SUMMARY

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

STUDY TITLE

Comparison of Amino Acid Sequence Similarity Between the ORF3 and ORF4 Sequences Present in Maize Line 1507 and Known Protein Allergens

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-2002-035

Comparison of Amino Acid Sequence Similarity Between the ORF3 and ORF4 Sequences Present in Maize Line 1507 and Known Protein Allergens

SUMMARY

The conceptual translations of two unintended open reading frames present in maize line 1507 (ORF3 and ORF4) were compared to those of a dataset containing known allergenic proteins to evaluate the allergenic potential of the hypothetical proteins. The allergen dataset was assembled and the similarity searches were peformed in accordance with the recommendations of the FAO/WHO Expert consultation on Allergenicity of Foods Derived from Biotechnology, January 2001. A description of the dataset construction is included in the report. Two different similarity comparison strategies were applied to each of the two theoretical peptide sequences. A search was conducted for short contiguous amino acid strings occurring in the Event 1507 ORF3 and ORF4 sequences that match identical strings in the allergen dataset. The two conceptual protein sequences were analyzed for the presence of contiguous strings of 8 or more amino acids matching identical sequences from the allergen dataset. Neither of the hypothetical protein sequences contains contiguous residue strings of eight or more amino acids that match identical sequences within the allergen dataset. The results of the similarity searches indicate the hypothetical ORF3 and ORF4 sequences do not contain contiguous residue strings of eight or more amino acids that match identical sequences within the allergen dataset.

Both the ORF3 and ORF4 coding sequences were also compared to the allergen dataset by generating a series of all-possible overlapping 80 residue fragments for each sequence and comparing each of these fragments to the dataset entries using the FastA algorithm. According to FAO/WHO recommendations allergenic potential must be considered in further detail when there is a greater than 35% identity between the protein of interest and a known allergen over the length of the 80 residue fragment. There were no sequence homologies greater than 35% over 80 residues for either the ORF3 or ORF4 conceptual protein sequences.

In conclusion, sequence similarity comparison of the conceptual translations of the ORF3 and ORF4 sequences in maize line 1507 to known allergens does not indicate these proteins are potential allergens.

Comparison of Amino Acid Sequence Similarity Between the ORF3 and ORF4 Sequences Present in Maize Line 1507 and Known Protein Allergens

DATA REQUIREMENTS

None

AUTHOR(S)

Robert F. Cressman, Donald S. Walters

STUDY COMPLETED ON

October 4, 2002

PERFORMING LABORATORY

DuPont Crop Genetics
DuPont Experiment Station
Wilmington, DE 19880

LABORATORY STUDY ID

PHI-2002-035

STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS

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

Title: Comparison of Amino Acid Sequence Similarity Between the ORF3 and ORF4 Sequences Present in Maize Line 1507 and Known Protein Allergens

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).*

Company: Pioneer Hi-Bred International, Inc. – A 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: Comparison of Amino Acid Sequence Similarity Between the ORF3 and ORF4 Sequences Present in Maize Line 1507 and Known Allergens

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

Study Sponsor
Larry Zeph, Ph.D.
Pioneer HiBred

Date 10/11/2002

Donald S Walters

Study Director/ Author
Donald S. Walters
DuPont Crop Genetics

10/4/02
Date

QUALITY ASSURANCE STATEMENT

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

Title: Comparison of Amino Acid Sequence Similarity Between the ORF3 and ORF4 Sequences Present in Maize Line 1507 and Known Protein Allergens

Study Initiation Date: 3/01/2002 Study Completion Date: 10/04/2002

NON-GLP STUDY

SIGNATURE PAGE

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

10/4/02
Date

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

10/4/02
Date

Lawrence Zeph
Sponsor
Larry Zeph, Ph.D.

10/11/2002
Date

Greg Ladics
Peer Reviewer
Greg Ladics, Ph.D.

10/8/02
Date

TABLE OF CONTENTS

ABSTRACT 10
INTRODUCTION 12

METHODS

ALLERGEN DATASET CONSTRUCTION AND UPDATE 12
SIMILARITY COMPARISON STRATEGIES 16

RESULTS

ORF3 SHORT LINEAR EPITOPE SEARCH 19
ORF3 80 RESIDUE FASTA SEARCH 20
ORF4 SHORT LINEAR EPITOPE SEARCH 20
ORF4 80 RESIDUE FASTA SEARCH 21
CONCLUSIONS 21
REFERENCES 22

APPENDIX

Output for ORF3 Short Linear Epitope Search 23
Excerpt of Output for ORF3 80 Residue FastA Search 38
Output for ORF4 Short Linear Epitope Search 54
Excerpt of Output for ORF4 80 Residue FastA Search 63

Comparison of Amino Acid Sequence Similarity Between the ORF3 and ORF4 Sequences Present in Maize Line 1507 and Known Protein Allergens

ABSTRACT

The conceptual translations of two unintended open reading frames present in maize line 1507 (ORF3 and ORF4) were compared to those of a dataset containing known allergenic proteins to evaluate the allergenic potential of the hypothetical proteins. The allergen dataset was assembled and the similarity searches were peformed in accordance with the recommendations of the FAO/WHO Expert consultation on Allergenicity of Foods Derived from Biotechnology, January 2001. A description of the dataset construction is included in the report. Two different similarity comparison strategies were applied to each of the two theoretical peptide sequences. A search was conducted for short contiguous 8 amino acid strings occurring in the Event 1507 ORF3 and ORF4 sequences that match identical strings in the allergen dataset. The results of the similarity searches indicate the hypothetical ORF3 and ORF4 sequences do not contain contiguous residue strings of eight or more amino acids that match identical sequences within the allergen dataset.

Both the ORF3 and ORF4 coding sequences were also compared to the allergen dataset by generating a series of all-possible overlapping 80 residue fragments for each sequence and comparing each of the fragments to the dataset entries using the FastA algorithm. According to FAO/WHO recommendations allergenic potential must be considered in further detail when there is a greater than 35% identity between the protein of interest and a known allergen over the length of the 80 residue fragment. There were no sequence homologies greater than 35% over 80 residues for either the ORF3 or ORF4 conceptual protein sequences.

In conclusion, sequence similarity comparison of the conceptual translations of the ORF3 and ORF4 sequences in maize line 1507 to known allergens does not indicate these proteins are potential allergens.

INTRODUCTION

Sequence analysis of maize line 1507 indicates the presence of two unintended open reading frames in the recombinant insert in addition to the intended insert sequences that code for the Cry1F and PAT proteins. Since open reading frame sequences possess the theoretical potential to be transcribed and translated into peptides, efforts were made to insure that the hypothetical peptides do not exhibit sequence similarity to known toxins or allergens. The amino acid sequences produced by conceptual translation of the ORF3 and ORF4 open reading frames are provided in the appendix. This report describes the methods and results of an investigation of potential sequence similarity between the conceptual translations of the open reading frames and a dataset containing known allergenic proteins. The schemes used in the construction of the allergenic protein dataset are described in the report. Two different approaches were used to evaluate the amino acid sequence similarity of the two conceptually translated peptide sequences with the allergenic protein dataset. The ORF3 and ORF4 hypothetical sequences were first compared to the allergen dataset to identify any contiguous 8 amino acid matches present within the dataset entries. Secondly, the two hypothetical peptide sequences were considered as a series of all-possible overlapping contiguous strings of 80 amino acids and compared to all possible 80 amino acid blocks within the allergen dataset for percent identity.

METHODS

PIONEER HIBRED ALLERGEN DATASET CONSTRUCTION AND UPDATE

Initial Dataset Construction

The Allergen Reference Dataset used in the similarity evaluations described in this report was initially populated in September 2000 following the procedures described below.

A preliminary list of allergenic protein sequences and their accession numbers was compiled from the following sources:

- Metcalfe, D. D., J. D. Astwood, et al. (1996). "Assessment of the Allergenic Potential of Foods Derived from Genetically Engineered Crop Plants." Critical Reviews in Food Science and Nutrition **36(S)**: S165-S186.
- Bush, R. K. and S. L. Hefle (1996). Food Allergens. Critical Reviews in Food Science and Nutrition. **36(S):** S119-S163.
- Gendel, S. M. (1998). Sequence Databases for Assessing the Potential Allergenicity of Proteins Used in Transgenic Foods. Advances in Food and Nutrition Research. S. L. Taylor. San Diego, Academic Press. **42:** 63-92.
- Gendel, S. M. (1998). The Use of Amino Acid Sequence Alignments to Assess Potential Allergenicity of Proteins Used in Genetically Modified Foods. Advances in Food and Nutrition Research. S. L. Taylor. San Diego, Academic Press. **42:** 45-62.
- Gendel, S. M. (Last update unknown). Web Pages: Biotechnology Information for Food Safety - Food Allergy. http://www.iit.edu/~sgendel/fa.htm

These sequences were considered to be fully curated and were treated as confirmed allergen sequences. Their accession numbers, however, were checked against the public datasets to identify potential errors from the literature. Duplicate entries were removed from the list and the remaining 998 accessions became the starting allergen sequence reference list. Note that more than one accession number may be associated with a single allergen, and therefore, the total number of allergens represented in these datasets is somewhat less than the number of sequences listed. Redundancy of sequence information in this dataset does not negatively impact its use.

The dataset of known allergens from the five published compilations was then supplemented through a search of Public-Domain Protein Databases. The GCG Wisconsin Package (Version 10.0 VAX Open VMS) was used for full-text queries of the following databases. "Lookup", an implementation of the SRS indexing system, was used to search the complete text annotation of the public datasets.

Public Datasets Queried	Release (GCG)
GenPept	117.0
PIR-Protein	64.0
Swiss-Prot	36.0
SP-TrEMBL	12.0

Each of the public datasets was searched for the term *allerg* as well as a number of other search terms derived from species names, protein names, and allergen names associated with the initial 998 accessions identified in the published compilations. Following compilation and inspection of the text annotations for the search results, duplicate records were removed and accessions whose annotations did not support their inclusion as known allergens were extirpated. Those accessions resulting from the public dataset search and confirmed to allergens were merged with the starting 998 accessions in the initial reference set resulting in a final allergen reference set containing 1,509 accessions (as of Sept. 2000).

Update of Allergen Reference Dataset

In order to incorporate recent allergen additions to the public databases the reference dataset was updated during March 2002 following the procedures recommended during the validation of the initial dataset. Authors of the initial September 2000 dataset concluded that protein sequence records that do not contain the term *allerg* in their text annotations cannot be confirmed as allergens without further examination of the scientific literature. Furthermore the authors concluded that greater than 99% of the sequences in the final September 2000 dataset were identifiable through full-text searching for the *allerg* term. Therefore it was concluded that a full-text search procedure for *allerg* is sufficient to identify all confirmable allergen protein sequences from the public databases and that periodic updates of the allergen reference dataset would be performed in this manner. Furthermore, it was concluded that updates to the reference list could be performed by targeting the search to those records added to the public databases subsequent to the last reference list update.

In the March 2002 update of the Pioneer Allergen Reference Dataset the search term *allerg* was used to screen the public datasets for the interval between the last update in September 2000 and the latest release dates as presented in the table below.

Table 1. Public datasets queried.

Public Dataset	Release	Interval
GenPept	117.0 – 128.0	4/15/2000- 2/13/2002
PIR-Protein	64.0–71.0	3/31/2000-12/31/2001
Swiss-Prot	36.0 – 40.0	7/1998- 12/2001
SP-TrEMBL	12.0 – 19.0	11/1999 – 12/2001

The GCG (Version 10.2 Unix) tool "Lookup" (an implementation of the SRS indexing system) was used to search the complete text annotations of the entries in PIR, SWISS-PROT and Sp-TrEMBL for the dataset release intervals indicated above.

Because the GenPept database was not accessible through "Lookup", a query was made directly through the Entrez server at NCBI (http://www.ncbi.nlm.nih.gov/Entrez /) for the 4/15/2000-2/13/2002 interval noted above. The search term *allerg* was modified to allerg* due to limitations in the query structure. The search of the GenPept database resulted in 2974 additional protein "hits", for a total of 3226 proteins (2974 genpept +252 swissprot-sptrembl- PIR).
The 3226 accessions identified from the text search were transferred to a VectorNTI (InforMax, Inc., 6010 Executive Blvd., North Bethesda, MD, USA) database on a PC to facilitate manual inspection of the records. VectorNTI was chosen because sequence descriptions and annotations are easily accessible through a Windows-like graphical interface, and because the database structure allows for easy sorting and import/export into several common formats.

The set of 3226 protein accessions were then sorted to separate true allergens from spurious search term hits. It was assumed that all accessions containing the term allergen in the definition field were true allergens. A search of the definition field identified 763 accessions containing the term "allergen" which were exported to a separate database for inclusion in the Allergen Reference Dataset. Other filters were employed to remove human proteins, hits arising from the presence of the *allerg* term in journal titles, and other accessions containing incidental occurrences of the search term. Following the removal of these spurious hits the remaining 890 proteins were subjected to manual examination and sorting into allergen and non-allergen categories based upon the information in the text annotation and referenced journal abstracts when necessary. Ultimately 165 of these protein accessions were deemed allergens, and added to the 763 accessions identified earlier brought the total number of allergen sequences from the updated search to 928. This group of 928 accessions contained a proportion of duplicated sequences files as well as some duplication of the accessions in the original September 2000 Reference Dataset which were then eliminated. Although most of the duplicated accessions were eliminated, there are undoubtedly duplicate protein sequences remaining in the reference dataset. Redundancy of sequence information in the Reference Dataset should not negatively impact its use.

The final Allergen Reference Dataset as updated March 2002 contained a total of 2033 protein accessions representing an increase of 524 sequences over the dataset constructed in September 2000.

SIMILARITY COMPARISON STRATEGIES

Employing similarity search strategies conforming to the recommendations of the Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology, January 2001, the amino acid sequences of ORF3 and ORF4 present in maize line 1507 were evaluated against the 2033 protein accessions in the Pioneer Allergen Dataset. Although the FAO/WHO 2001 recommendations suggest that the threshold for allergenic epitope screening be reduced to stings of 6 contiguous residue matches, there is strong evidence that contiguous 6 amino acid matches occur commonly between unrelated proteins and is not a reliable criterion for predicting allergenic potential (ILSI Protein Allergenicity SubCommittee Report, July 2001, Allergenicity Assessment for Foods Derived from Genetically Modified Crops; Hileman et al., 2002). Therefore, the ORF3 and ORF4 sequences were screened for 8 contiguous amino acid matches in the allergen database.

Traditional similarity searches employ identity matrices that award positive scores for conserved, non-identical amino acid residues between a query and a target protein. While very useful for finding regions of functionally significant similarity, this approach may overlook short stretches of identical amino acids. In order to identify short stretches of identical amino acids, a modified identity matrix was employed that eliminated positive scoring contributions from non-identical but conserved residues (Figure 1). Validation of this matrix for use in identifying short stretches of identity was performed as follows: An 8 residue peptide sequence was artificially inserted into the sequence of a non-allergen protein. The resulting protein was added to the newly updated Allergen dataset. An identical stretch of 8 amino acids was also inserted into the rubisco large subunit protein from maize. This new protein was then used as a query in both a FastA (Pearson and Lipman (Proc. Natl. Acad. Sci. USA 85; 2444-2448 (1988)) and SSearch (William Pearson's implementation of the method of Smith and Waterman (Advances in Applied Mathematics 2; 482-489 (1981)) in the GCG software suite (Wisconsin Genetics). The results using the two different algorithms with the common identity matrix were essentially identical and in both cases the eight amino acid "spiked" sequence was identified. Three other sequences containing identical strings of 6 amino acids were also identified from the validation search. When the identical searches were conducted with a conventional substitution matrix neither was able to identify the "spiked" sequence. As expected both FastA and SSearch results are sensitive to the threshold expectation (E) value specified in performing the search. As reported, the validation exercise was able to identify strings of 6 identical amino acids with the expectation value cut-off set at 10.0. Performing the search using a cutoff E value of 1.0, only the 8 residue identical match was identified. The more sensitive and conservative cutoff of E = 10.0 was used in the ORF3 and ORF4 comparisons to ensure that similarities of concern were identified.

It should be mentioned here that there are limitations in the use of the FastA algorithm in conducting searches for 8 residue identical strings. The FastA algorithm is designed to identify the single "best" alignment between the query sequence and the entries in the dataset being searched.

Therefore, in instances where there are multiple 8 residue matches between two proteins FastA will only report a single "best" match between the query sequence and the target accession in the allergen dataset. Also, in some instances residue matches with less than 8 contiguous identical matching strings, eg. 2 contiguous identical matching strings of 7 residues interrupted by a single residue mismatch, may produce higher Expect values than an 8 residue match. Again in this instance the 7 + 7 alignment would be returned by FastA in preference to any 8 residue strings

Figure 1. Modified Amino Acid Substitution Matrix *(allornone14.cmp)* for 8 Amino Acid Identical Strings

Amino acid	A	B	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	X	Y	Z	*
A	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
B	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
C	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
D	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
E	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
F	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
G	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
H	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
I	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
K	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
L	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
M	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
N	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
P	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-25	-200
Q	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-25	-200
R	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-25	-200
S	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-25	-200
T	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-25	-200
V	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-25	-200
W	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-25	-200
X	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-25	-200
Y	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-25	-200
Z	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	-25	5	-200
*	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	-200	5

RESULTS

ORF3 Short Linear Epitope Search

Sequence analysis of the recombinant DNA insert present in maize line 1507 has identified two unintended open reading frame regions coding 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 analysis of the maize line 1507 insert. The data indicate 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.

The translation of the ORF3 coding sequence was compared to the sequences in the Pioneer Allergen Database as previously described. The search results indicate there are no stretches of eight contiguous amino acids from the conceptually translated ORF3 sequence that are identical to strings found in any of the allergens in the dataset. The longest contiguous matching string present in the Pioneer allergen dataset was six residues. Although the FAO/WHO Consultation document recommends that matching residue strings of 6 or more amino acids may be an indication of allergenic potential, others (ILSI Protein Allergenicity SubCommittee Report, July 2001, Allergenicity Assessment for Foods Derived from Genetically Modified Crops; Hileman et al., 2002) have concluded that six amino acid matches occur too randomly and frequently to be useful as an indicator of allergenic potential. Therefore, the ORF3 and ORF4 sequences were screened for 8 contiguous amino acid matches in the allergen database. The output of the search is presented in the Appendix.

In conclusion, comprehensive evaluation of the ORF3 coding sequence present in maize line 1507 indicates there are no contiguous amino acid strings of 8 residues or longer that match identical sequences from the Pioneer allergen dataset. Consequently, there does not appear to be cause for concern for allergenic potential from the ORF3 coding sequence.

ORF3 80 Residue FastA Search

The ORF3 sequence representing a hypothetical peptide sequence of 250 amino acids was segmented into a series of all possible, overlapping 80 amino acid peptides using the MID function in Excel. The segmentation resulted in 171 peptide sequences of 80 residues each. The fragment list was placed into FastA format and subjected to FastA analysis against the allergen dataset described in the Methods section using the "fasta" command in GCG. The default word size (2), gap creation and extension penalties (12 and 2, respectively) and the default identity matrix (blosum50) were used in the searches. Because of the inability of the GCG FastA implementation to accept multiple sequences, a Perl script was used to submit the sequences one at a time to the program and concatenate the resulting output. The search was performed using an Expect cutoff value of E=1.0. Histogram output was suppressed to reduce file size.

None of the resulting comparisons of the 80 residue fragments to the public database sequences scored higher than the 35% identity threshold set forth by the FAO/WHO recommendations. The most relevant portion of the search output is provided in the Appendix.

ORF4 Short Linear Epitope Search

The ORF4 coding sequence present in maize line 1507 was compared to the sequences in the Pioneer Allergen Database as described earlier. A 210 amino acid sequence comprises the complete ORF4 open reading frame. The search results indicate no stretches of eight contiguous amino acids from the ORF4 coding region are identical to strings found in any of the allergens in the dataset. The output from the short linear epitope search is provided in the Appendix.

ORF4 Coding Sequence 80 Residue FastA Search

Following recommended procedures (FAO/WHO Expert consultation on Allergenicity of Foods Derived from Biotechnology, January 2001) the ORF4 coding sequence was segmented into a series of all possible, overlapping 80 amino acid peptides using the MID function in Excel. The 210 amino acid ORF4 sequence comprised the complete open reading frame present in maize line 1507. The 131 individual peptide sequences of 80 residues each were placed into FastA format and subjected to FastA analysis against the Pioneer allergen dataset using the "fasta" command in GCG. The default word size (2), gap creation and extension penalties (12 and 2, respectively) and the default identity matrix (blosum50) were used in the searches. A Perl script was used to submit the sequences one at a time into the program and concatenate the resulting output. This search returned results using a cutoff expectation value (E) of less than 1.0 with histogram output suppressed.

None of the resulting comparisons was greater than the 35% identity threshold set forth by the FAO/WHO 2001 recommendations. The highest identity matches observed were less than 30%. The most relevant portion of the search output is provided in the Appendix.

CONCLUSIONS

A search for short linear epitope matches between the ORF3 and ORF4 sequences and the Pioneer Hi-Bred allergen dataset indicates there are no contiguous sequences of 8 residue or greater shared by the two theoretical peptides and the allergen entries in the dataset. Similarly FastA amino acid sequence alignments between 80 residue fragements of the ORF3 and ORF4 open reading frames and the allergen dataset did not identify alignments of greater than 35% sequence identity across the 80 residue fragments. It may be concluded that the conceptual translations of the ORF3 and ORF4 open reading frames are of limited amino acid sequence similarity to known allergens. There is no indication from these results that either of the unintended open reading frames present in maize line 1507 is a potential allergen.

REFERENCES

Bush, R. K. and S. L. Hefle 1996. Food Allergens. Critical Reviews in Food Science and Nutrition. 36(S): S119-S163.

FAO/WHO Expert consultation on Allergenicity of Foods Derived from Biotechnology, January 2001 (http://www.fao.org/es/esn/gm/allergygm.pdf)

Gendel, S. M. 1998. Sequence Databases for Assessing the Potential Allergenicity of Proteins Used in Transgenic Foods. Advances in Food and Nutrition Research. S. L. Taylor. San Diego, Academic Press. 42: 63-92.

Gendel, S. M. 1998. The Use of Amino Acid Sequence Alignments to Assess Potential Allergenicity of Proteins Used in Genetically Modified Foods. Advances in Food and Nutrition Research. S. L. Taylor. San Diego, Academic Press. 42: 45-62.

Gendel, S. M. (Last update unknown). Web Pages: Biotechnology Information for Food Safety - Food Allergy. http://www.iit.edu/~sgendel/fa.htm

Hileman, R.E. Silvanovich, A. Goodman, R.E. Rice, A.R. Holleschak, G. Astwood, J.D. and S.L. Hefle. 2002. Bioinformatic Methods for Allergenicity Assessment Using a Comprehensive Allergen Database. Int Arch Allergy Immunol 128: 280-291.

ILSI Protein Allergenicity SubCommittee Report, July 2001, Allergenicity Assessment for Foods Derived from Genetically Modified Crops

Metcalfe, D. D., J. D. Astwood, et al. 1996. Assessment of the Allergenic Potential of Foods Derived from Genetically Engineered Crop Plants. Critical Reviews in Food Science and Nutrition 36(S): S165-S186.

Pearson W. and D. Lipman. 1988. Improved Tools for Biological Sequence Comparison. Proc. Natl. Acad. Sci. USA 85: 2444-2448

Rothbard, J.B. and M.L. Gefter. 1991 Interactions between immunogenic peptides and MHC proteins. Annu. Rev. Immunol. 9: 527.

William Pearson's implementation of the method of Smith and Waterman. 1981. Advances in Applied Mathematics 2; 482-489

APPENDIX

Output for ORF3 Short Linear Epitope Search

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
```

Results of "SSearch" Query for 1507 ORF3 Coding Region Employing a Modified Substitution Matrix (*allornone14.cmp)*

```
(Peptide) SSEARCH of: orf3.pep  from: 1 to: 250  August 9, 2002 17:08


 TO: allergen2:*  Sequences:      2,033  Symbols:       452,461

 Scoring matrix: allornone14.cmp
 Gap creation penalty: 12  Gap extension penalty: 2



Histogram Key:
 Each histogram symbol represents 8 search set sequences
 Each inset symbol represents 1 search set sequences
 z-scores computed from s-w scores

z-score obs    exp
        (=)    (*)

< 20     85      0:===========
  22      0      0:
  24      0      0:
  26      0      0:
  28      0      0:
  30      0      3:*
  32      0     10: *
  34      3     28:=  *
```

```
  36     15     58:==      *
  38     64     96:========    *
  40    182    133:================*======
  42    307    163:====================*==================
  44    422    180:======================*==============================
  46     83    183:===========           *
  48     18    175:===                  *
  50     29    160:====               *
  52     68    141:=========        *
  54    114    120:==============*
  56    140    100:============*=====
  58    101     82:==========*==
  60    163     67:========*============
  62     69     53:======*==
  64     17     43:===  *
  66      3     34:=   *
  68      6     26:=  *
  70     26     21:==*=
  72     50     16:=*=====
  74     21     13:=*=
  76     21     10:=*=
  78      0      8:*
  80      1      6:*
  82      0      5:*
  84      1      4:*
  86      1      3:*
  88      3      2:*
  90     18      2:*==
  92      0      1:*         :*
  94      0      1:*         :*
  96      0      1:*         :*
  98      0      1:*         :*
 100      2      0:=         *==
 102      0      0:          *
 104      0      0:          *
 106      0      0:          *
 108      0      0:          *
 110      0      0:          *
 112      0      0:          *
 114      0      0:          *
 116      0      0:          *
 118      0      0:          *
>120      0      0:          *

 Smith-Waterman (PGopt): reg.-scaled


The best scores are:                                    s-w    z-sc  E(1948)..

/wrk/ptmp/cressmrf/mydata/allergen2/q9s8f2.pep    Begin: 14  End: 19
! LOCUS Q9S8F2 39 aa 27-MAR-2002                         30   100.3     1.7
/wrk/ptmp/cressmrf/mydata/allergen2/q9s8f3.pep    Begin: 17  End: 22
! LOCUS Q9S8F3 42 aa 27-MAR-2002                         30    99.9     1.8
/wrk/ptmp/cressmrf/mydata/allergen2/gdb0_wheat.pep    Begin: 68  End: 73
! ID GDB0_WHEAT STANDARD; PRT; 251 AA.                   30    90.8     5.8
```

```
/wrk/ptmp/cressmrf/mydata/allergen2/m16060.pep    Begin: 68  End: 73
! LOCUS M16060_1 [WHTGLIGP]                              30    90.8     5.8
/wrk/ptmp/cressmrf/mydata/allergen2/ps0094.pep    Begin: 68  End: 73
! F1;PS0094 - gamma-gliadin precursor...                 30    90.8     5.8
/wrk/ptmp/cressmrf/mydata/allergen2/q41543.pep    Begin: 109  End: 114
! ID Q41543 PRELIMINARY; PRT; 279 AA.                    30    90.3     6.2
/wrk/ptmp/cressmrf/mydata/allergen2/gda3_wheat.pep    Begin: 108  End: 113
! ID GDA3_WHEAT STANDARD; PRT; 282 AA.                   30    90.3     6.3
/wrk/ptmp/cressmrf/mydata/allergen2/m11076.pep    Begin: 108  End: 113
! LOCUS M11076_1 [WHTGLIABH]                             30    90.3     6.3
/wrk/ptmp/cressmrf/mydata/allergen2/m13713.pep    Begin: 68  End: 73
! LOCUS M13713_1 [WHTGLGB]                               30    90.1     6.4
/wrk/ptmp/cressmrf/mydata/allergen2/gdbb_wheat.pep    Begin: 68  End: 73
! ID GDBB_WHEAT STANDARD; PRT; 291 AA.                   30    90.1     6.4
/wrk/ptmp/cressmrf/mydata/allergen2/eewtg.pep    Begin: 68  End: 73
! P1;EEWTG - gamma-gliadin B precurso...                 30    90.1     6.4
/wrk/ptmp/cressmrf/mydata/allergen2/gda6_wheat.pep    Begin: 113  End: 118
! ID GDA6_WHEAT STANDARD; PRT; 296 AA.                   30    90.0     6.5
/wrk/ptmp/cressmrf/mydata/allergen2/a27319.pep    Begin: 113  End: 118
! P1;A27319 - gliadin - wheat                            30    90.0     6.5
/wrk/ptmp/cressmrf/mydata/allergen2/s07361.pep    Begin: 113  End: 118
! P1;S07361 - alpha/beta-gliadin prec...                 30    90.0     6.5
/wrk/ptmp/cressmrf/mydata/allergen2/m16496.pep    Begin: 113  End: 118
! LOCUS M16496_1 [WHTGLNA]                               30    90.0     6.5
/wrk/ptmp/cressmrf/mydata/allergen2/x02538.pep    Begin: 113  End: 118
! LOCUS X02538_1 [TAGLIAG1]                              30    90.0     6.5
/wrk/ptmp/cressmrf/mydata/allergen2/ja0153.pep    Begin: 38  End: 43
! P1;JA0153 - gamma-gliadin precursor...                 30    89.9     6.5
/wrk/ptmp/cressmrf/mydata/allergen2/m36999.pep    Begin: 38  End: 43
! LOCUS M36999_1 [WHTGGLN]                               30    89.9     6.5
/wrk/ptmp/cressmrf/mydata/allergen2/gdbx_wheat.pep    Begin: 38  End: 43
! ID GDBX_WHEAT STANDARD; PRT; 302 AA.                   30    89.9     6.5
/wrk/ptmp/cressmrf/mydata/allergen2/e22364.pep    Begin: 114  End: 119
! P1;E22364 - alpha/beta-gliadin prec...                 30    89.6     6.8
/wrk/ptmp/cressmrf/mydata/allergen2/js0402.pep    Begin: 68  End: 73
! P1;JS0402 - gamma-gliadin precursor...                 30    89.5     6.9
/wrk/ptmp/cressmrf/mydata/allergen2/gdb2_wheat.pep    Begin: 68  End: 73
! ID GDB2_WHEAT STANDARD; PRT; 327 AA.                   30    89.5     6.9
/wrk/ptmp/cressmrf/mydata/allergen2/m16064.pep    Begin: 68  End: 73
! LOCUS M16064_1 [WHTGLIGY]                              30    89.5     6.9
\\End of List


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/q9s8f2.pep

LOCUS       Q9S8F2          39 aa                   27-MAR-2002
DEFINITION  GLY M IB ALLERGEN (FRAGMENT).
ACCESSION   Q9S8F2
SOURCE
  ORGANISM  Glycine max (Soybean).
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
SCORES   z-score: 100.3 E(): 1.7
>>/wrk/ptmp/cressmrf/mydata/allergen2/q9s8f2.pep          (39 aa)
 Z-score: 100.3 expect():  1.7
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (108-113:14-19)

              80        90       100       110       120       130
orf3.pep     GGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQR
                                           ||||||
q9s8f2.pep                    TRPSXPDLSIXLNILGGSLGTVDDXXALIGGLXDXXAIV
                                      10        20        30

             140       150       160       170       180       190
orf3.pep     KKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRY


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/q9s8f3.pep

LOCUS       Q9S8F3          42 aa                     27-MAR-2002
DEFINITION  GLY M IA ALLERGEN (FRAGMENT).
ACCESSION   Q9S8F3
SOURCE
  ORGANISM  Glycine max (Soybean).
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;


SCORES   z-score: 99.9  E(): 1.8
>>/wrk/ptmp/cressmrf/mydata/allergen2/q9s8f3.pep          (42 aa)
 Z-score: 99.9 expect():  1.8
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (108-113:17-22)

              80        90       100       110       120       130
orf3.pep     GGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQR
                                           ||||||
q9s8f3.pep                 ALITRPSXPDLSIXLNILGGSLGTVDDXXALIGGLXDXXAIV
                                   10        20        30        40

             140       150       160       170       180       190
orf3.pep     KKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRY


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/gdb0_wheat.pep

ID   GDB0_WHEAT     STANDARD;      PRT;   251 AA.
AC   P08079;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   01-NOV-1990 (Rel. 16, Last annotation update)
DE   GAMMA-GLIADIN PRECURSOR (FRAGMENT). . . .
```

```
SCORES   z-score: 90.8  E(): 5.8
>>/wrk/ptmp/cressmrf/mydata/allergen2/gdb0_wheat.pep      (251 aa)
 Z-score: 90.8 expect():  5.8
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:68-73)

                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
gdb0_wheat.p PQPHQPFSQQPQQTFPQPQQTFPHQPQQQFPQPQQPQQQFLQPQQPFPQQPQQPYPQQPQ
              40        50        60        70        80        90

gdb0_wheat.p QPFPQTQQPQQLFPQSQQPQQQFSQPQQQFPQPQQPQQSFPQQQPPFIQPSLQQQVNPCK
             100       110       120       130       140       150


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/m16060.pep

LOCUS       M16060_1 [WHTGLIGP]
DEFINITION  Wheat (T.aestivum) gamma-gliadin gene, partial cds;
            gamma-gliadin.
DATE        27-APR-1993
ACCESSION   M16060
NID . . .


SCORES   z-score: 90.8  E(): 5.8
>>/wrk/ptmp/cressmrf/mydata/allergen2/m16060.pep         (251 aa)
 Z-score: 90.8 expect():  5.8
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:68-73)

                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
m16060.pep   PQPHQPFSQQPQQTFPQPQQTFPHQPQQQFPQPQQPQQQFLQPQQPFPQQPQQPYPQQPQ
              40        50        60        70        80        90

m16060.pep   QPFPQTQQPQQLFPQSQQPQQQFSQPQQQFPQPQQPQQSFPQQQPPFIQPSLQQQVNPCK
             100       110       120       130       140       150


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/ps0094.pep

F1;PS0094 - gamma-gliadin precursor (clone pW10) - wheat (fragment)
C;Species: Triticum aestivum (common wheat)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 20-Aug-1999
C;Accession: PS0094
R;Scheets, K.; Rafalski, J.A.; Hedgcoth, C.; Soell, D.G.
Plant Sci. Lett. 37, 221-225, 1985 . . .


SCORES   z-score: 90.8  E(): 5.8
```

```
>>/wrk/ptmp/cressmrf/mydata/allergen2/ps0094.pep          (251 aa)
 Z-score: 90.8 expect():  5.8
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:68-73)

                     200       210       220       230       240       250
orf3.pep       RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                             ||||||
ps0094.pep     PQPHQPFSQQPQQTFPQPQQTFPHQPQQQFPQPQQPQQQFLQPQQPFPQQPQQPYPQQPQ
                 40        50        60        70        80        90

ps0094.pep     QPFPQTQQPQQLFPQSQQPQQQFSQPQQQFPQPQQPQQSFPQQQPPFIQPSLQQQVNPCK
               100       110       120       130       140       150


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/q41543.pep

ID   Q41543      PRELIMINARY;      PRT;   279 AA.
AC   Q41543;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GAMMA-GLIADIN (FRAGMENT). . . .


SCORES   z-score: 90.3  E(): 6.2
>>/wrk/ptmp/cressmrf/mydata/allergen2/q41543.pep          (279 aa)
 Z-score: 90.3 expect():  6.2
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:109-114)

                     200       210       220       230       240       250
orf3.pep       RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                             ||||||
q41543.pep     QPFPQTQQPQQPFFPQSQQPQQPFFPQPQQQFPQPQQPQQSFPQQQPSLIQQSLQQQLNPCK
               80        90        100       110       120       130

q41543.pep     NFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQCCQQLAQIPQQLQCAAIHSIVHSIIMQQ
              140       150       160       170       180       190


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/gda3_wheat.pep

ID   GDA3_WHEAT     STANDARD;      PRT;   282 AA.
AC   P04723;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ALPHA/BETA-GLIADIN A-III PRECURSOR (PROLAMIN). . . .


SCORES   z-score: 90.3  E(): 6.3
>>/wrk/ptmp/cressmrf/mydata/allergen2/gda3_wheat.pep      (282 aa)
```

```
 Z-score: 90.3 expect():  6.3
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:108-113)

                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
gda3_wheat.p QPQPFPPQLPYPQTQPFPPQQPYPQPQPQYPQPQQPISQQQAQQQQQQQQTLQQILQQQL
              80        90       100       110       120       130

gda3_wheat.p IPCRDVVLQQHNIAHASSQVLQQSSYQQLQQLCCQQLFQIPEQSRCQAIHNVVHAIILHH
             140       150       160       170       180       190


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/m11076.pep

LOCUS       M11076_1 [WHTGLIABH]
DEFINITION  Wheat (T.aestivum) alpha-/beta-gliadin class A-III mRNA, complete
            cds, clone pA1235;
            pre-alpha-/beta-gliadin A-III.
DATE        27-APR-1993
ACCESSION   M11076 . . .


SCORES   z-score: 90.3  E(): 6.3
>>/wrk/ptmp/cressmrf/mydata/allergen2/m11076.pep          (282 aa)
 Z-score: 90.3 expect():  6.3
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:108-113)

                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
m11076.pep   QPQPFPPQLPYPQTQPFPPQQPYPQPQPQYPQPQQPISQQQAQQQQQQQQTLQQILQQQL
              80        90       100       110       120       130

m11076.pep   IPCRDVVLQQHNIAHASSQVLQQSSYQQLQQLCCQQLFQIPEQSRCQAIHNVVHAIILHH
             140       150       160       170       180       190


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/m13713.pep

LOCUS       M13713_1 [WHTGLGB]
DEFINITION  Wheat (T.aestivum) gamma gliadin gene, complete cds;
            gamma-gliadin B precursor.
DATE        27-APR-1993
ACCESSION   M13713
NID . . .


SCORES   z-score: 90.1  E(): 6.4
>>/wrk/ptmp/cressmrf/mydata/allergen2/m13713.pep          (291 aa)
 Z-score: 90.1 expect():  6.4
```

```
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:68-73)

                    200       210       220       230       240       250
orf3.pep      RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                            ||||||
m13713.pep    LQPHQPFSQQPQQIFPQPQQTFPHQPQQQFPQPQQPQQQFLQPRQPFPQQPQQPYPQQPQ
               40        50        60        70        80        90

m13713.pep    QPFPQTQQPQQPFPQSKQPQQPFPQPQQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQC
              100       110       120       130       140       150


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/gdbb_wheat.pep

ID   GDBB_WHEAT     STANDARD;      PRT;   291 AA.
AC   P06659;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-NOV-1990 (Rel. 16, Last annotation update)
DE   GAMMA-GLIADIN B PRECURSOR. . . .


SCORES   z-score: 90.1  E(): 6.4
>>/wrk/ptmp/cressmrf/mydata/allergen2/gdbb_wheat.pep      (291 aa)
 Z-score: 90.1 expect():  6.4
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:68-73)

                    200       210       220       230       240       250
orf3.pep      RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                            ||||||
gdbb_wheat.p  LQPHQPFSQQPQQIFPQPQQTFPHQPQQQFPQPQQPQQQFLQPRQPFPQQPQQPYPQQPQ
               40        50        60        70        80        90

gdbb_wheat.p  QPFPQTQQPQQPFPQSKQPQQPFPQPQQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQC
              100       110       120       130       140       150


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/eewtg.pep

P1;EEWTG - gamma-gliadin B precursor - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C;Accession: A25632
R;Rafalski, J.A.
Gene 43, 221-229, 1986 . . .


SCORES   z-score: 90.1  E(): 6.4
>>/wrk/ptmp/cressmrf/mydata/allergen2/eewtg.pep           (291 aa)
 Z-score: 90.1 expect():  6.4
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
```

```
 (222-227:68-73)

                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
eewtg.pep    LQPHQPFSQQPQQIFPQPQQTFPHQPQQQFPQPQQPQQQFLQPRQPFPQQPQQPYPQQPQ
              40        50        60        70        80        90

eewtg.pep    QPFPQTQQPQQPFPQSKQPQQPFPQPQQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQC
             100       110       120       130       140       150


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/gda6_wheat.pep

ID   GDA6_WHEAT     STANDARD;      PRT;   296 AA.
AC   P04726;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ALPHA/BETA-GLIADIN CLONE PW1215 PRECURSOR (PROLAMIN). . . .


SCORES   z-score: 90.0  E(): 6.5
>>/wrk/ptmp/cressmrf/mydata/allergen2/gda6_wheat.pep      (296 aa)
 Z-score: 90.0 expect():  6.5
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:113-118)

                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
gda6_wheat.p QPQPFPPQLPYPQPPPFSPQQPYPQPQPQYPQPQQPISQQQAQQQQQQQQQQQQQQQQQQ
                    90       100       110       120       130       140

gda6_wheat.p ILQQILQQQLIPCRDVVLQQHNIAHARSQVLQQSTYQPLQQLCCQQLWQIPEQSRCQAIH
                   150       160       170       180       190       200


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/a27319.pep

P1;A27319 - gliadin - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 04-Mar-1988 #sequence_revision 04-Mar-1988 #text_change 03-Feb-1994
C;Accession: A27319
R;Reeves, C.D.; Okita, T.W.
Gene 52, 257-266, 1987 . . .


SCORES   z-score: 90.0  E(): 6.5
>>/wrk/ptmp/cressmrf/mydata/allergen2/a27319.pep         (296 aa)
 Z-score: 90.0 expect():  6.5
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:113-118)
```

```
                    200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
a27319.pep   QPQPFPPQLPYPQPPPFSPQQPYPQPQPQYPQPQQPISQQQAQQQQQQQQQQQQQQQQQQ
                    90       100       110       120       130       140

a27319.pep   ILPQILQQQLIPCRDVVLQQHNIAHARSQVLQQSTYQPLQQLCCQQLWQIPEQSRCQAIH
                   150       160       170       180       190       200


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/s07361.pep

P1;S07361 - alpha/beta-gliadin precursor (clone pW1215) - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 08-Jun-1994 #sequence_revision 01-Dec-1995 #text_change 20-Aug-1999
C;Accession: S07361
R;Sumner-Smith, M.; Rafalski, J.A.; Sugiyama, T.; Stoll, M.; Soell, D.
Nucleic Acids Res. 13, 3905-3916, 1985 . . .


SCORES   z-score: 90.0  E(): 6.5
>>/wrk/ptmp/cressmrf/mydata/allergen2/s07361.pep          (296 aa)
 Z-score: 90.0 expect():  6.5
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:113-118)

                    200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
s07361.pep   QPQPFPPQLPYPQPPPFSPQQPYPQPQPQYPQPQQPISQQQAQQQQQQQQQQQQQQQQQQ
                    90       100       110       120       130       140

s07361.pep   ILQQILQQQLIPCRDVVLQQHNIAHARSQVLQQSTYQPLQQLCCQQLWQIPEQSRCQAIH
                   150       160       170       180       190       200


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/m16496.pep

LOCUS        M16496_1 [WHTGLNA]
DEFINITION   Triticum urartu alpha/beta-type gliadin gene, complete cds;
             gliadin.
DATE         27-APR-1993
ACCESSION    M16496
NID . . .


SCORES   z-score: 90.0  E(): 6.5
>>/wrk/ptmp/cressmrf/mydata/allergen2/m16496.pep          (296 aa)
 Z-score: 90.0 expect():  6.5
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:113-118)
```

```
                    200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                            ||||||
m16496.pep   QPQPFPPQLPYPQPPPFSPQQPYPQPQPQYPQPQQPISQQQAQQQQQQQQQQQQQQQQQQ
                     90       100       110       120       130       140

m16496.pep   ILPQILQQQLIPCRDVVLQQHNIAHARSQVLQQSTYQPLQQLCCQQLWQIPEQSRCQAIH
                    150       160       170       180       190       200


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/x02538.pep

LOCUS       X02538_1 [TAGLIAG1]
DEFINITION  Wheat gene for alpha/beta-gliadin storage protein (pW1215);
            gliadin precursor.
DATE        30-MAR-1995
ACCESSION   X02538
NID . . .


SCORES   z-score: 90.0  E(): 6.5
>>/wrk/ptmp/cressmrf/mydata/allergen2/x02538.pep          (296 aa)
 Z-score: 90.0 expect():  6.5
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:113-118)

                    200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                            ||||||
x02538.pep   QPQPFPPQLPYPQPPPFSPQQPYPQPQPQYPQPQQPISQQQAQQQQQQQQQQQQQQQQQQ
                     90       100       110       120       130       140

x02538.pep   ILQQILQQQLIPCRDVVLQQHNIAHARSQVLQQSTYQPLQQLCCQQLWQIPEQSRCQAIH
                    150       160       170       180       190       200


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/ja0153.pep

P1;JA0153 - gamma-gliadin precursor - wheat
N;Alternate names: seed storage protein
C;Species: Triticum aestivum (common wheat)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 15-Nov-1996
C;Accession: JA0153
R;Scheets, K.; Hedgcoth, C. . . .


SCORES   z-score: 89.9  E(): 6.5
>>/wrk/ptmp/cressmrf/mydata/allergen2/ja0153.pep          (302 aa)
 Z-score: 89.9 expect():  6.5
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:38-43)
```

```
                      200       210       220       230       240       250
orf3.pep        RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                              ||||||
ja0153.pep      TILAMATTIATANMQVDPSGQVQWPQQQPFPQPQQPFCQQPQRTIPQPHQTFHHQPQQTF
                 10        20        30        40        50        60

ja0153.pep      PQPQQTYPHQPQQQFPQTQQPQQPFPQPQQTFPQQPQLPFPQQPQQPFPQPQQPQQPFPQ
                 70        80        90       100       110       120


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/m36999.pep

LOCUS        M36999_1 [WHTGGLN]
DEFINITION   Wheat gamma gliadin gene, complete cds;
             gamma gliadin precursor.
DATE         27-APR-1993
ACCESSION    M36999
NID . . .


SCORES   z-score: 89.9  E(): 6.5
>>/wrk/ptmp/cressmrf/mydata/allergen2/m36999.pep          (302 aa)
 Z-score: 89.9 expect():  6.5
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:38-43)

                      200       210       220       230       240       250
orf3.pep        RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                              ||||||
m36999.pep      TILAMATTIATANMQVDPSGQVQWPQQQPFPQPQQPFCQQPQRTIPQPHQTFHHQPQQTF
                 10        20        30        40        50        60

m36999.pep      PQPQQTYPHQPQQQFPQTQQPQQPFPQPQQTFPQQPQLPFPQQPQQPFPQPQQPQQPFPQ
                 70        80        90       100       110       120


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/gdbx_wheat.pep

ID   GDBX_WHEAT     STANDARD;      PRT;   302 AA.
AC   P21292;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-MAY-1991 (Rel. 18, Last annotation update)
DE   GAMMA-GLIADIN PRECURSOR. . . .


SCORES   z-score: 89.9  E(): 6.5
>>/wrk/ptmp/cressmrf/mydata/allergen2/gdbx_wheat.pep      (302 aa)
 Z-score: 89.9 expect():  6.5
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:38-43)
```

```
                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
gdbx_wheat.p TILAMATTIATANMQVDPSGQVQWPQQQPFPQPQQPFCQQPQRTIPQPHQTFHHQPQQTF
              10        20        30        40        50        60

gdbx_wheat.p PQPQQTYPHQPQQQFPQTQQPQQPFPQPQQTFPQQPQLPFPQQPQQPFPQPQQPQQPFPQ
              70        80        90       100       110       120


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/e22364.pep

P1;E22364 - alpha/beta-gliadin precursor (clone A1235) - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 15-Nov-1996
C;Accession: E22364
R;Okita, T.W.; Cheesbrough, V.; Reeves, C.D.
J. Biol. Chem. 260, 8203-8213, 1985 . . .


SCORES   z-score: 89.6  E(): 6.8
>>/wrk/ptmp/cressmrf/mydata/allergen2/e22364.pep          (320 aa)
 Z-score: 89.6 expect():  6.8
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:114-119)

                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
e22364.pep   QPQPFPPQLPYPQTQPFPPQQPYPQPQPQYPQPQQPISQQQAQQQQQQQQQQQQQQQTLQ
                  90       100       110       120       130       140

e22364.pep   QILQQQLIPCRDVVLQQHNIAHASSQVLQQSSYQQLQQLCCQQLFQIPEQSRCQAIHNVV
                 150       160       170       180       190       200


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/js0402.pep

P1;JS0402 - gamma-gliadin precursor - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Nov-1996
C;Accession: JS0402
R;Sugiyama, T.; Rafalski, A.; Soell, D.
Plant Sci. 44, 205-209, 1986 . . .


SCORES   z-score: 89.5  E(): 6.9
>>/wrk/ptmp/cressmrf/mydata/allergen2/js0402.pep          (327 aa)
 Z-score: 89.5 expect():  6.9
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:68-73)
```

```
                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
js0402.pep   PQLQQPLSQQPQQTFPQPQQTFPHQPQQQVPQPQQPQQPFLQPQQPFPQQPQQPFPQTQQ
              40        50        60        70        80        90

js0402.pep   PQQPFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQQPQQPFPQLQQPQQPFPQPQQQLPQ
             100       110       120       130       140       150


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/gdb2_wheat.pep

ID   GDB2_WHEAT     STANDARD;      PRT;   327 AA.
AC   P08453;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   01-NOV-1990 (Rel. 16, Last annotation update)
DE   GAMMA-GLIADIN PRECURSOR. . . .


SCORES   z-score: 89.5  E(): 6.9
>>/wrk/ptmp/cressmrf/mydata/allergen2/gdb2_wheat.pep       (327 aa)
 Z-score: 89.5 expect():  6.9
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:68-73)

                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
gdb2_wheat.p PQLQQPLSQQPQQTFPQPQQTFPHQPQQQVPQPQQPQQPFLQPQQPFPQQPQQPFPQTQQ
              40        50        60        70        80        90

gdb2_wheat.p PQQPFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQQPQQPFPQLQQPQQPFPQPQQQLPQ
             100       110       120       130       140       150


orf3.pep
/wrk/ptmp/cressmrf/mydata/allergen2/m16064.pep

LOCUS       M16064_1 [WHTGLIGY]
DEFINITION  Wheat (T.aestivum) gamma-gliadin gene, complete cds;
            gamma-gliadin.
DATE        27-APR-1993
ACCESSION   M16064
NID . . .


SCORES   z-score: 89.5  E(): 6.9
>>/wrk/ptmp/cressmrf/mydata/allergen2/m16064.pep           (327 aa)
 Z-score: 89.5 expect():  6.9
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (222-227:68-73)
```

```
                   200       210       220       230       240       250
orf3.pep     RLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
                                           ||||||
m16064.pep   PQLQQPLSQQPQQTFPQPQQTFPHQPQQQVPQPQQPQQPFLQPQQPFPQQPQQPFPQTQQ
              40        50        60        70        80        90

m16064.pep   PQQPFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQQPQQPFPQLQQPQQPFPQPQQQLPQ
             100       110       120       130       140       150


! Distributed over 1 thread.
!      Start time: Fri Aug  9 17:08:00 2002
! Completion time: Fri Aug  9 17:08:37 2002

! CPU time used:
!         Database scan:  0:00:15.6
! Post-scan processing:  0:00:00.5
!        Total CPU time:  0:00:16.1
! Output File: orf3.ssearch
```

Excerpt of Output for ORF3 80 Residue FastA Search

ORF3 80mers

```
>3ORF001
MLPLCAGQVRSQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGL
>3ORF002
LPLCAGQVRSQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLV
>3ORF003
PLCAGQVRSQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVR
>3ORF004
LCAGQVRSQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRV
>3ORF005
CAGQVRSQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVK
>3ORF006
AGQVRSQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKR
>3ORF007
GQVRSQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRK
>3ORF008
QVRSQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKK
>3ORF009
VRSQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKS
>3ORF010
RSQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSH
>3ORF011
SQVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHT
>3ORF012
QVPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTE
>3ORF013
VPTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTEL
>3ORF014
PTHGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELK
>3ORF015
THGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKI
>3ORF016
HGEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIF
>3ORF017
GEHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFS
>3ORF018
EHQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSG
>3ORF019
HQCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGD
>3ORF020
QCSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDI
>3ORF021
CSLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIH
>3ORF022
SLGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHF
>3ORF023
LGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFP
>3ORF024
GRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPE
```

```
>3ORF025
RIVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEE
>3ORF026
IVALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEA
>3ORF027
VALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEAD
>3ORF028
ALPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADK
>3ORF029
LPPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKI
>3ORF030
PPQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKIL
>3ORF031
PQTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILG
>3ORF032
QTPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGG
>3ORF033
TPLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGS
>3ORF034
PLIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSL
>3ORF035
LIQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLI
>3ORF036
IQRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIP
>3ORF037
QRESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPP
>3ORF038
RESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPE
>3ORF039
ESLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPER
>3ORF040
SLILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPERE
>3ORF041
LILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREK
>3ORF042
ILQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKK
>3ORF043
LQELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKD
>3ORF044
QELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDS
>3ORF045
ELHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSK
>3ORF046
LHILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKE
>3ORF047
HILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKES
>3ORF048
ILPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESK
>3ORF049
LPRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKK
>3ORF050
PRSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKR
>3ORF051
RSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRK
>3ORF052
```

```
SSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKN
>3ORF053
SSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNW
>3ORF054
SLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWV
>3ORF055
LKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVY
>3ORF056
KVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYV
>3ORF057
VLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQ
>3ORF058
LDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQR
>3ORF059
DNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRK
>3ORF060
NSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKK
>3ORF061
SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKF
>3ORF062
IIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFL
>3ORF063
IGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLK
>3ORF064
GVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKS
>3ORF065
VDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSK
>3ORF066
DTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKE
>3ORF067
TQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEK
>3ORF068
QLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKY
>3ORF069
LTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYC
>3ORF070
TKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCG
>3ORF071
KNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGY
>3ORF072
NTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYL
>3ORF073
TRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLS
>3ORF074
RSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSI
>3ORF075
SRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIR
>3ORF076
RLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRA
>3ORF077
LGGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAA
>3ORF078
GGLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAAD
>3ORF079
GLVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADM
```

```
>3ORF080
LVRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMA
>3ORF081
VRVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAA
>3ORF082
RVKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAV
>3ORF083
VKRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVC
>3ORF084
KRKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCD
>3ORF085
RKKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDI
>3ORF086
KKSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIV
>3ORF087
KSHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVN
>3ORF088
SHTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNH
>3ORF089
HTELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHY
>3ORF090
TELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYI
>3ORF091
ELKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIE
>3ORF092
LKIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIET
>3ORF093
KIFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETS
>3ORF094
IFSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETST
>3ORF095
FSGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTV
>3ORF096
SGDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVN
>3ORF097
GDIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNF
>3ORF098
DIHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFR
>3ORF099
IHFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRT
>3ORF100
HFPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTE
>3ORF101
FPEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEP
>3ORF102
PEEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQ
>3ORF103
EEADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQT
>3ORF104
EADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTP
>3ORF105
ADKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQ
>3ORF106
DKILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQE
>3ORF107
```

```
KILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEW
>3ORF108
ILGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWI
>3ORF109
LGGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWID
>3ORF110
GGSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDD
>3ORF111
GSLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDL
>3ORF112
SLIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLE
>3ORF113
LIPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLER
>3ORF114
IPPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERL
>3ORF115
PPEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQ
>3ORF116
PEREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQD
>3ORF117
EREKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDR
>3ORF118
REKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRY
>3ORF119
EKKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYP
>3ORF120
KKDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPW
>3ORF121
KDSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWL
>3ORF122
DSKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLV
>3ORF123
SKESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVA
>3ORF124
KESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAE
>3ORF125
ESKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEV
>3ORF126
SKKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVE
>3ORF127
KKRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEG
>3ORF128
KRKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGV
>3ORF129
RKNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVV
>3ORF130
KNWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVA
>3ORF131
NWVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAG
>3ORF132
WVYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGI
>3ORF133
VYVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIA
>3ORF134
YVQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAY
```

```
>3ORF135
VQRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYA
>3ORF136
QRKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAG
>3ORF137
RKKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGP
>3ORF138
KKFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPW
>3ORF139
KFLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWK
>3ORF140
FLKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKA
>3ORF141
LKSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKAR
>3ORF142
KSKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARN
>3ORF143
SKEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNP
>3ORF144
KEKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQ
>3ORF145
EKYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQP
>3ORF146
KYCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQ
>3ORF147
YCGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQ
>3ORF148
CGYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQP
>3ORF149
GYLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPT
>3ORF150
YLSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTN
>3ORF151
LSIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNG
>3ORF152
SIRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGI
>3ORF153
IRAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIY
>3ORF154
RAADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYL
>3ORF155
AADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLA
>3ORF156
ADMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLAT
>3ORF157
DMAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATS
>3ORF158
MAAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSL
>3ORF159
AAVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLD
>3ORF160
AVCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDH
>3ORF161
VCDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQ
>3ORF162
```

```
CDIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQS
>3ORF163
DIVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQST
>3ORF164
IVNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTL
>3ORF165
VNHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLV
>3ORF166
NHYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVV
>3ORF167
HYIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVF
>3ORF168
YIETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFV
>3ORF169
IETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVA
>3ORF170
ETSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVAL
>3ORF171
TSTVNFRTEPQTPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNPQPQQPTNGIYLATSLDHQSTLVVFVALS
```

ORF3 Search Results for Accessions Exhibiting Greater Than 20% Identity

```
(Peptide) FASTA of: 3orf014.pep  from: 1 to: 80  August 13, 2002 09:53

3ORF014

 TO: bob:*  Sequences:      2,033  Symbols:       452,461  Word Size: 2

 Databases searched:
   , Release 1.0, Released on 8May2002, Formatted on 8May2002

 Scoring matrix: GenRunData:blosum50.cmp
 Variable pamfactor used
 Gap creation penalty: 12  Gap extension penalty: 2


The best scores are:                      init1 initn   opt    z-sc E(2022)..

BOB:BAA22613    Begin: 75  End: 112
BOB:T05910    Begin: 76  End: 113
BOB:S20853    Begin: 86  End: 121
BOB:X62588    Begin: 86  End: 121
BOB:X13443    Begin: 43  End: 106
BOB:IAA_HORVU    Begin: 43  End: 106
BOB:S01655    Begin: 43  End: 106
BOB:KABOS2    Begin: 133  End: 185
BOB:CAS2_BOVIN    Begin: 133  End: 185
BOB:M16644    Begin: 133  End: 185
\\End of List
```

```
3orf014.pep
BOB:IAA_HORVU

IAA_HORVU ID   IAA_HORVU      STANDARD;      PRT;   147 AA.

SCORES   Init1: 38   Initn: 38   Opt: 63   z-score: 108.2 E(): 0.65
>>BOB:IAA_HORVU                                   (147 aa)
 initn:  38 init1:  38 opt:  63 Z-score: 108.2 expect(): 0.65
Smith-Waterman score: 63;    26.5% identity in 68 aa overlap
 (1-65:43-106)

                                                   10        20        30
3orf014.pep                                PTHGEHQCSLGRIVALPPQTPLIQRESLIL
                                           |    |   : |: :  |: |:: :|
IAA_HORVU    AVLLSVLAVAAATLESVKDECQLGVDFPHNPLATCHTYVIKRVCGRGPSRPMLVKERCC-
                   20        30        40        50        60        70

                        40        50        60        70        80
3orf014.pep  QELHILP---RSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELK
             :||  :|   |  :|::| :   || |   : :::|||
IAA_HORVU    RELAAVPDHCRCEALRILMD---GVRTPEGRVVEGRLGDRRDCPREEQRAFAATLVTAAE
                   80        90          100       110       120

IAA_HORVU    CNLSSVQAPGVRLVLLADG
             130       140


3orf014.pep
BOB:X13443

X13443 LOCUS       X13443_1 [HVAATI]

SCORES   Init1: 38   Initn: 38   Opt: 63   z-score: 108.2 E(): 0.65
>>BOB:X13443                                      (147 aa)
 initn:  38 init1:  38 opt:  63 Z-score: 108.2 expect(): 0.65
Smith-Waterman score: 63;    26.5% identity in 68 aa overlap
 (1-65:43-106)

                                                   10        20        30
3orf014.pep                                PTHGEHQCSLGRIVALPPQTPLIQRESLIL
                                           |    |   : |: :  |: |:: :|
X13443       AVLLSVLAVAAATLESVKDECQLGVDFPHNPLATCHTYVIKRVCGRGPSRPMLVKERCC-
                   20        30        40        50        60        70

                        40        50        60        70        80
3orf014.pep  QELHILP---RSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELK
             :||  :|   |  :|::| :   || |   : :::|||
X13443       RELAAVPDHCRCEALRILMD---GVRTPEGRVVEGRLGDRRDCPREEQRAFAATLVTAAE
                   80        90          100       110       120
```

```
X13443       CNLSSVQAPGVRLVLLADG
            130       140




3orf014.pep
BOB:S01655

S01655 P1;S01655 - alpha-amylase/trypsin inhibitor precursor - barley


SCORES   Init1: 38    Initn: 38    Opt: 63    z-score: 108.2 E(): 0.65
>>BOB:S01655                                              (147 aa)
 initn:  38 init1:  38 opt:  63 Z-score: 108.2 expect(): 0.65
Smith-Waterman score: 63;    26.5% identity in 68 aa overlap
 (1-65:43-106)

                                                   10        20        30
3orf014.pep                                PTHGEHQCSLGRIVALPPQTPLIQRESLIL
                                           |    |   : |: :  |: |:: :|
S01655       AVLLSVLAVAAATLESVKDECQLGVDFPHNPLATCHTYVIKRVCGRGPSRPMLVKERCC-
                   20        30        40        50        60        70

                        40        50        60        70        80
3orf014.pep  QELHILP---RSSSLKVLDNSIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELK
             :||  :|   |  :|::| :   || |   : :::|||
S01655       RELAAVPDHCRCEALRILMD---GVRTPEGRVVEGRLGDRRDCPREEQRAFAATLVTAAE
                   80        90          100       110       120

S01655       CNLSSVQAPGVRLVLLADG
            130       140
(Peptide) FASTA of: 3orf023.pep  from: 1 to: 80  August 13, 2002 09:53

3ORF023

 TO: bob:*  Sequences:      2,033  Symbols:       452,461  Word Size: 2

 Databases searched:
   , Release 1.0, Released on 8May2002, Formatted on 8May2002

 Scoring matrix: GenRunData:blosum50.cmp
 Variable pamfactor used
 Gap creation penalty: 12  Gap extension penalty: 2


The best scores are:                    init1 initn   opt    z-sc E(2023)..

BOB:BAA22613    Begin: 75  End: 112
BOB:T05910    Begin: 76  End: 113
BOB:Q41279    Begin: 55  End: 124
BOB:X91801    Begin: 55  End: 124
BOB:S65481    Begin: 55  End: 124
BOB:S20853    Begin: 86  End: 121
BOB:X62588    Begin: 86  End: 121
```

```
BOB:Q41196    Begin: 55  End: 124
BOB:PC1246    Begin: 55  End: 124
BOB:S54101    Begin: 55  End: 124
\\End of List


3orf023.pep
BOB:PC1246

PC1246 LOCUS       PC1246        145 aa                     27-MAR-2002


SCORES   Init1: 25    Initn: 25    Opt: 63    z-score: 107.0 E(): 0.76
>>BOB:PC1246                                              (145 aa)
 initn:  25 init1:  25 opt:  63 Z-score: 107.0 expect(): 0.76
Smith-Waterman score: 63;    25.0% identity in 72 aa overlap
 (9-80:55-124)

                                           10        20        30
3orf023.pep                        LGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDN
                                           || |  ||  |: |  : | :   | |
PC1246       QWLHKQAMQSGSGPSWTLDDEFDFEDDMENPQGPQ-QRPPLLQQCCNELHQEEPLCVCP-
                 30        40        50         60        70        80

             40        50        60        70        80
3orf023.pep  SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFP
             :: |::  : :::|::||   :   : :|:  :|::   |:|
PC1246       TLKGASKAVKQQVRQQLGQQGQQGPHLQHVISRIYQTATHLPKVCNIRQVSVCPFKKTMP
                   90       100       110       120       130       140

PC1246       GPS



3orf023.pep
BOB:Q41196

Q41196 ID   Q41196      PRELIMINARY;      PRT;   145 AA.


SCORES   Init1: 25    Initn: 25    Opt: 63    z-score: 107.0 E(): 0.76
>>BOB:Q41196                                              (145 aa)
 initn:  25 init1:  25 opt:  63 Z-score: 107.0 expect(): 0.76
Smith-Waterman score: 63;    25.0% identity in 72 aa overlap
 (9-80:55-124)

                                           10        20        30
3orf023.pep                        LGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDN
                                           || |  ||  |: |  : | :   | |
Q41196       QWLHKQAMQSGSGPSWTLDDEFDFEDDMENPQGPQ-QRPPLLQQCCNELHQEEPLCVCP-
                 30        40        50         60        70        80
```

```
             40        50        60        70        80
3orf023.pep  SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFP
             :: |::  : :::|::||   :   : :|:  :|::   |:|
Q41196       TLKGASKAVKQQVRQQLGQQGQQGPHLQHVISRIYQTATHLPKVCNIRQVSVCPFKKTMP
                   90       100       110       120       130       140

Q41196       GPS




3orf023.pep
BOB:S54101

S54101 LOCUS       S54101_1 {S54101}


SCORES   Init1: 25    Initn: 25    Opt: 63    z-score: 107.0 E(): 0.76
>>BOB:S54101                                                 (145 aa)
 initn:  25 init1:  25 opt:  63 Z-score: 107.0 expect(): 0.76
Smith-Waterman score: 63;    25.0% identity in 72 aa overlap
 (9-80:55-124)

                                           10        20        30
3orf023.pep                        LGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDN
                                           || |  ||  |: |  : | :   | |
S54101       QWLHKQAMQSGSGPSWTLDDEFDFEDDMENPQGPQ-QRPPLLQQCCNELHQEEPLCVCP-
                 30        40        50         60        70        80

             40        50        60        70        80
3orf023.pep  SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFP
             :: |::  : :::|::||   :   : :|:  :|::   |:|
S54101       TLKGASKAVKQQVRQQLGQQGQQGPHLQHVISRIYQTATHLPKVCNIRQVSVCPFKKTMP
                   90       100       110       120       130       140

S54101       GPS




3orf023.pep
BOB:Q41279

Q41279 ID   Q41279      PRELIMINARY;      PRT;   145 AA.


SCORES   Init1: 25    Initn: 25    Opt: 65    z-score: 109.9 E(): 0.52
>>BOB:Q41279                                                 (145 aa)
 initn:  25 init1:  25 opt:  65 Z-score: 109.9 expect(): 0.52
Smith-Waterman score: 65;    25.0% identity in 72 aa overlap
 (9-80:55-124)

                                           10        20        30
3orf023.pep                        LGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDN
                                           || |  ||  |: |  : | :   | |
Q41279       QWLHKQAMQSGSGPSWTLDDEFDFEDDMENPQGPQ-QRPPLLQQCCNELHQEEPLCVCP-
                 30        40        50         60        70        80
```

```
               40        50        60        70        80
3orf023.pep    SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFP
               :: |::  : :::|::||   :   : :|:  :|::   |:|
Q41279         TLKGASKAVKQQVRQQLGQQGQQGPQVQHVISRIYQTATHLPKVCNIPQVSVCPFKKTMP
                     90       100       110       120       130       140

Q41279         GPS

3orf023.pep
BOB:X91801

X91801 LOCUS       X91801_1 [SASIN3]


3orf023.pep
BOB:S65481

S65481 LOCUS       S65481        145 aa                     27-MAR-2002


SCORES   Init1: 25    Initn: 25    Opt: 65    z-score: 109.9 E(): 0.52
>>BOB:S65481                                              (145 aa)
 initn:  25 init1:  25 opt:  65 Z-score: 109.9 expect(): 0.52
Smith-Waterman score: 65;    25.0% identity in 72 aa overlap
 (9-80:55-124)

                                             10        20        30
3orf023.pep                          LGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDN
                                             || |  ||  |: |  : | :   | |
S65481         QWLHKQAMQSGSGPSWTLDDEFDFEDDMENPQGPQ-QRPPLLQQCCNELHQEEPLCVCP-
                   30        40        50         60        70        80

               40        50        60        70        80
3orf023.pep    SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFP
               :: |::  : :::|::||   :   : :|:  :|::   |:|
S65481         TLKGASKAVKQQVRQQLGQQGQQGPQVQHVISRIYQTATHLPKVCNIPQVSVCPFKKTMP
                     90       100       110       120       130       140

S65481         GPS



SCORES   Init1: 25    Initn: 25    Opt: 65    z-score: 109.9 E(): 0.52
>>BOB:X91801                                              (145 aa)
 initn:  25 init1:  25 opt:  65 Z-score: 109.9 expect(): 0.52
Smith-Waterman score: 65;    25.0% identity in 72 aa overlap
 (9-80:55-124)

                                             10        20        30
3orf023.pep                          LGRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDN
                                             || |  ||  |: |  : | :   | |
X91801         QWLHKQAMQSGSGPSWTLDDEFDFEDDMENPQGPQ-QRPPLLQQCCNELHQEEPLCVCP-
                   30        40        50         60        70        80
```

```
             40        50        60        70        80
3orf023.pep  SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFP
             :: |::  : :::|::||   :   : :|:  :|::   |:|
X91801       TLKGASKAVKQQVRQQLGQQGQQGPQVQHVISRIYQTATHLPKVCNIPQVSVCPFKKTMP
                   90        100       110       120       130       140

X91801       GPS


(Peptide) FASTA of: 3orf024.pep  from: 1 to: 80  August 13, 2002 09:53

3ORF024

 TO: bob:*  Sequences:      2,033  Symbols:       452,461  Word Size: 2

 Databases searched:
   , Release 1.0, Released on 8May2002, Formatted on 8May2002

 Scoring matrix: GenRunData:blosum50.cmp
 Variable pamfactor used
 Gap creation penalty: 12  Gap extension penalty: 2


The best scores are:                    init1 initn   opt    z-sc E(2025)..

BOB:Q41279    Begin: 55  End: 125
BOB:X91801    Begin: 55  End: 125
BOB:S65481    Begin: 55  End: 125
BOB:BAA22613    Begin: 75  End: 112
BOB:T05910    Begin: 76  End: 113
BOB:S20853    Begin: 86  End: 121
BOB:X62588    Begin: 86  End: 121
BOB:Q41196    Begin: 55  End: 125
BOB:PC1246    Begin: 55  End: 125
BOB:S54101    Begin: 55  End: 125
BOB:Q41281    Begin: 55  End: 125
BOB:X91798    Begin: 55  End: 125
BOB:S65478    Begin: 55  End: 125
\\End of List


3orf024.pep
BOB:Q41281

Q41281 ID   Q41281      PRELIMINARY;      PRT;   145 AA.


SCORES   Init1: 25    Initn: 25    Opt: 62    z-score: 105.0 E(): 0.98
>>BOB:Q41281                                                (145 aa)
 initn:  25 init1:  25 opt:  62 Z-score: 105.0 expect(): 0.98
Smith-Waterman score: 62;    23.3% identity in 73 aa overlap
 (8-80:55-125)
```

```
                                            10        20        30
3orf024.pep                        GRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDN
                                           || |  |:  |: |  : | :   | |
Q41281       QWLHKQAMQSGSGPSWTLDDEFDFEDDMENPQGPQ-QKPPLLQQCCNELHQEEPLCVCP-
                 30        40        50         60        70        80

              40        50        60        70        80
3orf024.pep  SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPE
             :: |::  : :::|::||   :   : :|:  :|::   |:|:
Q41281       TLKGASKAVKQQVRQQLGQQGQQGPQVQHVISRIYQTATHLPKVCNIPQVSVCPFKKTMP
                   90       100       110       120       130       140

Q41281       GPS



3orf024.pep
BOB:X91798

X91798 LOCUS       X91798_1 [SASIN5]


SCORES   Initl: 25    Initn: 25    Opt: 62    z-score: 105.0 E(): 0.98
>>BOB:X91798                                              (145 aa)
 initn:  25 initl:  25 opt:  62 Z-score: 105.0 expect(): 0.98
Smith-Waterman score: 62;    23.3% identity in 73 aa overlap
 (8-80:55-125)

                                            10        20        30
3orf024.pep                        GRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDN
                                           || |  |:  |: |  : | :   | |
X91798       QWLHKQAMQSGSGPSWTLDDEFDFEDDMENPQGPQ-QKPPLLQQCCNELHQEEPLCVCP-
                 30        40        50         60        70        80

              40        50        60        70        80
3orf024.pep  SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPE
             :: |::  : :::|::||   :   : :|:  :|::   |:|:
X91798       TLKGASKAVKQQVRQQLGQQGQQGPQVQHVISRIYQTATHLPKVCNIPQVSVCPFKKTMP
                   90       100       110       120       130       140

X91798       GPS



3orf024.pep
BOB:S65478

S65478 LOCUS       S65478        145 aa                           27-MAR-2002


SCORES   Initl: 25    Initn: 25    Opt: 62    z-score: 105.0 E(): 0.98
>>BOB:S65478                                              (145 aa)
 initn:  25 initl:  25 opt:  62 Z-score: 105.0 expect(): 0.98
Smith-Waterman score: 62;    23.3% identity in 73 aa overlap
 (8-80:55-125)
```

```
                                             10        20        30
3orf024.pep                          GRIVALPPQTPLIQRESLILQELHILPRSSSLKVLDN
                                            || |  |:  |: |  : | :   | |
S65478        QWLHKQAMQSGSGPSWTLDDEFDFEDDMENPQGPQ-QKPPLLQQCCNELHQEEPLCVCP-
                  30        40        50         60        70        80

               40        50        60        70        80
3orf024.pep   SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPE
              :: |::  : :::|::||   :   : :|:  :|::   |:|:
S65478        TLKGASKAVKQQVRQQLGQQGQQGPQVQHVISRIYQTATHLPKVCNIPQVSVCPFKKTMP
                    90       100       110       120       130       140

S65478        GPS



(Peptide) FASTA of: 3orf031.pep  from: 1 to: 80  August 13, 2002 09:53

3ORF031

 TO: bob:*  Sequences:      2,033  Symbols:       452,461  Word Size: 2

 Databases searched:
   , Release 1.0, Released on 8May2002, Formatted on 8May2002

 Scoring matrix: GenRunData:blosum50.cmp
 Variable pamfactor used
 Gap creation penalty: 12  Gap extension penalty: 2


The best scores are:                    init1 initn   opt    z-sc E(2024)..

BOB:Q41279    Begin: 55  End: 125
BOB:X91801    Begin: 55  End: 125
BOB:S65481    Begin: 55  End: 125
BOB:KABOS2    Begin: 133  End: 206
BOB:M16644    Begin: 133  End: 206
BOB:CAS2_BOVIN    Begin: 133  End: 206
\\End of List


3orf031.pep
BOB:KABOS2

KABOS2 P1;KABOS2 - alpha-s2-casein precursor - bovine


SCORES   Init1: 36    Initn: 36    Opt: 67    z-score: 106.5 E(): 0.81
>>BOB:KABOS2                                                (222 aa)
 initn:  36 init1:  36 opt:  67 Z-score: 106.5 expect(): 0.81
Smith-Waterman score: 67;    20.8% identity in 77 aa overlap
 (1-77:133-206)
```

```
                                                   10        20        30
3orf031.pep                                PQTPLIQRESLILQELHILPRSSSLKVLDN
                                           | || ::||:|  :| :    :::: : ::
KABOS2       FYQKFPQYLQYLYQGPIVLNPWDQVKRNAVPITPTLNREQLSTSEEN---SKKTVDMEST
                  110       120       130       140          150

                     40        50        60        70        80
3orf031.pep  SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILG
              ::   |:||:: ::||: | ::::: ::  |  :   ::  ::| |
KABOS2       EVFTKKTKLTEEEKNRLNFLKKISQRYQKFALPQYLKTVYQHQKAMKPWIQPKTKVIPYV
           160       170       180       190       200       210

KABOS2       RYL
           220


3orf031.pep
BOB:M16644

M16644 LOCUS       M16644_1 [BOVCSAS2A]


SCORES   Init1: 36    Initn: 36    Opt: 67    z-score: 106.5 E(): 0.81
>>BOB:M16644                                              (222 aa)
 initn:  36 init1:  36 opt:  67 Z-score: 106.5 expect(): 0.81
Smith-Waterman score: 67;    20.8% identity in 77 aa overlap
 (1-77:133-206)

                                                   10        20        30
3orf031.pep                                PQTPLIQRESLILQELHILPRSSSLKVLDN
                                           | || ::||:|  :| :    :::: : ::
M16644       FYQKFPQYLQYLYQGPIVLNPWDQVKRNAVPITPTLNREQLSTSEEN---SKKTVDMEST
                  110       120       130       140          150

                     40        50        60        70        80
3orf031.pep  SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILG
              ::   |:||:: ::||: | ::::: ::  |  :   ::  ::| |
M16644       EVFTKKTKLTEEEKNRLNFLKKISQRYQKFALPQYLKTVYQHQKAMKPWIQPKTKVIPYV
           160       170       180       190       200       210

M16644       RYL
           220


3orf031.pep
BOB:CAS2_BOVIN

CAS2_BOVIN ID   CAS2_BOVIN     STANDARD;      PRT;   222 AA.


SCORES   Init1: 36    Initn: 36    Opt: 67    z-score: 106.5 E(): 0.81
>>BOB:CAS2_BOVIN                                          (222 aa)
 initn:  36 init1:  36 opt:  67 Z-score: 106.5 expect(): 0.81
Smith-Waterman score: 67;    20.8% identity in 77 aa overlap
 (1-77:133-206)
```

```
                                                     10        20        30
3orf031.pep                                  PQTPLIQRESLILQELHILPRSSSLKVLDN
                                             | || ::||:|  :| :    :::: : ::
CAS2_BOVIN     FYQKFPQYLQYLYQGPIVLNPWDQVKRNAVPITPTLNREQLSTSEEN---SKKTVDMEST
                    110       120       130       140          150

                       40        50        60        70        80
3orf031.pep    SIIGVDTQLTKNTRSRLGGLVRVKRKKSHTELKIFSGDIHFPEEADKILG
                ::   |:||:: ::||: | ::::: ::  |  :   ::  ::| |
CAS2_BOVIN     EVFTKKTKLTEEEKNRLNFLKKISQRYQKFALPQYLKTVYQHQKAMKPWIQPKTKVIPYV
             160       170       180       190       200       210

CAS2_BOVIN     RYL
             220
```

Output for ORF4 Short Linear Epitope Search

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
```

Results of "SSearch" Query for 1507 ORF4 Coding Sequence Employing a Modified Substitution Matrix (*allornone14.cmp)*

```
(Peptide) SSEARCH of: 1507orf4.gp  from: 1 to: 210  May 20, 2002 12:50

LOCUS       1507\orf\4    210 aa                         20-MAY-2002
DEFINITION  Translation of a fragment of Insert in Event 1507-14.
KEYWORDS    TRANSLATED.
ORIGIN

 TO: allergen2:*  Sequences:      2,033  Symbols:       452,461

 Scoring matrix: allornone14.cmp
 Gap creation penalty: 12  Gap extension penalty: 2



Histogram Key:
 Each histogram symbol represents 8 search set sequences
 Each inset symbol represents 1 search set sequences
 z-scores computed from s-w scores

z-score obs    exp
        (=)    (*)

< 20     70      0:=========
  22      0      0:
  24      0      0:
  26      0      0:
  28      0      0:
```

```
  30       0      3:*
  32       0     10: *
  34       0     28:   *
  36      30     58:====   *
  38      62     96:========   *
  40     181    134:================*======
  42     299    164:====================*=================
  44     472    181:======================*====================================
  46     118    184:===============       *
  48      17    177:===                   *
  50      31    161:====                *
  52      39    142:=====            *
  54      22    121:===            *
  56      67    101:=========   *
  58     107     83:==========*===
  60     191     67:========*===============
  62     150     54:======*============
  64      93     43:=====*======
  66      13     34:==  *
  68       1     27:=  *
  70       2     21:= *
  72       2     16:=*
  74       0     13: *
  76      13     10:=*
  78      17      8:*==
  80      17      6:*==
  82       8      5:*
  84       1      4:*
  86       0      3:*
  88       0      2:*
  90       0      2:*
  92       0      1:*         :*
  94       0      1:*         :*
  96       1      1:*         :*
  98       9      1:*=        :*========
 100       0      0:          *
 102       0      0:          *
 104       0      0:          *
 106       0      0:          *
 108       0      0:          *
 110       0      0:          *
 112       0      0:          *
 114       0      0:          *
 116       0      0:          *
 118       0      0:          *
>120       0      0:          *

 Smith-Waterman (PGopt): reg.-scaled

The best scores are:                                     s-w     z-sc  E(1963)..

/wrk/ptmp/cressmrf/mydata/allergen2/x62588.pep    Begin: 261   End: 266
! LOCUS X62588_1 [TDLMWG]                                 30     98.5     2.2
/wrk/ptmp/cressmrf/mydata/allergen2/s20853.pep    Begin: 261   End: 266
```

```
! F1;S20853 - glutenin low molecular ...                 30    98.5     2.2
/wrk/ptmp/cressmrf/mydata/allergen2/gltc_wheat.pep    Begin: 263  End: 268
! ID GLTC_WHEAT STANDARD; PRT; 295 AA.                   30    98.3     2.2
/wrk/ptmp/cressmrf/mydata/allergen2/x51759.pep    Begin: 263  End: 268
! LOCUS X51759_1 [TDGLUT]                                30    98.3     2.2
/wrk/ptmp/cressmrf/mydata/allergen2/jn0696.pep    Begin: 263  End: 268
! P1;JN0696 - glutenin low molecular ...                 30    98.3     2.2
/wrk/ptmp/cressmrf/mydata/allergen2/gdb1_wheat.pep    Begin: 273  End: 278
! ID GDB1_WHEAT STANDARD; PRT; 304 AA.                   30    98.2     2.3
/wrk/ptmp/cressmrf/mydata/allergen2/m11077.pep    Begin: 273  End: 278
! LOCUS M11077_1 [WHTGLIGBA]                             30    98.2     2.3
/wrk/ptmp/cressmrf/mydata/allergen2/m11336_1.pep    Begin: 292  End: 297
! LOCUS M11336_1 [WHTGLIGBB]                             30    97.9     2.4
/wrk/ptmp/cressmrf/mydata/allergen2/glta_wheat.pep    Begin: 325  End: 330
! ID GLTA_WHEAT STANDARD; PRT; 356 AA.                   30    97.4     2.5
/wrk/ptmp/cressmrf/mydata/allergen2/x07747.pep    Begin: 325  End: 330
! LOCUS X07747_1 [TAGLUT]                                30    97.4     2.5
\\End of List


1507orf4.gp
/wrk/ptmp/cressmrf/mydata/allergen2/x62588.pep

LOCUS       X62588_1 [TDLMWG]
DEFINITION  T.durum gene for LMW glutenin.
DATE        16-APR-1992
ACCESSION   X62588
NID
ORGANISM    Triticum durum . . .


SCORES   z-score: 98.5  E(): 2.2
>>/wrk/ptmp/cressmrf/mydata/allergen2/x62588.pep          (285 aa)
 Z-score: 98.5 expect():  2.2
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (6-11:261-266)

                                             10        20        30
1507orf4.gp                          MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPV
                                          ||||||
x62588.pep    GQQPQQQVQQGTFLQPHQIAHLEVMTSIALRTLPTMCSVNVPLYSSTTSVPFGV
                     240       250       260       270       280

                40        50        60        70        80        90
1507orf4.gp   RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR


1507orf4.gp
/wrk/ptmp/cressmrf/mydata/allergen2/s20853.pep

F1;S20853 - glutenin low molecular weight chain precursor - durum wheat
 (fragment)
C;Species: Triticum durum (durum wheat)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 20-Aug-1999
C;Accession: S20853; S17355
```

```
R;d'Ovidio, R.; Tanzarella, O.A.; Porceddu, E. . . .


SCORES   z-score: 98.5  E(): 2.2
>>/wrk/ptmp/cressmrf/mydata/allergen2/s20853.pep          (285 aa)
 Z-score: 98.5 expect():  2.2
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (6-11:261-266)

                                              10        20        30
1507orf4.gp                           MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPV
                                           ||||||
s20853.pep   GQQPQQQQVQQGTFLQPHQIAHLEVMTSIALRTLPTMCSVNVPLYSSTTSVPFGV
                    240       250       260       270       280

                40        50        60        70        80        90
1507orf4.gp  RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR


1507orf4.gp
/wrk/ptmp/cressmrf/mydata/allergen2/gltc_wheat.pep

ID   GLTC_WHEAT     STANDARD;      PRT;   295 AA.
AC   P16315;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLUTENIN, LOW MOLECULAR WEIGHT SUBUNIT PTDUCD1 PRECURSOR. . . .


SCORES   z-score: 98.3  E(): 2.2
>>/wrk/ptmp/cressmrf/mydata/allergen2/gltc_wheat.pep      (295 aa)
 Z-score: 98.3 expect():  2.2
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (6-11:263-268)

                                              10        20        30
1507orf4.gp                           MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPV
                                           ||||||
gltc_wheat.p GQQPQQQQVLQGTFLQPHQIAHLEVMTSIALRTLPTMCSVNVPLYSSTTSVPFSVGTGVG
                  240       250       260       270       280       290

                40        50        60        70        80        90
1507orf4.gp  RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR

gltc_wheat.p AYL



1507orf4.gp
/wrk/ptmp/cressmrf/mydata/allergen2/x51759.pep

LOCUS       X51759_1 [TDGLUT]
DEFINITION  Triticum durum mRNA for glutenin;
            precursor peptide (AA -20 to 275).
```

```
DATE        23-MAR-1995
ACCESSION   X51759
NID . . .


SCORES   z-score: 98.3  E(): 2.2
>>/wrk/ptmp/cressmrf/mydata/allergen2/x51759.pep          (295 aa)
 Z-score: 98.3 expect():  2.2
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (6-11:263-268)

                                              10        20        30
1507orf4.gp                           MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPV
                                           ||||||
x51759.pep   GQQPQQQQVLQGTFLQPHQIAHLEVMTSIALRTLPTMCSVNVPLYSSTTSVPFSVGTGVG
                  240       250       260       270       280       290

                40        50        60        70        80        90
1507orf4.gp  RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR

x51759.pep   AYL


1507orf4.gp
/wrk/ptmp/cressmrf/mydata/allergen2/jn0696.pep

P1;JN0696 - glutenin low molecular weight chain precursor (clone pTdUCD1) -
 durum wheat
C;Species: Triticum durum (durum wheat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: JN0696; S08683
R;Cassidy, B.G.; Dvorak, J. . . .


SCORES   z-score: 98.3  E(): 2.2
>>/wrk/ptmp/cressmrf/mydata/allergen2/jn0696.pep          (295 aa)
 Z-score: 98.3 expect():  2.2
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (6-11:263-268)

                                              10        20        30
1507orf4.gp                           MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPV
                                           ||||||
jn0696.pep   GQQPQQQQVLQGTFLQPHQIAHLEVMTSIALRTLPTMCSVNVPLYSSTTSVPFSVGTGVG
                  240       250       260       270       280       290

                40        50        60        70        80        90
1507orf4.gp  RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR

jn0696.pep   AYL


1507orf4.gp
```

```
/wrk/ptmp/cressmrf/mydata/allergen2/gdb1_wheat.pep

ID   GDB1_WHEAT     STANDARD;      PRT;   304 AA.
AC   P04729;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GAMMA-GLIADIN B-I PRECURSOR. . . .


SCORES   z-score: 98.2  E(): 2.3
>>/wrk/ptmp/cressmrf/mydata/allergen2/gdb1_wheat.pep      (304 aa)
 Z-score: 98.2 expect():  2.3
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (6-11:273-278)

                                              10        20        30
1507orf4.gp                           MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPV
                                           ||||||
gdb1_wheat.p PQQQQQQQVLQGTFLQPHQIAHLEAVTSIALRTLPTMCSVNVPLYSATTSVPFGVGTGVG
                  250       260       270       280       290       300

                40        50        60        70        80        90
1507orf4.gp  RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR

gdb1_wheat.p AY



1507orf4.gp
/wrk/ptmp/cressmrf/mydata/allergen2/m11077.pep

LOCUS       M11077_1 [WHTGLIGBA]
DEFINITION  Wheat (T.aestivum) gamma-gliadin class B-I mRNA, complete cds,
            clone pB11-33;
            pre-gamma-gliadin B-I.
DATE        27-APR-1993
ACCESSION   M11077 . . .


SCORES   z-score: 98.2  E(): 2.3
>>/wrk/ptmp/cressmrf/mydata/allergen2/m11077.pep          (304 aa)
 Z-score: 98.2 expect():  2.3
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (6-11:273-278)

                                              10        20        30
1507orf4.gp                           MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPV
                                           ||||||
m11077.pep   PQQQQQQQVLQGTFLQPHQIAHLEAVTSIALRTLPTMCSVNVPLYSATTSVPFGVGTGVG
                  250       260       270       280       290       300

                40        50        60        70        80        90
1507orf4.gp  RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR
```

```
m11077.pep   AY



1507orf4.gp
/wrk/ptmp/cressmrf/mydata/allergen2/m11336_1.pep

LOCUS       M11336_1 [WHTGLIGBB]
DEFINITION  Wheat (T.aestivum) gamma-gliadin class B-I mRNA, clone pB312;
            gamma-gliadin.
DATE        27-APR-1993
ACCESSION   M11336
NID . . .


SCORES   z-score: 97.9  E(): 2.4
>>/wrk/ptmp/cressmrf/mydata/allergen2/m11336_1.pep        (323 aa)
 Z-score: 97.9 expect():  2.4
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (6-11:292-297)

                                              10        20        30
1507orf4.gp                           MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPV
                                           ||||||
m11336_1.pep PQQQQQQQVLQGTFLQPHQIAHLEAVTSIALRTLPTMCSVNVPLYSATTSVPFGVGTGVG
                   270       280       290       300       310       320

                40        50        60        70        80        90
1507orf4.gp  RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR

m11336_1.pep AY


1507orf4.gp
/wrk/ptmp/cressmrf/mydata/allergen2/glta_wheat.pep

ID   GLTA_WHEAT     STANDARD;      PRT;   356 AA.
AC   P10385;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLUTENIN, LOW MOLECULAR WEIGHT SUBUNIT PRECURSOR. . . .


SCORES   z-score: 97.4  E(): 2.5
>>/wrk/ptmp/cressmrf/mydata/allergen2/glta_wheat.pep      (356 aa)
 Z-score: 97.4 expect():  2.5
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (6-11:325-330)
```

```
                                              10        20        30
1507orf4.gp                          MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPV
                                           ||||||
glta_wheat.p QQLQQQLGQQPQQQQLAHQIAQLEVMTSIALRTLPTMCNVNVPLYETTTSVPLGVGIGVG
                300       310       320       330       340       350

                40        50        60        70        80        90
1507orf4.gp  RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR

glta_wheat.p VY



1507orf4.gp
/wrk/ptmp/cressmrf/mydata/allergen2/x07747.pep

LOCUS       X07747_1 [TAGLUT]
DEFINITION  Wheat gene for low molecular weight glutenin;
            LMW glutenin (AA 1-356).
DATE        12-SEP-1993
ACCESSION   X07747
NID . . .


SCORES   z-score: 97.4  E(): 2.5
>>/wrk/ptmp/cressmrf/mydata/allergen2/x07747.pep          (356 aa)
 Z-score: 97.4 expect():  2.5
Smith-Waterman score: 30;   100.0% identity in 6 aa overlap
 (6-11:325-330)

                                              10        20        30
1507orf4.gp                          MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPV
                                           ||||||
x07747.pep   QQLQQQLGQQPQQQQLAHQIAQLEVMTSIALRTLPTMCNVNVPLYETTTSVPLGVGIGVG
                300       310       320       330       340       350

                40        50        60        70        80        90
1507orf4.gp  RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR

x07747.pep   VY




! Distributed over 1 thread.
!      Start time: Mon May 20 12:49:25 2002
! Completion time: Mon May 20 12:50:08 2002

! CPU time used:
!        Database scan:  0:00:13.4
! Post-scan processing:  0:00:00.2
!       Total CPU time:  0:00:13.7
! Output File: 1507orf4.tightssearch
```

Excerpt of Output for ORF4 80 Residue FastA Search

ORF4 80mers

```
>4ORF001
MPGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTE
>4ORF002
PGSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTEL
>4ORF003
GSTLRTLPTLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELA
>4ORF004
STLRTLPTLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAV
>4ORF005
TLRTLPTLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVK
>4ORF006
LRTLPTLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKT
>4ORF007
RTLPTLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTG
>4ORF008
TLPTLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGE
>4ORF009
LPTLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQ
>4ORF010
PTLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQF
>4ORF011
TLSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFI
>4ORF012
LSMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQ
>4ORF013
SMVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQS
>4ORF014
MVSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSL
>4ORF015
VSAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLL
>4ORF016
SAPALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLR
>4ORF017
APALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRL
>4ORF018
PALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLN
>4ORF019
ALSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLND
>4ORF020
LSSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDK
>4ORF021
SSIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKK
>4ORF022
SIPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKK
>4ORF023
IPSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKI
>4ORF024
PSQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIF
>4ORF025
```

```
SQSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFV
>4ORF026
QSEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVN
>4ORF027
SEYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNM
>4ORF028
EYPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMV
>4ORF029
YPIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVE
>4ORF030
PIPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEH
>4ORF031
IPRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHD
>4ORF032
PRPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDT
>4ORF033
RPVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTL
>4ORF034
PVRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLV
>4ORF035
VRCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVY
>4ORF036
RCKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYS
>4ORF037
CKNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSK
>4ORF038
KNGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKN
>4ORF039
NGSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNI
>4ORF040
GSVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIK
>4ORF041
SVHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKD
>4ORF042
VHPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDT
>4ORF043
HPPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTV
>4ORF044
PPLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVS
>4ORF045
PLLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSE
>4ORF046
LLGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSED
>4ORF047
LGIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQ
>4ORF048
GIPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQR
>4ORF049
IPVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRA
>4ORF050
PVRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAI
>4ORF051
VRVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIE
>4ORF052
RVTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIET
```

```
>4ORF053
VTFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETF
>4ORF054
TFVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQ
>4ORF055
FVHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQ
>4ORF056
VHQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQR
>4ORF057
HQDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRV
>4ORF058
QDGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVI
>4ORF059
DGTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVIS
>4ORF060
GTAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISG
>4ORF061
TAAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGN
>4ORF062
AAADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNL
>4ORF063
AADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLL
>4ORF064
ADRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLG
>4ORF065
DRIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGF
>4ORF066
RIPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFH
>4ORF067
IPMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHC
>4ORF068
PMESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCP
>4ORF069
MESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPA
>4ORF070
ESKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAI
>4ORF071
SKIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAIC
>4ORF072
KIQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICH
>4ORF073
IQIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHF
>4ORF074
QIEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFI
>4ORF075
IEDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIV
>4ORF076
EDLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVK
>4ORF077
DLTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKI
>4ORF078
LTELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIV
>4ORF079
TELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVE
>4ORF080
```

```
ELAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEK
>4ORF081
LAVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKE
>4ORF082
AVKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEG
>4ORF083
VKTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGG
>4ORF084
KTGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGS
>4ORF085
TGEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSY
>4ORF086
GEQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYK
>4ORF087
EQFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKC
>4ORF088
QFIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCH
>4ORF089
FIQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHH
>4ORF090
IQSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHC
>4ORF091
QSLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCD
>4ORF092
SLLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDK
>4ORF093
LLRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKG
>4ORF094
LRLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGK
>4ORF095
RLNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKA
>4ORF096
LNDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAI
>4ORF097
NDKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIV
>4ORF098
DKKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVE
>4ORF099
KKKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVED
>4ORF100
KKIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDA
>4ORF101
KIFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDAS
>4ORF102
IFVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASA
>4ORF103
FVNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASAD
>4ORF104
VNMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADS
>4ORF105
NMVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSG
>4ORF106
MVEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGP
>4ORF107
VEHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPK
```

```
>4ORF108
EHDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKD
>4ORF109
HDTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDG
>4ORF110
DTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGP
>4ORF111
TLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPP
>4ORF112
LVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPP
>4ORF113
VYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPT
>4ORF114
YSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTR
>4ORF115
SKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRS
>4ORF116
KNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSI
>4ORF117
NIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIV
>4ORF118
IKDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVE
>4ORF119
KDTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEK
>4ORF120
DTVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKE
>4ORF121
TVSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKED
>4ORF122
VSEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDV
>4ORF123
SEDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVP
>4ORF124
EDQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPT
>4ORF125
DQRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTT
>4ORF126
QRAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTS
>4ORF127
RAIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSS
>4ORF128
AIETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSK
>4ORF129
IETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQ
>4ORF130
ETFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQV
>4ORF131
TFQQRVISGNLLGFHCPAICHFIVKIVEKEGGSYKCHHCDKGKAIVEDASADSGPKDGPPPTRSIVEKEDVPTTSSKQVD
```

ORF4 Search Results for Accessions Exhibiting Greater Than 25% Identity

```
(Peptide) FASTA of: 4orf072.pep  from: 1 to: 80  May 20, 2002 14:12

4ORF072

 TO: bob:*  Sequences:      2,033  Symbols:       452,461  Word Size: 2

 Databases searched:
   , Release 1.0, Released on 8May2002, Formatted on 8May2002

 Scoring matrix: GenRunData:blosum50.cmp
 Variable pamfactor used
 Gap creation penalty: 12  Gap extension penalty: 2


The best scores are:                    init1 initn   opt    z-sc E(2023)..

BOB:MPAA_CORAV    Begin: 20  End: 100
BOB:X70999    Begin: 21  End: 101
BOB:X70997    Begin: 21  End: 101
BOB:S30055    Begin: 21  End: 101
BOB:S30053    Begin: 21  End: 101
\\End of List


4orf072.pep
BOB:MPAA_CORAV

MPAA_CORAV ID   MPAA_CORAV     STANDARD;      PRT;   159 AA.


SCORES   Init1: 51    Initn: 51    Opt: 74    z-score: 122.6 E(): 0.1
>>BOB:MPAA_CORAV                                          (159 aa)
 initn:  51 init1:  51 opt:  74 Z-score: 122.6 expect():  0.1
Smith-Waterman score: 74;    26.8% identity in 82 aa overlap
 (1-80:20-100)

                                        10        20          30
4orf072.pep                     KIQIEDLTELAVKTGEQFIQSLLRL--NDKKKIFVNMVEHD
                                |  : |  :|  |:: | | |:  :  |     : |::  :
MPAA_CORAV   GVFNYEVETPSVIPAARLFKSYVLDGDKLIPKVAPQAITSVENVEGNGGPGTIKNITFGE
                     10        20        30        40        50        60

            40        50        60        70        80
4orf072.pep  TLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICH
                | | :|: |:| : :  |::  || |::|| :   :||
MPAA_CORAV   GSRY-KYVKERVDEVDNTNFTYSYTVIEGDVLGDKLEKVCHELKIVAAPGGGSILKISSK
                      70        80        90       100       110

MPAA_CORAV   FHAKGDHEINAEEMKGAKEMAEKLLRAVETYLLAHSAEYN
           120       130       140       150
```

```
4orf072.pep
BOB:X70999

X70999 LOCUS       X70999_1 [CACORAI5]


SCORES   Init1: 51    Initn: 51    Opt: 74    z-score: 122.6 E(): 0.1
>>BOB:X70999                                              (160 aa)
 initn:  51 init1:  51 opt:  74 Z-score: 122.6 expect():  0.1
Smith-Waterman score: 74;    26.8% identity in 82 aa overlap
 (1-80:21-101)

                                         10        20          30
4orf072.pep                      KIQIEDLTELAVKTGEQFIQSLLRL--NDKKKIFVNMVEH
                                 |  : |  :|  |:: | | |:  :  |      : |::
X70999       MGVFNYEVETPSVIPAARLFKSYVLDGDKLIPKVAPQAITSVENVEGNGGPGTIKNITFG
                     10        20        30        40        50        60

             40      50        60        70        80
4orf072.pep  DTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICH
             :    ||  :|: |:| : : |::  || |::|| :   :||
X70999       EGSRY-KYVKERVDEVDNTNFTYSYTVIEGDVLGDKLEKVCHELKIVAAPGGGSILKISS
                      70        80        90        100       110

X70999       KFHAKGDHEINAEEMKGAKEMAEKLLRAVETYLLAHSAEYN
          120       130       140       150       160


4orf072.pep
BOB:X70997

X70997 LOCUS       X70997_1 [CACORAI11]


SCORES   Init1: 51    Initn: 51    Opt: 74    z-score: 122.6 E(): 0.1
>>BOB:X70997                                              (160 aa)
 initn:  51 init1:  51 opt:  74 Z-score: 122.6 expect():  0.1
Smith-Waterman score: 74;    26.8% identity in 82 aa overlap
 (1-80:21-101)

                                         10        20          30
4orf072.pep                      KIQIEDLTELAVKTGEQFIQSLLRL--NDKKKIFVNMVEH
                                 |  : |  :|  |:: | | |:  :  |      : |::
X70997       MGVFNYEAETTSVIPAARLFKSYVLDGDKLIPKVAPQAITSVENVEGNGGPGTIKNITFG
                     10        20        30        40        50        60

             40      50        60        70        80
4orf072.pep  DTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICH
             :    ||  :|: |:| : : |::  || |::|| :   :||
X70997       EGSRY-KYVKERVDEVDNTNFTYSYTVIEGDVLGDKLEKVCHELKIVAAPGGGSILKISS
                      70        80        90        100       110
```

```
X70997       KFHAKGDHEINAEEIKGAKEMAEKLLRAVETYLLAHSAEYN
           120       130       140       150       160


4orf072.pep
BOB:S30055

S30055 LOCUS       S30055        160 aa                          27-MAR-2002


SCORES   Init1: 51    Initn: 51    Opt: 74    z-score: 122.6 E(): 0.1
>>BOB:S30055                                              (160 aa)
 initn:  51 init1:  51 opt:  74 Z-score: 122.6 expect():  0.1
Smith-Waterman score: 74;    26.8% identity in 82 aa overlap
 (1-80:21-101)

                                         10        20          30
4orf072.pep                      KIQIEDLTELAVKTGEQFIQSLLRL--NDKKKIFVNMVEH
                                 |  : |  :|  |:: | | |:  :  |     : |::
S30055       MGVFNYEAETTSVIPAARLFKSYVLDGDKLIPKVAPQAITSVENVEGNGGPGTIKNITFG
                     10        20        30        40        50        60

             40        50        60        70        80
4orf072.pep  DTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICH
             :   | | :|: |:| : :  |::  || |::|| :   :||
S30055       EGSRY-KYVKERVDEVDNTNFTYSYTVIEGDVLGDKLEKVCHELKIVAAPGGGSILKISS
                      70        80        90       100       110

S30055       KFHAKGDHEINAEEMKGAKEMAEKLLRAVETYLLAHSAEYN
           120       130       140       150       160


4orf072.pep
BOB:S30053

S30053 LOCUS       S30053        160 aa                          27-MAR-2002


SCORES   Init1: 51    Initn: 51    Opt: 74    z-score: 122.6 E(): 0.1
>>BOB:S30053                                              (160 aa)
 initn:  51 init1:  51 opt:  74 Z-score: 122.6 expect():  0.1
Smith-Waterman score: 74;    26.8% identity in 82 aa overlap
 (1-80:21-101)

                                         10        20          30
4orf072.pep                      KIQIEDLTELAVKTGEQFIQSLLRL--NDKKKIFVNMVEH
                                 |  : |  :|  |:: | | |:  :  |     : |::
S30053       MGVFNYEVETPSVIPAARLFKSYVLDGDKLIPKVAPQAITSVENVEGNGGPGTIKNITFG
                     10        20        30        40        50        60

             40        50        60        70        80
4orf072.pep  DTLVYSKNIKDTVSEDQRAIETFQQRVISGNLLGFHCPAICH
             :   | | :|: |:| : :  |::  || |::|| :   :||
S30053       EGSRY-KYVKERVDEVDNTNFTYSYTVIEGDVLGDKLEKVCHELKIVAAPGGGSILKISS
                      70        80        90       100       110
```

```
S30053       KFHAKGDHEINAEEMKGAKEMAEKLLRAVETYLLAHSAEYN
           120       130       140       150       160
```