

## FINAL REPORT

### Study Title

Comparison of Amino Acid Sequence Similarity of Cry1F and PAT Proteins to  
Known Allergen Proteins

### Author

Terry Meyer  
Research Manager

### Date of Report

August 19, 1999

### Performing Laboratory

Trait and Technology Development  
Pioneer Hi-Bred International, Inc.  
7300 NW 62<sup>nd</sup> Ave.  
Johnston, Iowa 50131-1004

### Study Number

PHI99-013

### STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS

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

These data are the property of Mycogen Corporation, and as such, are considered to be confidential for all purposes other than compliance with FIFRA §10. Submission of these data in compliance with FIFRA does not constitute a waiver of any right to confidentiality that may exist under any other statute or in any other country.

Company:

Company Agent:

Diane Shanahan

Date:

8/25/99



**SIGNATURES OF APPROVAL**

**Study Number:** PHI99-013

**Title:** Comparison of Amino Acid Sequence Similarity of Cry1F and PAT Proteins to Known Allergen Proteins

**Facilities:** Pioneer Hi-Bred International, Inc.  
7300 NW 62<sup>nd</sup> Ave.  
Johnston, Iowa 50131-1004

**Study Director:** Terry Meyer  
Manager, Immunochemistry  
Trait and Technology Development  
Pioneer Hi-Bred International, Inc.

**Contributors:** Larry Zeph  
Manager  
Regulatory Affairs  
Pioneer Hi-Bred International, Inc.

**Study Initiation Date:** April 12, 1999

**Records Retention:** All study specific raw data, protocols, final reports and facility records will be retained at Pioneer Hi-Bred International, Johnston, Iowa.

**Specimen Storage:** No specimens were collected as part of this study.

**Signatures of Approval:**

**Study Director** Terry Meyer 8/23/99 **Date**

**Sponsor** Lawrence Zeph 8/20/99 **Date**

## TABLE OF CONTENTS

Statement of No Data Confidentiality Claims.....	2
GLP Compliance Statement.....	3
Signatures of Approval.....	4
Table of Contents .....	5
Summary.....	6
Introduction.....	7
Methods.....	7
Results.....	8
Discussion.....	8
<b>Figures</b>	
1. Comparison of Truncated Cry1F Protein sequence to Known Allergens.....	10
2. Comparison of PAT Protein sequence to Known Allergens .....	17
<b>Appendices</b>	
1. Study Protocol PHI99-013 .....	<u>23</u>

## Summary

In assessing the risk of food proteins to human health, the potential risk of an allergic response to the protein in the human diet should be evaluated. Proteins can be introduced into transgenic plants to confer a variety of desirable traits. These proteins may currently be components of the human diet or they may present new dietary exposures. A comparison of the amino acid sequence of the introduced protein to known protein allergens can provide information on the potential of the protein to act as a food allergen. In this study, a scheme formulated by Metcalfe et al. (1996) has been used to assess similarity of the Cry1F and phosphinothricin acetyltransferase (PAT) proteins to known allergenic proteins. An immunologically significant sequence identity requires a match of at least eight contiguous identical amino acids and may necessitate further testing of the allergenic properties of the protein. No significant homology was detected for either the Cry1F or PAT proteins.

## Introduction

The human health safety assessment of proteins introduced into transgenic plants includes an evaluation of whether the protein can act as a food allergen in the human diet. An understanding of the source of the protein, and its characteristics, is valuable to the evaluation of allergenic potential.

As a first step toward identification of proteins in transgenic plant products that are potential allergens, a simple screening process can be utilized based on a database of all available sequenced allergen or toxin proteins (Metcalf et al., 1996). This initial step consists of an assessment of the amino acid similarity of the protein of interest to known allergenic proteins. Specifically, an immunologically significant sequence identity requires a match of at least eight contiguous identical amino acids. This criterion is based on evidence that T-cell epitopes of known protein allergens are usually 8 or more amino acid residues in length (O'Hehir et al., 1991). A protein identified as having significant sequence identity to a known allergen can be further assessed based on a decision-tree scheme for identifying the allergenic potential of foods. Thus, the amino acid homology screen provides an initial indication of the potential allergenicity or toxicity that can be addressed through additional testing of the protein.

The present study summarizes the analysis of the homology of an insect control protein and selectable marker protein expressed in plants developed for resistance to European Corn Borer. The two proteins are the truncated Cry 1F protein from *Bacillus thuringiensis* and phosphinothricin acetyltransferase (PAT) protein. Comparison to known allergens was performed utilizing a TFASTA algorithm that performs a Pearson and Lipman search for similarity between a query sequence and any group of sequences.

## Methods

The procedure for conducting the homology comparison is described in Pioneer Hi-Bred Intl. Inc. Standard Procedure PHIGLP-13A/01. The database was compiled using the Wisconsin Genetics Computer Group (GCG) sequence analysis computer program using the keyword "allergen" to search standard DNA and protein sequence databases.

## Results

Figure 1 shows the results of the comparison of the amino acid sequence of Cry1F protein to known protein allergens. The 15 most homologous database sequences are shown. A variety of plant, animal, and bacterial genes are listed among these scores. A significant homology is defined as a sequence identity of 8 or more contiguous amino acids. The results demonstrate that Cry1F protein shares no significant amino acid sequence homology with known allergen proteins.

Figure 2 shows the results of the comparison of the amino acid sequence of PAT protein to known protein allergens. The 14 most homologous sequences are shown. As expected, the results demonstrate that PAT protein shares no significant amino acid sequence homology with known allergen proteins.

## Discussion

In this study, a comparison of the amino acid sequence has been used to assess similarity of the Cry1F and PAT proteins to known allergenic proteins. An immunologically significant sequence identity requires a match of at least eight contiguous identical amino acids and may necessitate further testing of the allergenic properties of the protein. No immunologically significant sequence homology was detected for either the Cry1F or PAT proteins. These results were expected for the PAT protein because it has been the subject of previous safety assessments for genetically modified plants. The Cry1F protein would also not be expected to be related to common food allergens since this protein originates from a common soil bacterium. The results of this study confirm that these two proteins are not likely to be allergenic when present in the food supply.

### Literature Cited

Metcalf, D.D. 1996. Assessment of the Allergenic Potential of Foods Derived from Genetically Engineered Crop Plants. In: Critical Reviews in Food Science and Nutrition (F. Clydesdale, ed.), Vol. 36, Supplement 1996. pp. S165-S186.

O'Hehir, R.E., R.D. Garman, J.L. Greenstein, J.R. Lamb. 1991. The specificity and regulation of T-cell responsiveness to allergens. *Ann. Rev. Immunol.* 9:67-95.

### Figure 1. Comparison of Truncated Cry1F Protein sequence to Known Protein Allergens

TRANSLATE of: trunccry1f.dna check: 3573 from: 1 to: 1818  
generated symbols 1 to: 606.  
Confidential. Pioneer HiBred International, Inc. Terry Meyer.  
Truncated Cry1F Bt DNA from clone P8999. April 12, 1999.  
TO: allergans:\* Sequences: 201 Symbols: 182,832 Word Size: 2  
Databases searched:  
ALLERGANS, Release 1.0, Released on 0Feb96, Formatted on 0Feb96  
Searching all six frames.  
Scoring matrix: GenRunData:Blosum50.Cmp  
Variable pamfactor used  
Gap creation penalty: 16 Gap extension penalty: 2  
Histogram Key:  
Each histogram symbol represents 1 search set sequences  
z-scores computed from opt scores

z-score	obs (=)	exp (+)
< 20	0	0:
22	1	0: =
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	1: *
34	0	3: *
36	4	6: ===== *
38	8	10: ===== *
40	11	14: ===== *
42	12	17: ===== *
44	24	18: =====*=====
46	28	19: =====*=====
48	21	18: =====*=====
50	13	16: ===== *
52	9	14: ===== *
54	19	12: =====*=====
56	7	10: ===== *
58	7	8: =====*
60	5	7: ===== *
62	1	5: = *
64	10	4: =====*
66	6	3: =====*
68	1	3: = *
70	5	2: =====*
72	1	2: =*
74	2	1: *=
76	2	1: *=
78	2	1: *=
80	0	1: *
82	2	0: ==
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:
100	0	0:
102	0	0:
104	0	0:
106	0	0:
108	0	0:
110	0	0:
112	0	0:



Trunccrylf.Pep  
Allg:Ambambia  
LOCUS AMBAMB1A 1182 bp mRNA PLN 26-JUN-1992  
DEFINITION *Ambrosia artemisiifolia* major allergen (Amb a I.4) mRNA, complete cds.  
ACCESSION M80562  
KEYWORDS  
SOURCE *Ambrosia artemisiifolia* pollen cDNA to mRNA.  
SCORES Frame: (1) Init1: 65 Initn: 65 Opt: 69 z-score: 82.2 E(): 1.8  
38.7% identity in 31 aa overlap  
310 320 330 340 350 360  
Trunccrylf.P ETVRSQTVWGGHLVSSRNTAGNRINFPSPYGVFNPPGGAIWIADEDPRPFYRTLSDPVF/RG  
| | | | | | | | | | : : : : |  
Ambambia RGAKVEIINAGFAIYNVKNIIHNIIMHDIVVNPGLIKSHDGPVPRKGS DGDGDAIGISG  
150 160 170 180 190 200  
370 380 390 400 410 420  
Trunccrylf.P GFGNPHYVLGLRGVAFQQTGTNHTRTFRNSGTIDSLDEIPPQDNSGAPWNDYSHVLNHVT  
|  
Ambambia GSQIWIWDHCSLSKAVDGLIDAKHGSTHFTVSNCLFTQHQYLLLFWD FDERGMLCTVAFNK  
210 220 230 240 250 260

Trunccrylf.Pep  
Allg:Pjmajall  
LOCUS PJMAJALL 653 bp RNA PLN 18-SEP-1995  
DEFINITION *P. judaica* mRNA for major allergen.  
ACCESSION X77414  
KEYWORDS major allergen.  
SOURCE *Parietaria judaica*.  
ORGANISM *Parietaria judaica*.  
SCORES Frame: (3) Init1: 64 Initn: 64 Opt: 64 z-score: 79.5 E(): 2.5  
53.3% identity in 15 aa overlap  
420 430 440 450 460 470  
Trunccrylf.P NDYSHVLNHVTVRWPGEISGSDSWRAPMFSWTHRSATPTNTIDPERITQIPLVKAHTLQ  
| | : | | | | : : : : |  
Pjmajall RRETQRQGPQALRHRXQQAAPAHXROHGLQDSWSSATPTSSLSPSWSRHRGPKXSRPQST  
70 80 90 100 110 120  
480 490 500 510 520 530  
Trunccrylf.P SGTTVVRGPGFTGGDILRRTSGGPFAYTIVNINGQLPQRYRARIRYASTNLR IYVTVAG  
Pjmajall VGETPDYSSAPRTGKSLTQLSPLLDERGXYXALMSNDGGTVLXEVICNNYLN IYIYIYS  
130 140 150 160 170 180

Trunccrylf.Pep  
Allg:Pvprp2  
LOCUS PVPRP2 693 bp RNA PLN 05-MAY-1995  
DEFINITION *P. vulgaris* PvPR2 mRNA for pathogenesis-related protein 2 (PvPR2) of unknown function.  
ACCESSION X61364  
KEYWORDS pathogenesis-related protein 2; plant defence gene; PR protein.  
SOURCE kidney bean.  
SCORES Frame: (1) Init1: 49 Initn: 49 Opt: 63 z-score: 77.8 E(): 3.1  
28.6% identity in 56 aa overlap  
150 160 170 180 190 200  
Trunccrylf.P FTLTSPFIPLLSVYVQAANLHLSLLRDAVSFGQGMGLDIATVNNHYNRL----INLIHRY  
| | | | : : : : | | | | : | | |  
Pvprp2 RHHRPKGCXFLQECXNRXGQRWPRNHQEDLFRXGMDKVCVAQNRRNRXGXLGIQLQHRW  
50 60 70 80 90 100  
210 220 230 240 250 260  
Trunccrylf.P TKHCLDTYNQGLENLRGTNTRQWARFNQFRDLTLTLVLDIVALFPNYDVRTYPIQTSSQL  
: | | : : | : : : | : : :  
Pvprp2 -RCCLARHCR--EDQHRLOQTQXWSQWRFSCEVEHKISQQRRCSTQXRXAQSWXSQEXCSF  
110 120 130 140 150

Trunccrylf.Pep

Allg:Svs38

LOCUS SVS38 435 bp RNA PLN 20-JUN-1994

DEFINITION *S.vulgaris* S38 mRNA.

ACCESSION X76539

KEYWORDS S38 gene.

SOURCE *Syringa vulgaris*.

ORGANISM *Syringa vulgaris* . . .

SCORES Frame: (1) Init1: 60 Initn: 60 Opt: 60 z-score: 76.9 E(): 3.5  
47.8% identity in 23 aa overlap

Trunccrylf.P	LNNPEVEILNEERSTGRPLDISLSLTRFL	LLSEFVPGVGVAFGLFDLIWGFITPSDWSLF
		: : : : : : : : : : : : : : : :
Svs38	EDVPQPPVQPFHQVYCDTCRARFITESEFIPGAGIRLQCKDGEHGKITFTEIGYT	
	10 20 30 40 50 60 70	
Trunccrylf.P	LLQIEQLIEQRIETLERNRAITTLRGLADSYEYIEALREWEANPNNAQLREDVRIREFAN	
Svs38	RAEGLYSMLVEGDHKNFCEITLLSSSRKDCDEIPIEGWVKPSLKFILNTVNGTTRTINP	
	60 70 80 90 100 110	

Trunccrylf.Pep

Allg:Oeole1c

LOCUS OEOLE1C 435 bp RNA PLN 20-JUN-1994

DEFINITION *Olea europaea* OLE1c mRNA.

ACCESSION X76397

KEYWORDS pollen allergen; pollen protein.

SOURCE common olive.

ORGANISM *Olea europaea* . . .

SCORES Frame: (1) Init1: 55 Initn: 55 Opt: 60 z-score: 76.9 E(): 3.5  
48.3% identity in 29 aa overlap

Trunccrylf.P	PYNCLNPEVEILNEERSTGRPLDISLSLTRFL--LSEFVPGVGVAFGLFDLIWGFITP	
		: : : : : : : : : : : : : : : :
Oeole1c	EDVPQPPISQFYVQGVYCDTCRTRFITEFSEFIPGAGVRLQCKDGENGKITF	
	10 20 30 40 50 60	
Trunccrylf.P	SDWSLFLQIEQLIEQRIETLERNRAITTLRGLADSYEYIEALREWEANPNNAQLREDV	
Oeole1c	TEVGYTRAEGLYSMLIERDHKNFCEITLLSSSRKDCDEIPIEGWVKPSVKFILNTVNGT	
	60 70 80 90 100 110	

Trunccrylf.Pep

Allg:Pprphlplx

LOCUS PPRPHLPLX 1149 bp RNA PLN 30-AUG-1995

DEFINITION *P.pratense* mRNA for pollen allergen PhlpI.

ACCESSION Z27090

KEYWORDS PhlpI; pollen allergen.

SOURCE timothy grass.

ORGANISM *Phleum pratense* . . .

SCORES Frame: (1) Init1: 58 Initn: 58 Opt: 63 z-score: 74.6 E(): 4.7  
37.9% identity in 29 aa overlap

Trunccrylf.P	DNSGAPWNDYSHVLNHTFVRWPGEISGSDSWRAPMFSWTHRSATPTNTIDPERITQIPL	
		: : : : : : : : : : : : : : : :
Pprphlplx	KTRWLLPHRFCWSWRCSPPSSWALRMASPRS--PPARTSRPT	
	10 20 30 40	
Trunccrylf.P	VKAHTLQSGTIVVRGPGFTGGDILRRTSGGPPAYTIIVNINGQLPQRYRARIRYASTTNLR	
Pprphlplx	ATSGWTRRAGTASRRPPVPRTTAARAGTRTWTSPRSAAXPAAATPPSSSPAGAAAPASR	
	50 60 70 80 90 100	

Trunccrylf.Pep  
Allg:Arqarahi  
LOCUS ARQARAH1 2032 bp mRNA PLN 15-DEC-1994  
DEFINITION *Arachis hypogaea* (clone P41b) *Ara h I* mRNA, complete cds.  
ACCESSION L34402  
KEYWORDS allergen.  
SOURCE *Arachis hypogaea* (strain Florunner) seed cDNA to mRNA.  
ORGANISM *Arachis hypogaea* . . .  
SCORES Frame: (6) Init1: 47 Initn: 47 Opt: 65 z-score: 73.5 E(): 5.4  
40.5% identity in 42 aa overlap  
230 240 250 260 270 280  
Trunccrylf.P RFNQFRRLTLTVLDIVALFPNYDVRTYPIQTSSQLTREIYTSVIEDSPV SANIPNGFN  
: ||| ||| ||| ||| :||:|  
Arqarahi VASILSXKPKSLKGVDPSSLGFPPLDPLSQETLSQETTEIEIE--IEDEHSQSEIPSGFX  
30 40 50 60 70 80  
290 300 310 320 330 340  
Trunccrylf.P RAEFGVRPPHLMDFMNSLFVTAETVRSQTVWGGHLVSSRNTAGNRINFP SYGVFNPGGAI  
| :| :|  
Arqarahi XASQLVHPTQGM LNPSLASLSGLSHCPYHLQGRFCGCFQRKYRSQADGVRKRKRWLLDEL  
90 100 110 120 130 140

Trunccrylf.Pep  
Allg:Mpmyrpi  
LOCUS MPMYRPI 528 bp RNA INV 18-JUN-1993  
DEFINITION *M.pilosula* mRNA for major allergen Myr pI.  
ACCESSION X70256  
KEYWORDS IgE binding protein; major allergen.  
SOURCE Australian jumper ant.  
ORGANISM *Myrmecia pilosula* . . .  
SCORES Frame: (6) Init1: 49 Initn: 49 Opt: 57 z-score: 71.7 E(): 6.8  
42.1% identity in 19 aa overlap  
250 260 270 280 290 300  
Trunccrylf.P PNYDVRTYPIQTSSQLTREIYTSVIEDSPV SANIPNGFNRAEFGVRPPHLMDFMNSLFV  
: ||:| | :| | |  
Mpmyrpi IGIASFITLGNLTLATLGP SFLATLGITL PNIRANLPNTLPRPAPG SASPTASASPNASAK  
70 80 90 100 110 120  
310 320 330 340 350 360  
Trunccrylf.P TAETVRSQTVWGGHLVSSRNTAGNRINFP SYGVFNPGGAIWIADEDPRPFYRTLSDPVFV  
Mpmyrpi PTASDSGSAKSFAS TLGACTIVKTKIIVRVNKHDSFILXCSTATEF  
130 140 150 160 170

Trunccrylf.Pep  
Allg:Mmigm3lch  
LOCUS MMIGM3LCH 657 bp RNA ROD 19-SEP-1994  
DEFINITION *M.musculus* rearranged immunoglobulin kappa light chain variable region.  
ACCESSION Z37499  
KEYWORDS constant region; immunoglobulin kappa chain; variable region.  
SOURCE house mouse. . . .  
SCORES Frame: (2) Init1: 51 Initn: 51 Opt: 58 z-score: 71.6 E(): 6.9  
36.8% identity in 19 aa overlap  
270 280 290 300 310 320  
Trunccrylf.P LTR E IYTSVIEDSPV SANIPNGFNRAEFGVRPPHLMDFMNSLFVTAETVRSQTVWGGHL  
: ||| | | : : : : ||  
Mmigm3lch SSXXPSLHSPCLSVLEIKPPSLADLVRALYTLMETPIYIGTCRSQA  
330 340 350 360 370 380  
Trunccrylf.P VSSRNTAGNRINFP SYGVFNPGGAIWIADEDPRPFYRTLSDPVFVRGGFGNPHYVLGLRG  
Mmigm3lch CLQSSXSTXFPTDFLGSQTGSVAVDQGGI SHSRSAEWRLRIWEIFISALKVHMFLERSVEA  
50 60 70 80 90 100



Trunccrylf.Pep  
Allg:Meu08008  
LOCUS MEU08008 843 bp mRNA INV 11-JAN-1995  
DEFINITION *Metapenaeus ensis* tropomyosin mRNA, partial cds.  
ACCESSION U08008  
KEYWORDS  
SOURCE penaeid shrimps.  
ORGANISM *Metapenaeus ensis* . . .  
SCORES Frame: (5) Init1: 58 Initn: 58 Opt: 58 z-score: 70.0 E(): 8.4  
64.7% identity in 17 aa overlap

	240	250	260	270	280	290
Trunccrylf.P	RDLTLTVL	DIVALFP	NYDVRT	YPIQTSS	QLTREI	YTSSVIED
Meu08008	VRVLCISL	XASSRWL	AFSSSET	SRDFRLL	PTTRSS	SSSSST
	60	70	80	90	100	110
	300	310	320	330	340	350
Trunccrylf.P	RPPHLMDF	MNSLFV	TAE TVRS	QTVWGG	HVSSRN	TAGNRIN
Meu08008	SLRATSS	YFLSASS	ARKRASL	WFSRAS	MRSSSER	ERFSSTL
	120	130	140	150	160	170

Trunccrylf.Pep  
Allg:S72946  
LOCUS S72946 795 bp mRNA MAM 24-JAN-1995  
DEFINITION albumin=allergenic protein [dogs, salivary glands, mRNA Partial, 795 nt].  
ACCESSION S72946  
KEYWORDS  
SOURCE dogs salivary glands. . . .  
SCORES Frame: (5) Init1: 53 Initn: 53 Opt: 57 z-score: 69.1 E(): 9.5  
28.6% identity in 42 aa overlap

	270	280	290	300	310	320
Trunccrylf.P	IYTSSVIED	SPVSANI	PNGFNRAE	FVGRPPH	LDMFMS	LFVTAET
S72946	DRKSSAQD	ILSDSG	FLOHLV	PTFPPS	FLETST	TRVGVDT
	330	340	350	360	370	380
Trunccrylf.P	NTAGNRIN	FPPSYGV	FNPGGAI	WIADED	PRPFYR	TLSDPV
S72946	KSAFWKPY	SPSFSK	SSQFVLT	KFXGS	STRGLN	SSSTLAX
	50	60	70	80	90	100

! Distributed over 1 thread.  
! Start time: Mon Apr 12 10:38:54 1999  
! Completion time: Mon Apr 12 10:39:17 1999  
! CPU time used:  
! Database scan: 0:00:00.5  
! Post-scan processing: 0:00:00.7  
! Total CPU time: 0:00:01.3  
! Output File: Trunccrylf.Tfasta

## Figure 2. Comparison of PAT Protein sequence to Known Protein Allergens.

TRANSLATE of: pat.dna check: 2568 from: 1 to: 552  
generated symbols 1 to: 184.  
Confidential. Pioneer HiBred International, Inc. Terry Meyer.  
PAT DNA sequence from clone P8999. April 12, 1999.  
TO: allergans:\* Sequences: 201 Symbols: 132,832 Word Size: 2  
Databases searched:

ALLERGANS, Release 1.0, Released on 0Feb96, Formatted on 0Feb96  
Searching all six frames.  
Scoring matrix: GenRunData:Blosum50.Cmp  
Variable pamfactor used  
Gap creation penalty: 16 Gap extension penalty: 2  
Histogram Key:

Each histogram symbol represents 1 search set sequences  
z-scores computed from opt scores

z-score	obs	exp
	(=)	(*)
< 20	1	0:=
22	0	0:
24	0	0:
26	1	0:=
28	2	0:==
30	2	0:==
32	2	1:*=
34	4	3:==*=
36	3	6:=== *
38	9	10:=====*
40	14	14:=====*
42	9	17:===== *
44	23	18:=====*=====
46	14	19:===== *
48	19	18:=====*==
50	16	16:=====*
52	7	14:===== *
54	15	12:=====*====
56	10	10:=====*
58	8	8:=====*
60	8	7:=====*=
62	8	5:=====*=
64	5	4:=====*=
66	5	3:=====*=
68	5	3:=====*=
70	2	2:=====*
72	5	2:=====*=
74	0	1:=====*
76	1	1:=====*
78	1	1:=====*
80	0	1:=====*
82	1	0:=====
84	0	0:=====
86	0	0:=====
88	0	0:=====
90	0	0:=====
92	0	0:=====
94	1	0:=====
96	0	0:=====
98	0	0:=====
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:

Joining threshold: 36, opt. threshold: 36, opt. width: 16, reg.-scaled  
 The best scores are: frame init1 initn opt z-sc E(200)..  
 Allg:Meu08008  
 ! U08008 *Metapenaeus ensis tropomyosi...*(2) 38 38 66 95.3 0.34  
 Allg:Chclah4 Strand: -  
 ! X78223 *C.herbarum ClaH4 ribosomal p...*(6) 46 46 57 83.1 1.6  
 Allg:Chclah2 Strand: -  
 ! X78226 *C.herbarum ClaH2 mRNA for en...*(6) 58 58 59 78.6 2.9  
 Allg:Ppphlpi  
 ! X78813 *P.pratense mRNA for Phl pI a...*(3) 56 56 56 76.5 3.7  
 Allg:Pspalpr  
 ! X85187 *P.sativum mRNA for pollen al...*(1) 37 37 52 72.4 6.2  
 Allg:Ldlepidi  
 ! X81399 *L.destructor mRNA Lepd I all...*(3) 48 48 48 72.2 6.4  
 Allg:Cacorai5 Strand: -  
 ! X70999 *C.avellana Cor a I/5 mRNA. 1...*(4) 47 47 50 71.8 6.7  
 Allg:Lolg9i Strand: -  
 ! L13083 *Lolium perenne isoform 9 mRN...*(6) 44 62 54 71.5 7  
 Allg:S69138  
 ! S69138 *TcR V alpha=T-cell receptor ...*(3) 38 38 39 71.5 7  
 Allg:Cacorai11 Strand: -  
 ! X70997 *C.avellana Cor a I/11 mRNA. ...*(4) 47 47 50 71.3 7.2  
 Allg:S69142  
 ! S69142 *TcR V alpha=T-cell receptor ...*(3) 38 38 39 71.2 7.3  
 Allg:Hstcc4215  
 ! X86360 *H.sapiens mRNA for T cell re...*(2) 30 30 30 69.7 8.8  
 Allg:Pjmajall Strand: -  
 ! X77414 *P.judaica mRNA for major all...*(6) 49 49 49 69.6 8.8  
 Allg:Pspalpr  
 ! X85187 *P.sativum mRNA for pollen al...*(2) 41 41 50 69.1 9.5  
 \\ End of List

Pat.Pep  
 Allg:Meu08008  
 LOCUS MEU08008 843 bp mRNA INV 11-JAN-1995  
 DEFINITION *Metapenaeus ensis tropomyosin mRNA, partial cds.*  
 ACCESSION U08008  
 KEYWORDS .  
 SOURCE penaeid shrimps.  
 ORGANISM *Metapenaeus ensis* . . .  
 SCORES Frame: (2) Init1: 38 Initn: 38 Opt: 66 z-score: 95.3 E(): 0.34  
 25.5% identity in 55 aa overlap

		10	20	30	39
Pat.Pep		MSPERRPVEIRPATAADMAAVCDIVNHY-IETSTVNFRT			
		:    : : : :: : :			
Meu08008	LTAASSCSRRTWSALRNASTPPPPSWLRPPPTSPSACARCSRTAPFPMSAWTPWRTS				
	90 100 110 120 130 140				
	40 50 60 70 80 90				
Pat.Pep	PQTPQEWI----DDLRLQDRYPWLVAEVEGVVAGIAYAGPWKARNAYDWTVESTVYVSH				
	:    : :    :				
Meu08008	SRRPASWLRKPTGNTTRLFPVSWPWLRLPTLSVLKQVLRVLRVNRSSSLRKNCVSLATTXSL				
	150 160 170 180 190 200				

Pat.Pep  
 Allg:Chclah4  
 LOCUS CHCLAH4 639 bp RNA PLN 12-APR-1995  
 DEFINITION C.herbarum ClaH4 ribosomal protein P2 mRNA for minor allergen.  
 ACCESSION X78223  
 KEYWORDS ribosomal protein.  
 SOURCE Cladosporium herberum.  
 ORGANISM Cladosporium herberum . . .  
 SCORES Frame: (6) Init1: 46 Initn: 46 Opt: 57 z-score: 83.1 E(): 1.6  
 66.7% identity in 9 aa overlap

	150	160	170	180	
Pat.Pep	GYTARGTLRAAGYKHGGWHDVGFWRDPFELPAPPRPVRPVTQI				
					: : :
Chclah4	SLDQDFASNIHCRSSPLSRDRSPCRRTQPPSPQPSRPRGRPQRHHQRRQHRQRRHQ				
	50	60	70	80	90 100
Chclah4	HQERRRASQSRKRSARXCPCPRAPXGGAETAPRRRCQRSXGRSXCPCQQRGRSCPASRG				
	110	120	130	140	150 160

Pat.Pep  
 Allg:Chclah2  
 LOCUS CHCLAH2 1450 bp RNA PLN 12-APR-1995  
 DEFINITION C.herbarum ClaH2 mRNA for enolase.  
 ACCESSION X78226  
 KEYWORDS enolase.  
 SOURCE Cladosporium herberum.  
 ORGANISM Cladosporium herberum . . .  
 SCORES Frame: (6) Init1: 58 Initn: 58 Opt: 59 z-score: 78.6 E(): 2.9  
 52.6% identity in 19 aa overlap

	10	20	30	40	50
Pat.Pep	MSPERRPVEIRPATAADMAAVCDIVNHYIETSTVNFRTPEQTPOEWIDDLERLQDRYP				
					: : : :      : :
Chclah2	LSPSRSSQASWEPVEATEGTMARXRPVSVTMTSTVGFPPRESXTXREWILEIGMLAVSR				
	430	440	450	460	470 480
Pat.Pep	60	70	80	90	100 110
	WLVAEVEGVVAGIAYAGPWKARNAYDWTVESTVYVSHRHQRLGLGSTLYTHLLKSMEAQG				

Pat.Pep  
 Allg:Ppplpi  
 LOCUS PPHLPI 1066 bp RNA PLN 15-NOV-1994  
 DEFINITION P.pratense mRNA for Phl pI allergen.  
 ACCESSION X78813  
 KEYWORDS allergen; phlpI gene.  
 SOURCE timothy grass.  
 ORGANISM Phleum pratense . . .  
 SCORES Frame: (3) Init1: 56 Initn: 56 Opt: 56 z-score: 76.5 E(): 3.7  
 53.3% identity in 15 aa overlap

	140	150	160	170	180
Pat.Pep	RLHEALGYTARGTLRAAGYKHGGWHDVGFWRDPFELPAPPRPVRPVTQI				
					:       :
Ppplpi	RTRLRLLLXDVHQARGLLWRARGSPHERRQGAHRPLPLRPLRPRVVRGQEGRXAEAA				
	90	100	110	120	130 140
Ppplpi	QRRRAGAPVPARQVQVPGGHQGDLPREGVQPLPGAACEVVRXRRRRGGGGHQQEGGQGG				
	150	160	170	180	190 200

Pat.Pep  
Allg:Pspalpr  
LOCUS PSPALPR 822 bp RNA PLN 13-MAR-1995  
DEFINITION *P.sativum* mRNA for pollen allergen-like protein.  
ACCESSION X85187  
KEYWORDS  
SOURCE pea.  
ORGANISM *Pisum sativum* . . .  
SCORES Frame: (1) Init1: 37 Initn: 37 Opt: 52 z-score: 72.4 E(): 6.2  
30.4% identity in 23 aa overlap

30	40	50	60	70	80	89
Pat.Pep	ETSTVNFRTPEQPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNAYDWTVES					
	: : :   : :     : :					
Pspalpr	VLSLMAAAILLLTIASLTSLFAPTTARIPGVYTGPPWTSAHATFYGGSD					
	10	20	30	40	50	
90	100	110	120	130	140	149
Pat.Pep	TVVSHRHQRLGLGSTLYTHLLKSMEAQGFKSVVAVIGLPNDPSVRLHEALGYTARGTLR					
Pspalpr	ASGTMGGACGYGNLYSQGYGVNTAALSTALFNNGLSGACFELKCDQDPRWCNPGNPSIL					
	60	70	80	90	100	110

Pat.Pep  
Allg:Ldlepidi  
LOCUS LDLEPDI 420 bp RNA INV 28-FEB-1995  
DEFINITION *L.destructor* mRNA Lepd I allergen.  
ACCESSION X81399  
KEYWORDS allergen.  
SOURCE *Lepidoglyphus destructor*.  
ORGANISM *Lepidoglyphus destructor* . . .  
SCORES Frame: (3) Init1: 48 Initn: 48 Opt: 48 z-score: 72.2 E(): 6.4  
28.1% identity in 32 aa overlap

	10	20	30	40		
Pat.Pep	MSPERRPVEIRPATAADMAAVCDIVNHYIETSTVNFRTPEQPQE					
	: : :   : :     : :					
Ldlepidi	PSQEGRSSXLHLQRNHPSHHPQGGRRRHRADRRPRCHGLRYRPRTSRINLTXPNLTKPT					
	50	60	70	80	90	100
Pat.Pep	50	60	70	80	90	100
	WIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNAYDWTVESTVYVSHRHQRLGLGST					
Ldlepidi	WIFIVQSKFINKHNKFIQKLIKKKKKKKKST					
	110	120	130			

Pat.Pep  
Allg:Cacora15  
LOCUS CACORA15 619 bp RNA PLN 29-NOV-1993  
DEFINITION *C.avellana* Cor a I/5 mRNA.  
ACCESSION X70999 S56570  
KEYWORDS allergen.  
SOURCE *Corylus avellana*.  
ORGANISM *Corylus avellana* . . .  
SCORES Frame: (4) Init1: 47 Initn: 47 Opt: 50 z-score: 71.8 E(): 6.7  
23.6% identity in 72 aa overlap

	10	20	30			
Pat.Pep	MSPERRPVEIRPATAADMAAVCDIVNHYIET					
	:   : : :   : :   : : :					
Cacora15	HGRLWRGTCCXSSRWILHQGLPLSSARGRPSPACHPGHHPRSRCSMKSLCCQPHQPSPS					
	80	90	100	110	120	130
Pat.Pep	40	50	60	70	80	90
	STVNFRTPEQPQEWIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNAYDWTVESTV					
	: : :   : :     : :     : :					
Cacora15	RTCNGCLRORXYSXWPFQALHFLQRFQRWXKLEVQPL--GXAYRHQGHRTXTALPQGXRRR					
	140	150	160	170	180	190
Pat.Pep	100	110	120	130	140	150
	YVSHRHQRLGLGSTLYTHLLKSMEAQGFKSVVAVIGLPNDPSVRLHEALGYTARGTLRAA					
Cacora15	SQPRNXKHP					
	200					

Pat. Pep  
Allg:Lolg9i  
LOCUS LOLG9I 1272 bp mRNA PLN 16-FEB-1994  
DEFINITION *Lolium perenne* isoform 9 mRNA, complete cds.  
ACCESSION L13083  
KEYWORDS isoform 9.  
SOURCE *Lolium perenne* cDNA to mRNA.  
ORGANISM *Lolium perenne* . . .  
SCORES Frame: (6) Init1: 44 Initn: 62 Opt: 54 z-score: 71.5 E(): 7  
33.3% identity in 48 aa overlap  
100 110 120 130 140 150  
Pat. Pep R L G L G S T L Y T H L L K S M E A Q G F K S V V A V I G L P N D P S V R L H E A L G Y T A R G T L R A A G Y K H G G W  
Lolg9i R G R R W R Q W R R S C T R R R P C R X A A P R R R G W R R P X P P R R R --- A L R R G P R G C R R X R G A P R X G W  
210 220 230 240 250 260  
160 170 180  
Pat. Pep H D V G - F W Q R D F E L P A P P R P V R P V T Q I  
: | : : : | : | |  
Lolg9i R R R H R T W P R E W R - P G Q P R K R A L R R R R A W R X A G S C F D R R S R T R G P Q R R P C W W P C R R F R R T S  
270 280 290 300 310 320

Pat. Pep  
Allg:S69138  
LOCUS S69138 95 bp mRNA PRI 22-SEP-1994  
DEFINITION TcR V alpha-T-cell receptor alpha-chain (allergen-specific) [human, grass-sensitive individual VI 26, peripheral blood, mRNA Partial, 95 nt].  
ACCESSION S69138  
KEYWORDS . . . . .  
SCORES Frame: (3) Init1: 38 Initn: 38 Opt: 39 z-score: 71.5 E(): 7  
71.4% identity in 7 aa overlap  
140 150 160 170 180  
Pat. Pep D P S V R L H E A L G Y T A R G T L R A A G Y K H G G W H D V G F W Q R D F E L P A P P R P V R P V T Q I  
S69138 L S H L L L C S K Q R R X P V L F W D R D K F D G H S K Y P E  
10 20 30

Pat. Pep  
Allg:Cacoraill  
LOCUS CACORAI11 655 bp RNA PLN 18-JUL-1993  
DEFINITION *C. avellana* Cor a I/11 mRNA.  
ACCESSION X70997  
KEYWORDS allergen.  
SOURCE *Corylus avellana*.  
ORGANISM *Corylus avellana* . . .  
SCORES Frame: (4) Init1: 47 Initn: 47 Opt: 50 z-score: 71.3 E(): 7.2  
23.6% identity in 72 aa overlap  
10 20 30  
Pat. Pep M S P E R R P V E I R P A T A A D M A A V C D I V N H Y I E T  
Cacoraill H G H L W R G T C C X S S R W I L H Q G L P L S S A R G R P S P A C H P G H H P R S R C S C M X S L C C Q P H Q P S P S  
100 110 120 130 140 150  
40 50 60 70 80 90  
Pat. Pep S T V N F R T E P Q T P Q E W I D D L E R L Q D R Y P W L V A E V E G V V A G I A Y A G P W K A R N A Y D W T V E S T V  
Cacoraill R T C N G C L R Q R X Y S X W F Q A L H P L Q R F Q R W X X L E V Q P L - - G X A Y R H Q G H R T X T A L P Q G X R R W  
160 170 180 190 200  
100 110 120 130 140 150  
Pat. Pep Y V S H R H Q R L G L G S T L Y T H L L K S M E A Q G F K S V V A V I G L P N D P S V R L H E A L G Y T A R G T L R A A  
Cacoraill S Q P R N X K H P  
210

Pat. Pep  
Allg:S69142  
LOCUS S69142 98 bp mRNA PRI 22-SEP-1994  
DEFINITION TcR V alpha-T-cell receptor alpha-chain {allergen-specific} [human, grass-sensitive individual BAL 87, peripheral blood, mRNA Partial, 98 nt].  
ACCESSION S69142  
KEYWORDS . . . . .  
SCORES Frame: (3) Init1: 38 Initn: 38 Opt: 39 z-score: 71.2 E(): 7.3  
71.4% identity in 7 aa overlap  
140 150 160 170 180  
Pat. Pep DPSVRLHEALGYTARGTLRAAGYKHGGWHDVGFWRDQDFELPAPPRPVRPVTQI  
S69142 LSHLLCSKLNRRXPVLFWDRDKFDGHSKYPE  
10 20 30

Pat. Pep  
Allg:Hstcc4215  
LOCUS HSTCC4215 24 bp RNA PRI 21-APR-1995  
DEFINITION H.sapiens mRNA for T cell receptor, TCC 4:2.15.  
ACCESSION X86360  
KEYWORDS T-cell; T-cell receptor beta-chain.  
SOURCE human.  
ORGANISM Homo sapiens . . .  
SCORES Frame: (2) Init1: 30 Initn: 30 Opt: 30 z-score: 69.7 E(): 8.8  
60.0% identity in 5 aa overlap  
50 60 70 80 90 100  
Pat. Pep WIDDLERLQDRYPWLVAEVEGVVAGIAYAGPWKARNAYDWTVESTVYVSHRHQRLGLGST  
Hstcc4215 QPWXDRG  
110 120 130 140 150 160  
Pat. Pep LYTHLLKSMEAQGFKSVVAVIGLPNDPSVRLHEALGYTARGTLRAAGYKHGGWHDVGFWDQ

Pat. Pep  
Allg:Pjmajall  
LOCUS PJMAJALL 653 bp RNA PLN 18-SEP-1995  
DEFINITION P.judaica mRNA for major allergen.  
ACCESSION X77414  
KEYWORDS major allergen.  
SOURCE Parietaria judaica.  
ORGANISM Parietaria judaica . . .  
SCORES Frame: (6) Init1: 49 Initn: 49 Opt: 49 z-score: 69.6 E(): 8.8  
46.7% identity in 15 aa overlap  
70 80 90 100 110 120  
Pat. Pep EGVVAGIAYAGPWKARNAYDWTVESTVYVSHRHQRLGLGSTLYTHLLKSMEAQGFKSVVA  
Pjmajall REXCHYSLVNYIYIYXDNVICIXLLITQFHRXSLGLNNISLAHPRVEKAGLGFRCGG  
30 40 50 60 70 80  
130 140 150 160 170 180  
Pat. Pep VIGLPNDPSVRLHEALGYTARGTLRAAGYKHGGWHDVGFWRDQDFELPAPPRPVRPVTQI  
Pjmajall RNSNLGSLQPCFVGGITWARDGTMTERDWKGLPRNHSNCLAVHVDVNGRELAVNDAAVL  
90 100 110 120 130 140

```
Pat.Pep
Allg:Pspalpr
LOCUS      PSPALPR      822 bp      RNA      PLN      13-MAR-1995
DEFINITION P.sativum mRNA for pollen allergen-like protein.
ACCESSION  X85187
KEYWORDS   .
SOURCE     pea.
ORGANISM   Pisum sativum . . .
SCORES     Frame: (2) Init1: 41      Initn: 41      Opt: 50      z-score: 69.1  E(): 9.5
           24.2% identity in 33 aa overlap
           100      110      120      130      140      150
Pat.Pep     HQRLGLGSTLYTHLLKSMEAQGFKSVVAVIGLPNDPSVRLHEALGYTARGTLRAAGYKHG
           |  ::|:  |  |  :  :  ::  :|
Pspalpr     XSPTSRVQGISCAVXKKEPTQRGXQXAVTGGRGTGNPTQFLWARLSLLGSPAVTVVHPRHG
           190      200      210      220      230      240
           160      170      180
Pat.Pep     GWHVGVFWQRDFELPAPPRFVRPVTQI
           ||
Pspalpr     TWHHLIGSSDKPSQERISASEKIIKKKKK
           250      260      270
! Distributed over 1 thread.
! Start time: Mon Apr 12 10:44:55 1999
! Completion time: Mon Apr 12 10:44:59 1999
! CPU time used:
! Database scan: 0:00:00.4
! Post-scan processing: 0:00:00.7
! Total CPU time: 0:00:01.1
! Output File: Pat.Tfasta
```

**Appendix 1. Study Protocol PHI99-013.**