGCGTAGCGAGCGAGCACATTAGCGTTCTCCAGTATGGCTTGGTACGAA
GGAATGTTCTCGCCGATTTTGATGACGCAACGCCACTTGGCGAACTGGCAGCCTT
CCTTCTTGACGAGGCGCATCTCTCAGCCAGATCGTCGAGCCCCTGAGTCGTGCA
TTCGTCTCAGACCCAAAAAGTGGGACAACCCCTTGTCGACTTTGATGCCGGGG
ATGATTCCCTTCTTCTTGAGGACGTCGACGAACCGCGTGCCGTGCTCCGTCTTCTG
GAAGAGGGTCTCGTGGAAGAAAATGACGCCGGAGATGTTCTCCGCCACGAGGTT
ATCCGACGTGAAGAGGACTTGCCGGTACCGTCTTCGATTCTCTTCCGTGTTCTCCA
CGTTGATATCCTTGAAACGCTTTCCAATAGTGCCGGTTGATTATCCGCAGCCAG
AATCCCTTACCAGGAGCTACGATAGCCAGCGCGATTTCCTGAGCTCCTTTTGG
AGTTCTGGAGGGGGGTAATTGAAATACGTCGTCATTCTGAAACTATCAATCAGTA
GGAATACAGGCTGAGGTCCGAGGAGAGGTTGTTGGGA (SEQ ID NO: 103)

contig06101 length=1620 numreads=159

ATGGTACCAGCGCCCTCTATTTGACGAGAAACGACGAAACCTTCAGTATGGAAC
AACGACGGTTTAAATCTGTCATTTTCTCCCGACAGATTCTCATGAAACGAATCCT
GTTTACACTTTACAGGGTCAAAAAACCTCAACAATTGATCACAATTTTTTCTGG
ATCATTA AAAATGATCGTACATTTACGTATTACTAGATTGAACGAGGTTGAAATT
ACATATTATGATGAAGTTGTGAACATTTTAACATGTTTTAACAAAATGTTTAAGC
CAGTCTGTACGAACCCAAGCTATGAATTGCGGGGTCATGTTGGTTGTTTCGTGTG
TTGATTACTGTTGATGAGTGAGCAGCCTGAGACCTTCTGGTTCTTCACCTTCCACC
TCATCACGACACTACATCAGGAAGGGATGGAGAGAAGCCTCACGACAGCCTCAA
GCAGTTCACCCTTCAACCTGAACTAACGTGAGGTGACTTCATTTTTTGGTAGGTA
CCTACGTTTTTGTAGGTTACAGGCAGTATATTTTTAAAAATAAAACATTAATGTCA
TACGGGTGACCCAGGATGAGTCCAGTACGAGATTGTGAAGTCTTTTAGATCTTTC
GTCCAAAATATACGACGGTAGGTACCTAGTAGTAGTTAGAATAAATTA AAATTAT
AGAATTGGAAGTTATTGTAGTCAACTTGAATACGTTCACTAAAAACTACGTAGAA
TGGCTTGCTGCTAGGAGGTGCTGGTGCGCTCAGAGGGAGAGGACTGAGAAGCTG
CTGTCGAGGTTCTTGGGGACGCTGACGACGCGGATGGTGTGTTGGTGTAGCCAGAA

CCGGACTTCCAGCCGGTGAGGACGACGACTTGGTCGCCTTCCTTGAGGAAGCCCT
TCTTGGAGACGACCTGAAGGCCGTGGATGATCCTCACGTCGACGTCTTTCAACCA
GTCGGTGAGCGGCTCCTTGTCAAACACAACAGGGAGGATACCCCGGTAGAGATG
AGCTTGCCTGGCGACCTGTGGGTGACGGGTGACAGCGATGATGGGGCACCTCGG
CCTGTATTTGGCGACGAGGTGGGCAGAGCGACCACTGGTAGTGATGACCAAGAT
GGCGGACGCGAGAGCCTTGGTGGCTGCTTCGACAGCAGCGATGGCTACGCTGTG
AGTTGCATCGATGGGTGGAAGAACCTTTCCAGCCAAGTCAGAGAACAGCTGATT
GTGCCAAATAGCGGCTTCAGCCTCTTTACAGATGTTGGCCATGGTCTGGACGCAA
TCCAAGGGATACTCGCCCTTGGCCGTCTCTCCCGACAACATGACGCAGTCCGCGC
CGTCGAGGATGGCGTTGGCGACGTCAGAGCTTTCAGCTCGCGTGGGTTCGGGGCTT
CTTGATCATGGATTCCAACATTTGAGTGGCGCAAATGACTGGTTTTCTTGTATTAT
TACACTTGGCAATCATGGCTTTCTGGGCCAAGAAGACCTTTTCAGTTGGTATTTTC
AATACCAAGATCACCCCTAGCAACCATGATGCCATCGGCAGCTTCAATGATTTCA
TCCAAATTCTGCATGCCTTGATGATTTTCGATCTTGGCAATGATCAGGATGTTCTT
TCCAGCTTCGCCCAGAATTTCTCTTATTTCT (SEQ ID NO: 104)

contig04894 length=217 numreads=6

AATGTCTAGGGACGACCTTTTTACTGTCAACGCTTCGGTTCGTGAGAGATCTGGTA
ACCGCGTGTGCGTGTTTTTGCCCTAAAGCCCTCATCGGAATCATCTCTAATCCAAT
CAACAGCACCATTCCTCATCGCTGCTGAAGTCATGAACTGTCCAACGTGTACGAC
CCCAAGAGAATATTCGGGATCACGTCGCTGGATCTCGTAAGAGCCAGTACG (SEQ
ID NO: 105)

contig04593 length=245 numreads=2

TTGGATGGGTTTGGATATTGGAAACAAATCGATTGATTGTTTCTCCCAAGTCATTT
TAAGGTCCAAGACTATAGTTTGGAAACGGTCCCCCAGGTGTTTTTGAGTTCGAGAA
ATTCTCTCGAGGGACTCGAGGTATGTTGGACGCCGTCGTCGAAGCTACTTGCGAT
GGCGCGCTGACAATCATTGGAGGGGGCGACACGGCGACTGTGGCCAAGAAGTTC
AAAGCAGAAGATAAGGTATCTCACG (SEQ ID NO: 106)

contig03352 length=389 numreads=12

CTCAATGCTGGGGTTGCAGAGTTCAGGGATCTTTTCGTGCATGATGTTGTGGAAG
TGAGGGTGGAAACGGTACAGTGAACCATCGACTCCAACGGTGACAGAGGGGATG
TCCATCTTGTTGATGAGCGTGGCCACGCCCGCTGAGGCAAGGTAGGCTGCTCGCC
GGCTCACGCTTTCGCAGATGTACCTGATGTTGGAGCAGTCTTGGTCGGTGGGGTC
GAGGATGTTTCAAGTTTCCCAGAACGTTCCGGCACTGCTTGAAGACTCCTCTGGGA
TCATTTTCGATTTTCAGAAATGTATTCAGTGAGGAAGCTCCATCGCTCTTGAAGTTT
GGGGTTGATTTGACCTCCAAATAATACTTTGTCTTTGGTGAACCTTGATGATGAGA
AGTC (SEQ ID NO: 107)

contig00761 length=218 numreads=49

TGGCCACAACCGGGAAACAAGACATTCTGGATATCGTTCCCACCAAATCCACC
AAAGAAGCCCCATCATTCTCGTTCCAAGCGAGATGTTCAAGAACTGGTTGACCT
GTACCAAGATCCGAAATTCAGGGATTGAAACTCAACTGTAAAGACTGCTAATTA
TTAAGTTATTCGTTGTTATGTTCTGTTAGACACAAATAAAATAATTTAAATTA
(SEQ ID NO: 108)

contig00842 length=255 numreads=3

CCGTCAACTCCAACAGTGACGGACGGGATGTCCATCTTGTTGATGAGCGTTGCCA
CACCGGCGGACGCGAGGTAGGCCGCTCGCCTGCTCACACTTTCGCAGATGTATCT

GATGTTCTGAACAGTCTTGATCGGTTGGGTCGAGGATGTTCAATTTCCCCAGAACG
TTCCGACACTGCTTGAAAACACCTCTGGGATCACTTTCAATGTCAGAAACGTATT
CCGTAAGGAATCCCCATCGCTCCTGGAGCTTAGGG (SEQ ID NO: 109)

contig04919 length=243 numreads=4

GTGAGGTGATGGATCTTCAGCAAGGAGCTGCATTACCCCCGTCCATTAAAGTCTA
TGGCAGCACTGACTATGCAGTGACCAAGGACTCAAAGATTTGTATAATCACTGCA
GGCGCTCGGCAGAAGGAAGGCGAAACGCGCCTACAGCTGATTGAAAAGAACGTC
CAGATTTTGAAGCAAGTTGTGCCTGAGCTCCTCAAGTATAGTCCCAACGCCATCC
TACTGTTGGTAACCAACCCTGTTG (SEQ ID NO: 110)

contig01640 length=286 numreads=4

TGCTCGTGGTCTCCTGTTCTGTTGGGTCAGGCGTGAGCGGCGGTGAAGAGGGAGCC
AGGCGAGGACCTTCTCTCATGCCTGGAGGCGAACC AAAAGCCTGGCCTCATCTTG
CCCCTATTTTCCAGTCCATCAGTGCCAAAGTGAGGGGCGAATCCTGTTGCGACTG
GGTTGGCGATGGCGGCTCTGGTCATTTCTGTCAGATGGTCCACAATGGGATCGAG
TACGGCGACATGCAGCTGATCTGTGAGGCTTACCATCTCATGAAGGACCTTCTGG
GACTCGACCAC (SEQ ID NO: 111)

contig06111 length=1429 numreads=121

TCTCGAGTCGCCGGGGAGCCAACCACTTCAATCTTCACTTCAATCGTCACTTCAA
TCGCCACTTCCATCGCCAAAGGTCAAAGGTTCAACCATGAGGAATAAAGACGA
AATGGCGCTGGCGATCGATGCAGCGATCTCGCAGGAACTGCAGCCTTTGGTACAT
TCCGACCCGAATGTTCAAAATAAGCTAGAAGAGATTCTCCACTGCCTGAGGCTCT
CCGGTGATGACGTCATCCGCGTCGGGGACGTTTTTCGGAGAGGAAATGACTCGCG
GCATGAGAGACCAACCTTCGTCCCTGCGAATGGAAAATACGTACATTCTGAAAT
ATCCTGTTTCAAGAAGGCGAGTACTTGGGGCTGGACCTCGGAGGAACGAATTT
CCGAGTTCTCCTGGTGAACTGAAGGACGGCGAGGTGGTGGAGGAGAAGTTCT
CAGCTACGACGTCGACACGGAGACCCGTCAGGGCGAAGGCTCAAGACTCTTCGA
TTTCCTCGCTGACTGCGTCTCCGACTTCGCGCACAAAGAAGGCGTTGCAGACAAA
GGCCTCAACATGGGTTTTCACGTTTTCTTTTCCCATGCACCAGAAGTCCCTCCACAA
CGGGATTCTCACCCAATGGACCAAAAGCTTTTCCGCCACGAATGTCGTCGGTAAC
GACGTCGTCAAGATGCTCCAAGAAGCGCTGGAGAGGAGGGGCGACGTCAAAGTT
AACGTCAACGCCATCTTGAACGATACTACGGGGACGCTGTTGCAAGGCTGCGCG
ATGGACAGAAGAACGAGAGTCGGTCTGATCCTCGGAACAGGAAGTAACGCTGCT
TACCTCGAGAGAGCGGACCGCGTCTGCCATTGGGAATCAGAGAGACACGGCGAA
CAAGAGATTGTGATCGACATTGAATGGGGTGCGTTTCGGAGACAACGGCGTCATA
GATTTTCATCAAGTCACACTTCGACCGTCAGCTGGACAACGGTTCATTGCACATCA
ACTCCTTCTCGTTTGAAAAGTACATCAGCGGTTTGTACCTCGGCGAGCTGGTGAG
GATCGTCTTGCTCCACCTCGCCAACAACAGGTTGCTCTTCAACGGCATCCTAAGC
AAGGAAATTCAAGTCAAGGACTCATTTTTGTCTTCATACGTCTCAAGAGTCGAAA
AGGACAACGTAGACACCCAATCTCATGAGACTGAAACGTTGCTGAGGAAATTCA
ATTTGGATTTCTGCGATGATGATGTAGCTGTGATCAAATATGTCTGCGAGTTGAT
TTCCCTTAGAGCTTCTCTCTCGTGTGAGCTTGTGCGCCCGGCTATTGAAGCAAA
TGGACAAGCCCGACGTAACAATCGCCGTAGATGGATCTGTATTCAAATTCCACCC
TCGCTACCGGTTGTGGATGGAGAAGTTCATCAGGTTATTGGCACCCAGCCATAAA
TTCCAAT (SEQ ID NO: 112)

contig04328 length=244 numreads=5

CACGCTTACTTTCCCTGTCATTAACAACGCTCGTAACTGCATTTTCGCGCTTACTG
GTCAAGGAAAGGCAGAAATGGTTAAGAAAATACTGAAAAACAATGAAAACCTG
CCCGCAGGAATGGTGCTCCCTCGTGAAGGCGAAGTCTTATGGATTTTGGATAAAC
CAGCTGCATCTCTTCTGTAATCAGTAATTTGAATCATTTTACCCCCTAAACTCGAT
GGCAAACCGCACAAATGTGACAAA (SEQ ID NO: 113)

contig00122 length=246 numreads=2

TCAACAAAGCCCTGGAGAATGAGAAAATAAAATTGAGCGTGAATGATTTTCGTCA
TTAAAGCTACAGCTCTCGCATCCCTCAAAGTCAAAGAAGCTAATTCTGCGTGGAT
GGGAGGTTTCATTAGACAATATAACGGTGTAGATGTGAGTGTGGCTGTAAGCAC
GCCCAAAGGCCTGATCACGCCCATAGTATTCAACGCGCACATTAAGGGGCTTATC
GAAATTAGCACTGACGTCAAATCGCTCG (SEQ ID NO: 114)

contig01176 length=711 numreads=8

CTACGAATCCTTTGACTGGGAGCACGGTGTGCTCGTTGGGGCGGGAATGCGAAG
CGAAGCCACAGCAGCAGCTGAGCACCGAGGAAAGGAAATCATGCACGATCCCTT
CGCGATGAGACCGTTCTTTGGGTACAATTTTCGGCCACTACATGTCCCACTGGCTG
TCCAGCAGACTCGCCATGGAGATGTCAAGCTTCCCAAGATCTTCCACGTCAACT
GGTTCAGAAAAGATGCTGACGGAAAATTCCTCTGGCCTGGATTTCGGTGAGAACG
TCAGAGTTTTGGACTGGATTTTACGGCGAGTCGACAACGACTGTTCCATCATCGA
GCCAACACCAATAGGTTACGTGCCCAAAGAGGGGCAGCCTAAACCTGGATGGCTC
TCTCCTCGGCCAGACATGAAGAACTCCTCAGCGTTCCTAAAGACTTCTGGAAGA
AAGAAGCGGAGTCCATCGAGAAGTATTTGAAAGAACAAGTCGACAAAGACGTTC
CCGTGGAGGTTTTGAATCAGCTCAAAGCCCTCAAACCTGCGCCTCAGCTAAACATC
ATTGACAAGCGGCCATCATCTACTGAAGTCATCCATGGCTAATCGTATCAAATCC
AACTAATTTTCATTTTGTAGTAACTCTGTTCCTGTCTGTCTGTTATGGAACCTCA
ATTCACACAATAAAAACAACAGTTGTCAATCTGGTTACAATGGGAAAAAAAAAA
A (SEQ ID NO: 115)

contig02838 length=330 numreads=4

CGCGATGTTGTTTCGAGTTCCAGAGCGCGCTGAAGATGCGGTTGGCGAAGCGCGT
CGTGATGATGTTCTGCACCATCTCCTTGCCAAGGTAGTGGTCGATGCGGTAAATC
TGCGACTCGTCAAAGTAGGGCTCGATCGCCTTGACACAGCTCCGCGGAGCTCTGCG
TGTCGTGACCGAACGGCTTCTCAATAATCACACGAATCCACCCACCTCCTGAGG
CATCGCGCCACCGCGGATGCTGCCGCACGCCCCCGCGAAGGCGCTTGGCGGCAG
GGCAAGGTAGAAGAGGCGGTGCGCGCCCTTCTCGGGCCCCTCGAAGGCGTCCTC
CAG (SEQ ID NO: 116)

contig04580 length=248 numreads=2

GGTGAACCCAGCACCTGGCGCCACCCGCTGACAAGCACCAACGGATCGCCAGG
CTTGATGATGCCCTCCAATTTCCGGCATTGATTCCAAATTGAAGCCTCAAGTCA
ACATCCAAGCACCAAGTTTGGATTAGGAGCCACAATGTAGTGAAGGGGCCAACT
GCTCTGAACATGTGGAGTCTTCTGGCGACGCTGCCATACCTGGTGATGGCGATAA
CGGGGCACCTGGGACGGTACTTGGCCAAAT (SEQ ID NO: 117)

contig00231 length=596 numreads=15

AAGAAATGGTTCAGAATCTTATGGTTCTAAGATTTCGGCAATGGCATATTCAGTCC
GACGTGGAATCGCTTGTTTATTGCATCGATCCAGATCACTTTCAAAGAGCCTTTC
GGTACCGAAGGAAGAGGAGGTTACTTTGACGAATTTGGAATTATCAGAGACGTT

ATGCAGAACCACTTGGTTCAAATACTCACTCTGGTGGCAATGGAAAAGCCTGCTT
CGATACACCCTGATGATATTTCGAAATGAGAAGGTTAAAGTGTTAAAATGCATCA
AGCCTGTCCCTATCGATGATGTTGTGTTGGGGCAGTACGTCGGTAACCCGGATGG
GGAAGGAGAAGCCAAAATTGGTTACTTGGATGATCCCACCGTTCCAAAAGGATC
AACCACTCCCACGTTTTGCACTGCCATACTCAAATTGATAACGAAAGGTGGGAA
GGGGTCCCATTTCATATTACGCTGCGGGAAAGCACTCAACGAACGGAAAGCTGAG
ATCAGGATTCAATACAACGAAGTTCCCGGCGATATCTTTGGGGACAGTGCAAGA
GGAATGAACTGGTCATTTCGAGTGCAGCCAGGAGAAGCAGTCTATGTGAAAA
(SEQ ID NO: 118)

contig01580 length=309 numreads=7

ATCATCGCTGTCACCCGTCACCCACAGGTCGCCAGGCAAGCTCATCTCTACCGGG
GTATCCTCCCTGTTGTGTTTGACAGAGATCCTGCAAACGAGTGGATGAGCGATGT
AGAAGAGCGAATCAAATACGGTCTGACTGTGGGCTTCACGAGGGGGATCCTCCA
GAGTAACGACCCAGTCGTGGTCATGACCGGCTGGAAGCAGGGGGTTCGGCTTCAC
TAACACGCTACGCCTCGTGTACGTCAGGGACGACAAAGACAAAATGGGACTCTA
ATCATCTGGAAATTCTCAATAAAATCTGGTCAATTTT (SEQ ID NO: 119)

contig05630 length=638 numreads=13

GTGAGTTGAACGATTTTATTTTCAAACCTTAAATTCAAGATATAGAATAATTTGT
GACACGAGTAAAGTAGGCCGGGTACAGTAAGGTAACCAGTACGGATACCAAATT
GTACCTAACCAATAAACAATTTCGAAATTGAATCCACCTCTTATTTGCTTTTCTTGA
CGAATTCCACCCCTTTCTCAATGCTCTTCTGCAGTTCTTGTGCGGCGTTAGTAACT
AACTCTTGCTCATATTTGTTCAACTCTCCGTTTCCCAAGTTTTAGCTACACCGTC
TTTATTTAGCAAAAGTTTACTTGAGAAATGAGACACGCAATCGACGACGTTGGAA
GCGACGTAGGCCCAATTCGCAAACGTTTCGGTTCGCCTTTCAAAGCTTTGCACAGA
GAAAATGTGAATTTAGCTGCCGCATAGGCCATTGATAGAGTAGCTGAACCTGTTT
CCGCTTTAGCTTTTACCCTTCGGTTCCGGCTTCCTGGATCCTGTGGGTCAACTTG
GCCACTTCATCTTCGGAAAAGCAGAGGCCTGAGGGTTTAGTTTGTGATAAGACTG
GAATGATCGTCTTGCCCGCGTGTCTCCGATCACCGGAACCTTCAACGTCCGCAGG
GTCGAGGTTCTTCGCCTCGGCTATGAACGT (SEQ ID NO: 120)

contig03919 length=253 numreads=3

CGTGCAATTACAGCCTCAGGGTCGTTCGGAGTTGACATGGAAAATGGGAGCATTCA
CCACGCGAGCAACGTCTGTACAATACGGAGAGGAACGGGAGAACCTGGGATCGG
TAGTAAATCCAATCTGGTTGTTGACAACAATGTGGATGGTTCCGTGGGTTCGTGTA
GTCTGGGAGGTCAGACAAGTGGAAGTCTCAAACACGACGCCCTGACCACAGAA
AGCTGCGTCTCCGTGCAGAAGGATGGACATCACCTT (SEQ ID NO: 121)

contig01317 length=250 numreads=4

AGTATATTCGGATGGACATTGAAGTCGTTGGGAACTCGAACAGAGCTCAATCCA
ATAACTGTCAGCAAATCCTTGTCACCCCTGTATCCCAAGTTGTGATAGCACTTG
GGGCAGGTTTGTAAACCAGTCCATAATTGCCTTTAGGATCTGGTGGATAAAATTG
CTAGCATTCTCTAAAGCTGCAGTTAACTGCGCATGGTACTCATTTTTGATATCTTC
AATATCTTTTTTCAGTAATTAAACCACTCCT (SEQ ID NO: 122)

contig02816 length=256 numreads=11

AAGGTCTCATATTTGCTGATCGACCGACAGAATGCAGAGCCTTAACGAAATCGGT
TCCTTGTTGCAGTTTTTCTATAACTGGTTCTCCAACCTCTGGTCATAATCCTCATAG
AGGCTGCTACATAAGGCGAGTCCGTCACCTGAACACCGATTTTGGACATCGGACC

GCCAACTGGACCCATGGAGAAAGGGATGACGAACATTGTACGACCTTGCATGCA
ATTGGGGAATCGTTTCGTTAAGGCTCTGCATTCTGTC (SEQ ID NO: 123)

contig06121 length=594 numreads=439

CCAGAACAGGGCTCAGGCTGCAGTCGCCAACCGTCTGGGTGTGAGGATCAACCA
GGTGAAGAACGTCATCATCTGGGGCAATCACTCTTCCACCCAGTTCCTTGACGCT
CTCAACGCCTCTGTTCGTAGTTGAGGGTGCCACCAAGTCTGTCCCTCAAGCCCTCA
ACAACGACCCCTGGCTCAAGGAAGAATTTGTACGACTGTACAAAACCGTGGAG
CAGCTGTGATCGCGGCCAGGAAGATGTCTTCCGCCATGTCTGCGGCCAAAGCGG
CAGCAGACCACATGAAGGACATATGGAACGGAAGTCTCCCGGGACCTACACCT
CCATGGCTGTCTCTCTGATGGCTCTTACGGCGTCCCTGCGGATCTCATGTTCTCT
TTCCCCGTTACAATTGCCAACCAGAGTGGAAAATCGTCCAGGGACTCCCCATCG
ACGACTTCGCCCCGGGCTAAGCTGGACATCACCGCCAAGGAGCTCCTCGAAGAGA
AGGAGGATGCCGTTGCGCTCTGTTCAGGATTGACTAGACGTAAAGCCCAATCAAA
GGAGTAAATTGTAACCAACTCAAAGTATTAGTGCCATCGGACAAGTGG (SEQ ID
NO: 124)

contig02214 length=253 numreads=4

AGAGAATTTTGGGAGCGACCACAATCAAAGGTTTCCTGTAGTTTCTGACCATTTG
CCGCCGAAGGAGGTGGAAATACTGAGCGGGAGTAGTGGGGTTCACGATGTGGAT
ATTCACATCTTCTCCATCTGGGCGTACCTCAGATGAATCAGACAGCTGTAGAAAT
CTCTCCAATCTGCACGAAGTGTGTTCTGGACCAGCTCCATCGTATCCATGGGGTA
GCAACATGGTAAGACCACTCATCGTCATCCATTT (SEQ ID NO: 125)

contig06953 length=148 numreads=26

TGATGGCTCCTGCAGTTCTAATGAAAGATGCGAAGACCATGTGCGACTCCTTGCTG
AACTCCAAACTCCAAATCCTGTTTGTCTTTTTCAGACACGGCAGGAAGATCGACA
TGATGCCAGGAAGGTTGACACCTTTCTTGCTTCCAG (SEQ ID NO: 126)

contig01150 length=1037 numreads=14

TCCTTGATGAACACAGATCGGAGCTGACCGTTCATCTCGAAGAAGACTTCCCTTT
CTCCAGATGGCGTGAGGTCGTCAGCCGTCGCCAGACACTTGATGGAGAGGAGTT
TACCCTTCTCTAGTGGAACCTCGAACTCTTCACCGATCTTGGGGCCGGAGAGGAA
GGTCTGGTGTGCGAGCTTGTGCGACGGGGCCGAACTCGTCCCTGAAGTTGAGGAAG
TCGTGCGGTGACTGTGGGATAGAGGGCCGAGCTGATGACATCCTTATCGGTCACAA
ACGTGTGCTTCTCCTGCAGCGACTCTTTCAACTTTCCAAAGTCCAGAGGTTTGAG
AGACTCTCCAGGTCGCCCCGTCTATCCTTGGCATGTCTTTCAGCACCTGGGATCTG
AAGGGTTCTGGGAATCCTAGGTAGGGTTCGCCGATCGCACCCCTGCAGGAAGTCA
ACGACGGATTTGGGGAAGGACAAGTCCCCAGCGCGCTTGAGGACCTCCTCGGAG
CTCAACTTGTTCTGAACCATGAACTGAGCCAAATCTCCTACCACTTTG
GATGATGGAGTCACCTTGATGATGTCTCCGAGAATCAAGTTGGCTTCTCTGTAGG
CTTTCTTGACTTCCCTCGAACTGTTCTCCCAGACCGAGAGAGTAGGCTTGGAAGT
CAGGTTTCGTATACTGACCGCCAGGGATTTTCAATTCAGGTAGACGTCGGCGTTGCCA
GATTTTCATCGTGGTGGTGCATTCAAAGGGCGCGTAGAGCGTTCTAGTTTGTTC
AGTAAGCTGAGTACTCGGACACGTTCTCAGATCAAACCCTGTATCCAGCCTGGT
GCCAGCGAGAGAAGCGACGACGGCTCCCATGGAAGGCTGTGACGTCATCCCTGA
CATGGAGTCGACGGCTGTGTCCACTACGTCAGCGCCTGCATCGGCGCAGGCGAG
CATTGAAGCAACCCAGCGCCGCTTGTGTCTGCGTGTGGATGTGGAGGGGGAT
GTCTGGCACTGCGTCTTTCACAGCCTTCACCAGAAGTTTGGCCGCTTGCGGCTTC
AGGAG (SEQ ID NO: 127)

contig03691 length=251 numreads=3

AGAAACGATTACGAAGTCTGTCATGAAAACCAACCATCTCGTCTCAGTGGAACA
AGGTTGGCCTCAAAGCGGCGTCGGTGCTGAGATCTGCGCGAGGATCATGGAAAG
CGAAGCTTTCTTCCACCTAGACGCCCCGTCATCCGAATCACCGGTGCTGACCTT
CCAATGCCCTACAACCATCATCTGGAAGCACTCGCCCTCCCTCAAGTCAAACACG
TCGTAGAAACTGTCAAGAAAGTGCTCGGCGCGT (SEQ ID NO: 128)

Bt receptor:**contig04242 length=257 numreads=5**

TGAACAAGTGTTTTGCTCACTGTTATTACCAACAGACACCGTGATAAGCGGATAA
GATACTACTTGGATAATTTCCAAGACATGTTTCAGCGTGTTGGCAAGCGGAGACAT
CGTTTTGCGTAAGCACTTGGACTATGAGACGGAAGACACATACCTGTTCCAAGTG
TATGCCACTGATGGGACGATGAATGCTAGTGCAGCCGTTAACATAACGGTGCTG
AACGTGAACGATTGGGACCCCCGATTCAGGTTTCCTCA (SEQ ID NO: 129)

contig05324 length=1261 numreads=543

TCATTAAAAATTTACGGTTTAATAAAAAATTAATTCTGAATTTTCGAGAATATCCAA
CTAGGTAGTAAAGTTCAGTAGAAGTAGACAAGACGTTTTTCGATCGATAACGAGG
GTGACTTCGTCGTATCTACAGGACGGTAACTTTGATAGCTGCTCTGTGAGATTTG
CAGATGGCAGGCATTTCTCTCAGCAGGTTATCGACGTCGGTACTCAACAGGCAAA
TGTTGGCGAAGTTGTAAAAGTGTTGTAGGAAATATTCATTCTCATGCCCCGAAGCA
GAAAGATCTGAAGCGGTGAAGCAGCTTCTTGTAACAAGGGAAGATGTCCTTCCG
CAAGGCCAGCTTCTCCTCAGAGATTACAGACCTTACTTTATGGGAGTTTGCAGTA
GCAGCCATTACGAATTCATCAAACAATTGGTCACCTAATTTTCTGGCCTTGATAA
TTTCATCAACCTCGTTCTGTAACTTTTGTTTTTTCAGCCACATCGTTTGACTTGCTAA
TTTTATGCCGAAGGTAGTGCAAAGCAGCCTCTCTGCTGCTGACCGAAACAACATC
ACCGGGTACCGTTTCAGAATACGTTGGGTGCTTCTTGTTAGGCGTCAGCTCTTCC
GGCCCGAGTGAAAAATCGCCGTAGACCTGTACGTGGCTTTTCGTTACGGCGGTCC
TCACGTGTTGGTACTGGTGGAAAAACGTTTCCAGAGTGCCTGTCTTTTCCGAGTC
CATGTGCTCCATCCACCTGACGCTAAATTCATCTCCTAAGCAGGTGCCGTCATCG
CAGTACCATCCCCAAGAACTCTCGTCCCTAGCTGCAGCAGTCATCACGTAAATGT
CTGTATGTTCTGGCAAGATGTTGTGCGAACATAGATCCACTTTCACAAGCCTCAAC
ATAGAAGATAAACTTGTCGTACTTCTTCTTTCTACGCATCTGACTGAGAGTATTTT
CCAAATCGTCCGCCATCAGGTTGTCAATTGGGGAAACAAAGCATTCCAGTACCGCC
GTGGTCCGCGAAGTACACGAATAAGTTGGTGTTTTCATCACTTTCAATAACTCGT
CCGGTTCCAATTCCTTTTCATACCTGTTTTGTTTCCGCTGAGGATGTTTCAGGAAATT
TTGTGGCGTGAGATCATTTCTGTATAGTCAATTTTAACACCAGCGTAGACGTTG
GTTCCATTTGGTTTGTGATGATGACTCCAGGCGTCGGATTATCGGTATTATAGGC
GATGTCGTTGTACATCATCAATTATCTTTTCTGCTGGGATTCCTTC (SEQ ID
NO: 130)

contig05633 length=430 numreads=8

GAAGACCTGGTGGATGATGTAATTGGCCGCCAATCCGAAATCTCCCTTCGATATC
TGATAAATCGATACTCGACGCCATCTTGAACCTCAGTGCCTTCTCGATTCTGTCTTCC
GGAGGACAGTTTAGTTTTCTGTCGTAGAGGGGAGTCGAGTCGCCGAGGGACATGG
ATTGAGGGCTGTTGATGGTTCGAGGGGTCGACCGGGCTGAGCGAGGGGTGCCGG
AAACGCGGGGAGTGTTTCGTTCCGAGGTCGAAAGGATCGTATTTCCCTTGAGTCGG
TGGCAGTCGAGGGGATCTCCCGGTTTCGATCGTTCCCTTGAGATCCCATATTAGTC
GCTCTACCCCATCCTACTCTAGGAGACTGCATGCTGCGAGGACTGAAATCCACGT

CGTTGTACCGCGGAGTTCTCTGTTCCCTCCCGTCATCCCGAGGAGTG (SEQ ID NO: 131)

contig06107 length=1094 numreads=95

CTCCGTTTAGGGCTCTTCGTCTGTTTCGTGGTGACGGTGAGCGGTTTGGAACCGG
TGGAAGACACGAGGTATTATGTTGAACTCGTTAATTCAGCAAACACTACGTGGAC
GGCCGGTTACAACCTTTGAAAACCTCTGTTGAGTATCTGAGGCAATTCGTGGGC
ACTGAAGTCAACCAGTTCGAACGTATTTTGCCGAATTTTGAAAGTGCTGGTGGCC
ATGAAGACATCGATATCCCGACTAGTTTCGACGCAAGCAAGAAATGGCCCAAGT
GCGCTCACATCAAACACGTCAGAGACCAGAGTGCCTGCGGCTCTTGTTGGGCTGT
GTCTCTGGCGTTCGGTTTTGACTGACCGCCTGTGCATCGCCTCAGGAGGAACCTTC
AGGGCGCCAATGTCCGCGGAAACCTGATGTCATGTTGCTCTGAATGCAGTAATGG
CAACGGGTGTCAGGGTGGAAATCACGCTGTCGCTTGGGTCTACGCCAAAACCCA
AGGCATCGTTACTGGTGGGGACTACCACTCTCATATTGGTTGCCAACCTACCTA
CTTCAGCCTTGCGGACATTCTGCCGACTCCAAGCGTGTTTCCTGTGGAGAATTATT
GCCCTCAGTGAAGGCCCCAGTTTGCGAGAACACCAAGTGCTTGAACCCCTTTCTAC
AAGAAAAATACATTCTCCAAAGACCATAAATTCGTAAAAGAAGCCTATGCTATT
ATGAACGACGAAAAAGCAATTCAACGCGAAATCCTGCAGAGGGGACCTGTGGCT
GCAGCTTTCGTCGTTTACGCAGATTTCCCCTCCTATAAATCAGGTGTGTACCAGA
GAATGTCGGAGGCTGTGGTTGGAGGTCACGGCGTTAAAATCATCGGATGGGGAA
CTGAAGGGGGTGTGCGACTACTGGCTGGTTCGTCAACTCCTGGAACACCGACTGGG
GTGAGAAAGGGCTCTTCAAAATCCGCCGAATGACCGCAGGCGTCGATTTCGAAC
AGATGATCACTTTTGGCGTCGTATAATTTGTTACCGTTCTGTTATATAGGATTAGG
TACCAATGTTATCTCACAATTAATTTAATGTGAATGATATATATTATTTTAAAC
(SEQ ID NO: 132)

contig06354 length=444 numreads=275

TGAAAGTGTAGGCATTGGTAAGTTCAACTGCCTCGATAGGTGGAATATCCGTATG
GGCTGAGCAGCAATGGCACGTGATAATGTGTGTTTCCATCTTCAATGTTGAAAAC
GACTTCAACATAAGGATAAAACGTTTTGGCTCCCTTCGACGAGAAATACTGCTCA
GTCTCGAATTTTCATCTTATGGATTCCGGGAGAGATAGATCCAGGCTTGGTGAAAC
CAGGGAGGCGTCCGTCCGAATTCGTTTTGTCTTTGTTGATCTCCACCCAATGGTTC
CCAACGTCTGGTGCCATTTTGTAAGTGTAACAGACAGTCCAGCTGCTGGAAGGC
CAGAAGATGTGTCCAGAACGTGGCAAGTGATTGGAATCACGTTAGACGCCATGG
TGTTTTTTCACGGAAAGATGTTGGGAAATCACTTTGAGTCGTTCCCCGGCCGCTGA
CATA (SEQ ID NO: 133)

The foregoing is illustrative of the present invention, and is not to be construed as limiting thereof. The invention is defined by the following claims, with equivalents of the claims to be included therein.

THAT WHICH IS CLAIMED IS:

1. A transgenic plant that contains a polynucleotide construct that comprises a heterologous nucleotide sequence operably linked to a promoter that drives expression in the plant, said heterologous nucleotide sequence encoding an active agent toxic to an insect pest;

said active agent selected from the group consisting of: (i) an RNA that inhibits expression of a target gene in said insect pest, and reduces or inhibits the expression of the expression product (*e.g.*, protein, ribosomal RNA) (ii) an antibody that binds to the expression product of said target gene and inhibits the activity thereof, and (iii) an aptamer that binds to the expression product of said target gene and inhibits the activity thereof.

2. The plant of claim 1, wherein said insect pest is a member of the Heteroptera suborder of the Hemiptera order.

3. The plant of claim 2, wherein said insect pest is a member of the Miridae family.

4. The plant of claim 3, wherein said insect pest is a plant bug, leaf bug, or grass bug.

5. The plant of claim 2, wherein said insect pest is a member of the Pentatomoidea superfamily.

6. The plant of claim 5, wherein said insect pest is a shield bug or stink bug.

7. The plant of claim 1-6, wherein said target gene encodes a protein of said insect pest selected from the group consisting of: transcription factors, translation factors, microvilli proteins (*e.g.*, mucin, pseudomucin), electron transport proteins, *Bacillus thuringiensis* toxin (Cry toxin) receptors, insecticide target proteins (*e.g.*, acetylcholine esterase, GABA-gated chloride channels, sodium channels, nicotinic acetylcholine receptors, chloride channels, midgut membrane proteins, mitochondrial ATP synthase, citin biosynthesis enzymes, ecdysone receptors, mitochondrial complex I, III, and IV electron transporters, Mitochondrial complex I electron transporters, voltage-dependent sodium channels, acetyl CoA carboxylase, *etc.*), and ribosomal RNA.

8. The plant of claim 1-7, wherein said target gene is given in **TABLE 1** herein.

9. The plant of claim 1-8, wherein said plant is a monocot.

10. The plant of claim 9, wherein the plant is a monocot selected from the group consisting of wheat, oats, millet, barley, rye, maize, rice, buckwheat, sorghum, triticale, spelt, linseed, and sugar cane.

11. The plant of claim 1-8, wherein said plant is a dicot.

12. The plant of claim 11, wherein said dicot is potato, soybean, Brassica, sunflower, cotton, or alfalfa.

13. The plant of claim 12, wherein said plant is cotton.

14. The plant of claim 1-13, wherein said active agent is a dsRNA.

15. The plant of claim 14, wherein said dsRNA is a double-stranded molecule with a first strand consisting essentially of a ribonucleotide sequence which corresponds to a nucleotide sequence of the target gene and a second strand consisting essentially of a ribonucleotide sequence which is complementary to the nucleotide sequence of the target gene, wherein the first and the second ribonucleotide strands are separate complementary strands (*e.g.*, on the same molecule or on separate molecules) that hybridize to each other to form said double-stranded molecule, and the double-stranded molecule inhibits expression of the target gene.

16. The plant of claim 1-13, wherein said active agent is a secretory antibody.

17. The plant of claim 16, wherein said plant further comprises:

(a) plant cells containing nucleotide sequences encoding immunoglobulin heavy- and light-chain polypeptides, wherein each polypeptide contains a leader sequence forming a secretion signal; and

(b) immunoglobulin molecules encoded by said nucleotide sequences, wherein said leader sequence is cleaved from said immunoglobulin molecules following proteolytic processing.

18. The plant of claim 16, wherein said plant further comprises:

(a) plant cells containing nucleotide sequences encoding a biologically functional multimeric protein comprising at least two different polypeptides not naturally produced by plants, wherein each nucleotide sequence encoding a polypeptide of the multimeric protein encodes a leader sequence forming a secretion signal that is cleaved from said polypeptide following proteolytic processing; and

(b) biologically functional multimeric protein encoded by said nucleotide sequences and resulting from assembly of said at least two different polypeptides wherein the multimeric protein is an abzyme or immunoglobulin.

19. A transgenic seed of the plant of claim 1-18, wherein the seed comprises said heterologous polynucleotide.

20. A method for protecting a plant from an insect pest, said method comprising: introducing into said plant at least one polynucleotide construct that comprises a heterologous nucleotide sequence operably linked to a promoter that drives expression in the plant, wherein said nucleotide sequence encodes an active agent toxic to an insect pest; and wherein said plant is a plant of claim 1-18.

21. A method for protecting a plant from an insect pest, said method comprising administering to said insect pest an active agent selected from the group consisting of: (i) an RNA that inhibits expression of a target gene in said insect pest, and reduces or inhibits the expression of the expression product (*e.g.*, protein, ribosomal RNA), (ii) an antibody that binds to the expression product of said target gene and inhibits the activity thereof, and (iii) an aptamer that binds to the expression product of said target gene and inhibits the activity thereof.

22. The method of claim 21, wherein said administering step is carried out by oral or topical administering said active agent.

23. The method of claim 21, wherein said active agent is a dsRNA.

24. The method of claim 21, wherein said active agent is a secretory antibody.

25. The method of claim 21-24, wherein said target gene encodes a protein of said insect pest selected from the group consisting of: transcription factors, translation factors, microvilli proteins (*e.g.*, mucin, pseudomucin), electron transport proteins, *Bacillus thuringiensis* toxin (Cry toxin) receptors, insecticide target proteins (*e.g.*, acetylcholine esterase, GABA-gated chloride channels, sodium channels, nicotinic acetylcholine receptors, chloride channels, midgut membrane proteins, mitochondrial ATP synthase, citin biosynthesis enzymes, ecdysone receptors, mitochondrial complex I, III, and IV electron transporters, Mitochondrial complex I electron transporters, voltage-dependent sodium channels, acetyl CoA carboxylase, *etc.*), and ribosomal RNA.

26. The method of claim 21-25, wherein said target gene is given in **TABLE 1** herein.

27. The method of claim 21-26, wherein said administering step is carried out by applying said active agent to said plant.

28. The method of claim 27, wherein said applying step is carried out by applying an insecticidal bait composition to said plant, said insecticidal bait composition comprising said active agent.

29. An insecticidal composition comprising a carrier media and an active agent, wherein said active agent comprises a compound selected from the group consisting of: (i) an RNA that inhibits expression of a target gene in said insect pest, and reduces or inhibits the expression of the expression product (*e.g.*, protein, ribosomal RNA), (ii) an antibody that binds to the expression product of said target gene and inhibits the activity thereof, and (iii) an aptamer that binds to the expression product of said target gene and inhibits the activity thereof.

30. In an insecticidal bait composition comprising an insecticide and one or more of insect food, insect attractant, carrier, and an insecticidal active agent, the improvement comprising employing as said active agent a compound selected from the group consisting of: (i) an RNA that inhibits expression of a target gene in said insect pest, and reduces or inhibits the expression of the expression product (e.g., protein, ribosomal RNA), (ii) an antibody that binds to the expression product of said target gene and inhibits the activity thereof, and (iii) an aptamer that binds to the expression product of said target gene and inhibits the activity thereof.

31. The composition of claim 29-30, wherein said active agent is a dsRNA.

32. The composition of claim 29-30, wherein said active agent is a secretory antibody.

33. The composition of claim 29-32, wherein said target gene encodes a protein of said insect pest selected from the group consisting of: transcription factors, translation factors, microvilli proteins (e.g., mucin, pseudomucin), electron transport proteins, *Bacillus thuringiensis* toxin (Cry toxin) receptors, insecticide target proteins (e.g., acetylcholine esterase, GABA-gated chloride channels, sodium channels, nicotinic acetylcholine receptors, chloride channels, midgut membrane proteins, mitochondrial ATP synthase, citin biosynthesis enzymes, ecdysone receptors, mitochondrial complex I, III, and IV electron transporters, Mitochondrial complex I electron transporters, voltage-dependent sodium channels, acetyl CoA carboxylase, etc.), and ribosomal RNA.

34. The composition of claim 29-32, wherein said target gene is given in **TABLE 1** herein.

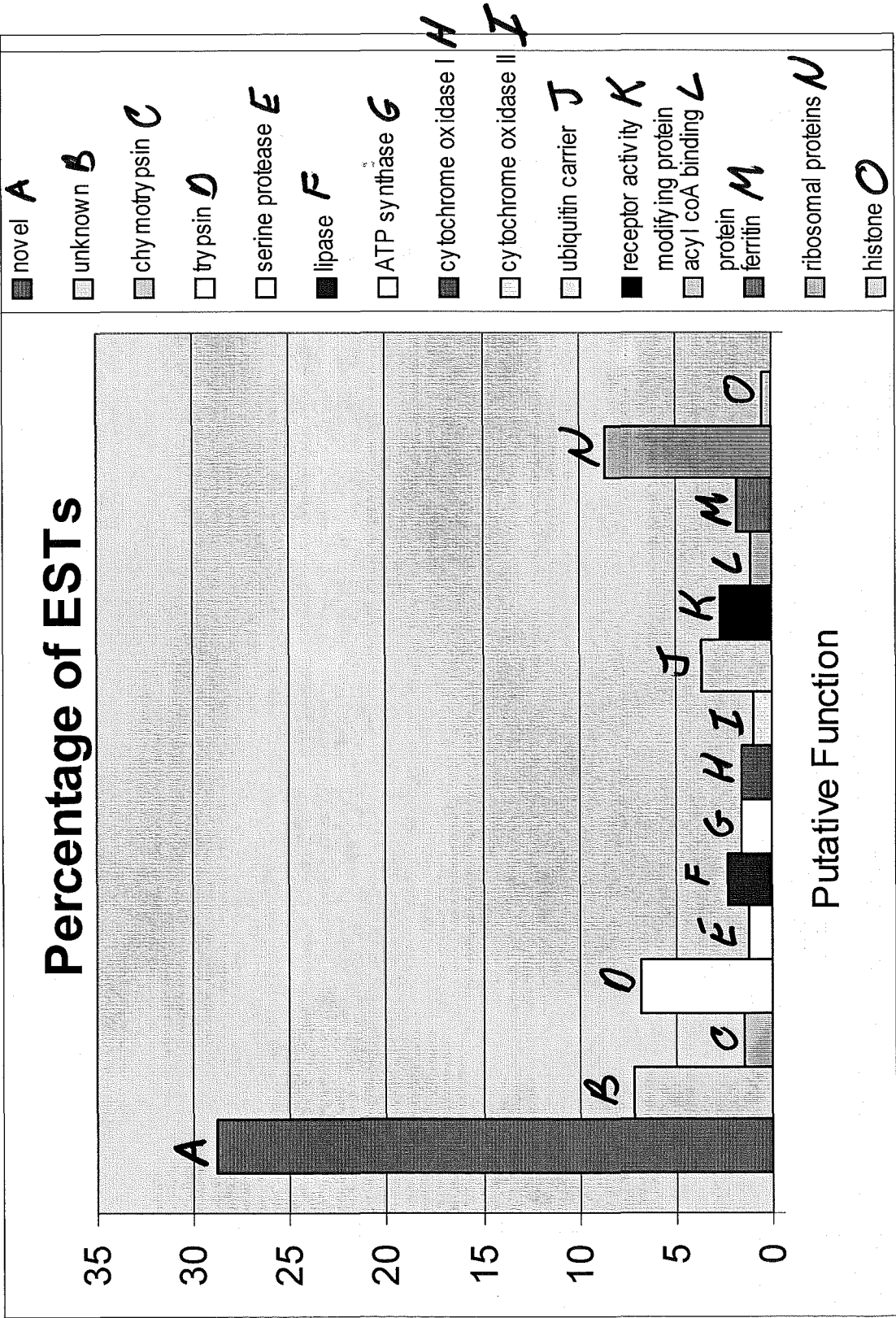


Fig. 1. Percentage of ESTs matching putative functions listed.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/36275

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 15/82 ; A01H 5/00 (2010.01)

USPC - 800/279, 800/302

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8): C12N 15/82 ; A01H 5/00 (2010.01)

USPC: 800/279, 800/302

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
800/278, 800/285, 800/287, 800/298, 800/295

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Electronic data base consulted: PubWest, DialogPRO--Chemical Engineering and Biotechnology Abstracts, INSPEC, NTIS (National Technical Information Service), PASCAL, Current Contents Search, MEDLINE. search terms: Transgenic plant, heterologous, polynucleotide, insect resistant, heteropte

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y	US 2008/0214443 A1 (Baum, et al.) 04 September 2008 (04.09.2008), para [0019], [0026], [0042], [0052], [0058], [0066], [0122], [0133].	1-7, 21-23, 25, 29-31 ----- 24, 32
Y	US 2003/0028913 A1 (Hein, et al.) 06 February 2003 (06.02.2003), para [0015].	24, 32
A	US 2009/0031457 A1 (Howe, et al.) 29 January 2009 (29.01.2009)	1-32

☐ Further documents are listed in the continuation of Box C.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

22 July 2010 (22.07.2010)

Date of mailing of the international search report

30 AUG 2010

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/36275

Box No. 1 Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:

a. (means)

☐

on paper

☒

in electronic form

b. (time)

☐

in the international application as filed

☒

together with the international application in electronic form

☐

subsequently to this Authority for the purposes of search

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/36275

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 8-20, 26-28, 33-34
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.