SUMMARY

(In accordance with 40 CFR part 152, this summary is available for public release after registration)

STUDY TITLE

Evaluation of the Sequence Similarities of the Unintended Open Reading Frames Present in Maize Line 1507 to the Public Protein Sequence Datasets

DATA REQUIREMENTS

None

AUTHOR(S)

Donald S. Walters, Robert F. Cressman

STUDY COMPLETED ON

October 4, 2002

PERFORMING LABORATORY

DuPont Crop Genetics
DuPont Experiment Station
Wilmington, DE 19880

LABORATORY STUDY ID

PHI- 2002-036

Evaluation of the Sequence Similarities of the Unintended Open Reading Frames Present in Maize Line 1507 to the Public Protein Sequence Datasets

SUMMARY

Two unintended open reading frames designated ORF3 and ORF4 are present in the maize line 1507 insert sequence. A global sequence similarity search for the conceptual translations of both of the sequences was conducted against the GenBank dataset in an effort to identify potential safety concerns by examining the characteristics of related proteins. The similarity search was conducted using the BLASTP 2.2.2 algorithm with a cutoff expectation (E) value of 1.0.

The ORF3 similarity search identified 45 protein accessions with an expect value less than 1. Eighteen of the accessions returned by the searches represent chloroplast RNA polymerase beta subunit (*rpoC2*) sequences from a wide variety of plants. A portion of the ORF3 sequence exhibits 100% identity to the maize chloroplast RNA polymerase beta subunit sequence. The remainder of the similarity hits all represent known or putative phosphinothricin acetyltransferases. None of the similar proteins returned by the search identify safety concerns that might arise from the potential expression of the ORF3 sequence in maize line 1507.

The results of the ORF4 protein search returned 17 protein accessions with sufficient sequence similarity to meet the cutoff expectation value of 1.0. The top 16 of these show a high degree of identity ($E<1x10^{-6}$) with the C-terminus of the ORF VI gene product (inclusion body matrix protein) from cauliflower mosaic virus (CaMV). Athough this gene product appears to play several roles in the viral life cycle, the region with similarity to ORF4 (residues 379-520) is not implicated in these functions (De Tapia et al., 1993). One of the 16 accessions is a closely related carnation etched ring virus matrix protein. The remaining accession is at a relatively higher E value (0.64) and represents a hypothetical protein conceptually translated from *Plasmodium falciparum* genome sequencing data. None of the similar proteins returned by the search identify safety concerns that might arise from the potential expression of the ORF4 sequence in maize line 1507.

STUDY TITLE

Evaluation of the Sequence Similarities of the Unintended Open Reading Frames Present in Maize Line 1507 to the Public Protein Sequence Datasets

DATA REQUIREMENTS

None

AUTHOR(S)

Donald S. Walters, Robert F. Cressman

STUDY COMPLETED ON

October 4, 2002

PERFORMING LABORATORY

DuPont Crop Genetics
DuPont Experiment Station
Wilmington, DE 19880

LABORATORY STUDY ID

PHI- 2002-036

STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS

Compound: B.t. Cry1F maize, Event 1507

Title: Evaluation of the Sequence Similarities of the Unintended Open Reading Frames Present in Maize Line 1507 to the Public Protein Sequence Datasets

No claim of confidentiality is made for any information contained in this study on the basis of its falling within the scope of FIFRA Section 10 (d)(1)(A)(B), or (C).*

Pioneer Hi-Bred International, Inc. – A

Company: Dupont Company

Company Agent: Angel Fuentes, Ph.D.

Title: Regulatory Scientist

Signature: Angel M. Fuentes

Date: 22nd October 2002

*In the United States, the above statement supersedes all other statements of confidentiality that may occur elsewhere in this report.

THIS DATA MAY BE CONSIDERED CONFIDENTIAL IN COUNTRIES OUTSIDE THE UNITED STATES.

STATEMENT OF COMPLIANCE WITH GOOD LABORATORY PRACTICE STANDARDS

Title: Evaluation of the Sequence Similarities of the Unintended Open Reading Frames Present in Maize Line 1507 to the Public Protein Sequence Datasets

Study Initiation Date:	March 1, 2002	Study Completion Date:	October 4, 2002
Experimental Start Date:	March 1, 2002	Experiment Termination Date:	August 19, 2002

This report represents data generated after the effective date of the EPA FIFRA Good Laboratory Practice Standards.

United States Environmental Protection Agency
Title 40 Code of Federal Regulations Part 160
FEDERAL REGISTER, August 17, 1989

Organisation for Economic Co-Operation and Development
ISBN 92-64-12367-9, Paris 1982

This study was not conducted according to the United States EPA FIFRA Good Laboratory Practice Standards (40 CFR Part 160). However, documentation of all critical data and quality control measures were used to ensure the integrity of the study results.

Lawrence Zeph — 10/11/2002

Study Sponsor
Larry Zeph, Ph.D.
Pioneer HiBred International, Inc.

Donald S Walters — 10/4/02

Study Director/ Author
Donald S. Walters
DuPont Crop Genetics

QUALITY ASSURANCE STATEMENT

Compound: *B.t.* Cry1F maize, Event 1507

Title: Evaluation of the Sequence Similarities of the Unintended Open Reading Frames Present in Maize Line 1507 to the Public Protein Sequence Datasets

Study Initiation Date: 3/01/2002 Study Completion Date: October 4, 2002

NON-GLP STUDY

SIGNATURE PAGE

Donald S Walters — 10/4/02

Author — Date
Donald S. Walters, Ph.D.

Robert F Cressman — 10/4/02

Co-Author — Date
Robert F. Cressman, M.S.

Lawrence Zeph — 10/10/2002

Sponsor — Date
Larry Zeph, Ph.D.

Greg Ladics — 10/8/02

Peer Reviewer — Date
Greg Ladics, Ph.D.

TABLE OF CONTENTS

Evaluation of the Sequence Similarities of the Unintended Open Reading Frames Present in Maize Line 1507 to the Public Protein Sequence Datasets

ABSTRACT

Two unintended open reading frames designated ORF3 and ORF4 are present in the maize line 1507 insert sequence. A global sequence similarity search for the conceptual translations of both of the sequences was conducted against the GenBank dataset in an effort to identify potential safety concerns by examining the characteristics of related proteins. The similarity search was conducted using the BLASTP 2.2.2 algorithm with a cutoff expectation (E) value of 1.0.

The ORF3 similarity search identified 45 protein accessions with an expect value less than 1. Eighteen of the accessions returned by the searches represent chloroplast RNA polymerase beta subunit (*rpoC2*) sequences from a wide variety of plants. A portion of the ORF3 sequence exhibits 100% identity to the maize chloroplast RNA polymerase beta subunit sequence. The remainder of the similarity hits represent known or putative phosphinothricin acetyltransferases. None of the similar proteins returned by the search identify safety concerns that might arise from the potential expression of the ORF3 sequence in maize line 1507.

The results of the ORF4 protein search returned 17 protein accessions with sufficient sequence similarity to meet the cutoff expectation value of 1.0. The top 16 of these show a high degree of identity ($E<1x10^{-6}$) with the C-terminus of the ORF VI gene product (inclusion body matrix protein) from cauliflower mosaic virus (CaMV). Athough this gene product appears to play several roles in the viral life cycle, the region with similarity to ORF4 (residues 379-520) is not implicated in these functions (De Tapia et al., 1993). One of the 16 accessions is a closely related carnation etched ring virus matrix protein. The remaining accession is at a relatively higher E value (0.64) and represents a hypothetical protein conceptually translated from *Plasmodium falciparum* genome sequencing data. None of the similar proteins returned by the search identify safety concerns that might arise from the potential expression of the ORF4 sequence in maize line 1507.

INTRODUCTION

The sequence similarity of the two unintended open reading frame sequences present in maize line 1507 was evaluated against proteins present in the public sequence databases. The similarity searches were conducted to identify possible toxicity, anti-nutritional, or other safety concerns arising from the potential expression of the two hypothetical proteins. Database annotations of the resulting group of proteins with measurable similarity to the open reading frame sequences were manually inspected to identify those that may be toxic to humans or livestock, or possess catalytic activities that could produce metabolites that could impact the safety or nutritional quality of food or feed derived from maize line 1507.

METHODS

DESCRIPTION OF IDENTIFIED ORF's IN MAIZE LINE 1507

Sequence analysis of the recombinant DNA insert present in maize line 1507 has identified two unintended open reading frame regions which code for peptide sequences greater than 200 amino acid residues. The two regions have been labeled ORF3 and ORF4 and fully described separately. The sequence data indicates the presence of an open reading frame in the insert of maize line 1507 which starts with an ATG start codon in a 5' Cry1F fragment and terminates within an inverted PAT fragment. This open reading frame has been designated ORF3 and codes for a conceptual protein of 250 amino acid residues presented along with the search output in the appendix of this report. The ORF3 sequence was subjected to a comprehensive BLAST search of the public sequence databases.

ORF4 is a 630 nt sequence begining in the Cry1F - ORF25 terminator, spanning 141 nt of ORF25, 56 nt of intervening sequence, and 433 nt of the CaMV 35S promoter. ORF4 is present in the transformation construct used to produce maize line 1507 and the sequence has been well confirmed in both the plant and the plasmid vector. The conceptually translated open reading frame encompassing 210 amino acid residues, was used for the subsequent BLAST searching. The ORF4 sequence is provided in the appendix.

SIMILARITY SEARCH METHODS

The similarity search was conducted using the BLAST algorithm (see BLAST output for reference). Sequence files comprising the conceptual translations for ORF3 and ORF4 from maize line 1507 were queried using the BLASTP 2.2.2 algorithm and the Genbank dataset (Genbank versions 129.0 release April 22, 2002 and 131.0 release August 15, 2002 respectively). This dataset incorporates entries from all non-redundant GenBank CDS translations (GenPept) including the RefSeq database; TrEMBL (release 20, March 2002 and release 21, June 2002 respectively); translated DDBJ (http://www.ddbj.nig.ac.jp/, release 49, Apr. 2002 and release 50, June 2002 respectively); Swiss-Prot (release 40, Oct 2001 and release 40.26 Aug. 13, 2002 respectively); PIR (release 72.01 April 19, 2002 and release 73.03 Aug. 16, 2002 respectively); Protein Research Foundation (PRF); and Protein Data Bank (PDB). A cutoff expectation (E) value of 1.0 was used to restrict the search to proteins with meaningful homology to the two proteins of interest. Although a statistically significant sequence similarity generally requires a match with an expectation value less than 0.01, a cutoff of E<1 insures that proteins with even limited similarity will not be overlooked in the search (Pearson, 2000). Low complexity filtering was turned off and the sequence output was set at 500 hits.

RESULTS

ORF3 BLAST SEARCH RESULTS

The results of the ORF3 BLAST searches identified 45 protein sequences with an E value less than 1. Nineteen of the 45 accesions represent chloroplast RNA polymerase beta subunit sequences from a wide variety of plants. This reflects the composition of the 5' region of the ORF3 sequence that has been identified as a fragment of the maize chloroplast *rpoC2* gene coding for a prokaryotic-like RNA polymerase (Hu and Bogorad, 1990). The four highest scoring hits represent chloroplast RNA polymerase beta prime or beta double-prime subunits from maize, wheat, and rice. Lower scoring hits include tobacco, Arabidopisis, belladonna, white mustard, spinach, Hooker's evening primrose, *Lotus japonicus*, garden pea, sorghum, and liverwort. Given that *rpoC2* is present in chloroplasts from major food and forage crops it is undoubtedly present when those plants are consumed for food. There appear to be no health or safety issues arising from the presence of the *rpoC2* sequence in maize line 1507.

The remaining aligments identified in the ORF3 similarity searches all occur to accessions identified as phosphinothricin acetyltransferase (PAT) enzymes. This is the result of an in-frame translation of a PAT gene fragment at the 3' end of ORF3, representing 67 amino acids. The accessions include genes from *Streptomyces coelicolor*, *Streptomyces hygroscopicus*, and several synthetic contructs. The properties of phosphinothricin acetyltransferase have been extensively characterized and are believed to pose no hazard to human or animal health (OECD Consensus Document on General Information Concerning the Genes and Their Enzymes that Confer Tolerance to Phosphinothricin Herbicide). The results of the ORF3 BLAST alignment show only alignments to the two component sequences *rpoC2* and PAT.

In conclusion, the public database BLAST alignment search of the hypothetical ORF3 peptide sequence did not identify similarities to protein toxins or enzymes such that its potential expression would pose discernible hazards to human or animal health. The output from the ORF3 BLAST analysis is presented in the Appendix.

ORF4 BLAST SEARCH RESULTS

A BLAST search of the ORF4 coding sequence returned 17 protein accessions with E values less than 1 under the search conditions described. Sixteen of these accessions belong to a family of inclusion body matrix proteins, also known as viroplasmin, from the caulimoviruses. These alignments exhibit almost perfect identity with residues 378 to 520 of the viral gene VI product. This similarity reflects the fact that the Cauliflower Mosaic Virus (CaMV) 35S promoter used to drive PAT gene expression in maize line 1507 includes this C-terminal region of viral gene VI. Fifteen of these sixteen inclusion body matrix protein accessions are from various strains of CaMV; the sixteenth and lowest scoring alignment within this group is to gene VI from the closely related Carnation etched ring virus.

Cauliomoviruses are double stranded circular DNA plant viruses that are aphid transmitted and cause mosaic and mottle patterns to appear on leaves. Two major RNA transcripts are associated with CaMV; a 19S mRNA and a more familiar polycistronic 35S transcript driven by the 35S promoter. ORF VI is one of six or more gene products encoded by the CaMV genome, and is translated into a protein of approximately 62 kD. This protein forms the majority of the electron dense inclusion bodies present in infected plant cells (http://www.mpl.ird.fr/genetrop-old/caulimoviridae/www.scripps.edu/resources/iltab/caulimoviridae/rtd03caulimo.html). The ORF VI gene product has been shown to act as a translational activator, facilitating translation of both the 35S transcript as well as other, non-CaMV polyscistronic mRNA's in plants (Bonneville, et al., 1989). This activation is believed to involve the binding of RNA by the protein (Cerritelli et al., 1998). ORF VI has also been implicated in host specificity of the virus. A thorough analysis using deletion mutants coupled with transient expression assays was used to assign functional properties to distinct domains in the ORF VI protein (Broglio, 1995). The results indicate that the C terminus of the ORF VI protein, representing amino acids 379-520 from the complete 520 residue sequence, does not play a central role in any of the ORF VI protein functions identified. This region corresponds nearly exactly to the relevant sequence of ORF VI that is present in the ORF4 sequence of maize line 1507.

The one remaining non-ORF VI accession returned by the ORF4 BLAST search exhibits a relatively high E value of 0.64 (ie. weak alignment) to the conceptual translation of a DNA sequence from *Plasmodium falciparum*. The accession appears to be a member of a multigene family of unknown function.

In conclusion, the public database BLAST alignment search of the hypothetical ORF4 peptide sequence did not identify similarities to protein toxins or enzymes such that it's expression would pose discernible hazards to human or animal health. The complete BLAST search output for the ORF4 sequence is presented in the Appendix.

REFERENCES

Bonneville J, Sanfacon H, Futterer J, and T Hohn. 1989. Posttranscriptional Trans-Activation in Caulilflower Mosaic Virus. Cell 59 (6): 1135-1144.

Broglio E. , 1995. Mutational Analysis of Cauliflower Mosaic Virus Gene VI: Changes in Host Range, Syptoms and Discovery of Transativation-Positive, Noninfectionus Mutants. Molecular Plant Microbe Interactions 8(5) 755-760.

Cerritelli S, Fedoroff O, Reid B, and R Crouch. 1998. A Common 40 Amino Acid Motif in Eukaryotic Rnases H1 and Caulimovirus ORF VI Proteins Binds to Duplex RNAs. Nucleic Acids Res. 26 (7): 1834-1840.

De Tapia, M,; Himmelbach, A, and T; Hohn, T,. 1993. Molecular dissection of the cauliflower mosaic virus translation transactivator. , The EMBO Journal, 12: (8): 3305-3314.

Hu J and L., Bogorad L., 1990. Maize chloroplast RNA polymerase: the 180-, 120-, and 38-kilodalton polypeptides are encoded in chloroplast genes. Proc Natl Acad Sci U S A 1990 Feb;87 (4): 1531-1535.

OECD Consensus Document on General Information Concerning the Genes and Their Enzymes that Confer Tolerance to Phosphinothricin Herbicide. 1999. ENV/JM/MONO(99)13. Environment Directorate-Joint Meeting of the Chemicals Committee and the Working Party on Chemicals. Series on Harmonization of Regulatory Oversight in Biotechnology No. 11

Pearson, W.R. 2000. Flexible sequence similarity searching with the FASTA3 program package. Methods Mol Biol 132: 185-219.

APPENDIX

ORF3 BLAST Search Output

Conceptual Translation of the ORF3 Sequence

```
  1  MLPLCAGQVR SQVPTHGEHQ CSLGRIVALP PQTPLIQRES LILQELHILP

 51  RSSSLKVLDN SIIGVDTQLT KNTRSRLGGL VRVKRKKSHT ELKIFSGDIH

101  FPEEADKILG GSLIPPEREK KDSKESKKRK NWVYVQRKKF LKSKEKYCGY

151  LSIRAADMAA VCDIVNHYIE TSTVNFRTEP QTPQEWIDDL ERLQDRYPWL

201  VAEVEGVVAG IAYAGPWKAR NPQPQQPTNG IYLATSLDHQ STLVVFVALS
```

BLASTP 2.2.3 [May-13-2002]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query=
(250 letters)

Database: /biolib/db/blast2/public/1/nr
1,034,565 sequences; 327,214,780 total letters

Searching.....done

Sequences producing significant alignments:	(bits)	E Value	%similar
Entrez Blink gi|11467184|ref|NP_043017.1| (NC_001666) RNA polymerase beta' su...	214	7e-55	(90%),
Entrez Blink gi|14017564|ref|NP_114251.1| (NC_002762) RNA polymerase beta' su...	207	9e-53	(88%),
Entrez Blink gi|11466779|ref|NP_039375.1| (NC_001320) RNA polymerase beta' su...	201	6e-51	(91%),
Entrez Blink gi|226691|prf||1603356R RNA polymerase beta'-2 [Oryza sativa]	201	6e-51	(91%),
Entrez Blink gi|8474524|sp|Q57146|PAT_STRVR PHOSPHINOTHRICIN N-ACETYLTRANSFER...	144	7e-34	(85%),
Entrez Blink gi|82232|pir||A05028 rpoC protein homolog - common tobacco chlor...	132	4e-30	(55%),
Entrez Blink gi|585922|sp|P38550|RPOD_TOBAC DNA-directed RNA polymerase beta"...	132	4e-30	(55%),
Entrez Blink gi|225275|prf||1211235K rpoC-like ORF 862 [Nicotiana tabacum]	132	4e-30	(55%),
Entrez Blink gi|11465944|ref|NP_054486.1| (NC_001879) RNA polymerase beta'' s...	132	4e-30	(55%),
Entrez Blink gi|7525023|ref|NP_051049.1| (NC_000932) RNA polymerase beta' sub...	130	2e-29	(52%),
Entrez Blink gi|20068321|emb|CAC88034.1| (AJ316582) RNA polymerase beta II su...	128	7e-29	(54%),

```
Entrez Blink gi|13449972|gb|AAK21981.1| (AY028212) phosphinothricin acetyltra...   126   3e-28 (71%),
Entrez Blink gi|225701|prf||1311256A gene bar,herbicide resistance [Streptomy...   126   3e-28 (71%),
Entrez Blink gi|6537290|gb|AAF15587.1|AF187951_1 (AF187951) promotes resistan...   126   3e-28 (71%),
Entrez Blink gi|114833|sp|P16426|PAT_STRHY PHOSPHINOTHRICIN N-ACETYLTRANSFERA...   126   3e-28 (71%),
Entrez Blink gi|12229937|sp|Q9THV5|RPOD_SINAL DNA-directed RNA polymerase bet...   124   7e-28 (51%),
Entrez Blink gi|11497514|ref|NP_054922.1| (NC_002202) RNA polymerase beta'' s...   119   2e-26 (52%),
Entrez Blink gi|13518329|ref|NP_084688.1| (NC_002693) RNA polymerase beta'' s...   118   7e-26 (57%),
Entrez Blink gi|13518438|ref|NP_084798.1| (NC_002694) RNA polymerase beta' su...   118   7e-26 (51%),
Entrez Blink gi|133447|sp|P12227|RPOD_PEA DNA-DIRECTED RNA POLYMERASE BETA" C...   110   1e-23 (51%),
Entrez Blink gi|2117915|pir||S30926 DNA-directed RNA polymerase (EC 2.7.7.6) ...    69   6e-11 (87%),
Entrez Blink gi|401021|sp|Q01923|RPOD_SORBI DNA-DIRECTED RNA POLYMERASE BETA"...    69   6e-11 (87%),
Entrez Blink gi|11466681|ref|NP_039277.1| (NC_001319) rpoC2 [Marchantia polym...    59   6e-08 (32%),
Entrez Blink gi|18860301|ref|NP_569619.1| (NC_003386) RNA polymerase subunit ...    55   7e-07 (35%),
Entrez Blink gi|17548618|ref|NP_521958.1| (NC_003296) PROBABLE PHOSPHINOTHRIC...    51   1e-05 (35%),
Entrez Blink gi|15673759|ref|NP_267934.1| (NC_002662) acyltransferase [Lactoc...    43   0.003 (29%),
Entrez Blink gi|21243039|ref|NP_642621.1| (NC_003919) phosphinothricin N-acet...    42   0.006 (48%),
Entrez Blink gi|15888215|ref|NP_353896.1| (NC_003062) AGR_C_1598p [Agrobacter...    41   0.011 (30%),
Entrez Blink gi|400898|sp|P31668|PAT_ALCFA PHOSPHINOTHRICIN N-ACETYLTRANSFERA...    41   0.011 (28%),
Entrez Blink gi|20521524|dbj|BAB91588.1| (AP005147) (44 pct identical to pir:...    40   0.018 (27%),
Entrez Blink gi|17988141|ref|NP_540775.1| (NC_003317) PHOSPHINOTHRICIN N-ACET...    39   0.041 (31%),
Entrez Blink gi|7480316|pir||T42042 phosphinothricin N-acetyltransferase (EC ...    39   0.069 (25%),
Entrez Blink gi|7239723|gb|AAA26766.2| (M37919) ORF X product (put.); putativ...    39   0.069 (25%),
Entrez Blink gi|15894145|ref|NP_347494.1| (NC_003030) Phosphinothricin acetyl...    39   0.069 (23%),
Entrez Blink gi|21221638|ref|NP_627417.1| (NC_003888) phosphinothricin acetyl...    39   0.069 (25%),
Entrez Blink gi|13473667|ref|NP_105235.1| (NC_002678) phosphinotricin acetylt...    38   0.091 (29%),
Entrez Blink gi|21231639|ref|NP_637556.1| (NC_003902) phosphinothricin N-acet...    38   0.12  (47%),
Entrez Blink gi|15801695|ref|NP_287713.1| (NC_002655) putative resistance pro...    37   0.20  (30%),
Entrez Blink gi|1742360|dbj|BAA15080.1| (D90784) Phosphinothricin acetyltrans...    37   0.26  (30%),
Entrez Blink gi|16129407|ref|NP_415965.1| (NC_000913) putative resistance pro...    37   0.26  (30%),
Entrez Blink gi|16125187|ref|NP_419751.1| (NC_002696) phosphinothricin N-acet...    36   0.45  (28%),
Entrez Blink gi|15601150|ref|NP_232781.1| (NC_002506) toxin resistance protei...    35   0.59  (29%),
Entrez Blink gi|3095168|gb|AAC38427.1| (AF055586) unknown [Vibrio cholerae]       35   0.59  (29%),
Entrez Blink gi|17986662|ref|NP_539296.1| (NC_003317) PHOSPHINOTHRICIN N-ACET...    35   0.77  (27%),
Entrez Blink gi|16764934|ref|NP_460549.1| (NC_003197) putative acyltransferas...    35   0.77  (37%),
```

```
>gi|11467184|ref|NP_043017.1| (NC_001666) RNA polymerase beta' subunit-2 [Zea mays]
 gi|133443|sp|P16025|RPOD_MAIZE DNA-DIRECTED RNA POLYMERASE BETA" CHAIN
 gi|66987|pir||RNZMB2 DNA-directed RNA polymerase (EC 2.7.7.6) beta'-2 chain - maize
          chloroplast
 gi|12482|emb|CAA35197.1| (X17318) RNA polymerase beta-2 subunit (AA 1-1527) [Zea mays]
 gi|902214|emb|CAA60278.1| (X86563) RNA polymerase beta' subunit-2 [Zea mays]
         Length = 1527

 Score =  214 bits (545), Expect = 7e-55
 Identities = 110/121 (90%), Positives = 114/121 (93%), Gaps = 2/121 (1%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   ILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF
Sbjct: 811 VDRFFFILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 870

Query: 96  SGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRA 155
           SGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKY  ++S+R
Sbjct: 871 SGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKY--FVSVRP 928

Query: 156 A 156
           A
Sbjct: 929 A 929
```

```
>gi|14017564|ref|NP_114251.1| (NC_002762) RNA polymerase beta' subunit-2 [Triticum
aestivum]
 gi|15214266|sp|Q9XPS9|RPOD_WHEAT DNA-directed RNA polymerase beta" chain
 gi|4958869|dbj|BAA78042.1| (AB027572) RNA polymerase subunit beta [Triticum aestivum]
 gi|13928197|dbj|BAB47026.1| (AB042240) RNA polymerase beta' subunit-2 [Triticum
aestivum]
          Length = 1479

 Score =  207 bits (527), Expect = 9e-53
 Identities = 105/119 (88%), Positives = 111/119 (93%), Gaps = 2/119 (1%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   ILQELHILPRSSSLK+LDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF
Sbjct: 774 VDRFFFILQELHILPRSSSLKILDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 833

Query: 96  SGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIR 154
           SGDIHFPEEADKILGG LIPPER+KKDSKESKKRKNWVYVQRKK LKSKEKY  ++S+R
Sbjct: 834 SGDIHFPEEADKILGGCLIPPERQKKDSKESKKRKNWVYVQRKKILKSKEKY--FVSVR 890


>gi|11466779|ref|NP_039375.1| (NC_001320) RNA polymerase beta' subunit-2 [Oryza sativa]
[Oryza
           sativa (japonica cultivar-group)]
 gi|133446|sp|P12093|RPOD_ORYSA DNA-DIRECTED RNA POLYMERASE BETA" CHAIN
 gi|66986|pir||RNRZC2 DNA-directed RNA polymerase (EC 2.7.7.6) beta'-2 chain - rice
           chloroplast
 gi|11973|emb|CAA33988.1| (X15901) RNA polymerase beta' subunit-2 [Oryza sativa
(japonica
           cultivar-group)]
          Length = 1513

 Score =  201 bits (511), Expect = 6e-51
 Identities = 103/112 (91%), Positives = 105/112 (92%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   ILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF
Sbjct: 792 VDRFFFILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 851

Query: 96  SGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKY 147
           SGDIHFPEEADKILGGSLIP EREKKDSKESKKR+NWVYVQ KK LKSKEKY
Sbjct: 852 SGDIHFPEEADKILGGSLIPLEREKKDSKESKKRENWVYVQWKKILKSKEKY 903


>gi|226691|prf||1603356R RNA polymerase beta'-2 [Oryza sativa]
          Length = 1000

 Score =  201 bits (511), Expect = 6e-51
 Identities = 103/112 (91%), Positives = 105/112 (92%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   ILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF
Sbjct: 792 VDRFFFILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 851
```

```
Query: 96  SGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKY 147
           SGDIHFPEEADKILGGSLIP EREKKDSKESKKR+NWVYVQ KK LKSKEKY
Sbjct: 852 SGDIHFPEEADKILGGSLIPLEREKKDSKESKKRENWVYVQWKKILKSKEKY 903


>gi|8474524|sp|Q57146|PAT_STRVR PHOSPHINOTHRICIN N-ACETYLTRANSFERASE (PPT N-
ACETYLTRANSFERASE)
           (PHOSPHINOTHRICIN-RESISTANCE PROTEIN)
 gi|80871|pir||JT0409 phosphinothricin N-acetyltransferase (EC 2.3.1.-) - Streptomyces
           viridochromogenes
 gi|295179|gb|AAA72709.1| (M22827) phosphinothricin-N-acetyltransferase [Streptomyces
           viridochromogenes]
 gi|581786|emb|CAA46314.1| (X65195) phosphinothricin-N-acetyltransferase [Streptomyces
           viridochromogenes]
          Length = 183

 Score =  144 bits (364), Expect = 7e-34
 Identities = 68/80 (85%), Positives = 72/80 (90%)

Query: 155 AADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYA 214
           AADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYA
Sbjct: 15  AADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYA 74

Query: 215 GPWKARNPQPQQPTNGIYLA 234
           GPWKARN       + +Y++
Sbjct: 75  GPWKARNAYDWTVESTVYVS 94


>gi|82232|pir||A05028 rpoC protein homolog - common tobacco chloroplast
          Length = 862

 Score =  132 bits (332), Expect = 4e-30
 Identities = 74/133 (55%), Positives = 93/133 (69%), Gaps = 6/133 (4%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   I +E+HILP SSS+ V +NSI+GVDTQ+T N RSR+GGLVRV+RKK   ELKIF
Sbjct: 134 VDRFFFIPEEVHILPGSSSIMVRNNSIVGVDTQITLNLRSRVGGLVRVERKKKRIELKIF 193

Query: 96  SGDIHFPEEADKIL--GGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSI 153
           SGDIHFP E DKI    G LIPP   K++SKESKK KNW+YVQR     SK+K+  ++ +
Sbjct: 194 SGDIHFPGETDKISRHTGVLIPPGTGKRNSKESKKVKNWIYVQR--ITPSKKKF--FVLV 249

Query: 154 RAADMAAVCDIVN 166
           R      + D +N
Sbjct: 250 RPVVTYEITDGIN 262


>gi|585922|sp|P38550|RPOD_TOBAC DNA-directed RNA polymerase beta" chain
          Length = 1388

 Score =  132 bits (332), Expect = 4e-30
 Identities = 74/133 (55%), Positives = 93/133 (69%), Gaps = 6/133 (4%)
```

```
Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   I +E+HILP SSS+ V +NSI+GVDTQ+T N RSR+GGLVRV+RKK   ELKIF
Sbjct: 660 VDRFFFIPEEVHILPGSSSIMVRNNSIVGVDTQITLNLRSRVGGLVRVERKKKRIELKIF 719

Query: 96  SGDIHFPEEADKIL--GGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSI 153
           SGDIHFP E DKI    G LIPP   K++SKESKK KNW+YVQR     SK+K+  ++ +
Sbjct: 720 SGDIHFPGETDKISRHTGVLIPPGTGKRNSKESKKVKNWIYVQR--ITPSKKKF--FVLV 775

Query: 154 RAADMAAVCDIVN 166
           R      + D +N
Sbjct: 776 RPVVTYEITDGIN 788


>gi|225275|prf||1211235K rpoC-like ORF 862 [Nicotiana tabacum]
          Length = 862

 Score =  132 bits (332), Expect = 4e-30
 Identities = 74/133 (55%), Positives = 93/133 (69%), Gaps = 6/133 (4%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   I +E+HILP SSS+ V +NSI+GVDTQ+T N RSR+GGLVRV+RKK   ELKIF
Sbjct: 134 VDRFFFIPEEVHILPGSSSIMVRNNSIVGVDTQITLNLRSRVGGLVRVERKKKRIELKIF 193

Query: 96  SGDIHFPEEADKIL--GGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSI 153
           SGDIHFP E DKI    G LIPP   K++SKESKK KNW+YVQR     SK+K+  ++ +
Sbjct: 194 SGDIHFPGETDKISRHTGVLIPPGTGKRNSKESKKVKNWIYVQR--ITPSKKKF--FVLV 249

Query: 154 RAADMAAVCDIVN 166
           R      + D +N
Sbjct: 250 RPVVTYEITDGIN 262


>gi|11465944|ref|NP_054486.1| (NC_001879) RNA polymerase beta'' subunit [Nicotiana
tabacum]
 gi|2924258|emb|CAA77410.1| (Z00044) RNA polymerase beta'' subunit [Nicotiana tabacum]
          Length = 1392

 Score =  132 bits (332), Expect = 4e-30
 Identities = 74/133 (55%), Positives = 93/133 (69%), Gaps = 6/133 (4%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   I +E+HILP SSS+ V +NSI+GVDTQ+T N RSR+GGLVRV+RKK   ELKIF
Sbjct: 664 VDRFFFIPEEVHILPGSSSIMVRNNSIVGVDTQITLNLRSRVGGLVRVERKKKRIELKIF 723

Query: 96  SGDIHFPEEADKIL--GGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSI 153
           SGDIHFP E DKI    G LIPP   K++SKESKK KNW+YVQR     SK+K+  ++ +
Sbjct: 724 SGDIHFPGETDKISRHTGVLIPPGTGKRNSKESKKVKNWIYVQR--ITPSKKKF--FVLV 779

Query: 154 RAADMAAVCDIVN 166
           R      + D +N
Sbjct: 780 RPVVTYEITDGIN 792


>gi|7525023|ref|NP_051049.1| (NC_000932) RNA polymerase beta' subunit-2 [Arabidopsis
thaliana]
```

```
 gi|6685905|sp|P56764|RPOD_ARATH DNA-directed RNA polymerase beta" chain
 gi|5881684|dbj|BAA84375.1| (AP000423) RNA polymerase beta' subunit-2 [Arabidopsis
thaliana]
          Length = 1376

 Score =  130 bits (326), Expect = 2e-29
 Identities = 70/133 (52%), Positives = 94/133 (70%), Gaps = 6/133 (4%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   I +E+HILP SS++ V + SIIGVDT+LT N RS++GGL+RV++KK   ELKIF
Sbjct: 662 VDRFFFIPEEVHILPESSAIMVQNYSIIGVDTRLTLNIRSQVGGLIRVEKKKKRIELKIF 721

Query: 96  SGDIHFPEEADKIL--GGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSI 153
           SGDIHFP++ DKI    G LIPP R KK+SKESKK KNW+YVQR      +K+K+  ++ +
Sbjct: 722 SGDIHFPDKTDKISRHSGILIPPGRGKKNSKESKKFKNWIYVQR--ITPTKKKF--FVLV 777

Query: 154 RAADMAAVCDIVN 166
           R      + D +N
Sbjct: 778 RPVATYEIADSIN 790


>gi|20068321|emb|CAC88034.1| (AJ316582) RNA polymerase beta II subunit [Atropa
belladonna]
          Length = 1389

 Score =  128 bits (321), Expect = 7e-29
 Identities = 73/133 (54%), Positives = 91/133 (67%), Gaps = 6/133 (4%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   I +E+HILP SSS+ V +NSI+GVDTQ+T N RSR+GGLVRV+RKK   ELKIF
Sbjct: 664 VDRFFFIPEEVHILPGSSSIMVRNNSIVGVDTQITLNLRSRVGGLVRVERKKKRIELKIF 723

Query: 96  SGDIHFPEEADKIL--GGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSI 153
           SGDIHFP E DKI    G LIPP   K +SKESKK K W+YVQR      SK+K+  ++ +
Sbjct: 724 SGDIHFPGETDKISRHTGLLIPPGTGKINSKESKKVKKWIYVQR--ITPSKKKF--FVLV 779

Query: 154 RAADMAAVCDIVN 166
           R      + D +N
Sbjct: 780 RPVVTYEITDGIN 792


>gi|13449972|gb|AAK21981.1| (AY028212) phosphinothricin acetyltransferase [synthetic
construct]
          Length = 184

 Score =  126 bits (316), Expect = 3e-28
 Identities = 62/87 (71%), Positives = 68/87 (77%), Gaps = 2/87 (2%)

Query: 156 ADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAG 215
           ADM AVC IVNHYIETSTVNFRTEPQ PQEW DDL RL++RYPWLVAEV+G VAGIAYAG
Sbjct: 17  ADMPAVCTIVNHYIETSTVNFRTEPQEPQEWTDDLVRLRERYPWLVAEVDGEVAGIAYAG 76

Query: 216 PWKARNPQPQQPTNGIYLATSLDHQST 242
           PWKARN       + +Y+  S  HQ T
Sbjct: 77  PWKARNAYDWTAESTVYV--SPRHQRT 101
```

```
>gi|225701|prf||1311256A gene bar,herbicide resistance [Streptomyces hygroscopicus]
          Length = 181

 Score =  126 bits (316), Expect = 3e-28
 Identities = 62/87 (71%), Positives = 68/87 (77%), Gaps = 2/87 (2%)

Query: 156 ADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAG 215
           ADM AVC IVNHYIETSTVNFRTEPQ PQEW DDL RL++RYPWLVAEV+G VAGIAYAG
Sbjct: 14  ADMPAVCTIVNHYIETSTVNFRTEPQEPQEWTDDLVRLRERYPWLVAEVDGEVAGIAYAG 73

Query: 216 PWKARNPQPQQPTNGIYLATSLDHQST 242
           PWKARN       + +Y+  S  HQ T
Sbjct: 74  PWKARNAYDWTAESTVYV--SPRHQRT 98


>gi|6537290|gb|AAF15587.1|AF187951_1 (AF187951) promotes resistance to glutamine
synthetase inhibitors
           [Activation-tagging vector pSKI015]
          Length = 183

 Score =  126 bits (316), Expect = 3e-28
 Identities = 62/87 (71%), Positives = 68/87 (77%), Gaps = 2/87 (2%)

Query: 156 ADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAG 215
           ADM AVC IVNHYIETSTVNFRTEPQ PQEW DDL RL++RYPWLVAEV+G VAGIAYAG
Sbjct: 16  ADMPAVCTIVNHYIETSTVNFRTEPQEPQEWTDDLVRLRERYPWLVAEVDGEVAGIAYAG 75

Query: 216 PWKARNPQPQQPTNGIYLATSLDHQST 242
           PWKARN       + +Y+  S  HQ T
Sbjct: 76  PWKARNAYDWTAESTVYV--SPRHQRT 100


>gi|114833|sp|P16426|PAT_STRHY PHOSPHINOTHRICIN N-ACETYLTRANSFERASE (PPT N-
ACETYLTRANSFERASE)
           (PHOSPHINOTHRICIN-RESISTANCE PROTEIN)
 gi|80811|pir||S08615 phosphinothricin N-acetyltransferase (EC 2.3.1.-) - Streptomyces
           hygroscopicus
 gi|47129|emb|CAA35093.1| (X17220) phosphinothricin acetyl transferase (AA 1-183)
           [Streptomyces hygroscopicus]
 gi|581681|emb|CAA29262.1| (X05822) bar gene product (AA 1 - 183) [Streptomyces
hygroscopicus]
 gi|2547092|gb|AAB81249.1| (AF013602) phosphinothricin acetyl transferase [synthetic
           construct]
 gi|4191251|emb|CAA77216.1| (Y18556) phosmothicin acetyl transferase [Cloning vector
pSLJ8313]
 gi|7453573|gb|AAF62892.1|AF218816_2 (AF218816) basta resistance protein [Cloning
vector pEGAD]
 gi|16903541|gb|AAL30502.1|AF404854_1 (AF404854) phosphinothricin acetyltransferase
[Binary vector
           pJawohl3-RNAi]
          Length = 183

 Score =  126 bits (316), Expect = 3e-28
```

```
 Identities = 62/87 (71%), Positives = 68/87 (77%), Gaps = 2/87 (2%)

Query: 156 ADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAG 215
           ADM AVC IVNHYIETSTVNFRTEPQ PQEW DDL RL++RYPWLVAEV+G VAGIAYAG
Sbjct: 16  ADMPAVCTIVNHYIETSTVNFRTEPQEPQEWTDDLVRLRERYPWLVAEVDGEVAGIAYAG 75

Query: 216 PWKARNPQPQQPTNGIYLATSLDHQST 242
           PWKARN       + +Y+  S  HQ T
Sbjct: 76  PWKARNAYDWTAESTVYV--SPRHQRT 100


>gi|12229937|sp|Q9THV5|RPOD_SINAL DNA-directed RNA polymerase beta" chain
 gi|5725463|emb|CAB48415.2| (AJ243754) RNA polymerase A beta prime prime subunit
[Sinapis alba]
          Length = 1384

 Score =  124 bits (312), Expect = 7e-28
 Identities = 68/133 (51%), Positives = 92/133 (69%), Gaps = 6/133 (4%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   I +E+HILP SS++ V + SIIGVDT++T N RS++GGL+RV+RKK   ELKIF
Sbjct: 667 VDRFFFIPEEVHILPESSAIMVENYSIIGVDTRITLNIRSQVGGLIRVERKKKRIELKIF 726

Query: 96  SGDIHFPEEADKIL--GGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSI 153
           SGDIHFP++ DKI    G LIPP R K +SKESK  KNW+YVQR     +K+K+  ++ +
Sbjct: 727 SGDIHFPDKTDKISRHSGILIPPGRGKTNSKESKIVKNWIYVQR--ITPTKKKF--FVLV 782

Query: 154 RAADMAAVCDIVN 166
           R      + D +N
Sbjct: 783 RPVATYEIADSIN 795


>gi|11497514|ref|NP_054922.1| (NC_002202) RNA polymerase beta'' subunit [Spinacia
oleracea]
 gi|133448|sp|P11704|RPOD_SPIOL DNA-directed RNA polymerase beta" chain
 gi|81503|pir||A29959 DNA-directed RNA polymerase (EC 2.7.7.6) beta'' chain - spinach
          chloroplast
 gi|7636095|emb|CAB88715.1| (AJ400848) RNA polymerase beta'' subunit [Spinacia
oleracea]
          Length = 1361

 Score =  119 bits (299), Expect = 2e-26
 Identities = 70/133 (52%), Positives = 89/133 (66%), Gaps = 8/133 (6%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   I +E+HIL  SSS+ V +NSIIGVDT +T NTRSR+GG+VRV+RKK   EL IF
Sbjct: 656 VDRFFFIPEEVHILAGSSSIMVRNNSIIGVDTWITLNTRSRIGGVVRVERKKKKIELTIF 715

Query: 96  SGDIHFPEEADKIL--GGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSI 153
           SGDIHFP E DKI    G LIPP R  K+SK+SK  K W+YVQR     +K+KY  ++ +
Sbjct: 716 SGDIHFPGETDKISRHSGILIPPSR--KNSKDSKNLKKWIYVQR--ITPTKKKY--FVLV 769

Query: 154 RAADMAAVCDIVN 166
           R      + D +N
Sbjct: 770 RPVVPYEITDGIN 782
```

```
>gi|13518329|ref|NP_084688.1| (NC_002693) RNA polymerase beta'' subunit [Oenothera
elata subsp.
           hookeri]
 gi|6723744|emb|CAB67153.1| (AJ271079) RNA polymerase beta'' subunit [Oenothera elata
subsp.
           hookeri]
          Length = 1386

 Score =  118 bits (295), Expect = 7e-26
 Identities = 67/117 (57%), Positives = 79/117 (67%), Gaps = 2/117 (1%)

Query: 30  PPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSH 89
           P     I R   I +E+HILP SSS+ V +NS+IGVDT++  NTRSR GGLVRV+RKK
Sbjct: 656 PKYQMTIDRFFFIPEEVHILPESSSIMVRNNSLIGVDTRIALNTRSRAGGLVRVERKKRG 715

Query: 90  TELKIFSGDIHFPEEADKIL--GGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSK 144
             L+IFSG IHFP E DKI    G LIPP   K++SKESKK KN +YVQR    K K
Sbjct: 716 IALQIFSGTIHFPGETDKISWDSGILIPPGTGKRNSKESKKWKNGIYVQRITPTKKK 772


>gi|13518438|ref|NP_084798.1| (NC_002694) RNA polymerase beta' subunit-2 [Lotus
japonicus]
 gi|13358979|dbj|BAB33196.1| (AP002983) RNA polymerase beta' subunit-2 [Lotus
japonicus]
          Length = 1332

 Score =  118 bits (295), Expect = 7e-26
 Identities = 69/134 (51%), Positives = 91/134 (67%), Gaps = 7/134 (5%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF 95
           + R   I QE+HIL  SSS+ V +NSIIGV+T +T N +SR+GGLVRV++ K   ELKIF
Sbjct: 635 VDRFFFIPQEVHILSESSSIMVRNNSIIGVNTPITLNKKSRVGGLVRVEKNKKKIELKIF 694

Query: 96  SGDIHFPEEADKILGGS--LIPPER-EKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLS 152
           SGDIHFP E DKI   S  LIPPE  +KK+SKESKK+ NW Y+Q      +K+KY  ++
Sbjct: 695 SGDIHFPGEIDKISQHSAILIPPEMVKKKNSKESKKKTNWRYIQ--WITTTKKKY--FVL 750

Query: 153 IRAADMAAVCDIVN 166
           +R   +  + D +N
Sbjct: 751 VRPVILYDIADSIN 764


>gi|133447|sp|P12227|RPOD_PEA DNA-DIRECTED RNA POLYMERASE BETA" CHAIN
 gi|81957|pir||S07137 DNA-directed RNA polymerase (EC 2.7.7.6) beta'-2 chain - garden
pea
           chloroplast (fragment)
 gi|829325|emb|CAA27545.1| (X03912) ORF with large areas of homology to RNA polymerase
subunit
           beta [Pisum sativum]
          Length = 1163

 Score =  110 bits (275), Expect = 1e-23
 Identities = 65/126 (51%), Positives = 86/126 (67%), Gaps = 6/126 (4%)
```

```
Query: 26  IVALPPQTPL-IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVK 84
           I  L P++ + + R   I +E+HILP+SSSL V +NS++G+ T +T N RSR+GGLVR+
Sbjct: 444 IKELKPKSQIQVDRFFFIPEEVHILPKSSSLMVRNNSLVGIGTPITFNIRSRVGGLVRLD 503

Query: 85  RKKSHTELKIFSGDIHFPEEADKILGGS--LIPPER-EKKDSKESKKRKNWVYVQRKKFL 141
           +KK   ELKIFSG+IHFP E DKI   S  LIPP   +KK   +SKK KNW+YVQ
Sbjct: 504 KKKKKIELKIFSGNIHFPGEMDKISRHSAILIPPGTVKKKKCNKSKKIKNWIYVQ--WIA 561

Query: 142 KSKEKY 147
            +K+KY
Sbjct: 562 TTKKKY 567


>gi|2117915|pir||S30926 DNA-directed RNA polymerase (EC 2.7.7.6) beta''-2 chain -
Sorghum
           chloroplast (strain KS 39A) (fragment)
 gi|18029|emb|CAA78708.1| (Z14984) rpoC2 [Sorghum bicolor]
          Length = 380

 Score = 68.6 bits (166), Expect = 6e-11
 Identities = 35/40 (87%), Positives = 36/40 (89%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRS 75
           + R   ILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRS
Sbjct: 341 VDRFFFILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRS 380


>gi|401021|sp|Q01923|RPOD_SORBI DNA-DIRECTED RNA POLYMERASE BETA" CHAIN
 gi|478854|pir||S30925 DNA-directed RNA polymerase (EC 2.7.7.6) beta'' chain - Sorghum
           chloroplast (strain Calico) (fragment)
 gi|18027|emb|CAA78707.1| (Z14983) rpoC2 [Sorghum bicolor]
 gi|445059|prf||1908376A RNA polymerase [Sorghum bicolor]
          Length = 435

 Score = 68.6 bits (166), Expect = 6e-11
 Identities = 35/40 (87%), Positives = 36/40 (89%)

Query: 36  IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRS 75
           + R   ILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRS
Sbjct: 396 VDRFFFILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRS 435


>gi|11466681|ref|NP_039277.1| (NC_001319) rpoC2 [Marchantia polymorpha]
 gi|133444|sp|P06274|RPOD_MARPO DNA-directed RNA polymerase beta" chain
 gi|66985|pir||RNLVC2 DNA-directed RNA polymerase (EC 2.7.7.6) beta'-2 chain -
liverwort
           (Marchantia polymorpha) chloroplast
 gi|11650|emb|CAA28063.1| (X04465) rpoC2 [Marchantia polymorpha]
          Length = 1386

 Score = 58.5 bits (140), Expect = 6e-08
 Identities = 49/150 (32%), Positives = 81/150 (53%), Gaps = 22/150 (14%)
```

```
Query: 42  ILQELHILPRS-SSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRK-KSHTELKIFSGDI 99
           I +E+++L +S SS+ + +N  I   T +T N RS   GLV++++K  ++ ELKI  G I
Sbjct: 650 IPEEVYVLTQSLSSVFIKNNKFIQAGTLITSNIRSNTNGLVKIQKKGNNNYELKILPGTI 709

Query: 100 HFPEEADKILG--GSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAA- 156
           ++P E  KI      LIPP ++  +  E    KNW Y+Q    + SKEK   ++ IR A
Sbjct: 710 YYPNETYKISKQISILIPPGKKLFNEFEC---KNWTYLQ--WIMPSKEK--PFVLIRPAV 762

Query: 157 --------DMAAVCDIV--NHYIETSTVNF 176
                   + + + D++  N  +E  T+N+
Sbjct: 763 EYKISKKLNKSTLFDLLKKNKKVEIKTINY 792


>gi|18860301|ref|NP_569619.1| (NC_003386) RNA polymerase subunit beta'' [Psilotum
nudum]
 gi|18389454|dbj|BAB84206.1| (AP004638) RNA polymerase subunit beta'' [Psilotum nudum]
          Length = 1408

 Score = 55.1 bits (131), Expect = 7e-07
 Identities = 34/95 (35%), Positives = 54/95 (56%), Gaps = 6/95 (6%)

Query: 45  ELHILPRSSS-LKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPE 103
           E++I+  SS+ + V +NSI+   TQ+T    S+LGGLV++K  +   E++I  G IH P+
Sbjct: 666 EIYIVHESSAYILVSNNSIVQAGTQITPTLTSQLGGLVQIKNIQKSFEIRILPGTIHHPK 725

Query: 104 EADKI--LGGSLIPPEREKKDSKESKKRKNWVYVQ 136
                I      LIPP +   D   + K  +W+Y+Q
Sbjct: 726 RIPSISKQNNMLIPPGQSVFD---NLKFDHWIYLQ 757


>gi|17548618|ref|NP_521958.1| (NC_003296) PROBABLE PHOSPHINOTHRICIN ACETYLTRANSFERASE
PROTEIN
           [Ralstonia solanacearum]
 gi|17430866|emb|CAD17548.1| (AL646078) PROBABLE PHOSPHINOTHRICIN ACETYLTRANSFERASE
PROTEIN
           [Ralstonia solanacearum]
          Length = 183

 Score = 50.8 bits (120), Expect = 1e-05
 Identities = 29/81 (35%), Positives = 41/81 (49%), Gaps = 1/81 (1%)

Query: 154 RAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDR-YPWLVAEVEGVVAGIA 212
           R AD+ A+  I  H++ T T +F  +P    E     + LQD   P+LVA+    + G A
Sbjct: 26  RLADLPAIAAIYAHHVRTGTASFEIDPPDLAEMTRRFQTLQDTGMPYLVAQTGRKLLGFA 85

Query: 213 YAGPWKARNPQPQQPTNGIYL 233
           YAGP +AR        + IYL
Sbjct: 86  YAGPHRARPAYRHTVEDSIYL 106


>gi|15673759|ref|NP_267934.1| (NC_002662) acyltransferase [Lactococcus lactis subsp.
lactis]
 gi|12724800|gb|AAK05875.1|AE006407_11 (AE006407) acyltransferase [Lactococcus lactis
subsp. lactis]
          Length = 187
```

```
 Score = 43.1 bits (100), Expect = 0.003
 Identities = 20/67 (29%), Positives = 36/67 (52%)

Query: 154 RAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAY 213
           + +D   + +I   Y+E + + F  E  T  E+   +E++  RYP++VA     + G AY
Sbjct: 7   KKSDAKRLLEIYKPYVEKTAITFEYEVPTIAEFEKRIEKIGSRYPYIVAIENDKIIGYAY 66

Query: 214 AGPWKAR 220
           AG ++ R
Sbjct: 67  AGAYRER 73


>gi|21243039|ref|NP_642621.1| (NC_003919) phosphinothricin N-acetyltransferase
[Xanthomonas
           axonopodis pv. citri str. 306]
 gi|21108549|gb|AAM37157.1| (AE011867) phosphinothricin N-acetyltransferase
[Xanthomonas
           axonopodis pv. citri str. 306]
          Length = 179

 Score = 42.0 bits (97), Expect = 0.006
 Identities = 20/41 (48%), Positives = 28/41 (67%)

Query: 197 YPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSL 237
           YP+LVAE++GVVAG AYA  ++AR        N IYL+ ++
Sbjct: 52  YPYLVAELDGVVAGYAYASAFRARAGYRWTVENSIYLSAAM 92


>gi|15888215|ref|NP_353896.1| (NC_003062) AGR_C_1598p [Agrobacterium tumefaciens]
[Agrobacterium
           tumefaciens str. C58 (Cereon)]
 gi|17934782|ref|NP_531572.1| (NC_003304) acetyltransferase [Agrobacterium tumefaciens
str. C58
           (U. Washington)]
 gi|15155865|gb|AAK86681.1| (AE008019) AGR_C_1598p [Agrobacterium tumefaciens str. C58
           (Cereon)]
 gi|17739251|gb|AAL41888.1| (AE009053) acetyltransferase [Agrobacterium tumefaciens
str. C58
           (U. Washington)]
          Length = 206

 Score = 41.2 bits (95), Expect = 0.011
 Identities = 24/79 (30%), Positives = 39/79 (48%), Gaps = 1/79 (1%)

Query: 156 ADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERL-QDRYPWLVAEVEGVVAGIAYA 214
           AD+ A+ +I    +     +F  +P       +    L +  YP+LVAE +G VAG AYA
Sbjct: 43  ADIPAITEIYRDAVLHGRASFEIDPPDVATMAERRRLLVEGNYPYLVAEHDGKVAGYAYA 102

Query: 215 GPWKARNPQPQQPTNGIYL 233
           G ++AR        + +Y+
Sbjct: 103 GAYRARPAYGATVEDSVYI 121
```

```
>gi|400898|sp|P31668|PAT_ALCFA PHOSPHINOTHRICIN N-ACETYLTRANSFERASE (PPT N-
ACETYLTRANSFERASE)
           (PHOSPHINOTHRICIN-RESISTANCE PROTEIN)
          Length = 197

 Score = 41.2 bits (95), Expect = 0.011
 Identities = 25/88 (28%), Positives = 42/88 (47%), Gaps = 1/88 (1%)

Query: 148 CGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDR-YPWLVAEVEG 206
           C    +R  D+ A+  I  H++ T T +F   P    E      ++ D   P+LVAE +G
Sbjct: 26  CTVRVVRDDDLPAITAIYAHHVRTGTASFEEVPPDDTEMRARCAKVLDAGLPYLVAERDG 85

Query: 207 VVAGIAYAGPWKARNPQPQQPTNGIYLA 234
            + G AYA  ++ R+       + +Y+A
Sbjct: 86  KLLGYAYATHYRPRSAYRFTLEDSVYIA 113


>gi|20521524|dbj|BAB91588.1| (AP005147) (44 pct identical to
pir:C87365[phosphinothricin
           N-acetyltransferase of Caulobacter crescentus])~53 pct
           identical to sp:PAT_ALCFA[phosphinothricin
           N-acetyltransferase of Alcaligenes faecalis] [Salmonella
           typhimurium]
          Length = 178

 Score = 40.4 bits (93), Expect = 0.018
 Identities = 23/84 (27%), Positives = 40/84 (47%), Gaps = 1/84 (1%)

Query: 160 AVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDR-YPWLVAEVEGVVAGIAYAGPWK 218
           A+ DI  H++   T +F TEP    E +  L++++++  PW+VA  E  V G  Y   ++
Sbjct: 13  AIRDIYAHHVLHGTASFETEPPDTHEMLTRLKKIRNQALPWVVALEEEKVIGYCYLTRYR 72

Query: 219 ARNPQPQQPTNGIYLATSLDHQST 242
            R        + IY+      + T
Sbjct: 73  ERYAYRHTLEDSIYIHPDAQRRGT 96


>gi|17988141|ref|NP_540775.1| (NC_003317) PHOSPHINOTHRICIN N-ACETYLTRANSFERASE
[Brucella
           melitensis]
 gi|17983898|gb|AAL53039.1| (AE009619) PHOSPHINOTHRICIN N-ACETYLTRANSFERASE [Brucella
           melitensis]
          Length = 179

 Score = 39.3 bits (90), Expect = 0.041
 Identities = 25/80 (31%), Positives = 37/80 (46%), Gaps = 1/80 (1%)

Query: 156 ADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDR-YPWLVAEVEGVVAGIAYA 214
           AD+  +  I    + T T ++  EP T  E         D+ +P LVAE +G V G AYA
Sbjct: 10  ADIETITAIYTQAVLTGTGSYEIEPPTMDEMAKRFAAFADQGFPILVAEADGRVLGYAYA 69

Query: 215 GPWKARNPQPQQPTNGIYLA 234
             ++ R        + IY+A
Sbjct: 70  SYFRVRPAYRWLAEDSIYIA 89
```

```
>gi|7480316|pir||T42042 phosphinothricin N-acetyltransferase (EC 2.3.1.-) -
Streptomyces
           coelicolor (fragment)
          Length = 98

 Score = 38.5 bits (88), Expect = 0.069
 Identities = 23/92 (25%), Positives = 39/92 (42%), Gaps = 24/92 (26%)

Query: 157 DMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEG---------- 206
           D+  + D+ NHY+  + + F TEP TP+E          R PWL++  E
Sbjct: 15  DLKPLTDLYNHYVRETPITFDTEPFTPEE----------RRPWLLSHPEDGPYRLRVATD 64

Query: 207 ----VVAGIAYAGPWKARNPQPQQPTNGIYLA 234
                + G A + P++A+          +Y+A
Sbjct: 65  AESQEILGYATSSPYRAKPAYATSVETTVYVA 96


>gi|7239723|gb|AAA26766.2| (M37919) ORF X product (put.); putative [Streptomyces
coelicolor]
          Length = 99

 Score = 38.5 bits (88), Expect = 0.069
 Identities = 23/92 (25%), Positives = 39/92 (42%), Gaps = 24/92 (26%)

Query: 157 DMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEG---------- 206
           D+  + D+ NHY+  + + F TEP TP+E          R PWL++  E
Sbjct: 16  DLKPLTDLYNHYVRETPITFDTEPFTPEE----------RRPWLLSHPEDGPYRLRVATD 65

Query: 207 ----VVAGIAYAGPWKARNPQPQQPTNGIYLA 234
                + G A + P++A+          +Y+A
Sbjct: 66  AESQEILGYATSSPYRAKPAYATSVETTVYVA 97


>gi|15894145|ref|NP_347494.1| (NC_003030) Phosphinothricin acetyltransferase
[Clostridium
           acetobutylicum]
 gi|15023752|gb|AAK78834.1|AE007601_5 (AE007601) Phosphinothricin acetyltransferase
[Clostridium
           acetobutylicum]
          Length = 196

 Score = 38.5 bits (88), Expect = 0.069
 Identities = 16/68 (23%), Positives = 37/68 (53%)

Query: 153 IRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIA 212
           ++  D   + +I   +I+ + + F  +  + Q++ + +  + ++Y +LV E++  VAG A
Sbjct: 7   VKKEDSEEILNIYKPFIQNTAITFDYDIPSIQKFTEKVSNISNKYAYLVCEIDEKVAGYA 66

Query: 213 YAGPWKAR 220
           YA  + R
Sbjct: 67  YASSFNER 74
```

```
>gi|21221638|ref|NP_627417.1| (NC_003888) phosphinothricin acetyltransferase
[Streptomyces
           coelicolor A3(2)]
 gi|2506900|sp|P21861|PAT_STRCO Phosphinothricin N-acetyltransferase (PPT N-
acetyltransferase)
           (Phosphinothricin-resistance protein)
 gi|80739|pir||JH0246 phosphinothricin N-acetyltransferase (EC 2.3.1.-) - Streptomyces
           coelicolor
 gi|153179|gb|AAA26705.1| (M62753) phosphorinothyrcin n-acetyltransferase [Streptomyces
           coelicolor]
 gi|7799523|emb|CAB90987.1| (AL355832) phosphinothricin acetyltransferase [Streptomyces
           coelicolor A3(2)]
          Length = 171

 Score = 38.5 bits (88), Expect = 0.069
 Identities = 23/92 (25%), Positives = 39/92 (42%), Gaps = 24/92 (26%)

Query: 157 DMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEG---------- 206
           D+  + D+ NHY+  + + F TEP TP+E          R PWL++  E
Sbjct: 16  DLKPLTDLYNHYVRETPITFDTEPFTPEE----------RRPWLLSHPEDGPYRLRVATD 65

Query: 207 ----VVAGIAYAGPWKARNPQPQQPTNGIYLA 234
                + G A + P++A+           +Y+A
Sbjct: 66  AESQEILGYATSSPYRAKPAYATSVETTVYVA 97


>gi|13473667|ref|NP_105235.1| (NC_002678) phosphinotricin acetyltransferase
[Mesorhizobium loti]
 gi|14024417|dbj|BAB51021.1| (AP003003) phosphinotricin acetyltransferase
[Mesorhizobium loti]
          Length = 194

 Score = 38.1 bits (87), Expect = 0.091
 Identities = 24/81 (29%), Positives = 39/81 (47%), Gaps = 1/81 (1%)

Query: 155 AADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQ-DRYPWLVAEVEGVVAGIAY 213
           AAD+  + +I    +   T ++  EP +  E       L    +P+LVAE +G V G AY
Sbjct: 11  AADLDTITEIYADAVTHGTASYELEPPSRAEMGTRFATLTAGGFPYLVAEKDGAVLGYAY 70

Query: 214 AGPWKARNPQPQQPTNGIYLA 234
           AG ++ R         + +Y+A
Sbjct: 71  AGAFRPRPAYRFIVEDSVYVA 91


>gi|21231639|ref|NP_637556.1| (NC_003902) phosphinothricin N-acetyltransferase
[Xanthomonas
           campestris pv. campestris str. ATCC 33913]
 gi|21113333|gb|AAM41480.1| (AE012327) phosphinothricin N-acetyltransferase
[Xanthomonas
           campestris pv. campestris str. ATCC 33913]
          Length = 175

 Score = 37.7 bits (86), Expect = 0.12
 Identities = 19/40 (47%), Positives = 24/40 (59%)
```

```
Query: 197 YPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATS 236
           YP+LVAE +G V G AYA  ++AR        N IYLA +
Sbjct: 52  YPYLVAERDGAVIGYAYASSYRARAGYRWTVENSIYLAAA 91


>gi|15801695|ref|NP_287713.1| (NC_002655) putative resistance protein [Escherichia coli
O157:H7
           EDL933]
 gi|15831306|ref|NP_310079.1| (NC_002695) putative resistance protein [Escherichia coli
O157:H7]
 gi|12515251|gb|AAG56327.1|AE005361_8 (AE005361) putative resistance protein
[Escherichia coli O157:H7
           EDL933]
 gi|13361518|dbj|BAB35475.1| (AP002557) putative resistance protein [Escherichia coli
O157:H7]
          Length = 172

 Score = 37.0 bits (84), Expect = 0.20
 Identities = 27/88 (30%), Positives = 41/88 (45%), Gaps = 3/88 (3%)

Query: 154 RAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLE-RLQDRYPWLVAEVEGVVAGIA 212
           R AD AA+ +I NH +  +    +  +        I   E R    YP LV+E +GVV G A
Sbjct: 7   RKADCAAIAEIYNHAVLYTAAIWNDQTVDADNRIAWFEARTIAGYPVLVSEEDGVVTGYA 66

Query: 213 YAGPWKARNPQPQQPTNGIYLATSLDHQ 240
             G W++ +        + +Y+    DHQ
Sbjct: 67  SFGDWRSFDGFRHTVEHSVYVHP--DHQ 92


>gi|1742360|dbj|BAA15080.1| (D90784) Phosphinothricin acetyltransferase (EC 2.3.1.-).
           [Escherichia coli]
 gi|1742364|dbj|BAA15083.1| (D90785) Phosphinothricin acetyltransferase (EC 2.3.1.-).
           [Escherichia coli]
          Length = 212

 Score = 36.6 bits (83), Expect = 0.26
 Identities = 27/88 (30%), Positives = 40/88 (44%), Gaps = 3/88 (3%)

Query: 154 RAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLE-RLQDRYPWLVAEVEGVVAGIA 212
           R AD AA+ +I NH +  +    +  +        I   E R    YP LV+E  GVV G A
Sbjct: 47  RKADCAAIAEIYNHAVLYTAAIWNDQTVDADNRIAWFEARTLAGYPVLVSEENGVVTGYA 106

Query: 213 YAGPWKARNPQPQQPTNGIYLATSLDHQ 240
             G W++ +        + +Y+    DHQ
Sbjct: 107 SFGDWRSFDGFRHTVEHSVYVHP--DHQ 132


>gi|16129407|ref|NP_415965.1| (NC_000913) putative resistance protein [Escherichia coli
K12]
 gi|8480658|sp|P76112|YNCA_ECOLI Hypothetical acetyltransferase yncA
 gi|7427901|pir||C64897 probable phosphinothricin N-acetyltransferase (EC 2.3.1.-) -
           Escherichia coli
 gi|1787719|gb|AAC74530.1| (AE000241) putative resistance protein [Escherichia coli
K12]
          Length = 172
```

```
 Score = 36.6 bits (83), Expect = 0.26
 Identities = 27/88 (30%), Positives = 40/88 (44%), Gaps = 3/88 (3%)

Query: 154 RAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLE-RLQDRYPWLVAEVEGVVAGIA 212
           R AD AA+ +I NH +  +   +  +       I   E R    YP LV+E  GVV G A
Sbjct: 7   RKADCAAIAEIYNHAVLYTAAIWNDQTVDADNRIAWFEARTLAGYPVLVSEENGVVTGYA 66

Query: 213 YAGPWKARNPQPQQPTNGIYLATSLDHQ 240
             G W++ +        + +Y+    DHQ
Sbjct: 67  SFGDWRSFDGFRHTVEHSVYVHP--DHQ 92


>gi|16125187|ref|NP_419751.1| (NC_002696) phosphinothricin N-acetyltransferase
[Caulobacter
           crescentus CB15]
 gi|13422207|gb|AAK22919.1| (AE005770) phosphinothricin N-acetyltransferase
[Caulobacter
           crescentus CB15]
          Length = 180

 Score = 35.8 bits (81), Expect = 0.45
 Identities = 23/80 (28%), Positives = 35/80 (43%), Gaps = 1/80 (1%)

Query: 156 ADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDR-YPWLVAEVEGVVAGIAYA 214
           AD+ A+  I    +      F   P    E     +   DR  P+LVAE+ G V G  YA
Sbjct: 8   ADIPAITAIYGWNVLNGLGTFEEVPPDAAEMARRRQGFLDRGLPYLVAELNGKVVGYCYA 67

Query: 215 GPWKARNPQPQQPTNGIYLA 234
           GP++ R        + +Y++
Sbjct: 68  GPFRLRAAYRYTVEDSVYVS 87


>gi|15601150|ref|NP_232781.1| (NC_002506) toxin resistance protein [Vibrio cholerae]
 gi|11257285|pir||H82466 toxin resistance protein VCA0387 [imported] - Vibrio cholerae
           (group O1 strain N16961)
 gi|9657787|gb|AAF96293.1| (AE004374) toxin resistance protein [Vibrio cholerae]
          Length = 169

 Score = 35.4 bits (80), Expect = 0.59
 Identities = 24/81 (29%), Positives = 35/81 (42%), Gaps = 4/81 (4%)

Query: 157 DMAAVCDIVNHYIETSTVNFRTEP---QTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAY 213
           D+A + DI N YIE +   F   P   +  +EW     R   +Y   VA   GV+ G  A
Sbjct: 10  DIAGITDIFNFYIEQTNARFEEFPFTLENREEWFSQFSRTA-KYQIYVAVENGVLQGFAC 68

Query: 214 AGPWKARNPQPQQPTNGIYLA 234
           +  ++A           +YLA
Sbjct: 69  SQKYRAIPAFDDTVEVSVYLA 89


>gi|3095168|gb|AAC38427.1| (AF055586) unknown [Vibrio cholerae]
          Length = 169

 Score = 35.4 bits (80), Expect = 0.59
```

```
 Identities = 24/81 (29%), Positives = 35/81 (42%), Gaps = 4/81 (4%)

Query: 157 DMAAVCDIVNHYIETSTVNFRTEP---QTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAY 213
           D+A + DI N YIE +    F    P    +  +EW      R    +Y    VA    GV+ G A
Sbjct: 10  DIAGITDIFNFYIEQTNARFEEFPFTLENREEWFSQFSRTA-KYQIYVAVENGVLQGFAC 68

Query: 214 AGPWKARNPQPQQPTNGIYLA 234
           +   ++A            +YLA
Sbjct: 69  SQKYRAIPAFDDTVEVSVYLA 89


>gi|17986662|ref|NP_539296.1| (NC_003317) PHOSPHINOTHRICIN N-ACETYLTRANSFERASE
[Brucella
           melitensis]
 gi|17982280|gb|AAL51560.1| (AE009480) PHOSPHINOTHRICIN N-ACETYLTRANSFERASE [Brucella
           melitensis]
          Length = 164

 Score = 35.0 bits (79), Expect = 0.77
 Identities = 23/84 (27%), Positives = 42/84 (49%), Gaps = 5/84 (5%)

Query: 156 ADMAAVCDIVNHYIETSTVNFR---TEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIA 212
           AD+ A+   I N   +E +    +      + +   ++W+++   R   D +P LVAE  EG V G A
Sbjct: 11  ADLPALLAIYNDAVENTLAIWNETLVDLENRRQWLENRNR--DGFPVLVAEREGQVVGYA 68

Query: 213 YAGPWKARNPQPQQPTNGIYLATS 236
             GP++             +Y+A++
Sbjct: 69  SYGPFRPFEGFRHSSELSVYVASN 92


>gi|16764934|ref|NP_460549.1| (NC_003197) putative acyltransferase [Salmonella
typhimurium LT2]
 gi|16420114|gb|AAL20508.1| (AE008770) putative acyltransferase [Salmonella typhimurium
LT2]
          Length = 171

 Score = 35.0 bits (79), Expect = 0.77
 Identities = 25/67 (37%), Positives = 36/67 (53%), Gaps = 5/67 (7%)

Query: 156 ADMAAVCDIVNHYI-ETSTV-NFRT-EPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIA 212
           AD AA+ +I NH +  T+ + N RT +      W +   + L   YP LV+E  GVV G A
Sbjct: 9   ADCAAITEIYNHAVLHTAAIWNDRTVDTDNRLAWYEARQLLG--YPVLVSEENGVVTGYA 66

Query: 213 YAGPWKA 219
             G W++
Sbjct: 67  SFGDWRS 73


  Database: /biolib/db/blast2/public/1/nr
    Posted date:  Aug 26, 2002 10:34 AM
  Number of letters in database: 327,214,780
  Number of sequences in database:  1,034,565

Lambda     K      H
   0.317    0.135    0.397
```

```
Gapped
Lambda      K       H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 165,778,036
Number of Sequences: 1034565
Number of extensions: 6749790
Number of successful extensions: 19986
Number of sequences better than  1.0: 45
Number of HSP's better than  1.0 without gapping: 28
Number of HSP's successfully gapped in prelim test: 17
Number of HSP's that attempted gapping in prelim test: 19939
Number of HSP's gapped (non-prelim): 47
length of query: 250
length of database: 327,214,780
effective HSP length: 118
effective length of query: 132
effective length of database: 205,136,110
effective search space: 27077966520
effective search space used: 27077966520
T: 11
A: 40
X1: 16 ( 7.3 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 41 (21.7 bits)
S2: 79 (35.0 bits)
```

ORF4 BLAST Search Ouptput

Conceptual Translation of the ORF4 Sequence

```
  1  MPGSTLRTLP TLSMVSAPAL SSIPSQSEYP IPRPVRCKNG SVHPPLLGIP
 51  VRVTFVHQDG TAAADRIPME SKIQIEDLTE LAVKTGEQFI QSLLRLNDKK
101  KIFVNMVEHD TLVYSKNIKD TVSEDQRAIE TFQQRVISGN LLGFHCPAIC
151  HFIVKIVEKE GGSYKCHHCD KGKAIVEDAS ADSGPKDGPP PTRSIVEKED
201  VPTTSSKQVD
```

1507 ORF 4 results E=1
BLASTP 2.2.2 [Jan-08-2002]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query=
(210 letters)

Database: /biolib/db/blast2/public/2/nr
919,285 sequences; 288,558,979 total letters

Searching.....done

Include the input sequence

Sequences producing significant alignments:	Score (bits)	E Value	
gi|9626944|ref|NP_056729.1| (NC_001497) inclusion body matrix pr...	289	1e-77	(84%),
gi|840739|emb|CAA55975.1| (X79465) ORF VI [Cauliflower mosaic vi...	282	1e-75	(81%),
gi|729799|sp|Q05651|IBMP_CAMVW INCLUSION BODY MATRIX PROTEIN (VI...	282	1e-75	(81%),
gi|124046|sp|P03558|IBMP_CAMVC INCLUSION BODY MATRIX PROTEIN (VI...	281	3e-75	(81%),
gi|124044|sp|P22547|IBMP_CAMV4 INCLUSION BODY MATRIX PROTEIN (VI...	279	1e-74	(80%),

```
gi|124049|sp|P18617|IBMP_CAMVP INCLUSION BODY MATRIX PROTEIN (VI...   276  1e-73 (80%),
gi|400039|sp|Q02954|IBMP_CAMVE INCLUSION BODY MATRIX PROTEIN (VI...   275  2e-73 (80%),
gi|1352411|sp|Q00957|IBMP_CAMVN INCLUSION BODY MATRIX PROTEIN (V...   275  3e-73 (80%),
gi|331555|gb|AAA21737.1| (M90543) inclusion body matrix protein ...   271  2e-72 (79%),
gi|124048|sp|P13218|IBMP_CAMVJ INCLUSION BODY MATRIX PROTEIN (VI...   268  3e-71 (78%),
gi|124047|sp|P03557|IBMP_CAMVD INCLUSION BODY MATRIX PROTEIN (VI...   264  5e-70 (77%),
gi|75493|pir||QQCV6 hypothetical protein 6 - cauliflower mosaic ...   259  9e-69 (76%),
gi|124045|sp|P16666|IBMP_CAMVB INCLUSION BODY MATRIX PROTEIN (VI...   226  9e-59 (69%),
gi|5002172|gb|AAD37342.1|AF140604_6 (AF140604) inclusion body pr...   222  2e-57 (65%),
gi|331576|gb|AAA66609.1| (M32811) matrix protein [Cauliflower mo...   125  4e-28 (91%),
gi|19919895|ref|NP_612578.1| (NC_003498) Inclusion body matrix p...    58  7e-08 (32%),
gi|16805108|ref|NP_473136.1| (NC_000521) PFC0005w (MAL3P8.1), va...    35  0.64 (28%),

>gi|9626944|ref|NP_056729.1| (NC_001497) inclusion body matrix protein [Cauliflower
mosaic
          virus]
 gi|124050|sp|P03559|IBMP_CAMVS INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|75495|pir||QQCV6S hypothetical protein 6 - cauliflower mosaic virus (strain
          Strasbourg)
 gi|58827|emb|CAA23461.1| (V00141) reading frame (VI) [Cauliflower mosaic virus]
         Length = 520

 Score =  289 bits (739), Expect = 1e-77
 Identities = 149/177 (84%), Positives = 153/177 (86%), Gaps = 4/177 (2%)

Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MESKIQIEDLTELAVKTGEQFIQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKVVPTESKAMESKIQIEDLTELAVKTGEQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 463

Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 210
           VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD
Sbjct: 464 VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 520


>gi|840739|emb|CAA55975.1| (X79465) ORF VI [Cauliflower mosaic virus]
         Length = 520

 Score =  282 bits (722), Expect = 1e-75
 Identities = 145/177 (81%), Positives = 151/177 (84%), Gaps = 4/177 (2%)

Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MESKIQIEDLTELAVKTGEQFIQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKVVPTESKAMESKIQIEDLTELAVKTGEQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 463

Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 210
            + VEKEGGSYKCHHCDKGKAIV+DASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD
Sbjct: 464 ERTVEKEGGSYKCHHCDKGKAIVQDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 520
```

```
>gi|729799|sp|Q05651|IBMP_CAMVW INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|331574|gb|AAA46360.1| (L09053) major inclusion body protein [Cauliflower mosaic
virus]
          Length = 520

 Score =  282 bits (722), Expect = 1e-75
 Identities = 144/177 (81%), Positives = 153/177 (86%), Gaps = 4/177 (2%)

Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MESKIQIEDLTELAVKTGEQFIQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKLVPTESKAMESKIQIEDLTELAVKTGEQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTL+YSKNIK+TVSEDQRAIETFQQRVISGNLLGFHCPAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEHDTLIYSKNIKETVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 463
```

```
Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 210
           VKIVEKEGG+Y+CHHCDKGKAIV+DASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD
Sbjct: 464 VKIVEKEGGTYQCHHCDKGKAIVKDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 520


>gi|124046|sp|P03558|IBMP_CAMVC INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|75494|pir||QQCV6C hypothetical protein 6 - cauliflower mosaic virus (strain CM1841)
          Length = 520

 Score =  281 bits (719), Expect = 3e-75
 Identities = 144/177 (81%), Positives = 151/177 (84%), Gaps = 4/177 (2%)

Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MESKIQIEDLTELAVKTGEQFIQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKVVPTESKAMESKIQIEDLTELAVKTGEQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 463

Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 210
            + VEKEGG+YKCHHCDKGKAIV+DASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD
Sbjct: 464 KRTVEKEGGTYKCHHCDKGKAIVQDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 520


>gi|124044|sp|P22547|IBMP_CAMV4 INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|331578|gb|AAA03526.1| (M23620) gene six protein [Cauliflower mosaic virus]
          Length = 520

 Score =  279 bits (714), Expect = 1e-74
 Identities = 142/177 (80%), Positives = 150/177 (84%), Gaps = 4/177 (2%)

Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MESKIQIEDLTELAVKTGEQFIQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKVVPTESKAMESKIQIEDLTELAVKTGEQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 463

Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 210
            + VEKEGG+YKCHHCDKGKAI++DAS DSGPKDGPPPTRSIVEKEDVPTTSSKQVD
Sbjct: 464 ERTVEKEGGTYKCHHCDKGKAIIQDASTDSGPKDGPPPTRSIVEKEDVPTTSSKQVD 520


>gi|124049|sp|P18617|IBMP_CAMVP INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|93549|pir||S11217 viroplasmin - cauliflower mosaic virus
 gi|58905|emb|CAA37853.1| (X53860) viroplasmin (AA 1-520) [Cauliflower mosaic virus]
          Length = 520

 Score =  276 bits (706), Expect = 1e-73
 Identities = 142/177 (80%), Positives = 150/177 (84%), Gaps = 4/177 (2%)
```

```
Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MESKIQIEDLTELAVKTGEQ IQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKVVPTESKAMESKIQIEDLTELAVKTGEQSIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRA+ETFQQRVISGNLLGFHCPAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAMETFQQRVISGNLLGFHCPAICHFI 463

Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 210
            + VEKEGGSYKCHHCDKGKAIV++ASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD
Sbjct: 464 ERTVEKEGGSYKCHHCDKGKAIVQNASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 520


>gi|400039|sp|Q02954|IBMP_CAMVE INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|484343|pir||JN0498 hypothetical 58K protein - cauliflower mosaic virus (isolate
BBC)
 gi|293186|gb|AAA62376.1| (M90542) inclusion body matrix protein [Cauliflower mosaic
virus]
          Length = 520

 Score =  275 bits (703), Expect = 2e-73
 Identities = 143/177 (80%), Positives = 149/177 (83%), Gaps = 4/177 (2%)

Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MESKIQIEDLTELAVKTGEQFIQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKVVPTESKAMESKIQIEDLTELAVKTGEQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVE DTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEDDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 463

Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 210
            + VEKEGGSYK HHCDKGKAIV+DASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD
Sbjct: 464 ERTVEKEGGSYKVHHCDKGKAIVQDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 520


>gi|1352411|sp|Q00957|IBMP_CAMVN INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|331572|gb|AAA46359.1| (M90541) inclusion body matrix protein [Cauliflower mosaic
virus]
 gi|445599|prf||1909346F inclusion body matrix protein [Cauliflower mosaic virus]
          Length = 520

 Score =  275 bits (702), Expect = 3e-73
 Identities = 142/177 (80%), Positives = 149/177 (83%), Gaps = 4/177 (2%)

Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G VH P   I + V+          +   MESKIQIEDLTELAVKTG QFIQSL
Sbjct: 348 PVWTIQGLVHKPRQVIEIGVS----KKVVPTESKAMESKIQIEDLTELAVKTGGQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCP+ICHF+
Sbjct: 404 LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPSICHFM 463
```

```
Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 210
           + VEKEGGSYK HHCDKGKAIV+DASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD
Sbjct: 464 ERTVEKEGGSYKVHHCDKGKAIVQDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 520


>gi|331555|gb|AAA21737.1| (M90543) inclusion body matrix protein [Cauliflower mosaic
virus]
          Length = 520

 Score =  271 bits (694), Expect = 2e-72
 Identities = 141/177 (79%), Positives = 148/177 (82%), Gaps = 4/177 (2%)

Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MESKIQIE LTELAVKTGEQFIQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKVVPTESKAMESKIQIEALTELAVKTGEQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 463

Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD 210
           + VEKEGGSYKCHHCDKGKAIV++ASADSGPKDG PPTRSIVEK DVPTTSSKQVD
Sbjct: 464 ERTVEKEGGSYKCHHCDKGKAIVQNASADSGPKDGSPPTRSIVEKGDVPTTSSKQVD 520


>gi|124048|sp|P13218|IBMP_CAMVJ INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|76778|pir||S06092 inclusion body matrix protein - cauliflower mosaic virus
 gi|58829|emb|CAA33037.1| (X14911) inclusion body matrix protein (AA 1-522)
[Cauliflower
           mosaic virus]
          Length = 522

 Score =  268 bits (685), Expect = 3e-71
 Identities = 140/179 (78%), Positives = 149/179 (83%), Gaps = 6/179 (3%)

Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MESKIQIEDLTELAVK+GEQFIQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKIVPTESKAMESKIQIEDLTELAVKSGEQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 463

Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSG--PKDGPPPTRSIVEKEDVPTTSSKQVD 210
           +K VEKEGG+YKCHHC+KGKAIV+DAS D G   KDGPPPTRSIVEKEDVPTTSSKQVD
Sbjct: 464 MKTVEKEGGAYKCHHCEKGKAIVKDASTDRGTTDKDGPPPTRSIVEKEDVPTTSSKQVD 522


>gi|124047|sp|P03557|IBMP_CAMVD INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|331548|gb|AAA46351.1| (M10376) inclusion body protein [Cauliflower mosaic virus]
          Length = 522

 Score =  264 bits (674), Expect = 5e-70
 Identities = 138/179 (77%), Positives = 147/179 (82%), Gaps = 6/179 (3%)
```

```
Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MES+IQIEDLTELAVKTGEQFIQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKVIPTESKAMESRIQIEDLTELAVKTGEQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTLVYSKNIK+T SEDQRAIETFQQRVISGNLLGFHCPAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEHDTLVYSKNIKETDSEDQRAIETFQQRVISGNLLGFHCPAICHFI 463

Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSG--PKDGPPPTRSIVEKEDVPTTSSKQVD 210
           +K VEKEGG+YKCHHCDKGKAIV+DASAD G   K GPPPTRSIVEKEDVP TSSKQVD
Sbjct: 464 MKTVEKEGGAYKCHHCDKGKAIVQDASADEGTTDKSGPPPTRSIVEKEDVPNTSSKQVD 522


>gi|75493|pir||QQCV6 hypothetical protein 6 - cauliflower mosaic virus (strain D/ H)
          Length = 522

 Score =  259 bits (663), Expect = 9e-69
 Identities = 137/179 (76%), Positives = 146/179 (81%), Gaps = 6/179 (3%)

Query: 34  PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL 93
           PV    G +H P   I + V+          +   MES+IQIEDLTELAVKTGEQFIQSL
Sbjct: 348 PVWTIQGLLHKPRQVIEIGVS----KKVIPTESKAMESRIQIEDLTELAVKTGEQFIQSL 403

Query: 94  LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI 153
           LRLNDKKKIFVNMVEHDTLVYSKNIK+T SEDQRAIETFQQRVISGNLLGFH PAICHFI
Sbjct: 404 LRLNDKKKIFVNMVEHDTLVYSKNIKETDSEDQRAIETFQQRVISGNLLGFHXPAICHFI 463

Query: 154 VKIVEKEGGSYKCHHCDKGKAIVEDASADSG--PKDGPPPTRSIVEKEDVPTTSSKQVD 210
           +K VEKEGG+YKCHHCDKGKAIV+DASAD G   K GPPPTRSIVEKEDVP TSSKQVD
Sbjct: 464 MKTVEKEGGAYKCHHCDKGKAIVQDASADEGTTDKSGPPPTRSIVEKEDVPNTSSKQVD 522


>gi|124045|sp|P16666|IBMP_CAMVB INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|76777|pir||JA0072 hypothetical protein 6 - cauliflower mosaic virus (strain Bari 1)
 gi|222013|dbj|BAA00241.1| (D00335) 522 a.a.(putative) [Cauliflower mosaic virus]
          Length = 522

 Score =  226 bits (577), Expect = 9e-59
 Identities = 124/178 (69%), Positives = 138/178 (76%), Gaps = 11/178 (6%)

Query: 34  PVRCKNGSVHPP--LLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQ 91
           PV    G +H P  ++ I V    V ++  A      MESK   EDL ELA KTGEQFIQ
Sbjct: 349 PVWQDQGLLHKPKHVIEIGVSKKIVPKESKA------MESKDHSEDLIELATKTGEQFIQ 402

Query: 92  SLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICH 151
           SLLRLNDKKKIFVN+VEHDTLVYSKN K+TVSEDQRAIETFQQRVI+ NLLGFHCP+ICH
Sbjct: 403 SLLRLNDKKKIFVNLVEHDTLVYSKNTKETVSEDQRAIETFQQRVITPNLLGFHCPSICH 462

Query: 152 FIVKIVEKEGGSYKCHHCDKGKAIVEDASADS--GPKDGPPPTRSIVEKEDVPTTSSK 207
           FI + VEKEGG+YKCHHCDKGKAIV+DASADS    K+GPP T + VEKEDV TTSSK
Sbjct: 463 FIKRTVEKEGGAYKCHHCDKGKAIVQDASADSKVADKEGPPLTTN-VEKEDVSTTSSK 519


>gi|5002172|gb|AAD37342.1|AF140604_6 (AF140604) inclusion body protein [Cauliflower
mosaic virus]
```

```
          Length = 524

 Score =  222 bits (566), Expect = 2e-57
 Identities = 120/183 (65%), Positives = 139/183 (75%), Gaps = 12/183 (6%)

Query: 34  PVRCKNGSVHPP--LLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQ 91
           PV    G +H P  ++ I V    V ++  A      M+SKIQ+EDLTELA KTGEQFIQ
Sbjct: 348 PVWTIEGLLHKPRHVIEIGVSKKVVPKESKA------MQSKIQMEDLTELATKTGEQFIQ 401

Query: 92  SLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICH 151
           SLL+LN+K KIFVNMVEHD LVYSKN+K+T +EDQRAIETFQ+RVISG+LLGFHCP+ICH
Sbjct: 402 SLLKLNEKNKIFVNMVEHDVLVYSKNLKETTAEDQRAIETFQKRVISGSLLGFHCPSICH 461

Query: 152 FIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKD--GPPPTRSIVEKEDV--PTTSSK 207
           FI + VEKEGG+Y+CH CDKGKAIVED  ADSGP D  G   T+  VEKED    TTSSK
Sbjct: 462 FIKRTVEKEGGAYQCHQCDKGKAIVEDKPADSGPADSLGSTTTKENVEKEDAVSTTTSSK 521

Query: 208 QVD 210
           QVD
Sbjct: 522 QVD 524


>gi|331576|gb|AAA66609.1| (M32811) matrix protein [Cauliflower mosaic virus]
          Length = 61

 Score =  125 bits (313), Expect = 4e-28
 Identities = 56/61 (91%), Positives = 59/61 (95%)

Query: 150 CHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQV 209
           CHFI + VEKEGGSYKCHHCDKGKAI++DASADSGPKDGPPPTRSIVEKEDVPTTSSKQV
Sbjct: 1   CHFIERTVEKEGGSYKCHHCDKGKAIIQDASADSGPKDGPPPTRSIVEKEDVPTTSSKQV 60

Query: 210 D 210
           D
Sbjct: 61  D 61


>gi|19919895|ref|NP_612578.1| (NC_003498) Inclusion body matrix protein [Carnation
etched ring
          virus]
 gi|124051|sp|P05401|IBMP_CERV INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN)
 gi|76776|pir||S00855 hypothetical protein 6 - carnation etched ring virus
 gi|58864|emb|CAA28361.1| (X04658) pot. ORF 6 (AA 1-496) [Carnation etched ring virus]
 gi|225357|prf||1301227F ORF 6 [Carnation etched ring virus]
          Length = 496

 Score = 57.8 bits (138), Expect = 7e-08
 Identities = 42/129 (32%), Positives = 60/129 (45%), Gaps = 8/129 (6%)

Query: 69  MESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRA 128
           ME+ +  E    +A +   +F Q++  +   +KIF+   +   L+YS + K+   ED  A
Sbjct: 376 MEAVVLKEQWQGIAEEKAIEFFQAIEDILSNEKIFIIYCDDRILIYSSSPKERTKEDLMA 435
```

```
Query: 129 IETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDG 188
           I  FQ  V S  LLGFH   IC ++ K     G  Y C    KGKA++       SGP
Sbjct: 436 ILNFQSEVSSCKLLGFHSDKICSYLNK-KASVGKPYSCP--QKGKAVI-----TSGPSFS 487

Query: 189 PPPTRSIVE 197
               T S  E
Sbjct: 488 VEDTLSDTE 496


>gi|16805108|ref|NP_473136.1| (NC_000521) PFC0005w (MAL3P8.1), var gene (3D7-varT3-1),
len: 2209
            aa; Similarity to P.falciparum PfEMP-1 gene. Part of a
            multi-gene family, donor sequence corrected [Plasmodium
            falciparum]
 gi|7264037|emb|CAB39115.2| (AL034560) PFC0005w (MAL3P8.1), var gene (3D7-varT3-1),
len: 2209 aa;
            Similarity to P.falciparum PfEMP-1 gene. Part of a
            multi-gene family, donor sequence corrected [Plasmodium
            falciparum 3D7]
          Length = 2209

 Score = 34.7 bits (78), Expect = 0.64
 Identities = 28/98 (28%), Positives = 37/98 (37%), Gaps = 24/98 (24%)

Query: 124  EDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDK---------GKA 174
             E++  +E  +Q  +     +G   PAIC  + +  EKE G      CDK         G
Sbjct: 1622 EEEELLEEDEQNTVGKEKVGNKAPAICGDVEEQKEKEEGD-----CDKAVTPDSDTGGNG 1676

Query: 175  IVEDA----------SADSGPKDGPPPTRSIVEKEDVP 202
               ED+           S D G    PPP  S  EK   P
Sbjct: 1677 EKEDSRSEEEEEVSGSGDQGSPPAPPPPESPQEKAPAP 1714


  Database: /biolib/db/blast2/public/2/nr
    Posted date:  May 12, 2002  8:16 PM
  Number of letters in database: 288,558,979
  Number of sequences in database:  919,285

Lambda     K      H
   0.316    0.134    0.387

Gapped
Lambda     K      H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 121,810,209
```

```
Number of HSP's that attempted gapping in prelim test: 15425
Number of HSP's gapped (non-prelim): 17
length of query: 210
length of database: 288,558,979
effective HSP length: 115
effective length of query: 95
effective length of database: 182,841,204
effective search space: 17369914380
effective search space used: 17369914380
T: 11
A: 40
X1: 16 ( 7.3 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 41 (21.6 bits)
S2: 77 (34.3 bits)
```

Note to Study File Phi 2002-036

From Study Director Donald Walters October 24, 2002

Donald Walters 10/24/02

Subsequent to the completion and signing of the report for study Phi 2002-036 two minor changes were made to page 10 of the report in order to correct minor inaccuracies. The changes are described below. Since no one single report has been produced on the sequence of the 1507 maize insert the wording was changed to correct the statement in the 2nd paragraph on page 10.

The text of the 2nd sentence of paragraph 2 page 10 of the report has been changed from:

The two regions have been labeled ORF3 and ORF4 and fully described in a separate report on the sequence of the 1507 maize insert.

and now reads:

The two regions have been labeled ORF3 and ORF4 and fully described separately.

Reevaluation of the source sequence revealed the minor correction needed in the 3rd paragraph statement regarding where the sequence found in 1507 originated.

The text of the 1st sentence of paragraph 3 page 10 of the report has been corrected from:

ORF4 is a 630 nt sequence begining in the Cry1F - ORF25 terminator, spanning 142 nt of ORF25, 55 nt of intervening sequence, and 431 nt of the CaMV 35S promoter.

and now reads:

ORF4 is a 630 nt sequence begining in the Cry1F - ORF25 terminator, spanning 141 nt of ORF25, 56 nt of intervening sequence, and 433 nt of the CaMV 35S promoter.

Neither of these changes affects the results or conclusions presented in report Phi 2002-036 and neither correction affects the validity of the rest of the report. Therefore as study director the revised text will be substituted on page 10 of the report and the resulting amended report is to be considered as the accepted, authentic final report for study Phi 2002-036. Both the uncorrected and corrected versions of page 10 are included with this note to the study file.

Early version DJW 10/24/02

INTRODUCTION

The sequence similarity of the two unintended open reading frame sequences present in maize line 1507 was evaluated against proteins present in the public sequence databases. The similarity searches were conducted to identify possible toxicity, anti-nutritional, or other safety concerns arising from the potential expression of the two hypothetical proteins. Database annotations of the resulting group of proteins with measurable similarity to the open reading frame sequences were manually inspected to identify those that may be toxic to humans or livestock, or possess catalytic activities that could produce metabolites that could impact the safety or nutritional quality of food or feed derived from maize line 1507.

METHODS

DESCRIPTION OF IDENTIFIED ORF's IN MAIZE LINE 1507

Sequence analysis of the recombinant DNA insert present in maize line 1507 has identified two unintended open reading frame regions which code for peptide sequences greater than 200 amino acid residues. The two regions have been labeled ORF3 and ORF4 and are fully described in a separate report on the sequence of the 1507 maize insert. The sequence data indicates the presence of an open reading frame in the insert of maize line 1507 which starts with an ATG start codon in a 5' Cry1F fragment and terminates within an inverted PAT fragment. This open reading frame has been designated ORF3 and codes for a conceptual protein of 250 amino acid residues presented along with the search output in the appendix of this report. The ORF3 sequence was subjected to a comprehensive BLAST search of the public sequence databases.

ORF4 is a 630 nt sequence begining in the Cry1F - ORF25 terminator, spanning 142 nt of ORF25, 55 nt of intervening sequence, and 431 nt of the CaMV 35S promoter. ORF4 is present in the transformation construct used to produce maize line 1507 and the sequence has been well confirmed in both the plant and the plasmid vector. The conceptually translated open reading frame encompassing 210 amino acid residues, was used for the subsequent BLAST searching. The ORF4 sequence is provided in the appendix.

Final Version
DOW 10/24/02

INTRODUCTION

The sequence similarity of the two unintended open reading frame sequences present in maize line 1507 was evaluated against proteins present in the public sequence databases. The similarity searches were conducted to identify possible toxicity, anti-nutritional, or other safety concerns arising from the potential expression of the two hypothetical proteins. Database annotations of the resulting group of proteins with measurable similarity to the open reading frame sequences were manually inspected to identify those that may be toxic to humans or livestock, or possess catalytic activities that could produce metabolites that could impact the safety or nutritional quality of food or feed derived from maize line 1507.

METHODS

DESCRIPTION OF IDENTIFIED ORF's IN MAIZE LINE 1507

Sequence analysis of the recombinant DNA insert present in maize line 1507 has identified two unintended open reading frame regions which code for peptide sequences greater than 200 amino acid residues. The two regions have been labeled ORF3 and ORF4 and fully described separately. The sequence data indicates the presence of an open reading frame in the insert of maize line 1507 which starts with an ATG start codon in a 5' Cry1F fragment and terminates within an inverted PAT fragment. This open reading frame has been designated ORF3 and codes for a conceptual protein of 250 amino acid residues presented along with the search output in the appendix of this report. The ORF3 sequence was subjected to a comprehensive BLAST search of the public sequence databases.

ORF4 is a 630 nt sequence begining in the Cry1F - ORF25 terminator, spanning 141 nt of ORF25, 56 nt of intervening sequence, and 433 nt of the CaMV 35S promoter. ORF4 is present in the transformation construct used to produce maize line 1507 and the sequence has been well confirmed in both the plant and the plasmid vector. The conceptually translated open reading frame encompassing 210 amino acid residues, was used for the subsequent BLAST searching. The ORF4 sequence is provided in the appendix.