Evolutionary Analysis of the CAP Superfamily of Proteins using Amino Acid Sequences

and Splice Sites

by

Anup Abraham

A Thesis Presented in Partial Fulfillment of the Requirements for the Degree Master of Science

Approved April 2016 by the Graduate Supervisory Committee:

Douglas E. Chandler, Chair Kenneth H. Buetow Robert W. Roberson

ARIZONA STATE UNIVERSITY

May 2016

ABSTRACT

Here I document the breadth of the CAP (Cysteine-RIch Secretory Proteins (CRISP), Antigen 5 (Ag5), and the Pathogenesis-Related 1 (PR)) protein superfamily and trace some of the major events in the evolution of this family with particular focus on vertebrate CRISP proteins. Specifically, I sought to study the origin of these CAP subfamilies using both amino acid sequence data and gene structure data, more precisely the positions of exon/intron borders within their genes. Counter to current scientific understanding, I find that the wide variety of CAP subfamilies present in mammals, where they were originally discovered and characterized, have distinct homologues in the invertebrate phyla contrary to the common assumption that these are vertebrate protein subfamilies. In addition, I document the fact that primitive eukaryotic CAP genes contained only one exon, likely inherited from prokaryotic SCP-domain containing genes which were, by nature, free of introns. As evolution progressed, an increasing number of introns were inserted into CAP genes, reaching 2 to 5 in the invertebrate world, and 5 to 15 in the vertebrate world. Lastly, phylogenetic relationships between these proteins appear to be traceable not only by amino acid sequence homology but also by preservation of exon number and exon borders within their genes.

TABLE OF CONTENTS

LIST OF FIGURES iv
LIST OF APPENDIX FIGURES vi
LIST OF SUPPLEMENTAL DATA TABLESvii
CHAPTER
1 INTRODUCTION
2 IN-DEPTH ANALYSIS OF CAP SUPERFAMILY PROTEINS
The CAP Superfamily Proteins: Domains, Sequences and Structural Relationships. 5
Bacterial, Plant, and Fungal SCP/PR Proteins
CAP Proteins of Insects: the Venom Allergen Subfamily14
The Cysteine-RIch Secretory Protein (CRISP) Subfamily 17
Truncated CRISP (Allurin)
The LCCL Domain-containing Subfamily
The GLIPR/GLIPR-like Subfamily
The GAPR Subfamily
The Peptidase Inhibitor Subfamily
3 EVOLUTION OF CAP SUPERFAMILY GENE STRUCTURE
Evolution of CAP Superfamily Gene Structure
Conservation of Exon Borders within CAP Subfamily Genes
Physical Clustering of CAP Subfamily Genes in the Genome
The Origins of CAP Subfamilies with Emphasis on the Evolution of CRISP Genes58
4 EVOLUTION OF CRISP GENE STRUCTURE

CHAPTER	Page
Evolution of the CAP/PR Domain	66
Addition of the Hinge Domain	69
Addition of the ICR Domain	71
5 CONCLUSION	75
BIBLIOGRAPHY	81
APPENDIX	
A FIGURES	91
B TABLES	142

LIST OF FIGURES

Fig	gure Page
1.	Subfamilies of the CAP Superfamily
2.	Domain Structure of CAP Proteins7
3.	ClustalW Amino Acid Alignment of the CAP/PR Domain in Bacterial SCP
	Proteins
4.	ClustalW Amino Acid Alignment of the CAP/PR Domain in Plant Pathogenesis-
	Related Proteins
5.	ClustalW Amino Acid Alignment of the CAP/PR Domain in Fungal CAP
	Superfamily Proteins
6.	Tertiary Structure of the Tomato P14a Pathogenesis-Related Protein
7.	ClustalW Amino Acid Alignment of Insect Antigen-Related Proteins 16
8.	ClustalW Amino Acid Alignment of Vertebrate CRISP2 and CRISP2-like
	Proteins
9.	Domain Organization and Tertiary Structure of a Full Length CRISP Protein:
	Venom CRISP from the Chinese Cobra
10	. ClustalW Amino Acid Alignment and Secondary Structure of Selected CRISP
	Proteins
11.	. ClustalW Amino Acid Alignment of Invertebrate CRISP2-like Proteins
12.	. Global Genome Relatedness between Selected Organisms in This Study
13.	. Genomic Exon Border Structure in Selected Vertebrate CRISP2 and CRISP2-like
	Genes

14. ClustalW Amino Acid Alignment for Exon Borders in Selected CRISP and
CRISP-like Genes
15. Genomic Exon Border Structure in Selected Invertebrate CRISP and CRISP-like
Amino Acid Sequences
16. Evolution of CAP Proteins is accompanied by an Increase of Exons Representing
the CAP/PR Domain
17. Phylogenetic Relationships of CAP Superfamily Proteins in the X. tropicalis
(Western Clawed Frog) Genome
18. Genomic Exon Border Structure in CAP Superfamily Proteins in the X. tropicalis
(Western Clawed Frog) Genome
19. Phylogenetic Relationships Of All CAP Superfamily Proteins in the Homo
sapiens (Human) Genome
sapiens (Human) Genome
20. Genomic Exon Border Structure In All CAP Superfamily Proteins in the Homo
20. Genomic Exon Border Structure In All CAP Superfamily Proteins in the Homo sapiens (Human) Genome
 20. Genomic Exon Border Structure In All CAP Superfamily Proteins in the Homo sapiens (Human) Genome
 20. Genomic Exon Border Structure In All CAP Superfamily Proteins in the Homo sapiens (Human) Genome
 20. Genomic Exon Border Structure In All CAP Superfamily Proteins in the Homo sapiens (Human) Genome
 20. Genomic Exon Border Structure In All CAP Superfamily Proteins in the Homo sapiens (Human) Genome
 20. Genomic Exon Border Structure In All CAP Superfamily Proteins in the Homo sapiens (Human) Genome

APPENDIX A FIGURES

Figure	Page
1. A. Inventory of CAP-related Proteins	in Selected Species92
B. Phylogenetic Tree vs Number of CA	AP Protein in Genome96
2. ClustalW Amino Acid Alignment and	Color-Coded Mapping of Exons for
Selected Sequences	
3. ClustalW Amino Acid Alignment and	Color-Coded Mapping of Exons for
Selected Invertebrate CAP Protein Sec	quences104
4. ClustalW Amino Acid Alignment and	Color-Coded Mapping of Exons for CAP
Superfamily Genes in X. tropicalis Ge	nome108
5. Pairwise Distance of <i>X. tropicalis</i> CAR	P Amino Acid Sequences – Estimated
Evolutionary Divergence Between Sec	quences115
6. ClustalW Amino Acid Alignment and	Color-Coded Mapping of Exons for CAP
Superfamily Genes in the Homo sapier	<i>ns</i> Genome116
7. Pairwise Distance of <i>Homo sapiens</i> Ca	AP Amino Acid Sequences – Estimated
Evolutionary Divergence Between Sec	quences124
8. ClustalW Amino Acid Alignment and	Color-Coded Mapping of Exons for
GAPR/GLIPR2 Sequences	
9. Evolution of the Pathogenesis-Related	(CAP/PR) Domain in the CRISP Protein
Lineage	

APPENDIX B TABLES

Table	Page
1.	Subfamilies of the CAP Superfamily143
2.	Domain Structure of CAP Proteins
3.	Accession ID and Organism Information for ClustalW Amino Acid Alignment of
	CAP/SCP Proteins in Bacteria145
4.	Accession ID and Organism Information for ClustalW Amino Acid Alignment of
	CAP/SCP Proteins in Plants
5.	Accession ID and Organism Information for ClustalW Amino Acid Alignment of
	CAP/SCP Proteins in Fungus147
6.	Accession ID and Organism Information for ClustalW Amino Acid Alignment of
	Insect Antigen-Related Proteins
7.	Accession ID and Organism Information for ClustalW Amino Acid Alignment of
	Vertebrate CRISP2 and CRISP2-like Proteins
8.	Accession ID and Organism Information for ClustalW Amino Acid Alignment
	and Secondary Structure of Chinese Cobra CRISP Protein151
9.	Accession ID and Organism Information for ClustalW Amino Acid Alignment of
	Invertebrate CRISP2-like Proteins
10	. Accession ID and Organism Information for Genomic Exon Border Structure In
	Selected Invertebrate CRISP and CRISP-like Amino Acid Sequences153
11	. Accession ID and Organism Information for Phylogenetic Relationships of All
	CAP Superfamily Proteins In X. tropicalis (Western Clawed Frog) Genome154

12.	Accession ID and Organism Information for Phylogenetic Relationships Of All
	CAP Superfamily Proteins In the Homo sapiens (Human) Genome155
13.	Accession ID and Organism Information for Evolution of Multi-Domain CAP
	Superfamily Proteins156
14.	Accession ID and Organism Information for ClustalW Amino Acid Alignment
	and Color-Coded Mapping of Exons for Selected Protein Sequences (Appendix 2)
15.	Accession ID and Organism Information for ClustalW Amino Acid Alignment
	and Color-Coded Mapping of Exons for Selected Invertebrate Sequences
	(Appendix 3)160
16.	Accession ID and Organism Information for ClustalW Amino Acid Alignment
	and Color-Coded Mapping of Exons for GAPR/GLIPR2 Sequences (Appendix 8)
17.	Accession ID and Organism Information for Evolution of the Pathogenesis-
	Related (CAP/PR) Domain in the CRISP Protein Lineage (Appendix 9)163

CHAPTER 1: INTRODUCTION

The CAP superfamily of proteins has members that are wide spread throughout prokaryotic and eukaryotic phyla, including those within the bacteria, plant, fungus, and animal kingdoms. The name CAP is an acronym derived from the three major protein families, which includes: Cysteine-RIch Secretory Proteins (CRISP), Antigen 5 of insect venoms (Ag5), and the Pathogenesis-Related 1 (PR) proteins of plants. Members of this superfamily of proteins contain a conserved domain known as the SCP (Sperm-Coating Protein) domain, since several important members were first discovered by their ability to bind to mammalian sperm [1, 2]. The CAP superfamily of proteins plays a wide array of functions in different organisms; in addition, many other members have not been functionally characterized.

Phylogenetic trees, constructed by Mega 6 neighboring joining alignment based on amino acid sequence [3-5] such as that in Figure 1, show that the CAP superfamily contains an number of subfamilies including the following: bacterial SCP domain containing proteins, pathogenesis-related (PR) proteins most commonly found in plants, fungal pathogenesis related (PRY) proteins, venom antigens (Ag) of insects, CRISP proteins of vertebrates, glioma pathogenesis related (GLIPR) proteins, golgi-associated pathogenesis-related (GAPR) proteins, and Cysteine-Rich LCCL domain-containing (CRISP LD) proteins. A few smaller subfamilies have been excluded from Figure 1, which include: peptidase inhibitor (PI) proteins, C-type lectin (CLEC) proteins, and HrTT proteins, plus numerous eukaryotic SCP domain-containing proteins that do not fall within a recognized subfamily. (Note: In this thesis I will use the term CAP/PR domain, consistent with the terminology used in most of the research literature. However, anyone searching data bases in regard to this domain should use the term SCP domain. This acronym originated because CAP domain containing proteins were initially discovered in sperm.)

Not only are CAP proteins widely distributed throughout the eukaryotic taxa, species spanning both vertebrate and invertebrate animal phyla exhibit in their genomes multiple CAP family genes. An inventory substantiating this observation is found in Appendix 1, which includes representative species whose genomes have been completely sequenced. The inventory was generated using currently available amino acid data on NCBI's protein database (www.ncbi.nlm.nih.gov/protein/) and GeneBank. Current information on each of these CAP genes ranges from genomic sequence only (coding for a hypothetical, predicted, or "uncharacterized" protein), to genomic and mRNA sequence, to fully sequenced and structurally characterized protein products. Nevertheless, it is clear from this inventory that the CAP superfamily of genes and their protein products have undergone an extensive evolutionary history that started in the bacterial and primitive eukaryotic world and has continued to expand to the present day.

It is the goal of this thesis to document the breadth of the CAP superfamily and to trace some of the major events in the evolution of this family with particular focus on vertebrate CRISP proteins. Specifically, I sought to study the origin of these CAP subfamilies using both amino acid sequence data and gene structure data, precisely the positions of exon/intron borders within their genes. Surprisingly, I find that the wide variety of CAP subfamilies present in mammals, where they were originally discovered and characterized, have distinct homologues in the invertebrate phyla contrary to the common assumption that these are vertebrate subfamilies. In addition, I document the

2

fact that primitive eukaryotic CAP genes contained only one exon, likely inherited from prokaryotic SCP-domain containing genes which were, by nature, free of introns. As evolution progressed, an increasing number of introns were inserted into CAP genes, reaching 2 to 5 in the invertebrate world, and 5 to 15 in the vertebrate world. Lastly, phylogenetic relationships between proteins appear to be traceable not only by amino acid sequence homology but also by preservation of exon number and exon borders within their genes.

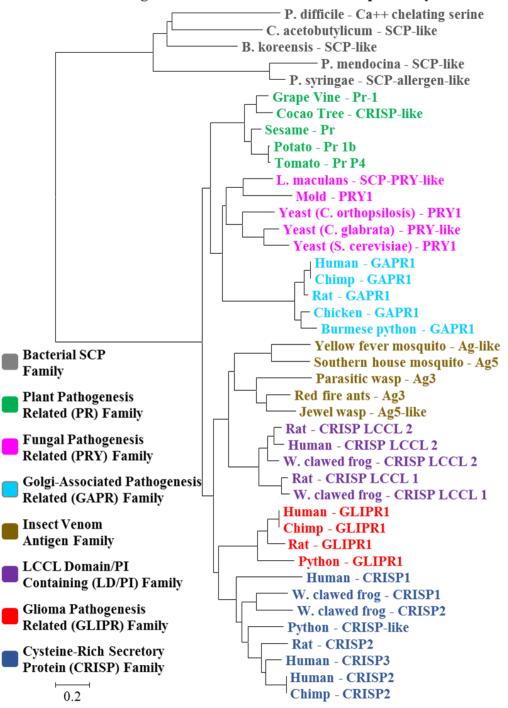


Figure 1. Subfamilies of the CAP Superfamily

Figure 1. MEGA7 ClustalW amino acid sequence alignment and Neighbor Joining Phylogeny Tree alignment of various CAP/PR domain-containing proteins. The figure presents a wide array of identified CAP/PR proteins in a large number of organisms. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 1.

CHAPTER 2: IN-DEPTH ANALYSIS OF CAP SUPERFAMILY PROTEINS

The CAP Superfamily Proteins: Domains, Sequences and Structural Relationships

All CAP superfamily proteins, by definition, have the CAP/PR domain consisting of about 160 residues typically found at the N-terminal (Figure 2). As a result, this domain is recognized as a conserved domain in NCBI databases

(www.ncbi.nlm.nih.gov/Structure/). More precisely, proteins of the CAP superfamily are characterized by four highly conserved "signature" sequences within the CAP/PR domain [6]. Of these highly conserved sequences, CAP1 is an eleven amino acid segment characterized by the following sequence: G H [Y or F] [S or T] Q [V or L] V W s s [S or T] (s = small residue). CAP2 is a twelve amino acid segment characterized by the following sequence: h h V C [N, H or Q] Y s P s G N h (h = hydrophobic residue). CAP3 is a five amino acid segment characterized by the following sequence: H N x x R (x = any residue). Finally, CAP4 is a four amino acid segment characterized by the following sequence: G [E or Q] N [I, L or V] [6]. In Figures 3, 4, and 5, these CAP signature sequences are boxed in the aligned sequences of bacterial, plant, and fungal CAP proteins, respectively.

Almost all members of the CAP superfamily are secretory glycoproteins and are stable over a wide range of conditions [7]. Currently, thousands of CAP superfamily proteins have been sequenced and a modest number have had their tertiary crystal structure resolved. In addition, a number of CAP superfamily proteins have had their biological functions defined. Pr-1 and Pr proteins express antifungal activity and play a role in pathogen resistance and wound-signaling in plants [8-10]. Ag5 is a major allergen of vespid venom, localized in the venom secretory ducts of stinging insects [11]. CRISP and CRISP-like proteins are expressed in the venom of insects and the reproductive track of vertebrates [12, 13]. Protease inhibitors, which are expressed in neuroblastoma and glioblastoma cell lines, are involved in trypsin inhibition [14].

Regardless of these diverse physiological functions and due to the high degree of conserved sequence motifs in the CAP superfamily, we hypothesize that each CAP subfamily shares a common point of origin. In this chapter we will define each CAP protein subfamily by its key characteristics, region of expression, cause of expression, and other relevant features. In Chapter three we will also take an in-depth look at the genome sequence and amino acid sequence data to identify evolutionarily conserved characteristics of the CAP superfamily of proteins. In particular, we will use currently available genome sequences and exon structure data to trace evolutionary relationships within the CAP superfamily. Exon border data can provide analytical information to extrapolate evolutionary associations between protein sequences within a genome and between species. Gene cluster data will be used to identify points where gene duplication has occurred, and this information will be used to infer the point of origin and diversification of protein function. Finally, I will focus on the evolution of vertebrate CRISP proteins, examining the origin of each of their three domains.

In order to analyze and evaluate evolutionary characteristics of CAP superfamily of proteins, I will assess CAP proteins in invertebrates and vertebrates. A wide array of organisms have been chosen for this analysis based on the availability of protein sequence, genome data, and scaffolding or genome assembly data.

6

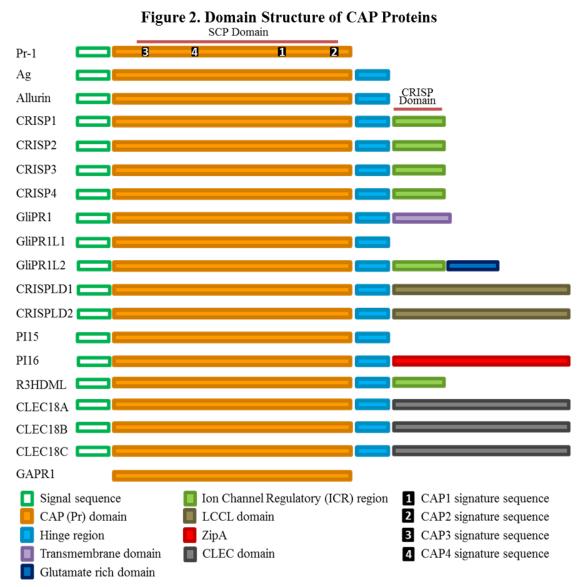


Figure 2. The domain structure of CAP superfamily of proteins after MEGA7 ClustalW alignment. All CAP proteins have the pathogenesis-related (CAP/PR) domain, which contains a SCP conserved domain (red line) and four CAP signature sequences. The four signature sequences, which are CAP1, CAP2, CAP3, and CAP4, are used to identify CAP superfamily proteins. Pr-1 has been labeled with these signature sequences; all subsequent CAP superfamily proteins have equivalent signature sequences in the same region, due to ClustalW alignment. Many but not all CAP proteins also contain a Hinge region (blue box). The CRISP subfamily contains an ICR domain (green box) in which is embedded a conserved Crisp domain (red line). Organisms and accession IDs of the amino acid sequences are available in supplemental data table 2.

Bacterial, Plant, and Fungal SCP/PR Proteins

CAP/PR related domains in their most primitive form occur in bacteria. While bacterial sequence information (Figure 3) show recognizable similarities to the consensus SCP sequence there are many residue sites that deviate from the consensus (consensus sequences shown of Figure 4 for reference), many of these even lying with the CAP signature regions (boxes in orange). Furthermore, unlike eukaryotic CAP superfamily proteins, those from bacteria lack any recognizable CAP2 signature sequence near their C-terminus. In addition, they do not contain the entire series of six conserved cysteine residues found in most CAP/PR domains but rather 2 to 4 of these evolutionarily advantaged cysteines. This suggests that primitive CAP proteins may have had a rather different tertiary structure than later CAP proteins or at least had a structure that was not as well stabilized by disulfide bonds. Whether this lack of cysteines stems from the early history of the earth when it lacked an oxidizing atmosphere is not clear. In contrast, the pathogenesis-related (Pr) proteins of plants, presumed descendants of bacterial SCP genes through phylogenetic analysis, have a distinct series of six cysteines all disulfidebonded. The NMR solution structure of recombinant tomato P14a protein (Figure 6), a prototype member of this group, demonstrates that these disulfide bonds (in red) link the central beta-sheet with the surrounding alpha helices thus creating a stable tertiary structure that is relatively heat and solvent resistant -a characteristic common to most CAP proteins [15].

Figure 3. ClustalW Amino Acid Alignment of the CAP/PR Domain in Bacterial
SCP Proteins

Common name - Protein name														1	Ami	no	Aci	1 S	equ	enc	e													
											C	A	P3																					
C. acetobutylicum - SCP-like	-	-	-	-	Q	V	Ι	Κ	L	v	Ν	S	Q	R	А	Κ	Y	-	-	-	-	-	-	-	-	-	G	L	A	Р	L	Κ	Α	Ν
P. syringae - SCP-Allergen-like																																	W	
P. mendocina - SCP-like	A	Е	G	Q	Ν	L	L	Q	R	Ι	Ν	A	A	R	Α	Q	А	R	Q	С	G	A	Q	Q	F	А	А	А	А	Р	L	Т	W	S
A. flavithermus - Membrane Protein	-																																W	
M. avium - SCP-like	-	-	-	-	Ι	L	Y	Ν	G	Ι	Ν	Q	L	R	Q	Α	С	G	-	-	-	-	-	-	-	-	-	-	-	Р	Ι	Α	Е	D
P. difficile 630 - Ca++ chelating serine	\mathbf{C}	F	Ε	L	Ε	Ι	Ι	D	L	V	Ν	S	V	R	Ν	Q	R	-	-	-	-	-	-	-	-	-	G	L	Ν	S	L	R	Y	S
B. koreensis - SCP-like	-	-	-	-	Q	V	A	Т	L	v	Ν	Q	Е	R	K	K	A	-	-	-	-	-	-	-	-	-	G	L	S	Р	L	Q	L	D
C. acetobutylicum - SCP-like	т	0	т	т	N	т	٨	P	v	v	s	P	D	м	м	м	ĸ	N	v	F	ç	ч	т	ç	D	т	F	G		P4		s	L	т
P. syringae - SCP-Allergen-like		-																															R	
P. mendocina - SCP-like																																	R	
A. flavithermus - Membrane Protein																																	R	
M. avium - SCP-like																																		
																																	R	
P. difficile 630 - Ca++ chelating serine																																	R	
B. koreensis - SCP-like	T	K	L	S	D	V	A	ĸ	Т	K.	S	K.	D	М	м	T	K.	G	Y	F	D	н	Q	S	Р	T	Ŷ	G	S	Р	F	D	Μ	M
C. acetobutylicum - SCP-like																																	Ν	
P. syringae - SCP-Allergen-like	E	L	A	G	Y	V	G	Q	Q	V	G	Е	Ν	Ι	Α	А	G	Q	-	-	-	D	S	A	R	Κ	V	V	D	G	W	L	L	S
P. mendocina - SCP-like	Ε	L	A	G	Y	G	G	G	R	Ι	G	Е	Ν	Ι	Α	А	А	L	-	-	-	D	Т	A	D	Κ	V	V	Ε	G	W	L	Α	S
A. flavithermus - Membrane Protein																																	Ν	
M. avium - SCP-like																																	Q	
P. difficile 630 - Ca++ chelating serine	Κ	R	Ε	G	Ι	V	Y	Т	S	A	G	Е	Ν	Ι	Α	Α	G	Q	-	-	-	Ι	Ν	A	Ι	Y	А	Η	Ε	Α	W	Μ	Ν	S
B. koreensis - SCP-like	K	Q	F	G	Ι	Т	Y	R	A	A	G	E	Ν	Ι	A	ĸ	G	Q	-	-	-	Q	Τ	P	E	E	V	Μ	Ν	S	W	М	Ν	S
							'Al												_	_														
C. acetobutylicum - SCP-like																																	Q	
P. syringae - SCP-Allergen-like																																	А	
P. mendocina - SCP-like																																	А	
A. flavithermus - Membrane Protein																																	R	
M. avium - SCP-like																																	G	
P. difficile 630 - Ca++ chelating serine	E	G	Η	R	Κ	Ν	I	L	G	-	Ν	Y	Ν	Ν	Ι	G	V	G	V	Ι	F	G	-	-	G	S	Y	Κ	Т	Υ	Υ	Т	Q	Ν
B. koreensis - SCP-like	S	G	Η	R	A	N	Ι	L	S	A	N	F	Т	A	Ι	G	V	G	F	V	K	G	-	-	S	Ν	G	Τ	Τ	Y	W	Т	Q	М
C. acetobutylicum - SCP-like	F	Ι	K	R	-														+	+														-
P. syringae - SCP-Allergen-like	F	G	Т	Q	Q															T														
P. mendocina - SCP-like			G																															
A. flavithermus - Membrane Protein			E																															
M. avium - SCP-like			T																1	Ť														
P. difficile 630 - Ca++ chelating serine			K																1	Ť														
B. koreensis - SCP-like	-		G																-	-	_	_												-

CAP signature sequences

Sites of non-aligned sequence omission

Figure 3. This figure provides MEGA7 ClustalW amino acid sequence alignment of CAP/SCP domain-containing proteins in bacteria. The CAP signature sequences are highlighted in orange. Unlike other CAP superfamily proteins, these bacterial proteins contain only three of the four CAP signature sequences. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 3.

Pr proteins contribute to the systemic acquired resistance and hypersensitive

responses in plants [10]. Systemic acquired resistance in plants is analogous to the innate

immune system found in animals. Tobacco and tomato plants expressing high levels of Pr proteins have significantly reduced disease symptoms when infected with several forms of fungus [8]. In addition, when researchers induce stress or wounds on tobacco leafs, Pr gene expression is elevated [9]. Indeed Pr protein expression has been identified as a defense response in many plants; they lower infection rates and hinder the spread of disease, in addition to inducing necrosis in neighboring cells to prevent the spread of infection [10]. This suggests that early forms of CAP proteins were secretory proteins, which played an important role in pathogen defense. However, it should be noted that plant Pr proteins pathogen defense function is not homologous to CAP proteins immune function in humans.

As shown in Figure 2, Pr-1 proteins have a single CAP/PR domain that exhibits all four CAP signature sequences (orange boxes, Figure 4). The SCP domain consensus sequence is highlighted green. This SCP 50% consensus sequence was generated by EMBL SMART (website: smart.embl-heidelberg.de). ClustalW alignment of all CAP amino acid sequences within *Solanum lycopersicum* (tomato) show 100% to 81% query coverage and between 97% and 36% identity (with an E value ranging from 5e-118 to 2e-25) (data not shown), a strong indication of homology due to gene duplication rather than independent origin. Additionally, ClustalW alignment of Pr-1 amino acid sequences in Figure 4 shows 100% to 96% query coverage and between 92% and 65% identity (with an E value ranging from 8e-113 to 3e-78) (data not shown). Also evident are six highly conserved cysteines (yellow highlight) which, as expected in secreted proteins, are disulfide-bonded.

10

				1	r a	th	υĘ	şei	ne	51	3-1																							
Common name - Protein name														Ar	nino	o A	cic	1 S	equ	ene	ce													
	_					_																												
Tomato - PR P4		м	ſG	L	F	N	T	S	L	L	L	т	C	L	м	v	L	А	T	F	н	S	C	E	A						h D			
Potato - PR 1b													C													-				-				
Cocao Tree - CRISP-like													C																					
Grape Vine - PR-1													C																					
Sesame - PR related													A																					
Peach - PR related													C																					
			C	CA1	P3																													
	x	Η	N	c	h	R	u	p	1	u	x	u	s	м	h	c	1	s	W	D	s	p	L	A	Q	s	t	A	Q	s	a	A	s	J
Tomato - PR P4	V	Η	N	D	A	R	А	Q	V	G	V	G	Ρ	Μ	S	W	D	A	Ν	L	A	S	R	A	Q	Ν	Υ	A	Ν	s	R	A	G	Ι
Potato - PR 1b								-					Ρ												-									
Cocao Tree - CRISP-like													Ρ																					
Grape Vine - PR-1													S																					
Sesame - PR related	Α	Η	N	A	A	R	А	Q	V	G	V	G	Ρ	Ι	Α	W	D	Е	K	V	A	A	F	A	R	Ν	Υ	V	Ν	Q	R	Ι	G	Ι
Peach - PR related	A	Η	N	A	A	R	А	Q	V	G	V	_	Р	_	_	W	D	Ρ	Ν	L	V	A	Y	A	Q	R	Υ	A	Ν	s	R	A	G	Γ
													CA		_																			
													E																					
Tomato - PR P4													E																-					
Potato - PR 1b													E																					
Cocao Tree - CRISP-like													E																					
Grape Vine - PR-1													E																					
Sesame - PR related													E																					
Peach - PR related	C	N	L	v	н	. 5	N	G	-	Ρ	Y	G	E	N	1	A	ĸ	S	1	G	D		S AF		1	A	A	V	N	L	r	v	G	ľ
	h	n		2		2		0					s		n		1	G	н		т				w			т	n	0	1	G	C	C
Tomato - PR P4	_							_				_	G	_		_						_						_		_	_	_		_
Potato - PR 1b										-			s																					
Cocao Tree - CRISP-like													S																					
Grape Vine - PR-1													G																					
Sesame - PR related													G																					
Peach - PR related													A																					
		-	-	-	-	-		-		-				_	CA	_		_		-						_					-	-		Ē
	s	s	p	С	s	s	s	h	р	a	1	V	С	p	Y	s	Р	s	G	Ν	h	h	u	x	x	x	x							
Tomato - PR P4													C																					
Potato - PR 1b	R	A	R	С	Ν	N	G	W	W	F	Ι	S	С	N	Y	D	Ρ	V	G	Ν	W	V	G	Q	R	Ρ	Y							
Cocao Tree - CRISP-like													С																					
Grape Vine - PR-1	R	V	Q	С	Ν	N	G	G	W	F	V	Т	С	N	Y	D	Ρ	Ρ	G	Ν	Y	V	G	Q	R	Ρ	Y							
Sesame - PR related	R	A	R	С	Ν	N	G	W	W	F	Ι	S	С	Ν	Y	D	Ρ	Ρ	G	Ν	Y	Ι	G	Q	R	Ρ	Y							
Peach - PR related	K	A	R	C	Т	'N	G	G	т	F	Т	G	C	NT	v	D	D	D	G	NT	т	D	C	\cap	D	D	v							

Figure 4. ClustalW Amino Acid Alignment of the CAP/PR Domain in Plant Pathogenesis-Related Proteins

CAP signature sequence

C Conserved cysteines

SCP consensus sequence

Figure 4. MEGA7 ClustalW amino acid sequence alignment of CAP/PR domaincontaining proteins in plants. The 50% consensus sequence of the SCP domain is highlighted in green. The CAP signature sequences are highlighted in orange. The highly conserved cysteine residues are highlighted in yellow with a red font color. Lower case abbreviations used in the SCP consensus sequence are as follows: x, any; s, small; u, tiny; h, hydrophobic; a, aromatic; l, aliphatic; o, alcohol; c, charged; p, polar; t, turnlike. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 4.

CAP proteins of fungi, exemplified by various examples presented in Figure 5, like the Pr proteins of plants, have one CAP/PR domain. Fungal PRY proteins share a notable degree of amino acid sequence homology and almost identical domain structure homology with plant Pr and bacterial SCP. However, in many cases they contain additional inserted sequences that make many of these proteins larger than their plant counterparts. Although the function of proteins belonging to the PRY subfamily is not known, it is speculated that they may play a role in host defense similar to the Pr proteins of plants [6]. ClustalW alignment of PRY amino acid sequences in Figure 5 show 92% to 59% query coverage and between 44% and 35% identity (with an E value ranging from 3e-34 to 4e-24) (data not shown). Like plant Pr proteins they contain cysteines at conserved CAP/PR sites, although frequently four rather than six.

Common name - Protein name													•	P	, 	~ ^	نہ ۱	1 S		1.000	00													
Common name - Protein name	_				_									AI	IUID	0 A	ACIO	12	equ	en	ce													
Mushroom (M. roreri) - Pr-1	-	-	-	-	-	т	р	L	L	F	D	E	Δ	G	E	F	F	D	G	D	м	E	T	D	S	Δ	L	F	E	D	-	w	к	(
Yeast (S. cerevisiae) - PRY1	-	S	т	т										S																				
Yeast (C. glabrata) - PRY-like														-																				
Yeast (C. orthopsilosis) - PRY1														s																				
Yeast (L. maculans) - SCP-PRY-like	0													-																				
Mold (C. militaris) - PRY1	-													-																				
Wold (C. Humaris) - FRTT	-	3		ľ	-	v		AF		v	Â	r	A	-	-	-	-	1	r	5	Y	Б	F	.5	1	5	ĸ	ĸ	1	v	-	1.	1	-
Mushroom (M. roreri) - Pr-1	0	v	V	R	0	н				R	Α	R	Y	G	A	Р	-	Ν	L	s	w	s	D	A	L	Y	Р	D	Т	A	R	Y	-	
Yeast (S. cerevisiae) - PRY1														K																				1
Yeast (C. glabrata) - PRY-like														Q																				
Yeast (C. orthopsilosis) - PRY1														Q																				
Yeast (L. maculans) - SCP-PRY-like														N																				
Mold (C. militaris) - PRY1														N																				
		•	2	- 1	-	-		-	-	-		×		- 1			_	CA	_	_		- 1		-	2			-			~	-	-	1
Mushroom (M. roreri) - Pr-1	-	A	G	0	С	K	F	0	н	s	Ν	G	G	G	K	Y					A	A	G	Т	G	G	F	s	s	G	L	K	s	7
Yeast (S. cerevisiae) - PRY1				_				-						G																				
Yeast (C. glabrata) - PRY-like														G																				
Yeast (C. orthopsilosis) - PRY1														G																				
Yeast (L. maculans) - SCP-PRY-like														G																				
Mold (C. militaris) - PRY1														G																				
																-	-					_	_	AP				-						F
Mushroom (M. roreri) - Pr-1	Μ	D	Е	A	s	к	Y	D	Y	Ν	ĸ	Ρ	G	F	s	Т	A	Т	G	н	F	Т	Q	v	v	w	K	s	s	K	Q	v	A	¢
Yeast (S. cerevisiae) - PRY1	Υ	Ν	Е	Ι	s	Ν	Y	D	F	s	Ν	Ρ	G	F	s	s	Ν	Т	G	Н	F	Т	Q	v	v	w	K	s	Т	Т	Q	V	G	C
Yeast (C. glabrata) - PRY-like	Y	D	Е	Ι	K	D	Υ	Ν	Υ	A	Ν	Ρ	G	F	s	Е	s	Т	G	н	F	Т	Q	v	v	w	K	s	s	Т	Κ	V	G	C
Yeast (C. orthopsilosis) - PRY1	Υ	D	Е	G	Е	Т	Y	Ν	Y	Т	-	-	-	A	A	Ν	Е	Y	Ν	н	F	Т	Q	v	v	w	K	G	s	Т	K	V	G	ç
Yeast (L. maculans) - SCP-PRY-like	G	Е	Е	R	Е	Е	Y	Ν	F	Ν	G	G	Q	F	s	s	s	Т	G	н	F	Т	Q	L	v	w	K	Ν	Т	Т	Q	V	G	ç
Mold (C. militaris) - PRY1														F																				
																		CA																
Mushroom (M. roreri) - Pr-1	Α	Ι	A	Ν	С	R	G	G	Q	Ρ	s	Κ	Y	Ι	V	С	R	Y	Т	Р	Р	G	Ν	F	A	G	R	F	A	Е	Ν	V	G	I
Yeast (S. cerevisiae) - PRY1														v																				
Yeast (C. glabrata) - PRY-like														v																				
Yeast (C. orthopsilosis) - PRY1														I																				
Yeast (L. maculans) - SCP-PRY-like														L																				
Mold (C. militaris) - PRY1														v																				
	-			-		-					-		-				_	-		-	-	-			•	-	-	-	-	-	-		-	f
Mushroom (M. roreri) - Pr-1	P	R	R	L	Q	A	D																											t
Yeast (S. cerevisiae) - PRY1		A																																Γ
Yeast (C. glabrata) - PRY-like		K																																Γ
Yeast (C. orthopsilosis) - PRY1	-		-																															t
Yeast (L. maculans) - SCP-PRY-like	A	V																																F
Mold (C. militaris) - PRY1							R																											t

Figure 5. ClustalW Amino Acid Alignment of the CAP/PR Domain in Fungal **CAP Superfamily Proteins**



C Conserved cysteines

Figure 5. MEGA7 ClustalW amino acid sequence alignment of CAP/PR domaincontaining proteins in fungi. The CAP signature sequences are highlighted in orange. The highly conserved cysteine residues are highlighted in yellow with a red font color. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 5.

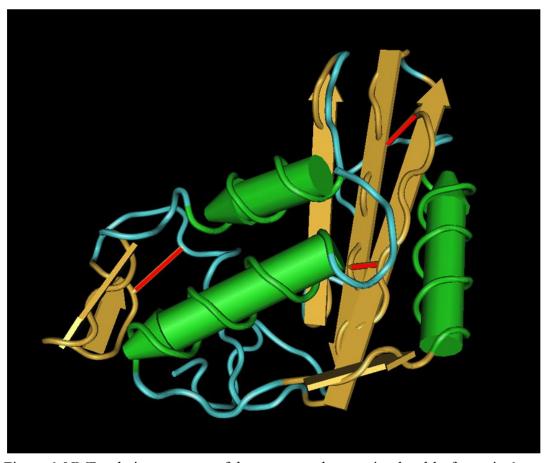


Figure 6. Tertiary Structure of the Tomato P14a Pathogenesis-Related Protein

Figure 6. NMR solution structure of the tomato pathogenesis related leaf protein 6 (accession ID: NP_001234314, PDB ID: 1CFE) [12]. The helices are represented by green rods, the β -strands are represented by tan arrows, and the polypeptide chain is represented in light blue. Additionally, the three disulfide bonds are represented in red. These disulfide bonds create a stable tertiary structure that is relatively heat and solvent resistant – a characteristic common to most CAP proteins.

CAP Proteins of Insects: the Venom Allergen Subfamily

Insect venom and saliva contain a series of antigenic proteins of which the best characterized are Allergen 5 (Ag5) and Allergen 3 (Ag3). They are highly immunogenic and are commonly associated with allergic response to insect bites [11, 16, 17]. These proteins are largely found in invertebrates and are rarely expressed in vertebrates. Ag5

and Ag-related (AgR) proteins have been identified in the saliva of ticks, sandflies, red fire ants, mosquitoes, honeybee, blood-feeding ticks, and many other species [18-22]. AgR genes are also expressed in the midgut of *Drosophila melanogaster* during late embryogenesis, larval, prepupal, and adult stages [23], although their function has not been identified. *Ancylostoma caninum* (hookworm) also expresses an AgR protein during its transition from the external, free-ranging stage to the parasitic stage in the host gut. Here, AgR is hypothesized to play an important role in inhibiting immune effector mechanisms by direct immunosuppression [24].

Sensitivity to insect allergen proteins can be extreme. The Ag5 proteins of Hymenopteran insects can elicit an allergic reaction in humans by elevation of both IgE and IgG antibody responses [11, 16, 17]. In many cases, a nanogram dose of insect venom can sensitize and provoke anaphylaxis [11, 17, 25].

The insect venom proteins, like the Pr proteins of plants, consist of only one domain – the CAP/PR domain (Figure 2) with typical CAP signature sequences (orange boxes, Figure 7). The proteins within this group are also highly homologous, 94% to 77% query coverage and between 63% and 48% identity (with an E value rage from 1e-90 to 6e-56) (data not shown). Likewise, their CAP/PR domain contains 6 highly conserved cysteines (yellow highlight, Figure 7) which are typically disulfide bond linked. Additional cysteines found near the N-terminal of some members are part of the signal sequence and therefore not found in the mature processed protein.

Common name - Protein name														Ar	nin	ь А	cid	l Se	qu	enc	e													
Parasitic wasp - Ag3	-	-	-	-	_	-	-	-	-	-	_	-	-	-	-	м	T	т	T.	S	ĸ	F	ĸ	C	_	_	-	-	_	-	-	-	-	_
Parasitic wasp - Ag5		-	-	м	Α	к	T	L	L	т	T									A												-	-	к
Jewel wasp - Ag5-like	Α																			A														
Jewel wasp - Ag3-like																				A														
Red fire ants - Ag3																				A														
Leafcutte bee - Ag3-like																				A														
Parasitic wasp - Ag3	-	-	-	-	-	-	-	E	S	w	Y	Р	s	s	S	C	G	Т	Ι	Q	s	N	Т	L	Т	N	E	E	Ι	Y	D	L	L	E
Parasitic wasp - Ag5	Т	Н	Т	М	С															Q														
Jewel wasp - Ag5-like																				Q														
Jewel wasp - Ag3-like																				Q														
Red fire ants - Ag3																				S														
Leafcutte bee - Ag3-like																				Ι														
				AI	_		1																							-				-
Parasitic wasp - Ag3	Ι	н	N	N	Κ	R	Α	F	V	A	s	G	S	Е	М	R	G	Т	Т	G	Р	0	Р	-	A	G	Ι	Ι	0	Ρ	L	0	W	D
Parasitic wasp - Ag5																				G														
Jewel wasp - Ag5-like																				G														
Jewel wasp - Ag3-like																				G														
Red fire ants - Ag3																				G														
Leafcutte bee - Ag3-like	A	н	N	Y	Y	R	А	R	V	Μ	Κ	G	E	Е	Κ	R	G	Ν	Ρ	G	Ρ	Q	Ρ	Ρ	A	s	Ν	Μ	Q	Κ	М	V	W	D
																														_	_	P4	-	
Parasitic wasp - Ag3	Y	Е	L	A	Κ	Ι	A	Q	R	W	A	D	Q	С	R	F	G	Η	D	Н	С	R	D	V	D	R	F	Р	V	G	Q	N	V	А
Parasitic wasp - Ag5																				K														
Jewel wasp - Ag5-like																				S														
Jewel wasp - Ag3-like																				-														
Red fire ants - Ag3																				A														
Leafcutte bee - Ag3-like																				s														
Parasitic wasp - Ag3	K	М	M	Н	s	A	G	Y	Т	Ι	K	L	s	D	L	V	Q	L	W	Y	D	E	v	Q	D	F	D	G	Q	s	v	Y	N	F
Parasitic wasp - Ag5																				Y														
Jewel wasp - Ag5-like																				Y														
Jewel wasp - Ag3-like																				Y														
Red fire ants - Ag3																				Y														
Leafcutte bee - Ag3-like																	E	Μ	W	Y														
	_												_				_	AP	-															
Parasitic wasp - Ag3																				W														
Parasitic wasp - Ag5																				W														
Jewel wasp - Ag5-like																				W														
Jewel wasp - Ag3-like																				W														
Red fire ants - Ag3	-	-																		W														
Leafcutte bee - Ag3-like	-	-	-	-	Т	G	Κ	D	L	S	Ν	V	G	Η	Y	Т	Q	L	V	W	A	Κ	S	Ν	R	L	G	С	G	Κ	Ι	Ι	Υ	Q

Figure 7. ClustalW Amino Acid Alignment of Insect Antigen-Related Proteins

CAP signature sequences

Sites of non-aligned sequence omission

Conserved cysteines

Figure 7.

Page 1 of 2

														CA	P2	2																
Parasitic wasp - Ag3																			NV													
Parasitic wasp - Ag5																			ΝV													
Jewel wasp - Ag5-like	D	-	G	Κ	F	Ν	Κ	F	Y	L	V	С	Ν	Y	G	P	S	G	ΝV	W	Ι	G	Е	Ρ	V	Y	Q	Т	R	-		
Jewel wasp - Ag3-like																			-													
Red fire ants - Ag3																			N													
Leafcutte bee - Ag3-like	Τ	-	D	A	W	K	Η	Υ	Y	V	V	С	Ν	Y	G	Р	G	G	N	V	L	Т	Q	Ρ	Ι	Υ	D	Ι	Κ	Κ		

Figure 7. ClustalW Amino Acid Alignment of Insect Antigen-Related Proteins

CAP signature sequences Sites of non-aligned sequence omission

C Conserved cysteines

Figure 7. MEGA7 ClustalW amino acid sequence alignment of antigen and antigen related proteins in insects (invertebrates). The CAP signature sequences are highlighted in orange. The eight highly conserved cysteines are highlighted in yellow with a red font color. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 6.

Page 2 of 2

The Cysteine-RIch Secretory Protein (CRISP) Subfamily

As represented in Figures 8 and 9 (and Figure 2), CRISP proteins have three domains – the CAP/PR domain at the N-terminal and the ion-channel regulatory (ICR) domain at the C-terminal (green, Figure 8), with a hinge region (blue, Figure 8) separating them. The Hinge region and the IRC domain together are sometimes referred to as a cysteine-rich domain (CRD) [26].

The CAP/PR domain is characterized by the usual four CAP signature sequences (orange, Figure 8) and there are sixteen conserved cysteine residues spanning the entire protein (yellow highlighted in Figure 8). ClustalW-aligned sequences of CRISP proteins from virtually all vertebrates share significant homology, with percent identity ranging from 40 to 90 and percent query cover ranging from 60 to 90 (Figure 8). Immediately obvious is the highly conserved positioning of the cysteines. As verified by the X-ray crystallographic structure of snake venom CRISP proteins, these conserved cysteines

form disulfide bonds (red lines, Figure 9) that not only stabilize the CRISP structure but also increase the likelihood that many CRISP protein have comparable tertiary structures. This likelihood is further increased by the fact that predicted alpha helix and beta sheet secondary structure is consistent over a wide range of CRISP protein sequences including the examples shown in Figure 10.

Surprisingly and almost completely unappreciated is the fact that CRISP-like proteins are also widely distributed throughout invertebrate phyla including worms, flies, mollusks and sea squirts (see Figure 11). This was discovered using NCBI's gene and protein database, Ensembl's database (http://uswest.ensembl.org/) and GeneBank. Like their vertebrate CRISP cousins, many of these proteins have CAP/PR domains (unshaded) and ICR domains (green) separated by a Hinge region (light blue). As expected, these invertebrate proteins have the four CAP signature sequences (orange) and six conserved cysteines in the CAP/PR domain and four and six conserved cysteines in the Hinge region and ICR domain, respectively.

The discovery of "vertebrate" CRISP and CRISP-like proteins in invertebrate phyla is contrary to current dogma and suggests that the origin of CAP proteins having multiple domains predated the divergence of vertebrates from invertebrates. This is not entirely surprising since even the most primitive of vertebrates such as lamprey, express CRISP proteins having the typical PR/Hinge/ICR domain organization like that of mammalian CRISP proteins (Figure 8).

18

Common name - Protein name														A	mir	10 A	Acid	1 Se	que	ence														
Lamprey - Unchar.	A	S	S	v	A	A	A	F	L	L	A	A	L	L	Р	A	G	А	S	A	W	D	L	v	Ν	D	W	R	s	L	ĸ	Т	V	Н
Spotted gar - Unchar.	F	Ľ	I	Ť	V	Н	V	F	L	F	S	L	P	L		ĸ		N			C	I	v	F	T	-	-	-	Ē	I	S	T	s	N
Zebrafish - CRISP3	-	-		H		Μ		Ā	L	V	Ī	Ĉ	F	L			L	Н				A	Ċ	S	v	S	G	-	-	V	C		Ē	L
Atlantic cod - CRISP3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I	č	P	T	D
Stickleback fish - CRISP3	-	-	-	-	-	М	F	Т	L	L	V	С	Ι	L		L				Н		Α	С	V	D	v	S	R	Ν	I	č		D	K
Tilapa fish- CRISP3	Т	-	L	Е	S	T	M		F	Y	Ť	F	P	L	L	C	Ā	-		L		A	V	Ļ	Q	v	P	G	T	L	G	V	T	T
Japanese rice fish - CRISP3	-	-	-	-	-	-		P	S	L	I	C	C	L		Ι				Y	-	-		-	~	-	-	-	-	I	C	P	Ē	D
Western clawed frog - CRISP2	-	М	L	Ι	G	Α	L	C	I	A	A	F	М		L	-	-	-		V	Е		A		Р	Р	F	S	s	I	s	Т	D	N
African clawed frog - CRISP2	M		L	I	A	V	L	G		М	T	L	M	I	Q	-	-			v	E	S	A	G	I	S	L	S	-	L	S	T	Y	I
Burmese python - CRISP-like		I	A	L	Ι	v	L	L	s	L	A	A	V	L	Q	-	-		S	s	G	-	-		D	G	F	A	A	E	S	Т	S	R
Anole lizard - Cysteine-rich P.	-	V		H		L	Y	I	C	V	I	I	Ľ	L	H	-	-	~		G			A		G	D	I	D	A	-	м	_	T	R
Softshell turtle - Unchar.	M		L	L	T	A	F	L	C	L	A	A	V	L	Q	-	-	-	T	T		Q	I		N		V	A	S	L	S	T	D	N
Falcon - Cysteine-rich P.	M		L	P	V	V	F	L	c	L	A	A		L	P	-	-	P	S	I	G	-	Ē		E	G	L	D	A	L	S	T	S	K
Golden manakin - CRISP2	M		L	P	V	V	F	L	c	L	T	A	V A	L	P	-	-	P	S	S	A		E	P A	G	G	V	D	A	L	S	T	S	R
	M		L	P	L	V	F	L	c	L	T			L	P	-	-					~			E						S			R
Woodpecker - CRSP2 Chielson Curtains rich P			L	P	A		F L	L	c	L		A	V					P		A		~	A		E	D	F	D	A	L	S	T T	S	-
Chicken - Cysteine-rich P.	M					V					A	A	V c	L	S	-	-			A	G	E	I			A	S	L	L	L			N	R
Platypus - CRISP1	I		C	L	T	V	A	A	I	L	P	I	S	I	L	- T	-	- T		V	G	P	P	R	V	P	Y	E	A	I	S	T	V	A
Armadillo - Cysteine-rich P.		D	L	L	R	L	V	V	F	L	A	A	I	L	L	T	S	L		A		G			P	G	F	S	A	L	L	T	T	Q
Mouse - CRISP2		A	W	F	Q		M		F	V	F	A	L	L	L	R	S		L	T		G			P	D	F	T	S	L	L		N	Q
Rat - CRISP2		A	W	F	Q	V	M	L	F	V	F	A	V	L	L	P	L		P	T		G			P	D	F	A	T	L	T		Ν	Q
Chimp - CRISP2		A	L	L	P	-	V	L	F	L	V	Т	V	L	L	P	S	L	P	A	E				P	A	F	Т	A	L	L	Т	Т	Q
Human - CRISP2		A		L	Р	-	V	L	F	L	V	Т	V	L	L	Ρ	S	-	Р	A		G				A	F	Т	A		L	Τ	Т	Q
Western clawed frog - Allurin	-	D	Т		Ν	F	Ι	Ι	С	Ι	S	A	L	F	Η	-	-	-	S	T				D		G	Τ		M		D		E	T
African clawed frog - Allurin		D	Т	-	Ν	F	Ι	L	С	Ι	S	A	V	Ι	Y	-	-	-	S	Т						Τ	G	S	Т	L	E		K	Т
Zebra Finch - Truncated CRISP	Μ	Ι	L	S	V	V	F	L	С	L	Т	Α	_		Р	-	-	Ρ	S	Т	G	Q	E	Т	E	Α	L	S	A	L	S	S	S	R
Lamprey - Unchar.	G	S	v	Q	Т	F	Ι	v	D	L	н	N			R	R	N	v	Т	Р	s	A	G	N	М	T	0	М	м	w	ĸ	ĸ	F	Α
Spotted gar - Unchar.	G	I		Q		E	I	v	K	_		N				R				P	_				M		_	M				T	E	A
Zebrafish - CRISP3	S	S	v	-	-	E	I	v	D	V		N		F	R		A			P		A			M			M		W		D		V
Atlantic cod - CRISP3	P	A	I	-	N	Q	I	v	D	A		N			R		A			P					M			M	I	W		N	n	V
Stickleback fish - CRISP3	P	A	I	Q	A	E	I	v I	D	V		N	A		R		A		-	г Р		A			M			M		Y	S		E	V
Tilapa fish- CRISP3	P T	T	E		N		I	V	N	-		N			R		N			P	T				M			M						
•	T	A	E V		A	E	I	v	D	N	Н	N	A		R		A		-	P	P					L			S	Y		E	E	A L
Japanese rice fish - CRISP3	-			Q						-								-										M			S		D	
Western clawed frog - CRISP2	V	T T	V V	T	Q	I	I	I V	N	A T		N			R		N N	A		P					M			M					~	A
African clawed frog - CRISP2	S			R	Q	T	I		D	-				Y	R					P			-		M			M					D	T
Burmese python - CRISP-like		E	K	-	K	E	I	V	D	H		N			R	R R		V		P	T T	A			M			M			H		D	A
Anole lizard - Cysteine-rich P.	A		-	~	K	E	I	V	D	K		N			R					P		A			M			M	E	W		N	E	
Softshell turtle - Unchar.	A		-	~	K	E	I	V		K		N		L	R		A			P	-				M			M			S	H	A	A
Falcon - Cysteine-rich P.		D	-	~	K		I	V	D	R		N	A		R		G			P	T				M			M			C	P	P	A
Golden manakin - CRISP2	A		-	~		L	I	V	D	R		N			R		G	-		P								M					R	A
Woodpecker - CRSP2			Q			L	Ι	V		K	Η	Ν		L	R	R		V		P		A			M			M		W			A	-
Chicken - Cysteine-rich P.		D	Q	~	K	L	I	V	D	K	H	N	A	L	R		R	V		P	P				M			M		W		P	Q	A
Platypus - CRISP1	-	R		Q	Q	E	Ι	V		K		Ν		L	R		L			Ρ		A						М					E	A
Armadillo - Cysteine-rich P.	Т	Q		~	K	E	Ι	V	Ν	K	Η	Ν	E	L	R		Ν			P		A			М			М		W			E	A
Mouse - CRISP2	L	Q	V	~	R	E	Ι	V		K	Η	Ν	Е	L	R	R				Р	Τ			D	Ι	L		Μ		W		Ι	Q	A
Rat - CRISP2	Ι	Q	V	Q	R	E	Ι	Ι	A	K	Η	Ν	Е	L	R		~			Р	Р			Ν	Ι	L		М			Ν	V	Q	A
Chimp - CRISP2	L	Q	V	~	R	E	Ι	V	Ν	Κ				L	R		A	V		Р	Р				М			Μ		W	S	R	E	V
Human - CRISP2	L	Q	V	Q	R	Ε	Ι	V	Ν	Κ	Η	Ν		L	R	Κ	A	V	S	Р	Р	А	S	Ν	М	L	Κ	Μ	E	W	S	R	Е	V
Western clawed frog - Allurin	-	-	-	Q	Ν	Y	L	V	D	L	Н	Ν	L	L	R	R	S	V	D	Р	Т	А	K	D	М	L	Κ	Μ	Е	W	S	Р	G	Α
African clawed frog-Allurin	-	-	-	Q	Ν	Κ	Ι	V	D	Η	H	N	Δ	L	R	R	S	v	D	Р	Т	Α	S	D	М	Κ	Κ	М	v	W	С	D	Р	Α
runcun duwed nog runum				ς.					~					-		1.	~	•	~	-	_								•	•••	~	~		

Figure 8. ClustalW Amino Acid Alignment of Vertebrate CRISP2 and CRISP2like Proteins

CAP signature sequences

ICR domain

Sites of non-aligned sequence omission

Hinge region

C Conserved cysteines

Figure 8. Page 1 of 5

Figure 8. ClustalW Amino Acid Alignment of Vertebrate CRISP2 and CRISP2like Proteins

Spectrodigar - Unchar. A A A A Q R W I N																															CA	P4	٦		
Sponted gar - Unchar. A A A A Q R W Y P P S S R I I Y Zdorafish - CRISP3 A A L S Q A W L M L H G P S S R M L N G F L G E N L H G P S S R H A Q A W D C S M H A Q A W D C V L H A D C S M H A S N A N A N A N A N A N A N A N A N A N A A N A A A A A	Lamprey - Unchar.	Y	L	Ν	A	L	s	W	A	Ν	Ν	С	Т	F	G	Н	S	Р	D	A	R	R	Κ	Т	Т	Q	W	Т	С				Ι	F	М
Zebrafia CRUSP3 A B A M M L M L M L M L M L M L M L M L M L M L M L M L M L M L M <	Spotted gar - Unchar.	Α	Α	Ν	Α	0	R	W	А	0	Т	С	S	М	Ν	Н	S	Р	D	S	S	R	R	Ι	Κ	-	S	Е	С	G	Е	Ν	L	Y	Μ
Admic od CRISP3 A																													Μ			Ν			Κ
ScickleachtshCRISP3 A A A D C V L A G P S T N C V L A V L C S M G P S S N N C C S N H N N C N <td>Atlantic cod - CRISP3</td> <td>Α</td> <td>Α</td> <td>L</td> <td>S</td> <td>0</td> <td>Α</td> <td>W</td> <td>L</td> <td>D</td> <td>Н</td> <td>С</td> <td>М</td> <td>L</td> <td></td> <td>Н</td> <td>G</td> <td>Р</td> <td>Р</td> <td>S</td> <td>S</td> <td>R</td> <td>Μ</td> <td>L</td> <td>Ν</td> <td>G</td> <td>Y</td> <td>Е</td> <td>L</td> <td>G</td> <td>Е</td> <td>Ν</td> <td>L</td> <td>F</td> <td>Y</td>	Atlantic cod - CRISP3	Α	Α	L	S	0	Α	W	L	D	Н	С	М	L		Н	G	Р	Р	S	S	R	Μ	L	Ν	G	Y	Е	L	G	Е	Ν	L	F	Y
Tabas b. CRISP3 E A I A Q R M S S R M I S G C G E N L Y S A Q S N M I N A C S S S N M I N A C S S S N M I N A N	Stickleback fish - CRISP3	Α	Α	s		-	Α	W	Α	D	Ν	С	V	L					Р	s	Т	R	Μ	L	Ν	G	Υ	Е	L	G	Е	Ν	L	F	Y
Japanee racefish - CRISP3 A V S A Q A W V D K C V L D H G P P S S R M L N G Y E L G E N L F Westem clawed frog - CRISP2 A I N A A S W A A T C S E F H S P S N Q R T I P G F G C G E N L Y Barnees python - CRISP1AEA A K N A K R W A D K C S F A H S P V N Q R T V Q E L Q C G E N L Y Schedul turle - Unchar. A K N A K R W A D K C S F A H S P V N Q R T V Q E L Q C G E N L Y Schedul turle - Unchar. A K N A K R W A N E C T I S H S P P A E R S T . S N G C G E N L Y Schedul turle - Unchar. A K N A K R W A N E C T I L S H S P P A E R S T . S N G C G E N L Y Schedul turle - Unchar. A K N A Q K W A N R C T T K H S P P A E R S T . S N G C G E N L Y Schedul turle - Unchar. A K N A Q K W A N R C T I L S H S P P A E R S T . S N G C G E N L Y Schedul turle - Unchar. A K N A Q K W A N Q C T L S H S P P A E R S T . T V G C G E N L Y Falcon - CRISP2 A N A Q K W A N Q C T L S H S P P A E R S T . T V G C G E N L Y Patypes - CRISP1 A V N A Q K W A N R C G S L S H S P P N R R T . N V L C G E N L Y Patypes - CRISP1 A Q N A R Y W A E K C E V Q H S P Q P G R Q I K I . N I K C G E N L Y Patypes - CRISP2 A N A A Q K W A N K C T L L H S S K D D R K I . N I K C G E N L Y Patypes - CRISP2 A A N A Q K W A N K C T L Q H S D Q P G R Q I K I . N I K C G E N L Y Patypes - CRISP2 A A N A Q K W A N K C T L Q H S D P E D R K I . N I K C G E N L Y Human - CRISP2 T T N A Q R W A N K C T L Q H S D P E D R K I . N I K C G E N L Y Human - CRISP2 T T N A Q R W A N K C T L Q H S D P E D R K I . N I K C G E N L Y Human - CRISP2 T T N A Q N W A N Q C Y A A K C V M Q H S S A T E R Q I Q S S T T C G G E N L Y Human - CRISP2 T T N A Q N W A N Q C Y A N K C T L Q H S D P E D R K I . N I K C G E N L Y Human - CRISP2 T T N A Q N W A N Q C Y A N K C T L Q H S D P E D R K I . N I K C G G E N L Y Human - CRISP3 S T D N A Q N W A N Q C Y V Q W Y N E V S N Y K Y P I G S G N G Q . N Spoteding - Allumin A L N A Y N W A N A Q C Y N A N Y C Y N W H S L V L R N I K N V V C G G S N I Y Cobrind- CRISP3 S T G N A Q N W A N Q C Y V Q Q W Y N E V I S R Y R Y R Y R G S G N G Q . N A	Tilapa fish- CRISP3	Ε										С										R	V	Ι				G	C	G	Е	Ν	L	Υ	Υ
Westem clawwed frog - CRISP2 A C C T F H S P N A C C T F S F F F F F T I C G C F F F P N A C G C T F F F N A A A A A A A A A A A A A A A A A <td< td=""><td>•</td><td>Α</td><td>V</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>Κ</td><td>С</td><td>V</td><td>L</td><td></td><td></td><td></td><td></td><td>Р</td><td></td><td></td><td>R</td><td>Μ</td><td>L</td><td>Ν</td><td></td><td></td><td>Е</td><td>L</td><td>G</td><td>Е</td><td>Ν</td><td>L</td><td>F</td><td>Y</td></td<>	•	Α	V								Κ	С	V	L					Р			R	Μ	L	Ν			Е	L	G	Е	Ν	L	F	Y
Adjican clawed forg CRISP2 A A K N A G N N A K N A K N A K N A K N A K N A K N A K N A K N A K N A K N A K N A K N A K N A K N A K N A K N A C T L S S R I N A Q K N A C C T L S S R N R K L N L N A L N N Q K N N Q K N N Q N N Q N N Q N N N N N N N N N N N N N </td <td>Westem clawed frog - CRISP2</td> <td>Α</td> <td>Ι</td> <td>Ν</td> <td></td> <td>-</td> <td>S</td> <td>W</td> <td>А</td> <td>А</td> <td>Т</td> <td>С</td> <td>S</td> <td>Е</td> <td>S</td> <td>Η</td> <td>S</td> <td>Р</td> <td>S</td> <td>D</td> <td>Κ</td> <td>R</td> <td>Т</td> <td>Ι</td> <td>Р</td> <td>G</td> <td>F</td> <td>G</td> <td>С</td> <td>G</td> <td>Е</td> <td>Ν</td> <td>L</td> <td>Y</td> <td>М</td>	Westem clawed frog - CRISP2	Α	Ι	Ν		-	S	W	А	А	Т	С	S	Е	S	Η	S	Р	S	D	Κ	R	Т	Ι	Р	G	F	G	С	G	Е	Ν	L	Y	М
Barmese pyhon - CRISP-ike A K R A K R A K K A K R A K K A K K A K <td>-</td> <td>Α</td> <td>0</td> <td>Ν</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>Α</td> <td></td> <td>С</td> <td></td> <td>F</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>Ν</td> <td>0</td> <td>R</td> <td>Т</td> <td>Ι</td> <td>Р</td> <td>G</td> <td>F</td> <td>Ν</td> <td>C</td> <td>G</td> <td>Е</td> <td>Ν</td> <td></td> <td>Υ</td> <td></td>	-	Α	0	Ν						Α		С		F						Ν	0	R	Т	Ι	Р	G	F	Ν	C	G	Е	Ν		Υ	
Anole Izrad - Cysteine-rich P A K N N R C T F K N S S R I I N G C T L S N P A F N N V Y L S N N N C T L S N N N V N V N V N V N V N V N V N V N V N N N V N N N V N N N V N N N N V N <t< td=""><td>-</td><td>Α</td><td>ĸ</td><td>Ν</td><td>Α</td><td>к</td><td>R</td><td>W</td><td>А</td><td>D</td><td>Κ</td><td>С</td><td>S</td><td>F</td><td>Α</td><td>Н</td><td>S</td><td>Р</td><td>V</td><td>Ν</td><td>ò</td><td>R</td><td>Т</td><td>V</td><td>G</td><td>Е</td><td>L</td><td>0</td><td>C</td><td>G</td><td>Е</td><td>Ν</td><td>L</td><td>F</td><td>М</td></t<>	-	Α	ĸ	Ν	Α	к	R	W	А	D	Κ	С	S	F	Α	Н	S	Р	V	Ν	ò	R	Т	V	G	Е	L	0	C	G	Е	Ν	L	F	М
Softshell turde - Unchar. A K N A K N A C T L S T S T C C G C G C G C G L S N N C C T L S N N N Q C T L S N N N Q C N N N Q Q N N N Q Q N N N Q Q N N N Q N N Q N N Q N N Q N N Q N N Q N N Q N N N Q N<		Α	K									С									-	R		Ι				-	С	G	Е	Ν	L	Y	М
Falcon - Cysteine-rich P. A T N Q C T L S H S P A S M R T T T Q C G I F Golden manakin - CRISP2 A N A Q K M N Q C A H S P D M R T - D V C G E N N Q V N Q N N N Q N N N Q S L N N R V N N Q S L N N N L L N	-	Α	K			_													Р				S	Т	-	S		G	C	G	Е	Ν			М
Golden manakin - CRISP2 A E N A Q K W N Q C T L S H S P P N L R R T - D V L C G E N I F N N A Q N N N Q V N <td>Falcon - Cysteine-rich P.</td> <td>Α</td> <td>Т</td> <td>Ν</td> <td>Α</td> <td>0</td> <td>Е</td> <td>W</td> <td>А</td> <td>Ν</td> <td>0</td> <td>С</td> <td>Т</td> <td>L</td> <td></td> <td>Н</td> <td>S</td> <td>Р</td> <td>Α</td> <td>S</td> <td>М</td> <td>R</td> <td></td> <td>Т</td> <td>-</td> <td>Т</td> <td>Т</td> <td>0</td> <td>C</td> <td>G</td> <td>Е</td> <td>Ν</td> <td>L</td> <td>F</td> <td>М</td>	Falcon - Cysteine-rich P.	Α	Т	Ν	Α	0	Е	W	А	Ν	0	С	Т	L		Н	S	Р	Α	S	М	R		Т	-	Т	Т	0	C	G	Е	Ν	L	F	М
Woodpecker - CRSP2 A N N A Q K W A D R C A L R H S T P D M R R T T V G C G E N L F V D M R R D K R T T V G C G E N L C G P N Q R R R R R R R R R R R N C G G N V I L R S N N R I N	•	Α	Е			-					-			L					Р						-		v	-	С	G	Е	Ν			М
Chicken - Cysteine-rich P. A V N A Q N A N Q C S H S P N Q R I G Q P C G Q P N Q R V N A N A Q N N N C C I L E N S V D N L C G D N L Q N		-																							-										Μ
Platypus - CRISP1 A Q N A R Y W A E K C E V Q H S P Q I K D H S C G E N L L Amadialo Cysteme-rich P. A N A Q K N N C I L E H S K D R K T N N C G E H S S T D R K C G E N N C I L E H S S T E D R T N I C G E N I I C G E N I N I N N N I I R N N I I N N I N N N I N N N I N I	•																								-									Y	
Annaddlo - Cysteine-rich P. A A N A V V N N C T L E H S D R T I N T C G E N D D R T I N T C G E N L U N D D R T N T N A Q N N C T L E H S S T D R T N T C G E N L N A Q N N C T L L N A Q N A Q N C T L Q N N C T L Q N N C C D N N N N N N N N N N N N N N N N N N											-																-								F
Mouse - CRISP2 T T N A Q K V A N C I L E H S S K D R K I R C G E N L Y Y R C G E N L V I L E D R K I N I K C G E N I V C G E N I N I N C G I L N A Q N N C T L R N N C I L R N																																		Y	
Rat - CRISP2 A A A Q K W A N C I L E H S S T E D R K I - N I K C G B N I V I V I V I V I V I V I V I V I V V I V V I V V V I V V V I V <t< td=""><td></td><td></td><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>Y</td><td></td></t<>						-																			-									Y	
Chimp - CRISP2 T T N A Q R W A N K C T L Q H S D P E D R K T S T R C G E N K T L Q N A Q R N N C C T L Q N N N C T L Q N N N K T L Q N		Α				-																							C					Y	
Human - CRISP2 T T N A Q R W A N K C T L Q H S D P E D R K T R C G E N A T C Q N A A K C V U S A T R C I N V C G E N I Y V C G E N I X V V C G E N I K D V V E V I R R I K D V V C G E N I K N V C I N V C I I N V I N I I I I N I I I N I I I N I I I I						~																			-								_	Y	
Westem clawed frog - Allurin A L N A Q N A A K C V M Q N Y V C G E N V V C G E N V V C G E N I V V V C G E N N V V C G E N N V V C G E N I V V C G E N I V V V C G E N V V C I I N V V V C G V V C G V V C G V V C G V V C G V <th< td=""><td>•</td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>_</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>Y</td><td></td></th<>	•	-													_										-									Y	
African clawed frog- Allurin A L N A Y N F A T Q C S M Y E R R I I K D V V C G E N I Y P I S V V I C G E N I V V C G E N I Y I V V I V V I V V I V V I V V I V V I V <th< td=""><td></td><td></td><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>_</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>V</td></th<>						-																				_									V
Zebra Finch - Truncated CRISP A E N A Q N A N Q C T L S P P L R R T N V L C G E N I F Q N V L C N V L C N V L C G E N I N V I S N V I N V I N V I N V I N V I N V V I N V I N V I N V I N V I N V V N V V N V V N V V I N V <t< td=""><td>-</td><td>-</td><td></td><td></td><td></td><td>_</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>~</td><td></td><td></td><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td>Y</td><td></td></t<>	-	-				_																	~						-					Y	
Lamprey - Unchar. S S A P F T W N Y N E V I S F Q Y D V G S T T T G - A Spotted gar - Unchar. S T D P S T W N E V I S W N K Y N G S V N G S V N G S V N G S V N G S V N G S V N G S V N G S V N G S V N G S V N G S V N A W N W N V N M W W N V N M N V N M N K N G N S V N <td>-</td> <td></td> <td>M</td>	-																																		M
Spotted gar - Unchar. S T D P S T W N E V I V S N Y K Y G V G S V N G Q S V V D A W H S V V A W H S V V N V K Y P I G S V D A Attantic cod - CRISP3 S S A L D W A V I A W H E V Q H Y Y P N G S I A W H V V A N V V V N N V V V N V V V V N V V V N V V N V V V N V N V N N V N			-							-		-	-	_	-			-	-		-			-				_		-	_		-	-	
Zebrafish - CRISP3 A T G I S S W T S V V D A W H S E V N N Y K Y P I G S I N G Q A A U V I T A W H N E V Q H Y P N G S G N A V V N A W H N E V Q H Y P N G S I N A W V N N V L Y P N G S I N N V N N V D F V D N G S I N A V N N N N N N N N N N N N N N N N	Lamprey - Unchar.	S	S	A	Р	F	Т	W	Ν	Y	A	Ι	R	D	W	Y	Ν	Е	V	Ι	S	F	Q	Y	D	V	G	S	Т	Т	Т	G	-	A	V
Atantic cod - CRISP3 S A L D W A T V I T A W H N E V Q H Y T Y P N G S G N A V V N A W H N E V Q H Y T Y P N G S G N A K V N A W H N V L V P N G S T N A V V A N V L V V N A V V D F S V G S T N A V V D F S V G S T N V V V N V N V N V N V N V N V N N N N N	Spotted gar - Unchar.	S	Т	D	Р	S	Т	W	Ν	Е	V	Ι	Q	S	W	Υ	Ν	Е	V	s	Ν	Υ	Κ	Y	G	V	G	S	V	Ν	G	G	-	V	Ι
Stickleback fish - CRISP3 S S S P Y P W T A V V N A W H S E V A N Y L Y P N G S T N A V V N A W H S E V A N Y L Y P N G S T N A V V D E V K D F S Y G S T N G G V N A W V V N A W V N A V V N A V V N D V G S T N G S T N G S T N D V S V N D V D V N D V D D D D D<	Zebrafish - CRISP3	А	Т	G	Ι	S	S	W	Т	S	V	V	D	А	W	Η	S	Е	V	Ν	Ν	Υ	Κ	Y	Р	Ι	G	S	Ι	Ν	G	Q	-	Α	Т
Tilapa fish- CRISP3 S S F Q N T W S N V I Q S W V D E V K D F S Y G G S V G G V V Japanese rice fish - CRISP3 S T S P M L W K S I I N A W Y K E V K D F Q V G V G V C V K D F Q V G V <th< td=""><td>Atlantic cod - CRISP3</td><td>S</td><td>Т</td><td>s</td><td>Α</td><td>L</td><td>D</td><td>W</td><td>Α</td><td>Т</td><td>V</td><td>Ι</td><td>Т</td><td>Α</td><td>W</td><td>Η</td><td>Ν</td><td>Е</td><td>V</td><td>Q</td><td>Η</td><td>Υ</td><td>Т</td><td>Y</td><td>Р</td><td>Ν</td><td>G</td><td>S</td><td>G</td><td>Ν</td><td>G</td><td>Ε</td><td>-</td><td>s</td><td>V</td></th<>	Atlantic cod - CRISP3	S	Т	s	Α	L	D	W	Α	Т	V	Ι	Т	Α	W	Η	Ν	Е	V	Q	Η	Υ	Т	Y	Р	Ν	G	S	G	Ν	G	Ε	-	s	V
Japanese rice fish - CRISP3 S T S P M L W K S I I N A W Y K E V S H F A F P N A S T N G K - P Westem clawed frog - CRISP2 A S Y P A S W E A V K G W Y S E F N M E W D F Q V G D P V Q D P V V Q V V Q S W V D E V D F N M V Q D V V Q S W V D E V D F N M V Q D V	Stickleback fish - CRISP3	S	s	s	Р	Υ	Р	W	Т	Α	V	V	Ν	А	W	Η	s	Е	V	Α	Ν	Υ	L	Y	Р	Ν	G	S	Т	Ν	Α	Κ	-	Т	V
Westem clawed frog - CRISP2 A S Y P A S W E A V K G W Y S E Y N D F Q Y G P K S P G L V African clawed frog - CRISP2 A T Y P A S W E A I Q A W Y S E G E F N M E W D F K D F K Y G V G D P L L P G D P V <td>Tilapa fish- CRISP3</td> <td>S</td> <td>S</td> <td>F</td> <td>Q</td> <td>Ν</td> <td>Т</td> <td>W</td> <td>S</td> <td>Ν</td> <td>V</td> <td>Ι</td> <td>Q</td> <td>S</td> <td>W</td> <td>Υ</td> <td>D</td> <td>Е</td> <td>V</td> <td>Κ</td> <td>D</td> <td>F</td> <td>S</td> <td>Y</td> <td>G</td> <td>V</td> <td>G</td> <td>S</td> <td>Ι</td> <td>Ν</td> <td>G</td> <td>G</td> <td>-</td> <td>V</td> <td>V</td>	Tilapa fish- CRISP3	S	S	F	Q	Ν	Т	W	S	Ν	V	Ι	Q	S	W	Υ	D	Е	V	Κ	D	F	S	Y	G	V	G	S	Ι	Ν	G	G	-	V	V
African clawed frog - CRISP2 A T Y P A S W E A I Q A W Y S E G E F N M E W D P L L P G D P Burmese python - CRISP-like S T H P Y S W T Q S W Y D E R K D F K Y G P I Q P N A V Anole lizard - Cysteine-rich P. S T A P N P V S D A V Q S W V E K K D F I Q A V Q S W V D E K D F K W Q D A V Q S V V D E K D F K W	Japanese rice fish - CRISP3	S	Т	s	Р	Μ	L	W	Κ	s	Ι	Ι	Ν	Α	W	Υ	Κ	Е	V	s	Η	F	Α	F	Р	Ν	Α	S	Т	Ν	G	Κ	-	Ρ	V
Burnese python - CRISP-like S T H P Y S W T A V T Q S W Y Q S W Y Q S W Y Q S W Y Q S W Y Q S W Y Q S W Y Q S W Y Q S W Y Q S W Y Q S W Y D E V D F E H G V G A V Q D V I Q S V Q D V I Q S V U D E V D E R K D F K V Q D A V D E V U D E V D E R K D F K V G A I Q	Westem clawed frog - CRISP2	Α	s	Υ	Р	Α	S	W	Е	Е	А	V	Κ	G	W	Υ	s	Е	Υ	Ν	D	F	Q	Y	G	V	G	Р	Κ	s	Р	G	L	V	Т
Anole lizard - Cysteine-rich P. S T A P N P N S D A V Q S W Y D E V H D F E H G Y Q D A V Softshell turtle - Unchar. S T K P H S W S D A I Q S W Y S E V H G Y G A I W Q D A V Golden manakin - CRISP2 S T A P F S W S D V I Q A W V D E K N F K T G A K T P D A V Q I Q I Q I Q I Q I Q I Q I Q I I I I I I	African clawed frog - CRISP2	Α	Т	Υ	Р	Α	S	W	Е	Е	А	Ι	Q	А	W	Υ	S	Е	G	Е	F	F	Ν	Μ	Е	W	D	Р	L	L	Р	G	D	Ρ	Q
Softshell turtle - Unchar. S T K P H S W S D A I Q S W Y S E V F V G T G P T T P D A V Falcon - Cysteine-rich P. S S A P F S W P D V I Q S W V D E K E D F K Y G T G A V G A V D V I Q S W V D E V D F K V D A V D E V I G A V D I Q N V I E V G T G A V D I V I Q I Q I Q I I Q I I D	Burmese python - CRISP-like	S	Т	Η	Ρ	Υ	S	W	Т	Α	V	Т	Q	S	W	Y	D	Е	R	Κ	D	F	Κ	Y	G	V	G	Р	Ι	Q	Ρ	Ν	Α	V	Т
Falcon - Cysteine-rich P. S S A P F S W P D V I Q S W P D V I Q S W P D V I Q S W P D V I Q S W P D V I Q S W P D V I Q S W P D V I Q S W P D V I Q Q W V D E K K Y G T G A V Golden manakin - CRISP2 S S A P F S W S D A I Q S V D E K N F K Y G A K T P G A N N D E K N F K Y G A	Anole lizard - Cysteine-rich P.	S	Т	Α	Р	Ν	Р	W	s	D	А	V	Q	S	W	Υ	D	Е	V	Η	D	F	Е	Η	G	Υ	G	А	Ι	W	Q	D	Α	V	V
Golden manakin - CRISP2 S T A P F S W S D V L Q A W Y D E E K N F K Y G T G A M M M Y D E E K N F K Y G T G A M M W D E E K N F K Y G T G A M M V D E V N V F E Y G T K G A M V D E V N V F E Y G V T A N A V Chicken - Cysteine-rich P. S L S W S T G I Q Y B S K N F K Y G A T K R G	Softshell turtle - Unchar.	S	Т	Κ	Р	Н	S	W	S	D	Α	Ι	Q	s	W	Υ	S	Е	V	Е	Ν	F	V	Y	G	Т	G	Р	Т	Т	Р	D	Α	V	Ι
Woodpecker - CRSP2 S S A P F P W S D A I Q S W V D E V N V F E Y G A V P G A V G A V F E V G V G A V Chicken - Cysteine-rich P. S T A P S S V S D S I Q A W F D E K D F K V G A V P G A V A V D E V N V F E Y G V G A V D E V N V F E V G V G A V V I I V I I V I I V I V I V I I V	Falcon - Cysteine-rich P.	S	s	Α	Ρ	F	S	W	Ρ	D	V	Ι	Q	S	W	Υ	D	Е	Κ	Е	D	F	Κ	Y	G	Т	G	Α	Κ	Т	Ρ	G	Α	V	Ι
Woodpecker - CRSP2 S S A P F P W S D A I Q S W V D E V N V F E Y G A V P G A V G A V F P V S D A I Q S W V D E V N V F E Y G V T P G A V Chicken - Cysteine-rich P. S S L S V S D A I Q I W V F E K Y G A T A N A V I I V I I K D F K D F K D G A D I I I I I I I I I I I I I I <th< td=""><td>Golden manakin - CRISP2</td><td>S</td><td>Т</td><td>Α</td><td>Р</td><td>F</td><td>S</td><td>W</td><td>s</td><td>D</td><td>V</td><td>L</td><td>Q</td><td>А</td><td>W</td><td>Υ</td><td>D</td><td>Е</td><td>Е</td><td>Κ</td><td>Ν</td><td>F</td><td>Κ</td><td>Y</td><td>G</td><td>Т</td><td>G</td><td>Α</td><td>Κ</td><td>Т</td><td>Κ</td><td>G</td><td>Α</td><td>М</td><td>Ι</td></th<>	Golden manakin - CRISP2	S	Т	Α	Р	F	S	W	s	D	V	L	Q	А	W	Υ	D	Е	Е	Κ	Ν	F	Κ	Y	G	Т	G	Α	Κ	Т	Κ	G	Α	М	Ι
Platypus - CRISP1 A T Y S L S W S D A I Q I W Y N E S K N F Q Y G A T K R G A M Amadilo - Cysteine-rich P. S S D P S S W S T G I Q V V E S K N F Q Y G A T K R G A M M M V	Woodpecker - CRSP2	S	s	Α	Р	F	Р	W	S	D	Α	Ι		S	W			Е	V	Ν	V	F	Е	Y	G	V	G	Е	Κ	Т	Р	G	Α	V	F
Platypus - CRISP1 A T Y S L S W S D A I Q I W Y N E S K N F Q Y G A T K R G A M Amadilo - Cysteine-rich P. S S D P S S T G I Q S S K N F Q Y G A T K R G A M Mouse - CRISP2 S T D P T L W S T V I Q S K N E N E N E N E N E N E N E N E N E N E N E N E N E N N N N N N N N N N N N N N N<	Chicken - Cysteine-rich P.	S	Т	Α	Р	S	S	W	S	D	S	Ι	0	А	W	F	D	Е	Е	Κ	D	F	Κ	Y	G	S	G	Α	Т	Т	Α	Ν	Α	V	Ι
Amadillo - Cysteine -rich P. S S D P S S V S T G I Q S V P N S N F I Y G V G P S S V V Mouse - CRISP2 S T D P I L W S T V I Q S V V V M S D V V I Q S W Y I N E N N N N N N N N N N N N N N N N	-	А		Υ	s	L	S	W	S	D			-																					М	V
	Armadillo - Cysteine-rich P.	S	S	D	Р	S	S	W	S	Т	G	Ι		S	W	Υ	D	Е	Ν	G	Ν	F	I	Y	G	V	G	Р	Κ	S	S	D	V	V	V
Rat - CRISP2 S T D P T S W R T V I Q S W Y E N F V F G V G A K - P N S A Chimp - CRISP2 S S D P T S W S S A I Q S W Y E N E N F V G V G A K - P N S A Chimp - CRISP2 S S D P T S W S S L D F V G V G N A N D I L D F V G N D N A N N S L D F V G N D N N D V I N D N N D N<		-		D	Р																														V
Chimp - CRISP2 S S D T S W S S A I Q S W D E S L D F V G V G P K S P N A V Human - CRISP2 S S D P I S W S S A I Q S W D E S L D F V G V G P N A V Human - CRISP2 S S D P S S S A I Q S W D E I L D F V G V G P N A V Western clawed frog - Allurin T T A K P D W A V N S V <th< td=""><td></td><td>_</td><td></td><td>D</td><td>Р</td><td></td><td></td><td>W</td><td></td><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td>V</td></th<>		_		D	Р			W					-																	-					V
Human - CRISP2 S S D P T S W S S A I Q S W Y D E I L D F V Y G V G P K S P N A V Western clawed frog - Allurin T T A K P D W A A V N S W F N D F T Y G V G P K S P N A V Western clawed frog - Allurin T T A K P D K A V N S W F N D F T Y G V G P N S D K - M Western clawed frog - Allurin T T A P D N N N D F T Y G V G N																																			V
Westem clawed frog - Allurin T T A K P D W A A A V N S W F N E R N D F T Y G V G P N S D K - M	•																																		v
		-											-																						İ
		-																																	I
Zebra Finch - Truncated CRISP S S A P F S W S D V L Q A W Y N E E K N F E Y G T G A K K K G A M	•																																		

CAP signature sequences

ICR domain

C Conserved cysteines

Sites of non-aligned sequence omission

Hinge region

Figure 8. Page 2 of 5

Figure 8. ClustalW Amino Acid Alignment of Vertebrate	CRISP2 and CRISP2-
like Proteins	

			CAP	1																	Г		CAP2	2
Lamprey - Unchar.	GΗ	Y 1	QL		VΥ	N	S F	Q	LC	Э <mark>С</mark>	A	N N	Y	С	Р	ĸΝ	G	V	A	K	Y			с Н
Spotted gar - Unchar.	DH	ΥΊ			VΥ	R	S N		ΙC		A١			С		N S	_	-				FΥ	v	сн
Zebrafish - CRISP3	GН	ΥΊ		v v	VΥ		S Y		VO		A١			С				-				FΥ	G	сн
Atlantic cod - CRISP3	GH	ΥΊ	οv	v v	V N	S	SW	0	VO	ЭC	GΝ	1 T		С	G		-	-	S	Κ	Υ	FΥ	G	<mark>c</mark> q
Stickleback fish - CRISP3	GH	Y 1	νõτ	v v	V N	S	SY	K	VC	ЭC	GN		L	С	Р	N -	-	-	G	Ι	Υ	FΥ	A	сн
Tilapa fish- CRISP3	GH	Y 1	νδι	v v	VΥ	R	S N	Q	ΙC	ЭC	ΑN	ſΑ	Υ	С	Р	N S	5 A	-	Υ	Κ	Υ	FΥ	v	сн
Japanese rice fish - CRISP3	GH	ΥΊ	νõτ	v v	V N	S	SY	~	VO		G١		L	С		N -		-	Ν			LΥ	G	сн
Western clawed frog - CRISP2	G H	Y 1	Q V	МV	VΥ	N	S Y	M	VC	ЪC	S V	/ S	Υ	С	Р	К -	S	Р	Y	Κ	Υ	FΥ	v v	<mark>c</mark> q
African clawed frog - CRISP2	GH	Y 1	QL	мv	VΥ	N	S Y	M	VC	ЪC	A١	/ S	Υ	С	Ρ	A -	N	Y	Y	Q	Υ	LΥ	v v	c q
Burmese python - CRISP-like	GH	Y 1	Q V	v v	VΥ	ĸ	S R	L	LC	ЭC	A	A	Κ	С	Р	AS	K	-	F	ĸ	Υ	ΥY	v v	СĤ
Anole lizard - Cysteine-rich P.	G H	Y 1	QV	v v	VΥ	R	SW	Κ	VC	ЪC	GI	A	R	С	Р	N K	R	-	Y	E	Ν	ΥY	v v	<mark>c</mark> q
Softshell turtle - Unchar.	G H	ΥΊ	Q V	v v	VΥ	Т	S Y	Q	VC	ЪС	A	A	V	С	L	E S	E	-	F	Ν	Υ	ΥY	v v	СН
Falcon - Cysteine-rich P.	GΗ	ΥΊ	QМ	V V	VΥ	N	S Y	Q	ΙC	ЭC	A١	/ A	Υ	С	Р	N S	R	-	Υ	Κ	Υ	FΥ	v v	<mark>c</mark> q
Golden manakin - CRISP2	G H	ΥΊ	QM	V V	VН	N	S Y	K	VA	A C	GI	. Α	S	С	S	N S	R	-	F	Κ	Υ	FΥ	v v	<mark>c</mark> Q
Woodpecker - CRSP2	G H	ΥΊ	Q V	v v	VН	S	S Y	L	LC	э <mark>С</mark>	G V	Λ	F	С	Ρ	NE	K	-	Υ	R	Υ	FΥ	v v	СН
Chicken - Cysteine-rich P.	G H	ΥΊ	QL	V V	VΥ	N	S Y	Q	VC	ЭC	A١	/ A	Υ	С	Ρ	ΕR	τ	-	F	K	Υ	FΥ	v v	СH
Platypus - CRISP1	S H	ΥΊ	QL	V V	v s	Т	S Y	Q	ΙC	ЭC	G A	A	Ν	С	Р	ΚE	K	S	F	Q	V	LF	I	<mark>c</mark> q
Armadillo - Cysteine-rich P.	G H	Y 1	QV	v v	VF	S	S Y	Ľ	VC	ЪC	G V	/ A	Υ	С	Р	NQ) E	Y	L	K	Υ	ΥY	v v	СН
Mouse - CRISP2	G H	ΥΊ	QL	v v	VΥ	s	SF	Κ	ΙC	ЪC	G	A	Υ	С	Ρ	ΝÇ) D	Ν	L	K	Υ	FΥ	v v	СH
Rat - CRISP2	GΗ	ΥΊ	QL	v v	VΥ	S	S F	Κ	VC	Э <mark>С</mark>	G١	/ A	Υ	С	Р	ΝÇ		Т	L	K	Υ	FΥ	v v	<mark>с</mark> н
Chimp - CRISP2	G H	Y 1	QL	v v	VΥ	s	ΤY	Q	VC	ЪC	G	A	Υ	С	Р	NQ		S	L	K	Υ	ΥY	v v	<mark>c</mark> q
Human - CRISP2	G H	ΥΊ	QL	v v	VΥ	s	ΤY	Q	VC	ЪC	G	A	Υ	С	Ρ	NQ	D	S	L	K	Υ	ΥY	v v	сõ
Western clawed frog - Allurin	GH	Y 1	οv	AV	V A	к	ΤY	L	LC	ЭC	GI	. Α	F	С	Р	GN	ΙY	-	Y	Ρ	Y	v s	I	СН
African clawed frog-Allurin	GH	Y 1	Q V	v v	V A	к	S Y	Ľ	LC	ЭC	A	N	F	С	Κ	ΕN	ΙK	-	Y	P	н	VF	v	сн
Zebra Finch - Truncated CRISP	G H	Y 1	QМ	v v	νн	N	S Y	K	ΙC	ЪC	GI	A	F	С	S	ΝΊ	K	-	Y	S	Υ	FΥ	v v	<mark>c</mark> q
		CA	P2																					
Lamprey - Unchar.	Y C	ΡΝ	IGN	Ll	ΤV		I N		ΡJ	D	L (τŧ	S	С		s <mark>c</mark>	P	К	Ν	C	V	ΝN	L	C D
Spotted gar - Unchar.	Y C	ΡΙ			QΥ		- E	_	ЪЛ	Κ	Α (S	С		D C	P	Ν	Α	_		DR	. L	C T
Zebrafish - CRISP3		D /	GN	FH	R Τ		- V	D	ΡΣ	7 0	1 (÷S	Р	C	Α	S C	P						г т .	CT
	ΥΥ	R A	GN		_		- v	_	P I	(S	г (-	\sim				Ν	Ν	С	E	DN	L	
Atlantic cod - CRISP3	Y Y Y Y	R A		Y	/ G	i -	- W	7 P	PJ	κ	E		A	c	D	s c	P	N N	N N	-		D N D K	L	C T
				Y	_	i -	- W - W	7 P 7 P		K K K	~ `			C C	D				N N N	С	E		L L L	
Atlantic cod - CRISP3	YY YY YC	R A R A P H	G N G N G N G N	Y V F H Y (VG RT QF	+ - - -	- W - W - T	7 P 7 P Q	Ь 7 Ь 7 Ь 2	K K K K K K	~ `	эт ЭР	A S S	C C C	D A G	s <mark>C</mark> s C D C	P P P P	Ν	11	C C	E E E	DK DK NN	L L	C T C T C T
Atlantic cod - CRISP3 Stickleback fish - CRISP3	YY YY	R A R A P H R A	G N G N	Y V F H Y C F H	VG RT QF KR	- - -	- W - W - T - W	7 P 7 P Q 7 P	P 7 P 7 P 7	K K K K K K K	E C V C	ЭТ ЭР ЭР ЭТ	A S S P	C C	D A G A	s c s c	P P P P P	N N N N	Ν	C C	E E E	D K	L L	C T C T C T C T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish- CRISP3	YY YY YC YY YC	R A R A P H R A P A	G N G N G N G N G N	Y V F H Y C F H L I	VG T F R S R S S	- - - T	- W - W - T - W M S	7 P 7 P Q 7 P 7 P T	Ь 7 Ь 7 Ь 2	K K K K K K K K	E C V C	θT θP θP θT	A S S P K	C C	D A G A A	s <mark>C</mark> s C D C	P P P P P	N N N	N A	C C C C C C C	E E E V	DK DK NN	L L L L L	C T C T C T C T C T C T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish- CRISP3 Japanese rice fish - CRISP3	Y Y Y Y Y C Y Y Y C Y C	R A R A P H R A P A T A	G N G N G N G N G N G N G N G N	Y Y F H Y C F H L I N V	/ G R T Q F K R O S / N	- - - - T	- W - W - T - W M S I T	7 P 7 P Q 7 P T T	b 2 b 2 b 2 b 2 b 2 b 2 b 2 b 2 b 2 b 2	K K K K K K K K K K K	E C V C K C V C	6 T 6 P 6 T 6 P 6 P	A S P K T	C C	D A G A G	s c s c D c s c D c D c	P P P P P P	N N N T G	N A D A A	C C C C C C	E E E V	DK DK NN DK	L L L L L	C T C T C T C T C T C T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish- CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2	Y Y Y Y Y C Y Y Y C Y C Y C	R A P H R A P A T A P A	G N G N G N G N G N G N G N G N	Y F H Y C F H L I N V F	/ G R T Q F C R O S / N I N	- - T T S	- W - W - T - W M S I T I A	7 P 7 P Q 7 P T T T . T	Ь 7 Ь 7 Ь 7 Ь 7	K K K K K K K K K K K K	E C V C K C V C T C	6 T 6 P 6 T 6 P 6 P	A S S P K	C C	D A G A G G	s c s c s c s c d c d c d c	P P P P P P P P	N N N T G S	N A D A A S	C C C C C C C C C C C C C C C C C C C	E E V D A	DK DK NN DK NC	L L L L L L L L	C T C T C T C T C T C T C T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish- CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burmese python - CRISP-like Anole lizard - Cysteine-rich P.	Y Y Y Y Y C Y Y Y C Y C Y C Y S	R A P H R A P A T A P A P A	G N G N G N G N G N G N G N G N G N	Y V F H Y C F H L I N V F L R H	/ G R T Q F K R O S / N I N K D	- - T T S L	- W - W - T - W M S I T I A L N	7 P 7 P 7 P 7 P 7 T 7 T 7 T 7 T	 b 2 c 4 b 2 c 4 /ul>	K K K K K K K K K K K K K K K K K K K	E (V (K (V (T (S (6 T 6 P 6 T 6 P 6 P 6 P 6 P	A S P K T P P	C C	D A G A G G G	s c s c s c b c d c d c d c d c d c	P P P P P P P P P P	N N N T G S K	N A D A A S H	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	E E V D A D	DK DK DK NC NC NC		C T C T C T C T C T C T C T C T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish- CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burmese python - CRISP-like	Y Y Y Y Y C Y C Y C Y C Y C Y S Y C	R A P H R A P A T A P A P A P A	G N G N G N G N G N G N G N G N G N	Y Y F H Y C F H L I N V F 1 R H Y H	/ G R T Q F K R D S / N I N K D R K	- - T T S L L	- W - W - T - W M S I T I A L N M K	 7 P 7 P 7 P 7 T 7 T 7 T 7 T 7 T 	b 2 b 2 b 2 b 2 b 2 b 2 b 2 b 2 b 2 b 2	K K K K K K K K K K K K K K K K K K K		6 T 6 P 6 T 6 P 6 P	A S P K T P	C C	D A G A G G G	s c s c s c s c d c d c d c	P P P P P P P P P P P P P P P P P P P	N N T G S K K	N A D A A S	000000000	E E V D A D	DK DK NN DK NC NC		C T C T C T C T C T C T C T C T C T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish- CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burmese python - CRISP-like Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P.	Y Y Y Y Y C Y Y Y C Y C Y C Y C Y S Y C Y C	R A P F P A P A P A P A P A P A P A	4 G N 4 G N	Y X F H Y C F H L I N V F 1 R H Y H I H	/ G R T Q F K R O S / N I N K D R K R S	- - - T T S L S	- W - W - T - W M S I T I A L N M K I P	 7 P 7 P 7 P 7 P 7 T 7 T 7 T 7 T 7 T 7 T 	 b 2 c 4 /ul>			5 T 5 P 5 T 5 P 5 P 5 P 5 P 5 P	A S P K T P A P	C C	D A G A G G G G G G			N N T G S K K D	N A D A S H A A	0000000000	E E D D D D D D D	DK DK DK NC NC NC NC		C T C T C T C T C T C T C T C T C T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish- CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burmese python - CRISP-like Anole lizard - Cysteine-rich P. Softshell turtle - Unchar.	Y Y Y Y Y C Y C Y C Y C Y C Y C Y C Y C	R A P F P A P A P A P A P A P A P M P M	4 G N 4 G N	Y (F F L N (F F F F F F F F F F	/ G R T Q F C R D S / N I N K D R K R S C S	- - T T S L S S S	- W - W M S I T I A L N M K I P M K	 7 P 7 P 7 P 7 T /ul>	 b b d d<		E C V C V C V C V C V C V C V C V C V C	5 T 5 P 5 P 5 P 5 P 5 P 5 P 5 F 7 F 7 F 7 F 7 F	A S P K T P A P P P		D A G A G G G G G G G G		P P P P P P P P P P P P P P P P P P P	N N N G S K K D N	N A D A A S H A A A A A	C C C C C C C C C C C C C C	E E V D D D D D E	D K D K D K N C N C N C N C D C		C T T C T T C C T T C C C C T T C C C C
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish- CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burmese python - CRISP-like Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P.	Y Y Y C Y Y Y C Y C Y C Y C Y C Y C Y C	R A P H R A P A T A P A P A P A P M P M P A	G N G N G N G N G N G N G N G N G N G N	Y Y F H L I N V F H F H F H F H F H F H F H F H F H F H	/ G R T Q F K R O S / N I N K D R K R S L S		- W - W M S I T I A L N M K I P M K I E	7 P 7 P Q 7 P T T T T T T T T T T K	 b b b d d<	X K X K X K X K X K X K X K X K X K X K		5 T 5 P 5 P 5 P 5 P 5 P 5 P 5 P 5 P	A S P K T P A P		D A G A G G G G G A	S C S C D C D C D C D C D C D C D C D C D C D C D C D C D C D C D C D C	P P P P P P P P P P P P P P P P	N N N T G S K K D N D	N A D A S H A A	C C C C C C C C C C C C C C C C C C C	E E V D D A D D E E	DK DK DK NC NC NC NC	L L L L L L L L L L L L L L L	C T T T T T T T T T T T T T T T T T T T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African clawed frog - CRISP2 Burnese python - CRISP2 Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P.	Y Y Y C Y Y Y C Y C Y C Y C Y C Y C Y C	R A P H R A P A T A P A P A P A P M P M P A	G N G N G N G N G N G N G N G N G N G N	Y Y F H F H L I N Y F H F H F H F H F H F H F H F H F H F H	/ G R T Q F K R N N N N N N N N N N N N N		- W - W M S I T I A L N M K I P M K I E I E	7 P 7 P Q 7 P T T T T T T T T T T K K T	 b b d d<				A S P K T P A P P P P P		D A G A G G G G G G A G G G G	s constants of the second seco	P P P	N N N T G S K K D N D S	N A D A S H A A A A A A A	C C C C C C C C C C C C C C C C C C C	E E V D D A D D E E D	D K D K D K N C N C N C N C D C		C T T T T T T T T T T T T T T T T T T T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Jiapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African clawed frog - CRISP2 Burnese python - CRISP1ike Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P. Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P. Platypus - CRISP1	Y Y Y C Y C Y C Y C Y C Y C Y C Y C Y C	R A P H R A P A P A P A P A P M P M P M P A H A	G N G N G N G N G N G N G N G N G N G N	Y Y F H Y (F H L I N V F 1 H F 1 H I I I I I I I I I I I I I I I I I I	/ G R T Q F K R O S / N I N K D K C D K K S L S I G P E		- W - W - T - W M S I T I A L N M K I P M K I E I E M N	7 P 7 P Q 7 P T T T T T T T T K K	 b 2 c 2 c 3 c 4 /ul>		E (V (V (V (V (V (V (V (V (V (H T P P P P P F K K Q P	A S P K T P A P P P P P P		D A G A G G G G G G G G G G G G G G G G	s constructions of the second	P P P P P P P P P P P P P P P P P P P	N N N T G S K K D N D S K	N A D A A S H A A A A A A N		E E V D D A D D E E D N	D K D K N N C N C N C N C N C N C N C N C N C		C C C C C C C C C C C C C C C C C C C
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African clawed frog - CRISP2 Burnese python - CRISP2 Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P.	Y Y Y C Y Y Y C Y C Y C Y C Y C Y C Y C	R A P H R A P A T A P A P A P A P M P M P A	G N G N G N G N G N G N G N G N G N G N	Y Y F H Y (F H L I N V F 1 H F 1 H I I I I I I I I I I I I I I I I I I	/ G R T Q F K R N N N N N N N N N N N N N		- W - W - T - W M S I T I A L N M K I P M K I E I E M N	7 P 7 P 7 P 7 P 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T	 b b d d<	κ κ	E (V (V (V (V (V (V (V (V (V (J T P P P P P F K K Q P	A S P K T P A P P P P P		D A G A G G G G G G G G G G G G G G G G	s constants of the second seco	P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P	N N N T G S K K D N D S	N A D A S H A A A A A A A		E E V D D A D D E E D N	D K D K D K N C N C N C N C R C		C T T T T T T T T T T T T T T T T T T T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burnese python - CRISP2-like Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P. Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P. Platypus - CRISP1 Armadillo - Cysteine-rich P. Mouse - CRISP2	Y Y Y Y Y C Y C Y C Y C Y C Y C Y C Y C	R A P H P A P A P A P A P A P A P A P A P A P A	G N G N G N G N G N G N G N G N G N G N	Y Y F I Y (F I I I I I I I I I I I I I I I I I I I	/ G R T Q F C R O S / N / N C D R K R S L S L S L S L S / N / N / N / N / N / N / N / N		- W - W - T - W M S I T I A L N M K I P M K I E I E I E M N K H K S	7 P 7 P Q 7 P T T T T T T T T T T T T T T T T T	 b b c d d<	K K K K Q K Q	E (((((((((((((((((((J T P P P P P F K K Q P T T	A S P K T P P P P P P P P P P		D A G A A G G G G G G A G A A A A	s c s c c s c c c c c c c c c c c c c c	P P P P P P P P P P P P P P P P P P P	N N N T G S K K D N D S K N N	N A D A A S H A A A A A A N H N		E E V D D A D D E E D N E E	D K D K D N C C C C C C C C C C C C C C C C C C C		C T T T T T T T T T T T T T T T T T T T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Jiapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burnese python - CRISP2-like Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P. Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P. Platypus - CRISP1 Armadillo - Cysteine-rich P. Mouse - CRISP2 Rat - CRISP2	Y Y Y Y Y C Y C Y C Y C Y C Y C Y C Y C	R A R A P H R A P A P A P A P A P M P M P M P M P M P M P M P M	G N G N G N G N G N G N G N G N G N G N	Y Y F H Y C F H L I F H F H F H F H F H F H I F H I F H I F H T F H T Y T F H T T F H T T T T T T T T T T T T T T T T T T T	/ G		- W - W - T - W M S I T I A I A K M K I P M K I E I E M N K K S K S K	7 P 7 P 7 P 7 P 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T	 b b c d d<	K K K K Q K Q	E (V (V (V (V (V (V (V (V (V (J P P P P P P P E K K Q P T T T	A S P K T P P P P P P P P P P P		D A G A A G G G G G G G A G G A A A A A	s c s c c s c s	P P P P P P P P P P P P P P P P P P P	N N N T G S K K D N D S K N N N	N A D A A S H A A A A A N H N N		EEVDDADDEEDNEED	D K D K N C N C N C N C N C N C N C N C N C N C		C T T T T T T T T T T T T T T T T T T T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Jiapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burnese python - CRISP2-like Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P. Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P. Platypus - CRISP1 Armadillo - Cysteine-rich P. Mouse - CRISP2 Rat - CRISP2 Chimp - CRISP2	Y Y Y Y Y C Y C Y C Y C Y C Y C Y C Y C	R A R A P I R A P	G N G N G N G N G N G N G N G N G N G N	Y Y F H Y C F H L I F H F H F H F H F H F H F H F H F H F H	/ G	 	- W - W - T - W M S I T I A N M K I P M K I E I E M N K K S K S K K	7 P 7 P 7 P 7 P 7 T 7 T 7 T 7 T 7 T 7 T 7 T 8 T 7	 b c d <lid< li=""> d d d<!--</td--><td>K K K K K K K K K K K K K K K K K K K</td><td>E V K V T S S A A C O O O O O O O O O O O O O O O O O</td><td>J T P P P P T P E K K Q P T T T T</td><td>A S P K T P P P P P P P P P P P</td><td></td><td>D A G A A G G G G G G A A A A A A A A</td><td>s c s c c s c c s c c c c c c c c c c c</td><td>P P P P P P P P P P P P P P P P P P P</td><td>N N N T G S K K D N D S K N N N D</td><td>N A D A A S H A A A A A N H N N D</td><td></td><td>EEEVDDADDEEDNEEDD</td><td>D K K N N C C C C N N C C C C N N C C C C</td><td></td><td>C C C T T T T T T T T T T T T T T T T T</td></lid<>	K K K K K K K K K K K K K K K K K K K	E V K V T S S A A C O O O O O O O O O O O O O O O O O	J T P P P P T P E K K Q P T T T T	A S P K T P P P P P P P P P P P		D A G A A G G G G G G A A A A A A A A	s c s c c s c c s c c c c c c c c c c c	P P P P P P P P P P P P P P P P P P P	N N N T G S K K D N D S K N N N D	N A D A A S H A A A A A N H N N D		EEEVDDADDEEDNEEDD	D K K N N C C C C N N C C C C N N C C C C		C C C T T T T T T T T T T T T T T T T T
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burnese python - CRISP2-like Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P. Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P. Platypus - CRISP1 Armadillo - Cysteine-rich P. Mouse - CRISP2 Rat - CRISP2 Chimp - CRISP2 Human - CRISP2	Y Y Y Y Y C Y C Y C Y C Y C Y C Y C Y C	R A R A P I R A P A P A P A P A P A P A P A P	G N G N G N G N G N G N G N G N G N G N	Y Y Y F I Y C F I L I I V Y F I I I I I I I I I I I I I I I I I I I	/ G		- W - W - T - W M S I T I A M K I P M K I E I E M N K S K S K S K N K N	7 P 7 P Q Q 7 P T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T	 b c d <lid< li=""> d d d d<!--</td--><td>K K K Q Q Q C H C Q</td><td>E (((((((((((((((((((</td><td>J P P P P P P P E K K Q P T T T T T</td><td>A S P K T P P P P P P P P P P P P P</td><td></td><td>D A G A A G G G G G G G A A A A A A A A</td><td>s c s c s c s c c s c c c c c c c c c c</td><td></td><td>N N N T G S K K D N D S K N N N D D</td><td>N A D A A S H A A A A A N H N N D D</td><td></td><td>EEEVDDADDEEDNEEDDD</td><td>D K K N N N N O O O O N N N O O O O O N N O O O O O O N N O O O O O N N O O O O O N N N C O O O O</td><td></td><td>C C C C C C C C C C C C C C C C C C C</td></lid<>	K K K Q Q Q C H C Q	E (((((((((((((((((((J P P P P P P P E K K Q P T T T T T	A S P K T P P P P P P P P P P P P P		D A G A A G G G G G G G A A A A A A A A	s c s c s c s c c s c c c c c c c c c c		N N N T G S K K D N D S K N N N D D	N A D A A S H A A A A A N H N N D D		EEEVDDADDEEDNEEDDD	D K K N N N N O O O O N N N O O O O O N N O O O O O O N N O O O O O N N O O O O O N N N C O O O O		C C C C C C C C C C C C C C C C C C C
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Jiapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burnese python - CRISP1ike Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P. Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P. Platypus - CRISP1 Armadillo - Cysteine-rich P. Mouse - CRISP2 Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Western clawed frog - Allurin	Y Y Y Y Y C Y C Y C Y C Y C Y C Y C Y C	R A A P I A	G N G N G N G N G N G N G N G N G N G N	Y F I F I F I F I F I F I F I F I F I F I	7 G R T Q F C R O S 7 N 1 N C N 7 N 1 N C N 7 N 7 N 1 N 7 N 7 N 7 N 7 N 7 N 7 N 7 N 7		- W - W - T - W M S I T I A I M K M K I P M K I P M K I E I E M N K K S K S K K N K I N I I I I I I I I I I I I I I I I I I	7 P 7 P 7 P 7 P 7 P 7 P 7 T	 b 1 b 2 c 1 b 2 c 1 /ul>	K K K C Q C Q C Q C Q C Q C Q C Q C Q C	E V K V T S S A A E C V K Q Q Q A	J P P P P P P P E K K Q P T T T T T E	A S P K T P P P P P P P P P P P P W		D A G A A G G G G G G A A A A A A A A A	s c s c c s c c s c c c s c c c c c c c		N N N T G S K K D N D S K N N N D D E	N A D A A S H A A A A A N H N N D D S		EEEVDDADDEEDNEEDDDE	D D N N N N N N D R N N N N K K D		C C C C C C C C C C C C C C C C C C C
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burmese python - CRISP1ke Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P. Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P. Platypus - CRISP1 Armadillo - Cysteine-rich P. Mouse - CRISP1 Armadillo - CSP2 Chimp - CRISP2 Human - CRISP2 Western clawed frog - Allurin	Y Y Y Y Y C Y C Y C Y C Y C Y C Y C Y C	R A A P I A P A A P I A P A A A A A A A A A A A A A A A A A A A A	G N G N G N G N G N G N G N G N G N G N	Y Y F H Y (Y - 1 F H I I V Y F H I I I I I I I I I I I I I I V V V V V	7 G R T Q F C R O S 7 N 1 N C D 8 K 8 S 1 G 9 E 7 N 7 M 7 M 7 M 7 M 7 M 7 M 7 M 7 M	 	- W - W - T - W M S I T I A I N M K I P M K I P M K I E I E M N K K S K K S K K N K K N K I K V P	7 P Q 7 P T T T T T T T T T T T T T T T T T T T	A definition of the second system of the second	K K K C Q C Q C Q C Q C Q C Q C Q C Q C	E V K V T S S A A E C V K Q Q Q A	J P P P P P P P E K K Q P T T T T T	A S P K T P P P P P P P P P P P W W		D A G A A G G G G G G G A A A A A A A A	s c s c s c s c c s c c c c c c c c c c		N N N T G S K K D N D S K N N N D D E E	N A D A A S H A A A A A N H N N D D		EEEVDDADDEEDNEEDDDED	D D N N N N N N N N N N N N K K D D		C C C C C C C C C C C C C C C C C C C
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Jiapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burnese python - CRISP1ike Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P. Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P. Platypus - CRISP1 Armadillo - Cysteine-rich P. Mouse - CRISP2 Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Western clawed frog - Allurin	Y Y Y Y Y C Y C Y C Y C Y C Y C Y C Y C	R A A P I A P A A P I A P A A A A A A A A A A A A A A A A A A A A	G N G N G N G N G N G N G N G N G N G N	Y Y F H Y (Y - 1 F H I I V Y F H I I I I I I I I I I I I I I V V V V V	7 G R T Q F C R O S 7 N 1 N C D 8 K 8 S 1 G 9 E 7 N 7 M 7 M 7 M 7 M 7 M 7 M 7 M 7 M	 	- W - W - T - W M S I T I A I M K M K I P M K I P M K I E I E M N K K S K S K K N K I N I I I I I I I I I I I I I I I I I I	7 P Q 7 P T T T T T T T T T T T T T T T T T T T	b d d d d d d d d d d d d d d d d d d d	K K K K K K K K K K K K K K K K K K K	E V K V T S S A A E E A E V K Q Q Q A E	J P P P P P P P E K K Q P T T T T T E	A S P K T P P P P P P P P P P P P W		D A G A A G G G G G G A G G A A A A A A	s c s c c s c c s c c c s c c c c c c c		N N N T G S K K D N D S K N N N D D E	N A D A A S H A A A A A N H N N D D S		EEEVDDADDEEDNEEDDDED	D D N N N N N N D R N N N N K K D		C C C C C C C C C C C C C C C C C C C
Atlantic cod - CRISP3 Stickleback fish - CRISP3 Tilapa fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - CRISP2 African dawed frog - CRISP2 Burmese python - CRISP1ke Anole lizard - Cysteine-rich P. Softshell turtle - Unchar. Falcon - Cysteine-rich P. Golden manakin - CRISP2 Woodpecker - CRSP2 Chicken - Cysteine-rich P. Platypus - CRISP1 Armadillo - Cysteine-rich P. Mouse - CRISP1 Armadillo - CSP2 Chimp - CRISP2 Human - CRISP2 Western clawed frog - Allurin	Y Y Y Y Y C Y C Y C Y C Y C Y C Y C Y C	R A P H P A P A P A P A P A P A P A P A P A P A	G N N G N N N G N N N G N N N G N N N G N N N N G N	Y F F F F F F F F F F F F F F F F F F F	7 G R T Q F C R O S 7 N 1 N C D 8 K 8 S 1 G 9 E 7 N 7 M 7 M 7 M 7 M 7 M 7 M 7 M 7 M	 	- W - W - W M S I T I A M S I T I A M K I P M K I E I E M N K K S K K S K K N K I K K V P M K	7 P Q 7 7 P 7 P 7 T 1 T 1 T 2 T 4 T 5 T 6 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 7 T 8 T 9 T 9 T 1 T 1 T 1 T 1 T 1 T 1 T	A definition of the second system of the second		E V K V T S S A A E E A E V K Q Q Q A E E	J J J J J J J J J J J J J J J J J J J	A S P K T P P P P P P P P P P P W W		D A G A A G G G G G G G A A A A A A A A	s c c s c c s c c s c c c s c c s c c s c c s c c s c c s c c s c c s c c s c c s c c s c c s c c s c s c c s c s c c s c s c c s c s c c s c s c c s		N N N T G S K K D N D S K N N N D D E E	N A D A A S H A A A A A N H N N D D S S		EEEVDDADDEEDNEEDDDED	D D N N N N N N N N N N N N K K D D		C C C C C C C C C C C C C C C C C C C

C Conserved cysteines

Hinge region

Sites of non-aligned sequence omission

Figure 8. Page 3 of 5

Figure 8. ClustalW Amino Acid Alignment of Vertebrate CRISP2 and CRISP2-
like Proteins

]																				
Lamprey - Unchar.	Ν	Р	С	K	Y	K	Ν	G	Y	A	Ν	С	A	A	М	V	A	Q	Y	Т	С	K	Ν	A	Ι	Ι	L	K	Ν	С	Р	A	S	С
Spotted gar - Unchar.	Ν	Р	С	Р	Y	Ι	Ν	Q	Y	Т	Ν	С	Р	S	L	R	Q	Е	Y	G	С	s	Ν	Α	Q	v	Q	S	W	С	L	Α	S	С
Zebrafish - CRISP3	Ν	Α	С	Р	Y	Ι	Ν	G	F	V	Ν	С	D	Α	L	к	Ā	Κ	L	Т	С					v			G	С	Р	Α	S	С
Atlantic cod - CRISP3	Ν	Р	С	Q	Y	Ι	Ν	R	Y	L	Ν	С	Р	Α						G	С	G	Ν	Н	w	v	Ν	L	w	С	L	Α	Е	С
Stickleback fish - CRISP3	Ν	Р	С	Р	Y	Ι	Ν	к	Y	L	Ν	С	Р	S	L	к	Т	L	V	G	С	S	Ν	Ν	L	v	S	s	W	С	Р	Α	S	С
Tilapa fish- CRISP3	Ν	Р	С	Р	F	S	D	к	Y	Ν	Ν	С	Р	Е	L	к	0	0	w	G	С	s	н	Р	D	v	Α	s	w	С	Р	Α	s	С
Japanese rice fish - CRISP3	Ν	Р	С	Р	Y	Ι	Ν	R	F	L	Ν	С		Α						G	С	Q	Ν	Р	R	v	s	S	w	С	Р	Α	s	С
Western clawed frog - CRISP2	Ν	Y	С	Р	Y	Q	D	L	Y	S	Ν	С	К	Т	Y	Α	s	Y	С	Ν	Т	F	Р	-	Т	Ι	R	D	G	С	К	Α	Т	С
African clawed frog - CRISP2	Ν	Y	С	Р	Y	0	D	L	Y	Т	G			Ν					С	Ν	Ι	Ι	Р	_	М	I	м	D	G	С	R	G	Т	С
Burmese python - CRISP-like	Ν	Р	С	Т	н	Е	D	Т	Y	s	Ν	С	Ν	Α	L	V	К	Е	н	к	С	Q	Ι	Р	W	Ι	R	к	S	С	Р	Α	S	С
Anole lizard - Cysteine-rich P.	Ν	Р	С	Ν	Y	D	D	Ν	Y	s	Ν	С		Ν							С		I	s	L	М	к	2	D	С	Е	А	s	С
Softshell turtle - Unchar.	Ν	Р	С	Ν	Y	v	D	v	F	s	Ν	С	Р	Α	L	к	s	F	F	G	С	s	н	Р	L	v	к	к	к	С	Р	Α	s	С
Falcon - Cysteine-rich P.	Ν	Р	С	к	Y	к	D	s	F	Е	Ν	С				к				Т	С			s		м	к	-	к	С	Р	Α	Т	С
Golden manakin - CRISP2	Ν	Р	С	к	Y	0	D	F	L	G	Ν	С		Ν						G	С	s	н	s	L	v	к	0	к	С	Р	А	Т	С
Woodpecker - CRSP2	Ν	Р	С		Y			н		А				Ν						G	С			Р			к	È	Ν	С		А		С
Chicken - Cysteine-rich P.	Ν	Р	С	к	Y	R	D	v	Y	s	Ν			Е						G	с	Е	н	s	F	I	к	Т	Ν	С	L	Α	s	С
Platypus - CRISP1	Ν	Р	С	Р	F		D	Ν			Ν					Ι				G		К			Е	L	Т	Е	Α	Y		-	-	-
Armadillo - Cysteine-rich P.		S	С	E	Y		D	A	L		N			Q							c		Н			L	к		К			А	Т	С
Mouse - CRISP2	-	ŝ	c		F	E		L	L		N		E						A			ĸ							ĸ			A		C
Rat - CRISP2	-	ŝ	č					L	L		N		E			ĸ				G		ĸ				L	ĸ	A	ĸ		· · · ·	A		C
Chimp - CRISP2	-	ŝ	c			õ				s				s							č					L			ĸ					C
Human - CRISP2	-	ŝ			Ŷ		D			ŝ				s												Ĺ			ĸ			A		č
Western clawed frog - Allurin	-	N		-		-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-
African dawed frog- Allurin	-	ĸ		ĸ		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Zebra Finch - Truncated CRISP	-	Т	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Zeolu Filen - Huleated ereist																								-										
Lamprey - Unchar.	Ν	С	Q	G	R	Ι	-	-	-	-	-	-																						
Spotted gar - Unchar.	Ν	C	Т	Ν	К	Ι	Ι	-	-	-	-	-																						
Zebrafish - CRISP3	L	С	D	Ν	К	Ι	Ι	Р	Ι	Α	Κ	К																						
Atlantic cod - CRISP3	Κ	C	Р	Ν	Е	Ι	Ι	-	-	-	-	-																						
Stickleback fish - CRISP3	Q	С	Ι	D	Е	Ι	Ι	-	-	-	-	-																						
Tilapa fish- CRISP3	Κ	C	Т	Ν	Q	Ι	Ι	-	-	-	-	-																						
Japanese rice fish - CRISP3	Κ	C	S	S	Е	Ι	Ι	-	-	-	-	-																						
Western clawed frog - CRISP2	L	С	Т	Ν	-	Q	Ι	Ι	-	-	-	-																						
African clawed frog - CRISP2	L	С	R	Ν	Ν	Q	Ι	Ι	-	-	-	-																						
Burmese python - CRISP-like	F	С	Т	Т	Е	Ι	Κ	-	-	-	-	-																						
Anole lizard - Cysteine-rich P.	Κ	C	Т	Т	Е	Ι	Q	-	-	-	-	-																						
Softshell turtle - Unchar.	L	С	Т	Т	Е	Ι	Κ	-	-	-	-	-																						
Falcon - Cysteine-rich P.	Κ	C	Т	Т	Е	Ι	Ι	-	-	-	-	-																						
Golden manakin - CRISP2	Κ	С	Т	Т	Q	Ι	Ι	-	-	-	-	-																						
Woodpecker - CRSP2	R	С	Т	Т	Q	Ι	Ι	-	-	-	-	-																						
Chicken - Cysteine-rich P.	R	С	L	S	Е	Ι	Ι	-	-	-	-	-																						
Platypus - CRISP1	-	-	-	-	-	-	-	-	-	-	-	-																						
Armadillo - Cysteine-rich P.	Ν	С	Е	D	Κ	Ι	F	-	-	-	-	-																						
Mouse - CRISP2	L	С			Κ	Ι	Н	-	-	-	-	-																						
Rat - CRISP2	L	С	E	D	K	Ι	Н	-	-	-	-	-																						
Chimp - CRISP2	L	С			Κ	Ι		-	-	-	-	-																						
Human - CRISP2	L					Ι		-	-	-	-	-																						
Western clawed frog - Allurin	-	-	-		-		-	-	-	-	-	-																						
African clawed frog-Allurin	-	-	-	-	-	-	-	-	-	-	-	-																						
Zebra Finch - Truncated CRISP	-	-	-	-	-	-	-	-	-	-	-	-																						
CAP signati	ire		ea	116	ene	ces	2					T	C	R (do	m	ai	n																
				•															_1'			1							:-	.:				
C Conserved of	•	ste	n	es									sit	es	0	í n	101	n-	alı	gn	lec	1 S	sec	ln	er	ICE	e 0	m	1S	S1C	m			
Hinge region	n																																	
Figure 8.																																		
Page 4 of 5																																		

Page 4 of 5

Figure 8. ClustalW Amino Acid Alignment of Vertebrate CRISP2 and CRISP2like Proteins

Figure 8. MEGA7 ClustalW amino acid sequence alignment of vertebrate CRISP2 and CRISP2-like proteins. The CAP signature sequences are highlighted in orange, the hinge region is highlighted in blue, and the ICR domain is highlighted in green. The highly conserved cysteines are highlighted in yellow with a red font color. CRISP proteins have sixteen highly conserved cysteines; six of the cysteines are located in the CAP/PR domain, four are located in the Hinge region, and six more are located in the ICR domain. These highly conserved cysteines form eight disulfide bonds (shown in Figure 9) improving the structural stability of the protein's tertiary structure. In addition, CRISP proteins can be traced back to early vertebrates and share a high degree of amino acid sequence conservation between species. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 7. Page 5 of 5

Figure 9. Domain Organization and Tertiary Structure of a Full Length CRISP Protein: Venom CRISP From the Chinese Cobra

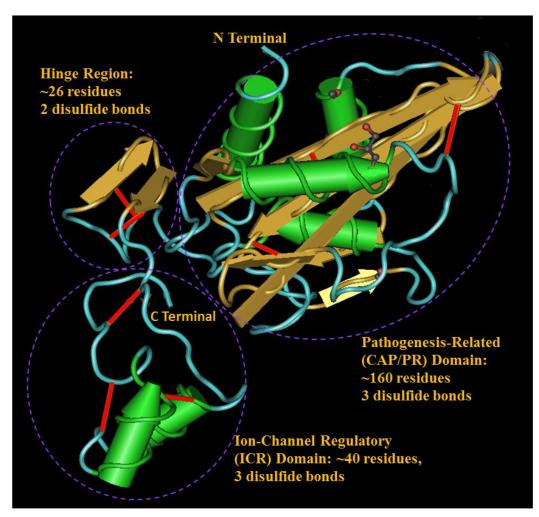


Figure 9. The tertiary structure of the Chinese cobra cysteine-rich venom protein natrin-1 (CRVP1) (accession ID: Q7T1K6.1, PDB ID: 3MZ8) [88]. The helices are represented by green rods, the β -strands are represented by tan arrows, and the polypeptide chain is represented in light blue. Additionally, the eight disulfide bonds are represented in red. The CAP/PR domain has seven cysteines, of which six are involved in three disulfide bonds; the Hinge region has four cysteines and two disulfide bonds; and the ICR domain has six cysteines and three disulfide bonds. These disulfide bonds create a stable tertiary structure that is relatively heat and solvent resistant – a characteristic common to most CAP proteins.

Figure 10. ClustalW Amino Acid Alignment and Secondary Structure of Selected CRISP Proteins

Common name - Protein name												Ar	nino	Ac	id S	leai	ien	ce												
																													α	1
Secondary Structure					_	_																								
Chinese cobra - CRISP-natrin																														
Lamprey - Unchar.	A S																													
Tilapa fish - CRISP3	TN																													
	M I M I																													
Falcon - Cysteine-rich P. Human - CRISP2																														VQ
African clawed frog - Allurin	MD																													
Secondary Structure			1	11		χ1	Ľ		1 0			1	1			U	D	D	-	-		βı		5	_	Ľ		α2		1 4
Chinese cobra - CRISP-natrin	ΚE	E I	V	D	L	ΗI	Ν	S I	LR	R	R	V	S I	P 7	ΓА	S	Ν	Μ	L	K	М	E	w	Y	Ρ	Е	Α	А	s	ΝA
Lamprey - Unchar.	TF	I	V	D	L	H	Ν	D	LR	R	N	V	ΤI	P S	S A	G	Ν	Μ	L	Q	M	M	W	K	K	Е	A	Y	L	ΝA
Tilapa fish - CRISP3	ΝE	Ι	V	Ν	K	H	N	A]	LR	R	N	V	Q]	P 7	ΓА	S	Ν	Μ	L	K	Μ	N	W	Ν	N	Е	A	Е	A	ΤА
Burmese python - CRISP-like	ΚE	Ι	V	D	Η	H	N	NI	LR	R	Т	V	S I	P 7	ΓА	S	Ν	Μ	L	R	М	E	W	Η	Ρ	D	A	А	K	ΝA
Falcon - Cysteine-rich P.		_																												ΝA
Human - CRISP2		_																												ΝA
African clawed frog - Allurin	NK	_		D	Η	Η	Ν	A]	LR	R	S			P 7	ΓА	-		Μ	-		Μ	V	W	С			A	A	L	NA
Secondary Structure		α2											α3			β2	2		β3		1				β-	4				
Chinese cobra - CRISP-natrin	ΕR	t W	А	Ν	Т	C	s	Ll	NH	I S	Ρ	D	ΝI	LI	R V	L	Е	G	Ι	Q	C	G	Е	s	I	Υ	Μ	S	s	ΝA
Lamprey - Unchar.	LS	W	А	Ν	Ν	С	Т	F	GH	I S	Ρ	D	AI	RI	R K	T	Т	Q	W	Т	C	G	Ε	Ν	I	F	Μ	S	s	AP
Tilapa fish - CRISP3																														FQ
Burmese python - CRISP-like																														
Falcon - Cysteine-rich P.																														A P
Human - CRISP2																														DP
African clawed frog - Allurin	YN	I F	А	Т	Q	С	S	_		S	L	V	EI		-	Ι	K	E			C	G	E	Ν	Ι	Y	Μ		_	AK
								(1/2						- 1	35				B6										$\alpha 5$	•
Secondary Structure								α								•			po											
Secondary Structure Chinese cobra - CRISP-natrin	R T							L١	NE					NI	F V				G	A								G		
Chinese cobra - CRISP-natrin Lamprey - Unchar.	R T F T	W	N	Y	A	Ι	R	L V D V	V H V Y	N	Е	V	IS	N H S H	FV FQ	Y	D	V	G G	A S	Т	Т	-	Т	G	A	V	G G	н	ΥΊ
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3	R 1 F 1 N 1	W W	N S	Y N	A V	I I	R Q	L \ D \ S \	W H W Y W Y	N D	E E	V V	I S K I	N H S H D H	FV FQ FS	Y Y	D G	V V	G G G	A S S	T I	T N	-	T G	G G	A V	V V	G G G	H H	Y T Y T
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like	R 1 F 1 N 1 Y S	C W C W	N S T	Y N A	A V V	I I T	R Q Q	L V D V S V S V	W H W Y W Y W Y	N D D	E E E	V V R	I S K I K I	N H S H D H D H	FV FQ FS FK	Y Y Y	D G G	V V V	G G G G	A S S P	T I I	T N Q	- - P	T G N	G G A	A V V	V V T	G G G G	H H H	Y T Y T Y T
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P.	R 1 F 1 N 1 Y S F S	W W W W	N S T P	Y N A D	A V V	I I T I	R Q Q Q	L V D V S V S V S V	W H W Y W Y W Y W Y	N D D D	E E E E	V V R K	I 9 K 1 K 1 E 1	N H S H D H D H D H	FV FQ FS FK FK	Y Y Y Y	D G G G	V V V T	G G G G G	A S S P A	T I I K	T N Q T	- - P P	T G N G	G G A A	A V V	V V T I	G G G G G	H H H H	Т Ү Ү Т Ү Т Ү Т Ү Т
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2	R T F T N T Y S F S T S	W W W W	N S T P S	Y N A D S	A V V A	I I T I I	R Q Q Q Q	L V D V S V S V S V S V	W H W Y W Y W Y W Y W Y	Y N Y D Y D Y D Y D	E E E E	V V R K I	I S K I K I E I L I	N H S H D H D H D H D H D H	F V F Q F S F K F K F V	Y Y Y Y Y	D G G G G	V V V T V	G G G G G	A S S P A P	T I K K	T Q T S	- P P P	T G N G N	G G A A A	A V V V	V V T I V	G G G G G G	H H H H	Y T Y T Y T Y T Y T Y T
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Allurin	R 1 F 1 N 1 Y S F S T S S D		N S T P S	Y N A D S	A V V A	I I T I I	R Q Q Q Q	L V D V S V S V S V S V	W H W Y W Y W Y W Y W Y	Y N Y D Y D Y D Y D Y N	E E E E E	V V R K I	I S K I K I E I L I	N H S H D H D H D H D H D H	F V F Q F S F K F K F V	Y Y Y Y Y	D G G G G	V V V T V	G G G G G	A S S P A P	T I K K	T Q T S	- P P P	T G N G N	G G A A A	A V V V P	V V T I V I	G G G G G G	H H H H	Т Ү Ү Т Ү Т Ү Т Ү Т
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure	R T F T N T Y S F S T S S C α5		N S T P S S	Y N A D S T	A V V A V	I T I I I I	R Q Q Q D	L V D V S V S V S V S V	W H W Y W Y W Y W Y W Y	γ D D D D N β	E E E E 7	V V R K I R	I S K I E I L I S I	N H S H D H D H D H D H D H D H	F V F Q F S F K F K F V F A	Y Y Y Y Y Y	D G G G G	V V T V K	G G G G G G G	A S S P A P K	T I K K I	T Q T S S	- P P P	T G N G N D	G A A A K	A V V V P	V V T I V I 88	G G G G G G	H H H H H	Y T Y T Y T Y T Y T Y T
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Allurin Secondary Structure Chinese cobra - CRISP-natrin	R T F T N T Y S F S T S S C Q I		N S T P S S S	Y N D S T	A V V A V	I T I I I T	R Q Q Q D Y	L V D V S V S V S V S V S V	W H W Y W Y W Y W Y W Y A C	N D D D D C D C	E E E E 7	V V R K I R V	I S K I E I L I S I S S	N H S H D H D H D H D H D H V (F V F Q F S F K F K F K F V F A	Y Y Y Y Y Y	D G G G G	V V T V K	G G G G G G S	A S S P A P K	T I K K I	T Q T S S S	- P P -	T G N G N D	G G A A K Y	Α V V V P β	V V T I V I S 8	G G G G G G Q	H H H H H Y	Y T Y T Y T Y T Y T Y T Y T
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar.	R T F T N T Y S F S T S S C Q I Q I		N S T P S S S W W	Y N A D S T Y Y	A V V A V Q N	I I I I I I I I I I I I I I I I I I I	R Q Q Q D Y F	L V D V S V S V S V S V R Z	W F. W Y W Y W Y W Y W Y A C L C	 N D D D C C C C 	E E E F 7 7	V V R K I R V V V	I S K I E I L I S I S X N X	N H S H D H D H D H D H D H D H C	F V F Q F S F K F K F K F K F A	Y Y Y Y Y Y K	D G G G G V	V V T V K - P	G G G G G G S S G	A S S P A P K V	T I K K I W A	T Q T S S S K	- P P - Y Y	T G N G N D F L	G A A A K Y Y	A V V V P β V V V	V V T I V I S 8 C C	G G G G G G G H	H H H H H Y Y	Y T Y T Y T Y T Y T Y T C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3	R T F T N T Y S F S T S S C Q I Q I Q L Q V		N S T P S S S W W W W	Y N A D S T Y Y Y Y	A V V A V R	I I I I I I I I I I I I I I I I I I I	R Q Q Q D Y F N	L V D V S V S V S V S V S V R 2 Q 1	W H W Y W Y W Y W Y W Y A C L C	V N V D V D V D V D V D V D V D V D V D V V D V V D V V D V D	E E E E 7 A A A A	V V R K I R V V V M	I S K I E I L I S I S X N X A X	H I I I I I I I I I I I I I I I I I I I	F V F Q F S F K F K F K F V F A	Y Y Y Y Y Y S K N	D G G G G G N	V V T V K - P -	G G G G G G S S	A S S P A P K V A	T I K K I W A Y	T Q T S S K K K	- P P - Y Y Y Y	T G N D F L F	G G A A K K Y Y Y Y	A V V V P β V V V V V	V V T V I S 8 C C C C	G G G G G G G G H H	H H H H H Y Y Y Y	Y T Y T Y T Y T Y T Y T C P C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Allurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like	R T F T N T Y S F S T S S C Q I Q I Q L Q V Q V	V V V V V V V V V V	N S T P S S S W W W W W W	Y N A D S T T Y Y Y Y Y Y	A V V A V R K	I 1 I 1 I 1 I 1 I 1 I 1 I 1 I 1	R Q Q Q D Y F N R	L V D V S V S V S V S V S V S V S V L 1	W H W Y W Y W Y W Y W Y W Y L C L C	 N D D D D D C C C C C C C 	E E E E 7 A A A A A	V V R K I R V V V M A	I S I K I E I L I S I S I N S A Y A H		F V F Q F S F K F K F K F V F A C P C P C P	Y Y Y Y Y Y X A	D G G G G N	V V T V K - P -	G G G G G G S S S	A S S P A P K V A K	T I K K I W A Y F	T Q T S S K K K K	- P P P - Y Y Y Y Y Y	T G N D F L F Y	G A A A K Y Y Y Y Y Y	A V V V P β V V V V V V V V V	V V T I V I S 8 C C C C C	G G G G G G G H H H	H H H H H Y Y Y Y Y	Y T Y T Y T Y T Y T Y T C P C P C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P.	R T F T Y S F S T S S D Q I Q I Q V Q V Q V	W W W W W W W W W W W W V V V V V V V V	N S T P S S S W W W W W W	Y N A D S T Y Y Y Y Y Y Y	A V V A V R K N	I 1 I 1 I 1 I 1 I 2 I 2 S 2 S 3 S 3	R Q Q Q D Y F N R Y Y	L V D V S V S V S V S V S V S V L 1 Q 1 L 1	W H W Y W Y W Y W Y W Y W Y A C L C L C I C	Υ N Υ D	E E E E 7 A A A A A A A	V V R K I R V V V M A V	I S I K I E I L I S I N S N X A X A A		F V F V F S F S F K F K F K F V F A P P P P P P P P P	Y Y Y Y Y Y S K N A N	D G G G G N	V V T V K - P - -	G G G G G G G G S S S S	A S S P A P K V A K R	T I K K I W A Y F Y	T Q T S S K K K K K	- P P P - - Y Y Y Y Y Y Y	T G N D F L F Y F	G A A A K Y Y Y Y Y Y Y	A V V V P β V V V V V V V V V V	V V T V I S 8 C C C C C C C	G G G G G G G G G G G H H H Q	H H H H H Y Y Y Y Y Y	Y T Y T Y T Y T Y T Y T Y T C P C P C P C P C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Allurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2	R T F T N T Y S F S T S S T Q I Q I Q V Q V Q V Q V Q V Q L	V V V V V V V V V V V V V V V V V V V	N S T P S S S W W W W W W W W W	Y N A D S T Y Y Y Y Y Y Y Y	A V V A V A V R K N S	I 1 I 1 I 1 I 1 I 1 I 1 I 1 I 1	R Q Q Q D Y F N R Y Y Y	L V D V S V S V S V S V S V S V L 1 Q 1 Q 1 Q 1 Q 1 Q 1	W F W Y W Y W Y W Y W Y W Y U Y U Y U C	Y N Y D	E E E T A A A A A A G	V V R I R V V V M A V I I	I S K I K I E I S I S V A V A V A V A V A V		F V F Q F S F K F K F K F K F A F A P C P P C P C P	Y Y Y Y Y Y X N N	D G G G G G V N Q	V V T V K - P - - -	G G G G G G G S S S S S D	A S S P A P K V A K R S	T I K K I W A Y F Y L	T Q T S S K K K K K K K	- P P P - · · · · · · · · · · · · · · ·	T G N D F L F Y F Y F Y	G A A K Y Y Y Y Y Y Y Y Y	A V V V P β V V V V V V V V V V V V	V V T V I S 8 C C C C C C C C C	G G G G G G H H H Q Q	H H H H H H Y Y Y Y Y Y Y	Y T Y T Y T Y T Y T Y T Y T C P C P C P C P C P C P C P C P C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin	R T F T N T Y S F S S D Q I Q L Q V Q V Q V Q V Q V Q V	V V V V V V V V V V V V V V V V V V V	N S T P S S S W W W W W W W W W	Y A D S T Y Y Y Y Y Y Y Y A	A V V A V A V R K N S K	I 1 I 1 I 1 I 1 I 1 I 1 I 1 I 1	R Q Q Q D Y F N R Y Y Y	L V D V S V S V S V S V S V S V L 1 Q 1 Q 1 Q 1 Q 1 Q 1	W F W Y W Y W Y W Y W Y W Y U Y U Y U Y U Y U C L C	Ν Δ <t< td=""><td>E E E F A A A A A A G A</td><td>V V K I R V V V M A V I Y</td><td>I S K I K I E I L I S I S X A X A X A X A X N I</td><td></td><td>F V F Q F S F K F K F K F K F K F V F A P C P C P C P C P C P C R</td><td>Y Y Y Y Y Y Y S K N A N N E</td><td>D G G G G G V</td><td>V V T V K - P - - - -</td><td>G G G G G G G G S S S S S S N</td><td>A S S P A P K V A K R S K</td><td>T I K K I V A Y F Y L Y</td><td>T Q T S S K K K K K K F</td><td>- P P P - - - Y Y Y Y Y Y Y H</td><td>T G N D F L F Y F Y V</td><td>G G A A K Y Y Y Y Y Y Y F</td><td>A V V V V P β V V V V V V V V V V V V V</td><td>V V T I V I S 8 C C C C C C C C C C</td><td>G G G G G G H H H Q Q H</td><td>H H H H H H Y Y Y Y Y Y Y</td><td>Y T Y T Y T Y T Y T Y T Y T C P C P C P C P C P C P</td></t<>	E E E F A A A A A A G A	V V K I R V V V M A V I Y	I S K I K I E I L I S I S X A X A X A X A X N I		F V F Q F S F K F K F K F K F K F V F A P C P C P C P C P C P C R	Y Y Y Y Y Y Y S K N A N N E	D G G G G G V	V V T V K - P - - - -	G G G G G G G G S S S S S S N	A S S P A P K V A K R S K	T I K K I V A Y F Y L Y	T Q T S S K K K K K K F	- P P P - - - Y Y Y Y Y Y Y H	T G N D F L F Y F Y V	G G A A K Y Y Y Y Y Y Y F	A V V V V P β V V V V V V V V V V V V V	V V T I V I S 8 C C C C C C C C C C	G G G G G G H H H Q Q H	H H H H H H Y Y Y Y Y Y Y	Y T Y T Y T Y T Y T Y T Y T C P C P C P C P C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin	R T F T N T Y S F S T S S C Q I Q V Q V Q V Q V Q V Q V Q V Q V Q V	V V V V V V V V V V V V V V V V V V V	N S T P S S S W W W W W W W W W W W	Y A D S T Y Y Y Y Y Y Y A	A V V A V A V R K N S K	I 1 I 1 I 1 I 1 I 1 I 1 I 1 I 1	R Q Q Q D Y F N R Y Y Y	L V D V S V S V S V S V S V S V L 1 Q 1 Q 1 Q 1 Q 1 Q 1	W H W Y W Y W Y W Y W Y W Y U Y U Y C L C L C L C L C L C L C L C L C L C L	$ \begin{array}{c} & \mathbf{N} \\ & \mathbf{N} \\ & \mathbf{D} \\ & $	E E E E F A A A A A A A A A A S	V R K I R V V V M A V I Y Of	I S K I K I E I S I S V A V A V A V A V A V		F V F Q F S F K F K F K F K F K F V F A P C P C P C P C P C P C R	Y Y Y Y Y Y Y S K N A N N E	D G G G G G V	V V T V K - P - - - -	G G G G G G G G S S S S S S N	A S S P A P K V A K R S K	T I K K I V A Y F Y L Y	T Q T S S K K K K K K F	- P P P - - - Y Y Y Y Y Y Y H	T G N D F L F Y F Y V	G G A A K Y Y Y Y Y Y Y F	A V V V V P β V V V V V V V V V V V V V	V V T I V I S 8 C C C C C C C C C C	G G G G G G H H H Q Q H	H H H H H H Y Y Y Y Y Y Y	Y T Y T Y T Y T Y T Y T Y T C P C P C P C P C P C P C P C P C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin CAP signatu C Conserved of	R T F T N T Y S F S S T S S C U U Q V Q V Q V Q V Q V Q V Q V Q V Q V Q V	V V V V V V V V V V V V V V V V V V V	N S T P S S S W W W W W W W W W W W	Y A D S T Y Y Y Y Y Y Y A	A V V A V A V R K N S K	I 1 I 1 I 1 I 1 I 1 I 1 I 1 I 1	R Q Q Q D Y F N R Y Y Y	L V D V S V S V S V S V S V L 1 Q 1 Q 1 L 1 Q 1 Q 1 L 1	W H W Y W Y W Y W Y W Y W Y U Y U Y C L C L C L C L C L C L C L C L C L C L	$ \begin{array}{c} & \mathbf{N} \\ & \mathbf{D} \\ & $	E E E E A A A A A A A A A C A eli	V R K I R V V V M A V V I Y Of	I S K I K I E I L I S I S X A X A X A X A X N I		F V F Q F S F K F K F K F K F K F V F A P C P C P C P C P C P C R	Y Y Y Y Y Y Y S K N A N N E	D G G G G G V	V V T V K - P - - - -	G G G G G G G G S S S S S S N	A S S P A P K V A K R S K	T I K K I V A Y F Y L Y	T Q T S S K K K K K K F	- P P P - - - Y Y Y Y Y Y Y H	T G N D F L F Y F Y V	G G A A K Y Y Y Y Y Y Y F	A V V V V P β V V V V V V V V V V V V V	V V T I V I S 8 C C C C C C C C C C	G G G G G G H H H Q Q H	H H H H H H Y Y Y Y Y Y Y	Y T Y T Y T Y T Y T Y T Y T C P C P C P C P C P C P C P C P C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin CAP signatu C Conserved of Hinge region	R T F T N T Y S F S S T Q I Q I Q V Q V Q V Q V Q V Q V Q V Q V Q V C V S S C C S S S S S S S S S S S S S S S S	V V V V V V V V V V V V V V V V V V V	N S T P S S S W W W W W W W W W W W	Y A D S T Y Y Y Y Y Y Y A	A V V A V A V R K N S K	I 1 I 1 I 1 I 1 I 1 I 1 I 1 I 1	R Q Q Q D Y F N R Y Y Y	L V D V S V S V S V S V S V L 1 Q 1 Q 1 L 1 Q 1 Q 1 L 1	W H W Y W Y W Y W Y W Y W Y U Y U Y C L C L C L C L C L C L C L C L C L C L	$ \begin{array}{c} & \mathbf{N} \\ & \mathbf{D} \\ & $	E E E E A A A A A A A A A C A eli	V R K I R V V V M A V V I Y Of	I S K I K I E I L I S I S X A X A X A X A X N I		F V F Q F S F K F K F K F K F K F V F A P C P C P C P C P C P C R	Y Y Y Y Y Y Y S K N A N N E	D G G G G G V	V V T V K - P - - - -	G G G G G G G G S S S S S S N	A S S P A P K V A K R S K	T I K K I V A Y F Y L Y	T Q T S S K K K K K K F	- P P P - - - Y Y Y Y Y Y Y H	T G N D F L F Y F Y V	G G A A K Y Y Y Y Y Y Y F	A V V V V P β V V V V V V V V V V V V V	V V T I V I S 8 C C C C C C C C C C	G G G G G G H H H Q Q H	H H H H H H Y Y Y Y Y Y Y	Y T Y T Y T Y T Y T Y T Y T C P C P C P C P C P C P C P C P C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin CAP signatu C Conserved of	R T F T N T Y S F S S T Q I Q I Q V Q V Q V Q V Q V Q V Q V Q V Q V C V S S C C S S S S S S S S S S S S S S S S	V V V V V V V V V V V V V V V V V V V	N S T P S S S W W W W W W W W W W W	Y A D S T Y Y Y Y Y Y Y A	A V V A V A V R K N S K	I 1 I 1 I 1 I 1 I 1 I 1 I 1 I 1	R Q Q Q D Y F N R Y Y Y	L V D V S V S V S V S V S V L 1 Q 1 Q 1 L 1 Q 1 Q 1 L 1	W H W Y W Y W Y W Y W Y W Y U Y U Y C L C L C L C L C L C L C L C L C L C L	$ \begin{array}{c} & \mathbf{N} \\ & \mathbf{D} \\ & $	E E E E A A A A A A A A A C A eli	V R K I R V V V M A V V I Y Of	I S K I K I E I L I S I S X A X A X A X A X N I		F V F Q F S F K F K F K F K F K F V F A P C P C P C P C P C P C R	Y Y Y Y Y Y S K N A N N E	D G G G G G V · N · · Q ·	V V T V K - P - - - -	G G G G G G G G S S S S S S N	A S S P A P K V A K R S K	T I K K I V A Y F Y L Y	T Q T S S K K K K K K F	- P P P - - - Y Y Y Y Y Y Y H	T G N D F L F Y F Y V	G G A A K Y Y Y Y Y Y Y F	A V V V V P β V V V V V V V V V V V V V	V V T I V I S 8 C C C C C C C C C C	G G G G G G H H H Q Q H	H H H H H H Y Y Y Y Y Y Y	Y T Y T Y T Y T Y T Y T Y T C P C P C P C P C P C P C P C P C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin CAP signatu C Conserved of Hinge region	R T F T N T Y S F S S T Q I Q I Q V Q V Q V Q V Q V Q V Q V Q V Q V C V S S C C S S S S S S S S S S S S S S S S	V V V V V V V V V V V V V V V V V V V	N S T P S S S W W W W W W W W W W W	Y A D S T Y Y Y Y Y Y Y A	A V V A V A V R K N S K	I 1 I 1 I 1 I 1 I 1 I 1 I 1 I 1	R Q Q Q D Y F N R Y Y Y	L V D V S V S V S V S V S V L 1 Q 1 Q 1 L 1 Q 1 Q 1 L 1	W H W Y W Y W Y W Y W Y W Y U Y U Y C L C L C L C L C L C L C L C L C L C L	$ \begin{array}{c} & \mathbf{N} \\ & \mathbf{D} \\ & $	E E E E A A A A A A A A A C A eli	V R K I R V V V M A V V I Y Of	I S K I K I E I L I S I S X A X A X A X A X N I		F V F Q F S F K F K F K F K F K F V F A P C P C P C P C P C P C R	Y Y Y Y Y Y S K N A N N E	D G G G G G V · N · · Q ·	V V T V K - P - - - -	G G G G G G G G S S S S S S N	A S S P A P K V A K R S K	T I K K I V A Y F Y L Y	T Q T S S K K K K K K F	- P P P - - - Y Y Y Y Y Y Y H	T G N D F L F Y F Y V	G G A A K Y Y Y Y Y Y Y F	A V V V V P β V V V V V V V V V V V V V	V V T I V I S 8 C C C C C C C C C C	G G G G G G H H H Q Q H	H H H H H H Y Y Y Y Y Y Y	Y T Y T Y T Y T Y T Y T Y T C P C P C P C P C P C P C P C P C P C P
Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin Secondary Structure Chinese cobra - CRISP-natrin Lamprey - Unchar. Tilapa fish - CRISP3 Burmese python - CRISP-like Falcon - Cysteine-rich P. Human - CRISP2 A frican clawed frog - Alhurin CAP signatu C Conserved C Hinge region ICR domain	R T F T N T Y S F S S T Q I Q I Q V Q V Q V Q V Q V Q V Q V Q V Q V C V S S C C S S S S S S S S S S S S S S S S	V V V V V V V V V V V V V V V V V V V	N S T P S S S W W W W W W W W W W W	Y A D S T Y Y Y Y Y Y Y A	A V V A V A V R K N S K	I 1 I 1 I 1 I 1 I 1 I 1 I 1 I 1	R Q Q Q D Y F N R Y Y Y	L V D V S V S V S V S V S V L 1 Q 1 Q 1 L 1 Q 1 Q 1 L 1	W H W Y W Y W Y W Y W Y W Y U Y U Y C L C L C L C L C L C L C L C L C L C L	$ \begin{array}{c} & \mathbf{N} \\ & \mathbf{D} \\ & $	E E E E A A A A A A A A A C A eli	V R K I R V V V M A V V I Y Of	I S K I K I E I L I S I S X A X A X A X A X N I		F V F Q F S F K F K F K F K F K F V F A P C P C P C P C P C P C R	Y Y Y Y Y Y S K N A N N E	D G G G G G V · N · · Q ·	V V T V K - P - - - -	G G G G G G G G S S S S S S N	A S S P A P K V A K R S K	T I K K I V A Y F Y L Y	T Q T S S K K K K K K F	- P P P - - - Y Y Y Y Y Y Y H	T G N D F L F Y F Y V	G G A A K Y Y Y Y Y Y Y F	A V V V V P β V V V V V V V V V V V V V	V V T I V I S 8 C C C C C C C C C C	G G G G G G H H H Q Q H	H H H H H H Y Y Y Y Y Y Y	Y T Y T Y T Y T Y T Y T Y T C P C P C P C P C P C P C P C P C P C P

Figure 10. ClustalW Amino Acid Alignment and Secondary Structure of Selected CRISP Proteins

Secondary Structure																									β5				βŧ	5				
Chinese cobra - CRISP-natrin	s	G	N	F	Q	G	K	Т	A	Т	Ρ	Y	Κ	L	G	Ρ	Ρ	С	G	D	С	Ρ	\mathbf{S}	Α	С	D	N	G	L	С	Т	Ν	Ρ	С
Lamprey - Unchar.	М	G	N	L	Ν	Т	R	Ι	N	R	Ρ	Y	D	L	G	Т	S	С	A	\mathbf{S}	С	Ρ	K	N	С	V	N	N	L	С	D	Ν	Ρ	С
Tilapa fish - CRISP3	Ρ	G	N	-	-	Y	Q	F	Т	Q	Ρ	Y	Κ	Κ	G	Ρ	S	С	G	D	С	Ρ	N	A	С	Е	N	N	L	С	Т	Ν	Ρ	С
Burmese python - CRISP-like	A	G	N	F	Ι	Ν	s	Ι	A	Т	Ρ	Y	K	S	G	Ρ	Ρ	С	G	D	С	Ρ	\mathbf{S}	\mathbf{S}	С	А	Ν	G	L	С	Т	Ν	Ρ	С
Falcon - Cysteine-rich P.	М	G	N	I	R	s	s	Ι	Ρ	Т	Ρ	Y	Κ	Е	G	Е	Ρ	С	G	D	С	Ρ	D	A	С	D	Ν	G	L	С	Т	Ν	Ρ	С
Human - CRISP2	A	G	N	N	Μ	Ν	R	K	Ν	Т	Ρ	Y	Q	Q	G	Т	Ρ	С	A	G	С	Ρ	D	D	С	D	K	G	L	С	Т	Ν	s	С
A frican clawed frog - Allurin	Μ	G	N	Μ	D	Е	s	V	Ρ	R	Ρ	Y	E	E	G	Е	W	С	A	\mathbf{S}	С	Ρ	E	\mathbf{S}	С	D	D	Κ	L	С	D	W	Κ	Ρ
Secondary Structure											α	6											α	7					0	1 8				
Chinese cobra - CRISP-natrin	Т	Ι	Y	N	Κ	L	Т	N	С	D	s	L	L	K	Q	\mathbf{S}	\mathbf{S}	С	Q	D	D	W	Ι	K	\mathbf{S}	N	С	Ρ	А	s	С	F	С	R
Lamprey - Unchar.	K	Y	K	N	G	Y	A	N	С	A	A	М	V	A	Q	Y	Т	С	K	N	A	I	I	L	K	N	С	Ρ	A	s	С	Ν	С	Q
Tilapa fish - CRISP3	Ρ	F	\mathbf{S}	D	K	Y	N	N	С	Ρ	E	L	K	Q	Q	W	G	С	\mathbf{S}	Н	Р	D	V	A	\mathbf{S}	W	С	Ρ	A	s	С	K	С	Т
Burmese python - CRISP-like	Т	Н	Е	D	Т	Y	\mathbf{S}	N	С	Ν	A	L	V	K	Е	Н	K	С	Q	Ι	Ρ	W	I	R	K	\mathbf{S}	С	Ρ	A	s	С	F	С	Т
Falcon - Cysteine-rich P.	K	Y	Κ	D	s	F	E	N	С	R	E	L	Κ	Т	L	L	Т	С	s	N	s	L	М	K	K	-	С	Ρ	A	Т	С	K	С	Т
Human - CRISP2	Q	Y	Q	D	L	L	\mathbf{S}	N	С	D	S	L	Κ	Ν	Т	A	G	С	Е	Н	E	L	L	K	Е	Κ	С	Κ	A	Т	С	L	С	E
African clawed frog - Allurin	Κ	E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Secondary Structure																																		
Chinese cobra - CRISP-natrin	N	K	Ι	Ι																														
Lamprey - Unchar.	G	R	I	-																														
Tilapa fish - CRISP3	N	Q	Ι	Ι																														
Burmese python - CRISP-like	Т	Е	Ι	Κ																														
Falcon - Cysteine-rich P.	Т	E	Ι	Ι																														
Human - CRISP2	N	Κ	Ι	Y																														
African clawed frog - Allurin	-	-	-	-																														
CAP signatu	ıre	e s	eq	ln	en	ce	s				Si	ite	s	of	'n	or	ı-a	lig	gn	ec	l s	ec	qu	en	ice	e o	m	is	sic	on				
C Conserved c	ys	ste	in	es	\$						α	-h	eli	x																				
Hinge region	n									Þ	β-	-sł	ne	et																				
ICR domain	l																																	

Figure 10. MEGA7 ClustalW amino acid sequence alignment and secondary structure predictions for the Chinese cobra cysteine-rich venom protein natrin-1 (CRVP1) (accession ID: Q7T1K6.1, PDB ID: 3MZ8) [88] and other selected CRISP proteins. Alpha helices are represented by green rods, the β -strands are represented by tan arrows. Additionally, the CAP signature sequences are highlighted in orange, the Hinge region is highlighted in blue, and the ICR domain is highlighted in green. The highly conserved cysteines are highlighted in yellow with a red font color. The CAP/PR domain has seven cysteines, of which six are involved in three disulfide bonds; the hinge region has four cysteines and two disulfide bonds; and the ICR domain has six cysteines and three disulfide bonds. These disulfide bonds create a stable tertiary structure that is relatively heat and solvent resistant – a characteristic common to most CAP proteins. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 8. Page 2 of 2

Figure 11. ClustalW Amino Acid Alignment of Invertebrate CRISP2-like Proteins

Common name - Protein Name														An	nino	ь А	cic	1 S	eqı	uen	ice													
Round worm - SCL5	-	-	М	н	Т	s	Ι	L	L	L	Т	A	L	C	A	G	A	Y	A	Q	F	s	A	N	G	Q	A	A	I	L	N	-	-	-
Fruit fly - CG42780																																	С	
Parasitic wasp - serotriflin-like																																	-	
Jewel wasp - CRISP2-like																																	-	
Fire ant - CRISP1-like																																	-	
Dwarf honey bee - CRISP2-like																																	-	
Leaf cutter bee - CRISP2-like																																	-	
Snail - CRISP-like																																	-	
Sea squirt (C. s.) - Unchara.	_																																-	
Sea squirt (C. i.) - CRISP3-like	м	Ν	F	Q	F	G	L	Κ	V	L	F	F	Ι	F	A	Ĺ	s	V	Ι	s	D	A	Ν	Ι	L	N	Е	L	Κ	Н	Е	-	-	-
• • •																															AI			
Round worm - SCL5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	Η	N	Т	L	R	s	R	Ι	A	Κ	G	Т	Y	V	A	Κ	G
Fruit fly - CG42780	V	N	G	s	F	G	S	s	С	Ρ	K	D	A	Т	V	Ι	L	Ν	L	G	D	K	Ν	A	L	Ι	K	A	Н	N	L	V	R	Q
Parasitic wasp - serotriflin-like	-	-	-	L	Y	G	D	Κ	Ι	Ρ	Т	R	V	L	Т	Ρ	s	Ν	K	Κ	V	Q	R	K	Ι	V	L	Υ	н	N	F	F	R	Т
Jewel wasp - CRISP2-like	-	-	-	L	Υ	G	D	S	Ι	Ρ	R	R	V	L	A	Т	S	D	K	Κ	V	Q	R	K	Ι	V	L	Υ	н	N	F	F	R	S
Fire ant - CRISP1-like	-	-	-	L	Y	G	D	R	V	S	V	S	S	Ι	Т	Т	Т	Ν	s	Κ	V	R	Q	K	Ι	V	D	Т	н	N	Y	F	R	Т
Dwarf honey bee - CRISP2-like	-	-	-	L	Y	G	Е	R	V	Ρ	V	Κ	L	L	R	Т	s	Ν	s	Κ	V	R	Q	K	Ι	V	D	Т	н	N	Y	L	R	Т
Leaf cutter bee - CRISP2-like	-	-	-	L	Y	G	D	R	V	Ρ	V	S	L	Ι	R	Т	s	s	s	Κ	V	R	Q	K	Ι	V	D	Т	н	N	Y	F	R	Т
Snail - CRISP-like																																	R	
Sea squirt (C. s.) - Unchara.	-	-	-	Η	R	A	K	R	A	Ρ	G	D	L	L	Е	Т	L	Т	E	A	Е	K	Ι	Е	Ι	V	R	L	Н	Ν	Q	Y	R	s
Sea squirt (C. i.) - CRISP3-like	-	-	-	R	F	R	R	S	L	A	G	A	L	A	V	N	L	Т	E	D	E	Ι	Т	Η	F	V	S	E	Η	N	K	Y	R	R
Round worm - SCL5	Т	A	K	Р	A	A	s	D	М	L	K	М	K	w	D	A	Т	V	A	A	s	A	Q	A	Y	A	N	K	С	Р	Т	G	Н	s
Fruit fly - CG42780	K	W	Α	s	Т	A	С	Κ	М	A	K	М	Е	W	Ν	Κ	D	L	Е	K	L	A	Ι	L	N	A	Κ	Т	С	L	М	G	Η	D
Parasitic wasp - serotriflin-like	H	V	Т	Ρ	Т	A	Α	Ν	М	L	A	М	Κ	W	Η	s	G	A	A	K	A	A	Q	R	W	A	Е	Α	С	Y	A	L	Т	Η
Jewel wasp - CRISP2-like																																	Т	
Fire ant - CRISP1-like	Q	V	Κ	Ρ	Т	A	A	Ν	М	L	Ι	М	K	W	Н	s	G	L	A	K	A	A	Q	R	w	A	D	R	С	L	G	L	V	Η
Dwarf honey bee - CRISP2-like	Q	V	Κ	Ρ	Ρ	A	A	Ν	М	L	A	Μ	Κ	W	Η	Ρ	G	L	A	K	A	A	Q	R	W	A	N	R	С	L	G	L	V	Η
Leaf cutter bee - CRISP2-like																																	V	
Snail - CRISP-like																																	Η	
Sea squirt (C. s.) - Unchara.	R	A	Ν	-	-	A	s	D	М	L	A	Μ	s	W	D	Ν	Т	Ι	A	Q	F	A	s	N	Υ	Ι	s	Ν	С	D	F	A	Η	s
Sea squirt (C. i.) - CRISP3-like	Μ	D	Ρ	-	-	A	s	N	Μ	R	L	Μ	A	W	_	_	_	-	A	A	V	A	K	A	Y	A	Ν	K	С	L	W	K	Η	S
Round worm - SCL5	G	Δ	Δ	G	T	G	F	N	T	v	w	v	w	т		CA A		_	т	N	T	р	0	F	G	Δ	т	G	-	_	S	Δ	A	w
Fruit fly - CG42780																			_				-										K	
Parasitic wasp - serotriffin-like																																	T	
Jewel wasp - CRISP2-like																																	T	
Fire ant - CRISP1-like																																	M	
Dwarf honey bee - CRISP2-like																																	M	
Leaf cutter bee - CRISP2-like																																	M	
Snail - CRISP-like	S	G	R	R	G	F	P	D	T.	P	T.	G	× S	V	G	0	N	v	A	w	S	S	G	ч D	T.	Т	F	T.	G	A	V	0	M	w
Sea squirt (C. s.) - Unchara.																																	A	
Sea squirt (C. i.) - CRISP3-like																																	N	
	1	-	1		1	1	n.	1										Ъ	11	1.41	11	1	11	1	5	1	v	19	Ъ	1	1	1	14	••

CAP signature sequences

Conserved cysteines

ICR domain

Sites of non-aligned sequence omission

Hinge region

Figure 11. Page 1 of 3

Figure 11. ClustalW	Amino Acid Alignment of Invertebrate	CRISP2-like Proteins
---------------------	--------------------------------------	-----------------------------

																				Ţ					C	A	P1					Ľ		
Round worm - SCL5	E	K	E	F	0	D	Y	G	W	S	S	N	Т	L	S	М	s	L	F	I	G	Н	A	Т				W	/ A	K	T	Ν	L	Ι
Fruit fly - CG42780																															s			
Parasitic wasp - serotriflin-like																															Т			
Jewel wasp - CRISP2-like																															' S			
Fire ant - CRISP1-like																															Ъ			
Dwarf honey bee - CRISP2-like																															' T			
Leaf cutter bee - CRISP2-like																															Т			
Snail - CRISP-like																															A			
Sea squirt (C. s.) - Unchara.																															T			
Sea squirt (C. i.) - CRISP3-like																															s			
																			ſ						CA					_	_	1	-	
Round worm- SCL5	G	С	G	V	K	N	С	G	K	D	Т	Ν	-	-	G	F	Ν	K	V	Т	V	V	С	Q	Y	K	P	Q	G	N	Y	L	Ν	Q
Fruit fly - CG42780																															I			
Parasitic wasp - serotriflin-like	G	С	G	W	A	Κ	С	Ν	G	Т	R	G	Ρ	R	G	Q	Ρ	Y	F	S	Y	V	С	N	Y	С	Ρ	A	G	N	Y	Ρ	D	R
Jewel wasp - CRISP2-like	G	С	G	W	A	Е	С	D	G	K	R	G	Ρ	R	G	V	Ρ	Υ	F	s	Y	V	С	N	Y	С	Ρ	A	G	N	R	Е	D	A
Fire ant - CRISP1-like	G	С	G	V	\mathbf{S}	Η	С	Т	G	G	Κ	G	Ρ	L	G	Κ	D	F	Y	М	Y	V	С	N	Y	A	Р	S	G	N	Y	Κ	G	R
Dwarf honey bee - CRISP2-like	G	С	G	V	\mathbf{S}	Η	С	Т	G	G	R	G	Ρ	L	G	Κ	D	F	F	М	Y	V	С	N	Y	A	Ρ	S	G	N	Y	Κ	G	R
Leaf cutter bee - CRISP2-like	G	С	G	V	\mathbf{S}	Η	С	Т	G	\mathbf{S}	R	G	Ρ	L	G	K	D	F	F	М	Y	V	С	N	Y	A	Ρ	S	G	N	Y	Κ	D	R
Snail - CRISP-like																																		M
Sea squirt (C. s.) - Unchara.	G	С	G	А	A	F	С	R	R	А	Т	s	Т	V	G	Y	-	R	L	L	V	s	С	N	Y	G	P	A	G	N	F	L	Ν	N
Sea squirt (C. i.) - CRISP3-like	G	С	G	A	A	F	С	Q	Η	Q	F	A	S	Η	-	-	-	s	Ι	Y	V	V	C	N	Y	G	P	S	G	N	Y	L	D	W
Round worm- SCL5	N	Ι	Y	Т	S	G	Т	Т	С	S	K	С	P	S	G	Т	S	C	E	A	A	Т	G	L	С	A	-	-	-	-	-	-	-	-
Fruit fly - CG42780	V	Υ	E	Е	С	А	K	A	G	Ι	Е	С	A	K	G	Ι	D	Q	Κ	Υ	Р	Ρ	L	С	А	Κ	Ι	-	-	-	-	-	-	-
Parasitic wasp - serotriflin-like																																		W
Jewel wasp - CRISP2-like	L	А	E	Ρ	Y	A	М	G	Е	S	С	\mathbf{S}	A	С	G	G	Н	С	R	L	G	K	L	С	K	N	A	C	Ρ	W	۲A	D	L	W
Fire ant - CRISP1-like	L	G	L	Ρ	Y	V	A	G	K	Ρ	С	\mathbf{S}	М	С	K	D	н	С	Т	Q	D	A	L	С	Т	N	A	C	Y	Н	T	D	L	W
Dwarf honey bee - CRISP2-like																																		W
Leaf cutter bee - CRISP2-like	L	G	Q	Ρ	Y	V	A	G	K	Р	С	S	М	С	Ν	G	Н	С	Т	R	N	K	L	С	Т	N	A	C	Y	Y	T	D	L	W
Snail - CRISP-like	G	D	E	Ρ	Y	Q	L	G	R	Ρ	С	\mathbf{S}	Q	С	R	S	s	С	Q	Н	Ι	R	G	\mathbf{S}	Q	G	R	W	/ G	S	F	Ν	Ι	Ν
Sea squirt (C. s.) - Unchara.	Т	A	Q	Ρ	\mathbf{S}	R	V	G	R	Ρ	F	E	R	Κ	G	A	Ρ	С	S	Q	С	A	Q	-	Q	D	Т	C	V	D) -	-	-	-
Sea squirt (C. i.) - CRISP3-like	N	Т	G	-	-	R	V	G	K	P	Y	-	K	V	G	S	R	C	S	s	C	Т	D	s	Т	D	V	С	S	D	Q	Т	Τ	T
Round worm- SCL5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fruit fly - CG42780	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Parasitic wasp - serotriflin-like	Α	N	С	R	Q	L	Y	Е	Т	W	Ρ	Ν	W	L	С	Е	S	D	Т	Q	Q	G	н	E	R	R	Q	F	С	R	A	Т	C	R
Jewel wasp - CRISP2-like	Α	N	С	Q	Q	L	R	A	Т	W	P	G	W	L	С	D	Т	D	Т	Е	Q	G	R	E	R	R	Q	F	С	R	A	Т	C	R
Fire ant - CRISP1-like	\mathbf{S}	N	С	A	E	L	v	R	Т	F	G	s	w	V	С	Е	Т	D	Т	Е	Е	G	R	E	R	R	K	F	С	G	A	Т	C	Ν
Dwarf honey bee - CRISP2-like	S	N	С	V	E	L	A	R	М	F	R	s	W	V	С	E	Т	Ν	Т	S	Е	G	R	E	R	R	Κ	F	С	G	A	Т	C	Ν
Leaf cutter bee - CRISP2-like	s	N	С	A	E	L	V	R	Т	F	R	A	w	V	С	Е	Т	N	T	K	Е	G	R	E	R	R	K	F	С	G	A	Т	C	Ν
Snail - CRISP-like																																		Α
Sea squirt (C. s.) - Unchara.																			-													-	-	-
Sea squirt (C. i.) - CRISP3-like	Т	N	Y	V	Ρ	F	W	Т	N	V	Ρ	F	W	F	Y	Ι	Т	N	N	G	Q	R	L	s	Ρ	R	Q	Ι	Q	Ν	L	L	Ν	Q
																					-													

CAP signature sequences

C Conserved cysteines

ICR domain Sites of non-aligned sequence omission

Hinge region

Figure 11. Page 2 of 3

Round worm - SCL5	-	-	-	-	-	-	-	-	-	-		
Fruit fly - CG42780	-	-	-	-	-	-	-	-	-	-		
Parasitic wasp - serotriffin-like	C	R	D	K	I	V	-	-	-	-		
Jewel wasp - CRISP2-like	C	D	G	K	I	Ι	-	-	-	-		
Fire ant - CRISP1-like	C	Κ	D	K	I	Y	Y	Η	Η	Е		
Dwarf honey bee - CRISP2-like	С	Е	G	K	I	Y	Y	Н	Н	G		
Leaf cutter bee - CRISP2-like	C	Κ	G	K	I	Y	Y	Н	Н	G		
Snail - CRISP-like	L	R	R	Р	Q	Η	W	С	Q	Y		
Sea squirt (C. s.) - Unchara.	-	-	-	-	-	-	-	-	-	-		
Sea squirt (C. i.) - CRISP3-like	V	S	W	L	Ν	G	Q	Ν	W	Ρ		
CAP signature s	eq	ue	nc	ces	5						CR domain	
C Conserved cyste	eine	es									ites of non-aligned sequence omissio	m

Figure 11. ClustalW Amino Acid Alignment of Invertebrate CRISP2-like Proteins

Figure 11. MEGA7 ClustalW amino acid sequence alignment of invertebrate CRISP2like proteins. The CAP signature sequences are highlighted in orange, the Hinge region is highlighted in blue, and the ICR domain is highlighted in green. The highly conserved cysteines are highlighted in yellow with a red font color. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 9. Page 3 of 3

Hinge region

Mammalian CRISP1: In many vertebrates and all mammals, spermatozoa leaving the testes are not competent to fertilize an egg. Mammalian sperm mature in the epididymis and reach a capacitated state in the female genital track. After capitation, with the help of chemoattractants released by the egg's cumulus oophorus, sperm will locate the egg and bind to the zona pellucida, thereby triggering the acrosome reaction and release of proteases that dissolve the zona pellucida. This clears a path for sperm delivery and fusion of the sperm with the egg plasma membrane [27].

In mice, CRISP1, a 32 kDa protein expressed in the epididymis, plays an important role in fertilization. CRISP1 binds to the sperm surface in the epididymis, migrates to an equatorial position during sperm capacitation and subsequently appears to play a role in both sperm-zona pellucida interaction and gamete fusion. In addition, sperm from

CRISP1^{-/-} knockout mice, exhibit lower levels of protein tyrosine phosphorylation during capacitation and significantly reduced ability to penetrate both intact zona pellucida and to fertilize zona pellucida-free eggs during in vitro fertilization [28].

Similarly, in humans CRISP1 is a major glycoprotein expressed in the epididymis and participates in sperm-egg fusion [29]. Its epididymal origin underlies the fact that CRISP1 is absent in the seminal plasma of individuals with obstructive azoospermia (OA) – a blockage of extratesticular ducts [30] – and its absence serves a good indicator for OA in males.

Mammalian CRISP2: Mouse CRISP2, also known as Tpx-1, is only expressed in the testis, where it has been associated with mediating the binding of spermatogenic cells to Sertoli cells [31]. In addition, CRISP2 has been localized to the sperm acrosomal granule, a single secretory granule that undergoes exocytosis during the acrosome reaction [32]. Antibody binding experiments have strongly supported the involvement of CRISP2 released at that time in sperm-oocyte binding [33-35]. Indeed, antibody inhibition of CRISP2 significantly decreases the sperm's ability to penetrate the zona pellucida [32, 35]. CRISP2 has been implicated in the initiation of Ca²⁺ fluxes observed during sperm capacitation. The ion channel regulatory domain has been hypothesized to activate ryanodine receptor 1 (RyR1) and inhibit RyR2 when applied to the cytoplasmic domain of the receptor. When applied to the luminal domain, CRISP2 can promote the activation of both RyRs [32]. The location of RyRs in smooth endoplasmic reticulum at the neck of the sperm raises the possibility that CRISP2 is involved with sperm motility and/or the acrosomal reaction.

30

NMR solution structure of the CRD domain of CRISP2 shows a high structural homology to other CAP family proteins such as snake venom CRISP proteins and invertebrate toxins whose structure has been determined by X-ray crystallography [6, 32] (see Figure 9). Therefore, my analysis of CAP family gene structure in the following chapter will focus first on genes that code for CRISP2 or CRISP-2 like proteins over a wide range of species.

Mammalian CRISP3: The mRNA for CRISP3, first identified in the mouse salivary gland as an androgen-dependent transcript, is widely expressed in the plasma, pancreas, prostate, and B-cells [36-39]. CRISP3 appears to play a role in immunological responses. Recent studies have shown that CRISP3 can influence Hepatitis C virus (HCV) resistance; at the early phase of infection, the presence of CRISP3 limited HCV replication in culture medium [38].

In addition, CRISP3 expression is upregulated in prostate cancer cells [39, 40]. In fact, the strong expression of CRISP3 in prostate is a good indicator of advanced tumor stages and a high Gleason score [41]. Further studies have shown that β microseminoprotein (MSMB) forms a non-covalent complex with CRISP3 in seminal fluid and serum. MSMB it is now be considered as a potential biomarker for prostate cancer. In women, its expression is significantly reduced in ovarian invasive neoplasms; whereas CRISP3 expression is elevated [41, 42]. Unlike CRISP1 and CRISP2, CRISP3 has not yet been implicated in sperm-egg interaction despite its presence in semen.

Mammalian CRISP4: Mammalian CRISP4 is a unique and somewhat controversial protein. CRISP4 is unique because it has only been identified and annotated in two species - mouse and rat; whereas it is controversial because mouse and rat CRISP4

(rather than mouse and rat CRISP1) share a higher sequence, exon structure, and predicated protein structural homology to CRISP1 of other mammalian species [6, 43, 44]. For example, the sequence homology between mouse CRISP4 and human CRISP1 is 59%, considerably higher than the 40% homology between mouse CRISP1 and human CRISP1 [43].

Furthermore, the CRISP4 signal sequence suggests that it is secreted into the epididymal lumen and interacts with sperm [43]. Later studies have identified that CRISP4 is expressed in the caput and corpus of the epididymis, which is similar to CRISP1 expression in other species [44].

Truncated CRISP (Allurin)

Allurin: Allurin, also known as CRISP A, is a 184-amino acid sperm chemoattractant protein from *Xenopus* egg jelly [12]. In *Xenopus laevis*, allurin is exclusively expressed in the female oviduct and is produced and secreted in the oviduct in a region-specific manner [45]. Allurin is expressed in the first third of the pars convoluta and secreted by the superficial ciliated epithelial cell layer, where it is brushed onto the egg surface along with other jelly components as the egg passes [45]. Subsequently, as *Xenopus* eggs are spawned into pond water, the jelly layer swells and releases small diffusible proteins including allurin. Allurin, binds to the sperm surface and is hypothesized to regulate flagellar calcium signaling thereby orienting and guiding the sperm up the allurin gradient, ultimately leading it to the egg [46].

32

The LCCL Domain-containing Subfamily

Mammalian CRISPLD1: CRISPLD1, which is also known as CAPLD1, CRISP LCCL1, CocoaCRISP, and CRISP10, has not been extensively studied or characterized. CRISPLD1 contains three domains: the CAP/PR domain, the Hinge region, and an LCCL domain. The LCCL (Limulus clotting factor C, Coch05b2 (Cochlin) and Lg11) domain is generally found in extracellular proteins in conjunction with other modular domains, like CAP/PR and C-type lectin modules [47].

CRISPLD1 and CRISPLD2 interact with folate pathway genes [48]. Variation in CRISPLD1 is considered to be a contributing factor for Non-Syndromic Cleft Lip with or without cleft Palate (NSCLP) through its interaction with CRISPLD2 and the folate pathway genes [48]. LCCL domain-containing proteins may also be involved in antibody independent host defense, via triggering anti-microbial activity [47].

Mammalian CRISPLD2: CRISPLD2, which is also known as the CAPLD2 and Late Gestational Lung protein I (LGL1), is expressed in human, rat and mouse fetal lungs during late gestation, where it plays a crucial role in the regulation of mesenchymalepithelial interaction during formation of alveoli [49, 50]. Indeed, disruption of CRISPLD2 mRNA results in inhibition of normal lung branching morphogenesis and results in dilated distal lung buds [49].

CRISPLD2 shares very high homology to CRISPLD1, even though they have unique expression patterns and functional characteristics [51]. Knockout studies of CRISPLD1 in mice show a complex respiratory phenotype including delayed histological maturation, goblet cell hyperplasia, fragmented elastin fibers, and elevated expression of T_H2

cytokines [52]. CRISPLD2 haploinsufficiency may lead to lung disease in and to increased risk for late-onset respiratory disease [52].

CRISPLD2 is also expressed in the craniofacies of developing mouse embryos and is potentially involved in non-syndromic cleft lip with or without cleft palate (NSCLP) in Chinese, Caucasians, South Americans, and Hispanics [51, 53]. However SNPs in CRISPLD2 alone may not lead to NSCLP (this study focused solely on Italian populations) [54].

The GLIPR/GLIPR-like Subfamily

Mammalian GLIPR1: GLIPR1 (Glioma Pathogenesis Related protein 1), also known as RTVP-1 (Related to Testis specific, Vespid and Pathogenesis related-1), was originally discovered in glioblastoma multiforme/astrocytoma due to elevated expression in brain tumors [55]. In addition, GLIPR1 expression is significantly increased in acute myeloid leukemia bone marrow samples, whereas it is markedly reduced in acute lymphoblastic leukemia, and slightly decreased in chronic lymphocytic leukemia and chronic myelocytic leukemia relative to normal levels [55, 56]. GLIPR1 can serve as an indicator for human myelomonocytic differentiation and various other types of cancers and tumors due to significantly altered expression patterns [56-59]. Though proliferation of GLIPR1 has been associated with Wilms' tumors [59], glioblastoma [58], and myeloid leukemia [56], GLIPR1 expression appears to suppress prostate cancer [57, 60]. Its gene has been identified as a *p53* target gene, which is widely associated with tumor suppression, cell cycle arrest, DNA repair, senescence, cell differentiation, and apoptosis [61, 62]. GLIPR1 is also involved in the endoplasmic reticulum (ER) secretory protein pathway, and affects G protein signaling and cell cycle regulation. GLIPR1 is likely a transmembrane protein of the ER which aids in budding of transport vesicles destined for the Golgi. GLIPR1 RNA knockdown studies have demonstrated down regulation of protein synthesis of products related to the ER-to-Golgi vesicle-mediated transport [56].

GLIPR1 arises from a single, well supported cluster composed of three distinct subclades – GLIPR1, GLIPR1L1 and GLIPR1L2 [6]. As shown in Figure 2, GLIPR1 is the shortest of these proteins, containing a signal sequence, a CAP/PR domain, Hinge region, and a transmembrane domain. GLIPR1 is widely expressed in many organisms, including vertebrates and invertebrates.

Mammalian GLIPR1L1: The GLIPR1L1 (Glioma Pathogenesis Related protein 1 Like 1) gene is recognized as a *p53* target gene in mammals, and its expression has high tissue-specificity to the testis. GLIPR1L1 shares a very high sequence homology to GLIPR1 and GLIPR1L2. The N-terminus signal peptide and the extracellular protein signature motifs suggest that GLIPR1L1 is located on the surface of the cell membrane or is secreted [61].

Mammalian GLIPR1L2: The GLIPR1L2 (Glioma Pathogenesis Related protein 1 Like 2) gene is also recognized as a *p53* target gene, and is highly expressed in the testis, but is also expressed at lower levels in a wide array of tissue types, including the prostate and the bladder [60, 63]. GLIPR1L2 shares a very high sequence homology to GLIPR1 and GLIPR1L1. The presence of a C-terminus membrane-spanning domain suggests that GLIPR1L2 is also a transmembrane protein of the ER [61]. Thus, GLIPR1, GLIPR1L1, and GLIPR1L2 expression can all be induced by p53 and these proteins are considered to be tumor suppressors with apoptotic function [60, 63]. GLIPR1's pro-apoptotic function arises from its role in increasing the production of reactive oxidative species and in activating the c-jun-NH₂ kinase pathway via the apoptosis signal-related kinase and the mitogen-activated protein kinase [63]. In addition to this, there is a strong correlation between the expression of GLIPR1, GLIPR1L1, and GLIPR1L2 and down regulation of human sarcoma, lymphoma, prostate, bladder, lung, and colon cancer cell lines [60-63].

The GAPR Subfamily

GAPR-1 (Golgi-associated Plant Pathogenesis Related Protein 1), which is also known as GLIPR2, RTVP-1 and COL4A3, has a high expression in immune-related tissues and cells, especially in monocytes, leukocytes, lung, spleen, and embryonic tissues [64]. Therefore, GAPR-1 is believed to play an important role in the innate immune system of mammals, similar to the anti-fungal and serine protease activity associated with the Pr-1 protein in plants [64, 65]. However, it should be noted that the similarity of function between GAPR-1 and Pr-1 is merely a coincidence rather than a conserved function. GAPR-1 is localized to lipid-enriched microdomains in the Golgi complex of mammalian cells and is tightly bound to the cytosolic leaflet of the Golgi membrane [64]. GAPR-1 was initially believed to be a non-secretory protein [64, 66] since it appears to lack a conventional N-terminal signal sequence; however recent studies show that GAPR-1 can be secreted as well [67]. GAPR-1 is also found enriched in the lumen of small prostasomes, an array of membrane vesicles produced by the prostatic epithelium [68]. Prostasomes are thought to help regulate sperm motility through interaction with Ca²⁺ delivery signaling tools [69], and may play a role in stimulating the acrosome reaction [70], and protecting the sperm cells from immune attack within the female reproductive track [71].

Despite the similarity in nomenclature between GLIPR1, GLIPR1L1, GLIPR1L2, and GLIPR2 (an alternative name and misnomer for GAPR-1), GAPR-1 does not have any phylogenetic, structural, or functional similarity with the others [6]. In addition, GAPR-1 protein does not contain a predicted signal sequence, Hinge domain or third domain as do the others [6, 64]. It has been previously suggested that GAPR-1 may be the most primitive CAP family protein sharing high amino acid sequence homology with invertebrate venom proteins, the Pr-1 proteins found in plants, and the PRY1 proteins found in yeast *Saccharomyes cervisae* [72]. As previously discussed, Pr-1 proteins serve an immune function plants and similarly GAPR1 appears to serve a function in the mammalian immune system; this may suggest a link between plant and mammalian immune systems, however no data have provided a sufficient link as of yet.

Though GAPR-1 is viewed as a CAP superfamily protein with Pr-1 ancestry [7, 66, 67, 73], proteomic analysis of more highly evolved vertebrate "GLIPR-2", using amino acid and secondary structure data, suggests that the latter is an extension of the original GAPR lineage.

The Peptidase Inhibitor Subfamily

Mammalian PI15: Peptidase Inhibitor 15 (PI15), also known as 25 kDa trypsin inhibitor, SugarCRISP, P25TI, and CRISP8, was first identified in human glioblastoma cells [74]. In humans, PI15 is expressed in the mammary gland, prostate, salivary gland, thyroid gland, brain, placenta, and lymphocytes [14, 75]. PI15 has a low affinity to trypsin, compared to other trypsin inhibitors; similar to GLIPR1, it is highly expressed in human neuroblastoma and gliolastoma cell lines [14, 74]. Elevated levels of PI15 are found in the prostatic secretions of individuals with prostate cancer [76].

In rats, reduced PI15 expression can result in increased protease activity in the aorta resulting in ruptures of the internal elastic lamina of the abdominal aorta and iliac arteries [77]. In the developing chicken embryo (stage 18), PI15 expression can also be observed in emerging lung buds, dorsal pancreatic mesoderm, and the gut; during stage 21 of the developing embryo, PI15 appearance can be observed in the anterior and posterior necrotic zones in the limb bud [75]. The timing and location of these expression patterns during embryogenesis suggest that PI15 may be involved in regulation of protease action during tissue remodeling. However, much remains to be done as the regulatory and signaling pathways for PI15's actions are still undefined.

Mammalian PI16: Peptidase Inhibitor 16 (PI16), also known as PSP94 (prostate secretory protein of 94 amino acids), CRISP9, β -microseminoprotein, PIP (Prostatic Inhibin Peptide), and protease inhibitor 16, is a major component of semen [78]. However, PI16 is found in a wide verity of tissues, including prostate, small intestine, colon, peripheral blood leukocytes, pituitary gland, parathyroid gland, tonsil, kidney, stomach, liver, and the Leydig cells within the testis [78]. PI16 has been implicated as a

38

modulator of circulating FSH (Follicle-Stimulating Hormone) levels in sheep seminal plasma [79], a competitive inhibitor of sperm motility by obstructing the activity of Na⁺, K⁺-ATPase [80], an immunoglobulin binding factor in female reproductive tract [81], a promoter of prostate cancer cell apoptosis [82, 83], a regulator of calcium levels during hypercalcaemia of malignancy [83], and an inhibitor of cardiomyocyte growth [84].

CHAPTER 3: EVOLUTION OF CAP SUPERFAMILY GENE STRUCTURE

Evolution of CAP Superfamily Gene Structure

The CAP/PR domain is expressed in prokaryotes and eukaryotes organisms with a high degree of identity and coverage. This suggests that all CAP proteins, through the CAP/PR domain, share a common ancestor and possibly a single point of origin. As presented in the table of Appendix 1A and Appendix 1B, a large array of expressed and hypothetical CAP-related protein sequences are present in organisms representing a wide range of kingdoms and phyla. The abundance and range of CAP/PR-related sequences in an organism does not correlate with genome size, protein count, or other genomic characteristics. This follows the generally observed evolutionary rule (known as the 'Cvalue paradox' and 'N-value paradox') that the amount of genetic material in an organism does not necessarily relate to increased complexity or gene expression [85, 86]. Nevertheless, with increasing organism complexity, CAP/PR proteins are observed to take on additional domains suggesting an increase in both their structural and functional repertoire (see Figure 2). In prokaryotes and early eukaryotes, the CAP superfamily proteins are essentially one domain proteins. Subsequently, invertebrate round worms and insects develop CAP proteins to which the Hinge region/domain has been added. Subsequently, in later invertebrates the CAP/PR domain and Hinge region are usually expressed with a third domain: an ion channel regulatory (ICR) domain, a transmembrane domain, a LCCL domain, a CLEC domain, a glutamate rich domain, or a ZipA domain.

In order to place the evolutionary history of CAP proteins and genes in perspective, I have chosen a set of representative species based on 1) their coverage of all major phyla

in the eukaryotic world with emphasis on vertebrates, and 2) species whose genome sequences are essentially complete and present in databases accessible through the NCBI, Ensemble, Xenbase, and Santa Cruz Genome Browser websites. A catalogue of these species and proteins, representing prokaryotes, plants, fungi, invertebrates and vertebrates including fish, amphibian, reptiles, birds and mammals, can be found in Appendix 1 and the Supplementary Data tables while their overall taxonomic relatedness can be visualized in the phylogenetic tree of Figure 12. The tree was generated using phyloT software implemented within the Interactive Tree of Life (iTOL) v2 website. Tree construction with phyloT relies on genome-wide sequence data from NCBI, phylogenetic analysis, and morphological studies [87, 88]. The tree was rooted using outgroup rooting network.

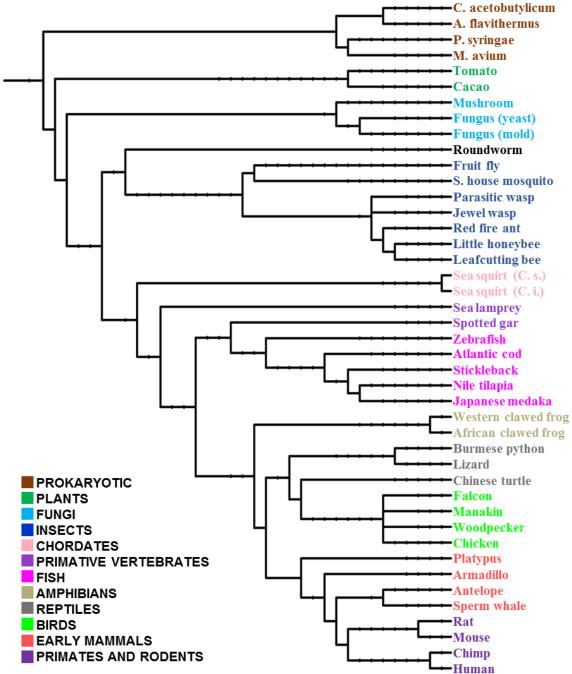


Figure 12. Global Genome Relatedness Between Selected Organisms in This Study

Figure 12. This rooted taxonomic tree represents the relationship between many organisms used in this dissertation. This tree of life was generated using phyloT implemented in the Interactive Tree of Life (iTOL) v2. phyloT creates phylogenetic trees based on NCBI taxonomy databases, which rely on sequence data, phylogenetic analysis, and morphological studies [84,85].

Conservation of Exon Borders within CAP Subfamily Genes

Given evidence of CAP/PR domain relatedness in phylogenetic trees (Figure 1) I sought to obtain further evidence for gene evolution by looking at conservation of exon borders. Using genomic sequences, I tallied the number of exons present in each CAP gene and sequence information at exon/intron borders to determine to what extent these had been conserved. The strongest conservation of gene structure came in the vertebrate CRISP genes within which the number of exons remained invariant at seven in every class of vertebrate (Figure 13 and Appendix 2). Furthermore, amino acid sequences found at exon/intron borders remained strikingly consistent as shown in Figures 13 and 14. As shown in Figure 14, there is a high degree of conservation of border position in the amino acid sequence, most notable at the vertebrate exon borders between exons 2 and 3, 3 and 4, 4 and 5, 5 and 6, and 6 and 7. In fact, amino acid codons that contain an internal ligation site between exons are also highly conserved as indicated by white boxed amino acids at the exon borders between exons 3 and 4, 5 and 6, and 6 and 7 in Figure 14.

In contrast, exon numbers and borders underwent marked changes in earlier stages of evolution leading up to CRISP genes as demonstrated in Figure 15. Pr proteins in plants and CRISP-like proteins of low homology in fungi, harboring a single CAP/PR domain (see Figure 4 and Figure 5), are coded for by genes that have only one exon (Figure 15). Some genes coding for CRISP-like proteins of fungi do exhibit 2 or 3 exons but the exon pattern is quite different from those of CRISP genes described above (e.g. mold in Figure 15).

Exon 1 Exon 2 Exon 3 Exon 4 Exon 5 Exon 6 Exon 7
Lamprey: CRISP-like
Gar: CRISP-like
Zebra fish: CRISP3
Atlantic cod: CRISP3
Stickleback: CRISP3
Tilapia: CRISP3
Medaka: CRISP3
W. clawed frog: allurin 💦 💼 💼 💼 💼 🖬 🖬 🖬 🖬 👘 🖬 👘
A. clawed frog: allurin
Burmese python: CRISP
Anole lizard: C-rich P.
Turtle: C-rich P.
Falcon: C-rich P.
Gold manakin: CRISP2
Woodpecker: CRISP2
Chicken: C-rich P.
Armadillo: C-rich P.
Platypus: CRISP1 [] [_] [
Mouse: CRISP2
Rat: CRISP2
Chimp: CRISP2
Human: CRISP2
Domain Organization Image: CAP Domain Signal CAP Domain
Figure 13.

Figure 13. Genomic Exon Border Structure in Selected Vertebrate CRISP2 and CRISP2-like Genes

Figure 13. Page 1 of 2

Figure 13. Genomic Exon Border Structure In Selected Vertebrate CRISP2 and CRISP2-like Genes

Figure 13. Genomic exon border structure and alignment of CRISP2 and CRISP2-like genes. Proteins sequences were aligned using ClustualW. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online databases for specific CAP genes. Color-coding is used to indicate exon order and relative length. Exon borders were located almost identically for all vertebrate genes coding for CRISP 2/CRISP 2-like proteins. Organisms and accession IDs of the amino acids are available in supplemental data table 7. Page 2 of 2

Figure 14. ClustalW Amino Acid Alignment for Exon Borders in Selected CRISP and CRISP-like Genes

	EAULI 1/2	Exon 2/3	Exon 3/4	border	Exon 4/5	Exon 5/6	Exon 6/7	
	border	border	border	invertebrate	border	border	border	
Common name - Protein name			Amino Ac	Acid Sequence				
Parasitic wasp - Serotriftin-fike	VLTP	-\\- L A M K W H -\\-	LHIDFGSC	-\\- Q V P W F F A -\\-	YTQMVW	G N Y	-\\- <mark>G L <mark>C</mark> K N S <mark>C</mark> -\\-</mark>	KIV
Jewel wasp - CRISP2-like	LAT	-\\- L V M K W H -\\-	TDFGS <mark>C</mark>	-\\- R V P W F F A -\\-	YTQMVW	P A G N R	-\\- G L C K N A C -\\-	KII
Fire ants - CRISP1-fike	SITT	-//- L I M K W H -//-	Y L D F G Q S	Ч	YTQVVW	GNY	-\\- D L C T N A C -\\-	KIY
Dwarf honey bee - CRISP-like	P T P R L Y G E R V P V K L L R T	-//- L A M K W H -//-	YLDFGQS	LWNF	YTQIAW	A P S G N Y	L C	
Leaf cutter bee - CRISP2-fike	NKPTPRLYGDRVPVSLIRT	-\\- <mark>L V M K W H</mark> -\\-	LYLDFGQS	-\\- Q T L W N F A -\\-	YTQIAW	GNY	-\\- N L <mark>C</mark> T N A <mark>C</mark> -\\-	K I Y
								_
ichar.	A G A S A W D L V N D W K S L K T	LQMMWR	V T T K M T C	4	T A T A M	I C L M C N L	N L CUN N	_
Gar - Unchar.	EIST	-\\- L K M G W N -\\-	R R I K T S E <mark>C</mark>	-\- D S T W N E V -\-	YTQVW	-\\- <mark>Y C</mark> P L G N Y .	-//- R L <mark>C</mark> T N P <mark>C</mark> -//-	KII
Zebra fish - CRISP3	V C T	-\\- L K M S W S -\\-	RMLNGYEM	-\- I S S W T S V -\-	Y T Q V V W	-\\-YYRAGNF	L C	KII
Cod - CRISP3		-//- L K M I W S -//-	YEL	-\- A L D W A T V -\-	YTQVVW	-//- Y Y R A G N Y -	-//- K L C T N P C -//-	EII
Stickleback - CRISP3	I C P	-\\- L L M S Y S -\\-	RMLNGYEL	-//- P Y P W T A V -//-	YTQVVW	-\\-YYRAGNF	-//- K L C T N P C -//-	EII
Tilapa fish - CRISP3	L L C A L G L T A V L Q V P G T L G V -	-\\- L K M N W N -\\-	R V I S T S G C	-//- V N Z W Z N V -//-	Y T Q V V W	-\\- Y <mark>C P</mark> P G N Y .	-//- N L C T N P C -//-	ΙIδ
Medaka - CRISP3	LHIDFTY I <mark>C</mark> P	-\\- L M M S Y S -\\-	GYEL	-//- P M L W K S I -//-	YTQVW	-\\-Y <u>Y</u> RAGNF.	-//- K L C T N P C -//-	
Western clawed frog - Allurin	TPM	-\\- L K M E W S -\\-	V C	-\\- K P D W A A A -\\-	Y T Q V A W	-\\- <mark>Y</mark> C P M G N M .	-//- K L C T S N P -//-	1 1 1
African clawed frog - Allurin	YSTYGDDT <mark>GST</mark>	-\\- K K M V W C -\\-	DVVC	-//- K S D W S T V -//-	Y T Q V V W	-\\- Y G P M G N M	-//- K L C D W K P -//-	•
Burme se python - CRISP-like	QQ S S G D D G - F A A E S T -	-\\- L R M E W H -\\-	ЕLQ <mark>C</mark>	-//- P Y S W T A V -//-	Y T Q V V W	-\\- Y <mark>C P</mark> A G N F -	ΝP	EIK
Anole lizard -Cysteine-rich P.	HQ SGIKAEGD - IDALMT	-\\- L K M E W D -\\-	G K K C	-\\- P N P W S D A -\\-	Y T Q V V W	-\\- Y S P A G N R	-//- G L C T N P C -//-	EIQ
Softshell turtle - Unchar.	SLST	-//- L R M E W S -//-	RST-STGC	-\\- P H S W S D A -\\-	Y T Q V V W	GNY	-//- G L C T N P C -//-	EIK
Falcon - Cysteine-rich P.	P S I G E E P E G - L D A L S T	-\\- L K M A W C -\\-	RRT-TTQ <mark>C</mark>	-\\- P F S W P D V -\\-	Y T Q M V W	G N I	-\\- G L C T N P C -\\-	
Golden manakin - CRISP2	PP SSAQEAGG - VDALST	-\\- M K M E W C -\\-	RRT-DVL <mark>C</mark>	-\- P F S W S D V -\-	Y T Q <mark>M V W</mark>	GNL	G L <mark>C</mark>	
Woodpecker - CRISP2	P S A G Q A P E D - F D A L S T	щ	V G C	PFPW	Y T Q V V W	N L	G L <mark>C</mark> T N P <mark>C</mark>	ΔII
-rich P.	LLST	-//- L R M E W S -//-	Q P C	-\\- B S X S D S -\\-	YTQLVW	-\\- Y <mark>C P</mark> A G N I -	G L C	
Platypus - CRISP1	PISILGVGPPRVPYEAIST	-\\- L K M E W N -\\-	RQIKDHS <mark>C</mark>	-\- S L S W S D A -\-	ΥΤQ <mark>ĽVW</mark>	-\\- Y C H A G N D -	-//- R L C T N P C -//-	- - -
Armadillo - Cysteine-rich P.	LTSLPADGKDPG - FSALLT	-//- L K M E W S -//-	RKT-NTKC	-\\- P S S W S T G -\\-	ΥΤQ <mark>VVW</mark>	-N-YCPAGNV	-//- G L C T N S C -//-	KIF
Antelope - CRISP2	LPSFPTEGKDPS - FTALIT	-\\- L K M E W S -\\-	RKT-STKC	-\\- P M A W S D A -\\-	YTQLVW	-\\- Y <mark>C</mark> P A G N N .	-//- G L C T N T C -//-	KIY
Sperm Whale - CRISP2	L P S F P T E A K D P S - F T A L L T	-\\- L K M E W N -\\-	RKT-STK <mark>C</mark>	-\\- P T A W S D A -\\-	ΥΤQ <mark>LVW</mark>	-\\- Y <mark>C</mark> P A G N N -	-//- G L C T N S C -//-	KIY
Mouse - CRISP2	LRSPLTEGKDPD - FTSLLT	-//- L K M E W S -//-	RKI-NIR <mark>C</mark>	-\\- P T L W S T V -\\-	Y T Q L V W	-\\- Y <mark>C</mark> P M G N N .	-//- G L C T N S C -//-	KIH
Rat - CRISP2	LPLPPTEGKDPD - FATLTT	-\\- L K M E W N -\\-	RKI-NIK <mark>C</mark>	-\\- P T S W R T V -\\-	ΥTQLVW	G N N	-//- G L C T N S C -//-	
Chimp - CRISP2	L P S L P A E G K D P A - F T A L L T	-\\- L K M E W S -\\-	T R C	-/- B T S W S S A -/-	YTQLVW	CPAGNN	-//- G L <mark>C</mark> T N S <mark>C</mark> -//-	KIY
Human - CRISP2	L P S L P A E G K D P A - F T A L L T	-\\- L K M E W S -\\-	RKT-STR <mark>C</mark>	-\\- P T S W S S A -\\-	ΥΤQ <mark>ĽVW</mark>	-\\- Y C P A G N N .	-//- G L <mark>C</mark> T N S <mark>C</mark> -//-	KIY

Figure 14. ClustalW Amino Acid Alignment for Exon Borders in Selected CRISP and CRISP-like Genes

Figure 14. Exon borders in selected CRISP and CRISP-like genes. Protein sequences were aligned using MEGA7 ClustalW. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online databases for specific CAP genes. Color-coding superimposed on the sequences is used to represent different exons and exon borders are indicated by a change in color. White (not highlighted) residues represent intra-codon borders. Exon borders within both insect CRISP genes (upper panel) and vertebrate CRISP genes (lower panel) are identical between members of the same taxonomic group. In addition, exon borders 2/3 and 5/6 are consistent between insect CRISP genes and vertebrate CRISP genes. In contrast, other borders are located differently in the two taxonomic groups. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 7.

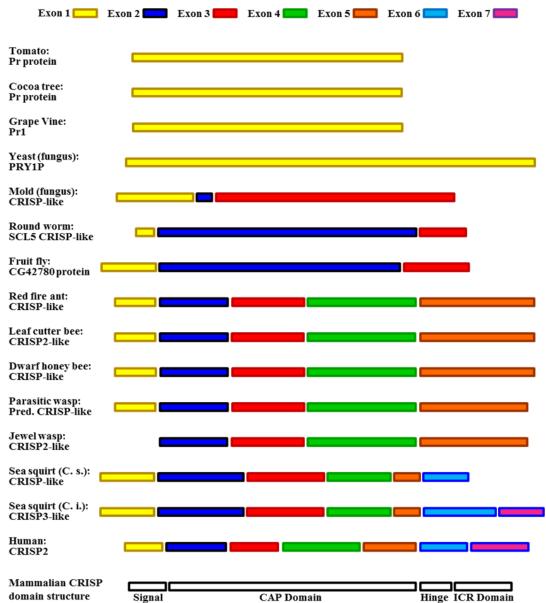


Figure 15. Genomic Exon Border Structure In Selected Invertebrate CRISP and CRISP-like Amino Acid Sequences

Figure 15. Genomic exon border structure of selected CAP proteins in invertebrates and human CRISP2 (for reference). Proteins sequences were aligned using ClustualW. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online databases for specific CAP genes. Color-coding is used to indicate exon order and relative length. The consistent exon numbers and borders seen in CRISP proteins of vertebrates (e.g. human CRISP 2, bottom entry) is not identical to that seen in invertebrates. Exon numbers appear to increase during evolution and some exon borders remain consistent while others either shift or appear for the first time. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 10. As one progresses into invertebrate CAP genes, for example the insect venom genes, exon numbers become variable from 1 to 5. This observation of exon number variability is also true for other invertebrate phyla including round worms, mollusks, echinoderms and urochorates (see Appendix 3). Finally, arriving at vertebrates, most CAP genes exhibit five or more exons. As shown in Figure 16, this overall increase in exon number results not just from the addition of new domains, but more pointedly from the number of exons representing the CAP/PR domain. Exons coding for this domain alone usually number from 4 to 6 in vertebrate CAP genes (Figure 16).

Figure 16. Evolution of CAP Proteins is Accompanied by an Increase of Exons Representing the CAP/PR Domain

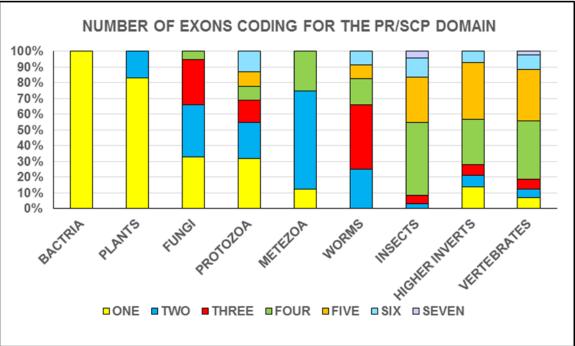


Figure 16. Evolution of CAP Genes is accompanied by an increase of exons coding for the CAP/PR domain. Ten to twenty-five CAP genes were sampled in each taxonomic group, for a total of 148 genes. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online databases for specific CAP genes.

A second observation is that the conservation of exon borders seen in the CRISP genes (Figures 13 and 14) occurs only within the CRISP subfamily. Exon borders and numbers differ markedly between CAP subfamilies.

To demonstrate this observation we will use the western clawed frog (*Xenopus tropicalis*), whose genome presents fourteen CAP superfamily genes. In Figure 17, the phylogenetic tree represents the nearest neighbor joining alignment of all fourteen western clawed frog CAP superfamily amino acid sequences; the tree has been divided into three clusters: cluster 1, cluster 2, and cluster 3. On the phylogenic tree, the evolutionary history was inferred using the UPGMA method [89]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method [4] and are in the units of the number of amino acid substitutions per site. Evolutionary analyses were conducted in MEGA6 [3].

Further analysis of the Open Reading Frames (ORF), Coding DNA Sequence (CDS), and amino acid sequence was used to generate Figure 18 and Appendix 4, which present the genomic exon composition of these western clawed frog CAP superfamily genes. By correlation of the data in Figures 17 and 18, we find that individual clusters present a high degree of conservation between exon borders. For instance, all eight genes within cluster 1 have nearly identical exon borders (first eight entries, Figure 18 and Appendix 4). Likewise, in cluster 2, representing the Peptidase Inhibitor/LCCL domain-containing subfamily, there is a high degree of exon border alignment in the CAP/PR domain and the Hinge region (entries nine through twelve, Figure 18 and Appendix 4), albeit distinctly different from the CRISP subfamily at the exon 3/exon 4 and exon 6/exon 7 boundaries. Finally, the two members of cluster 3, representing the GLIPR and GAPR subfamilies, have completely different exon borders from the first two clusters (entries thirteen and fourteen, Figure 18) in their CAP/PR domain. Further demonstration of different levels of homology within clusters and between clusters is demonstrated quantitatively in Appendix 5A using the Poisson correction model [4] in MEGA6 (Estimates of Evolutionary Divergence between Sequences) [3]. As shown in Appendix 5, within cluster 1, 2, and 3, the average number of amino acid substitution scores are 0.816, 0.547, and 0.913 respectively; however, between all three clusters, the average amino acid substitution score is 1.365, which is significantly higher.

To confirm these findings a similar analysis was done on the fifteen CAP protein present in the human genome. In Figure 19, all human CAP superfamily protein sequences have been aligned using neighbor joining alignment; eleven sequences have been divided into four clusters, based on physical clustering of genes (to be discussed below). Indeed, cluster 1, representing C-type lectin domain containing proteins, exhibited consistent exon borders within the subfamily (entries 1-3, Figure 20). However, these borders differed at multiple sites from cluster 2, the CRISP proteins (entries 4-6, Figure 20) even within the CAP/PR domain alone. Exon borders in cluster 3, the GLIPR subfamily (entries 8-10, Figure 20), differed from both cluster 1 and 2, yet were internally consistent. Finally, cluster 4, representing LCCL domain-containing proteins, is again internally consistent yet different from other subfamilies in exon structure. These findings on homology are further documented in Appendix 7A. As shown in Appendix 7, within clusters 1, 2, 3, and 4 the average amino acid substitution scores are 0.017, 0.543, 0.754, and 0.596 respectively; however, between all four clusters, the amino acid substitution score is 1.496, which is significantly higher.

While differences in exon borders between genes of different CAP subfamilies likely represent changes in gene structure that have occurred since the advent of subfamilies during invertebrate evolution, there are also certain exon borders that have remained constant within multiple subfamilies. Examples include the exon borders indicated by labelled arrows (A, B, and C) in Figures 18 and 20 and appear to represent the earliest exon borders that arose before the evolution of CAP subfamilies. Thus exon borders may provide clues as to the sequence of steps in CAP protein evolution, a topic that will be addressed at length in Chapter 4. But first, we will document the presence of physical clustering of CAP genes in the genomes of both vertebrates and invertebrates, an observation that suggests that CAP gene evolution was likely accompanied by multiple instances of gene duplication.

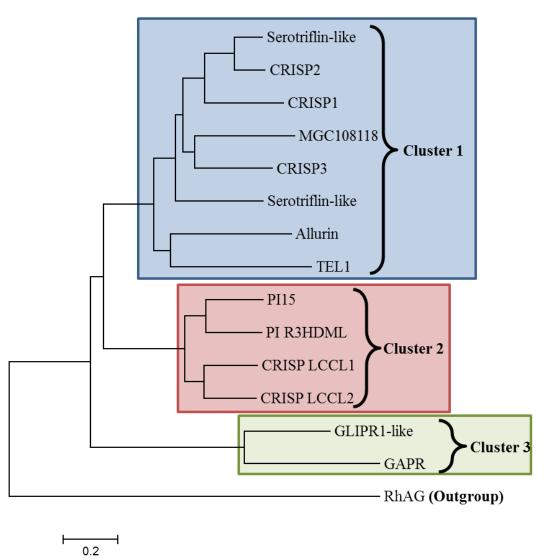


Figure 17. Phylogenetic Relationships of CAP Superfamily Proteins in the X. tropicalis (Western Clawed Frog) Genome

Figure 17. Phylogenetic relationships of CAP superfamily proteins in the *X. tropicalis* (Western Clawed Frog) genome. Protein sequences were aligned using MEGA7 ClustalW sequence alignment. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 15 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 152 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. The tree shows the presence of three cluster of CAP proteins in the frog genome corresponding to CAP subfamilies of proteins. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 11.

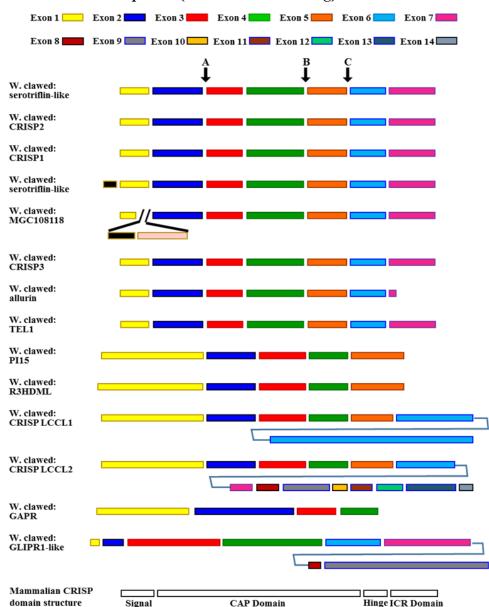


Figure 18. Genomic Exon Border Structure in CAP Superfamily Proteins in the *X. tropicalis* (Western Clawed Frog) Genome

Figure 18. Genomic exon border structure of CAP superfamily genes in the *X. tropicalis* (western clawed frog) genome. Protein sequences were aligned using MEGA7 ClustalW sequence alignment. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online databases for specific CAP genes. Color-coding is used to indicate exon order and relative length. Exon borders labeled A, B, and C (arrows) are found at identical positions in many of the vertebrate CAP subfamilies suggesting that these borders were established early in evolution, before divergence of these CAP subfamilies. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 11.

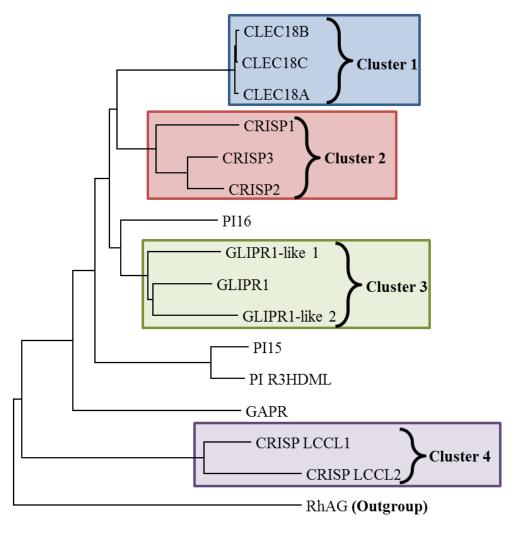


Figure 19. Phylogenetic Relationships Of All CAP Superfamily Proteins in the *Homo sapiens* (Human) Genome

Figure 19. Phylogenetic relationships of CAP superfamily proteins in the *Homo sapiens* (human) genome. Protein sequences were aligned using MEGA7 ClustalW sequence alignment. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 16 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 108 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. The tree shows the presence of four clusters of CAP proteins in the human genome corresponding to CAP subfamilies of proteins. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 12.

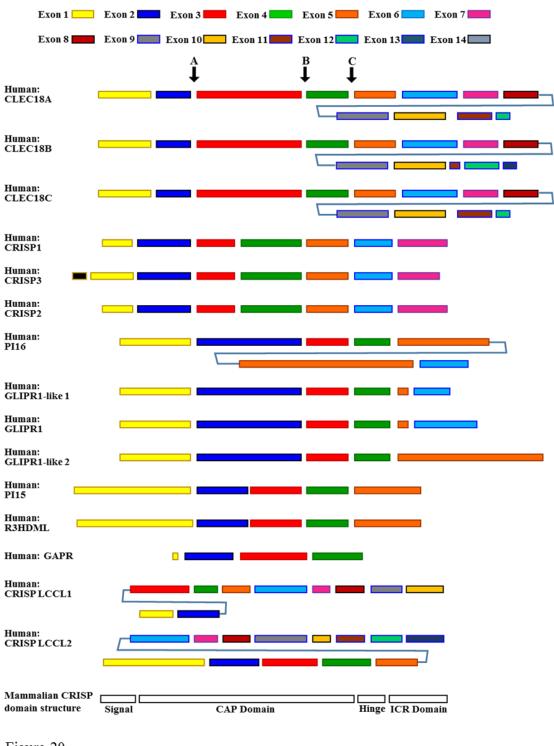


Figure 20. Genomic Exon Border Structure In All CAP Superfamily Proteins in the *Homo sapiens* (Human) Genome

Figure 20. Page 1 of 2

Figure 20. Genomic Exon Border Structure In All CAP Superfamily Proteins in the *Homo sapiens* (Human) Genome

Figure 20. Genomic exon border structure of CAP superfamily genes in the *Homo sapiens* genome. Protein sequences were aligned using MEGA7 ClustalW sequence alignment. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online databases for specific CAP genes. Color-coding is used to indicate exon order and relative length. Exon borders labeled A, B, and C (arrows) are found at identical positions in many of the vertebrate CAP subfamilies suggesting that these borders were established early in evolution, before divergence of these CAP subfamilies. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 12. Page 2 of 2

Physical Clustering of CAP Subfamily Genes in the Genome

Similar exon borders within CAP subfamilies suggested that these genes may have arisen by gene duplication. If so, I would hypothesize that genes belonging to the same cluster/subfamily might be physically clustered on the genome. In order to test this hypothesis, specific regions within the frog and human genomes were inspected in regard to gene order using the NCBI genome browser (www.ncbi.nlm.nih.gov/gene/). As shown in Figure 21 (bottom two entries), cluster 2 (CRISP subfamily) and cluster 3 (GLIPR subfamily) were each found to be clustered, albeit on different chromosomes. Similarly, in the frog genome the entire cluster 1 (CRISP subfamily) is found clustered on the same chromosome (sixth entry from top, Figure 21).

Additionally, it should be noted that these three cluster-specific gene clusters are found throughout Reptilia and Amphibia classes, therefore they are not simply unique to the western clawed frog (data not shown). Some of the species that present the three cluster-specific CAP gene clusters include: *Anolis carolinensis* (green anole), *Gekko japonicas* (Japanese gecko), *Pelodiscus sinensis* (Chinese soft-shelled turtle), *Alligator* *sinensis* (Chinese alligator), and *Alligator mississippiensis* (American alligator). Due to the lack of sufficient genomic data and gene assembly data, further analysis was not conducted.

As expected, similar to early chordates, gene clusters can be found in other later chordates as well. For example, peptidase inhibitor gene clusters can be found in most vertebrates, CRISP subfamily gene clusters are found in most mammals (e.g. mouse and human genomes (entry 9 and 10, Figure 21).

Consequently, many CAP gene clades are composed of genes clustered on the genome that share high sequence homology and share exon border homology (Figures 18 and 20 and Appendices 4 and 6). Physical clustering as evidence of gene duplication extends far back to the early stages in CAP gene evolution. For example, one domain CAP proteins, such as the pathogenesis-related (Pr) proteins of plants and the venom antigen (Ag) proteins of insects are both coded for by genes found in clusters (e.g., see first entry, Figure 21). Likewise, two domain CAP proteins, such as the SCL proteins of *C. elegans*, composed of a CAP/PR domain plus a Hinge region, are also coded for by genes found in multiple clusters. These findings all support the conclusion that CAP superfamily proteins are a result of gene duplication in early ancestors.

The Origins of CAP Subfamilies with Emphasis on the Evolution of CRISP Genes

Above, I made the surprising observation that, contrary to current thinking, CRISP proteins can be found in the invertebrate world as well as the vertebrate world. This suggests that CAP subfamilies (e.g. the CRISP, GLIPR, LD and PI subfamilies) thought

to be associated primarily vertebrates actually had their evolutionary origins in a common ancestor. In order to gain evidence for this, more detailed attention was focused on certain invertebrate phyla in regard to the presence of qualified subfamily members or precursors of subfamily members. Representative model organisms were chosen from the insects, echinoderms, and the urochordates for further study.

The primary tool was construction of a nearest neighbor-joining phylogenetic tree. For the phylogenetic tree, the evolutionary history was inferred using the UPGMA method [89]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method [4] and are in the units of the number of amino acid substitutions per site. Evolutionary analyses were conducted in MEGA6 [3]. As shown in Figure 22, separate clusters are present for the GAPR, LD, CRISP and GLIPR subfamilies. Within each cluster (excepting LCCL containing) are invertebrate proteins, clearly related to their vertebrate counterparts. The GAPR subfamily appears to be the most "primitive" and earliest to diverge, as they still bear noticeable sequence similarity to the PYR proteins of fungi as indicated in annotations of the GAPR subfamily. Thus, one might hypothesize that this subfamily diverged not too long after the branching off of the fungi from the main animal kingdom progression toward the invertebrates. In fact, if one adds to the alignment, the present-day PYR1 protein sequence from S. cerevisiae, this sequence nests within the GAPR subfamily (data not shown) largely because of similar CAP/PR domain sequences. This early divergence of the GAPR subfamily is stable to wide manipulation of the alignment membership.

59

Likewise, the CRISP subfamily also has clear membership from insect species whose ancestors were among the early invertebrates. The CRISP cluster was also stable to variations in alignment membership. The positioning of insect cysteine-rich venom proteins (CRVP) within the CRISP arborization correlates with the fact that they display an ICR domain with the 6 conserved cysteines characteristic of bonified mammalian CRISP proteins (see Figure 11).

In the GLIPR cluster of Figure 22, we find multiple members from the urochordate and molusc genomes. Although GLIPR and LCCD domain proteins are maintained in separate clusters in this particular alignment, they often intermix in their arborization as membership is changed. This suggests that these subfamilies are more closely related to one another and may have developed later during invertebrate evolution than did the CRISP and GAPR subfamilies.

For the CRISP subfamily the sequence relatedness data presented here is corroborated by evidence that multiple exon borders have identical placement within both insect and vertebrate genes, as shown in Figure 14 and Appendices 2 and 3. Is this the case for other subfamilies such as the GAPR family? To answer this question, key GAPR proteins from both invertebrate and vertebrate taxa were examined for consistency (or lack of consistency) in their exon borders. As shown in Appendix 8, the exon border structure of GAPR subfamily has conserved characteristics throughout its lineage. As one crosses the invertebrate-vertebrate threshold going from urochordates (e.g. sea squirts) to lower vertebrates (e.g. gars and bony fishes) to advanced vertebrates (e.g. mammals) one observes the exon 2/exon 3 border and the exon 4/exon 5 borders to be identical. Thus, genes in both the CRISP and GLIPR sub families appear to evolve increasing numbers of

60

exons and new exon borders as invertebrates gain complexity, but settle in to a relatively consistent gene structure during the invertebrate-vertebrate transition and as vertebrate evolution progresses.

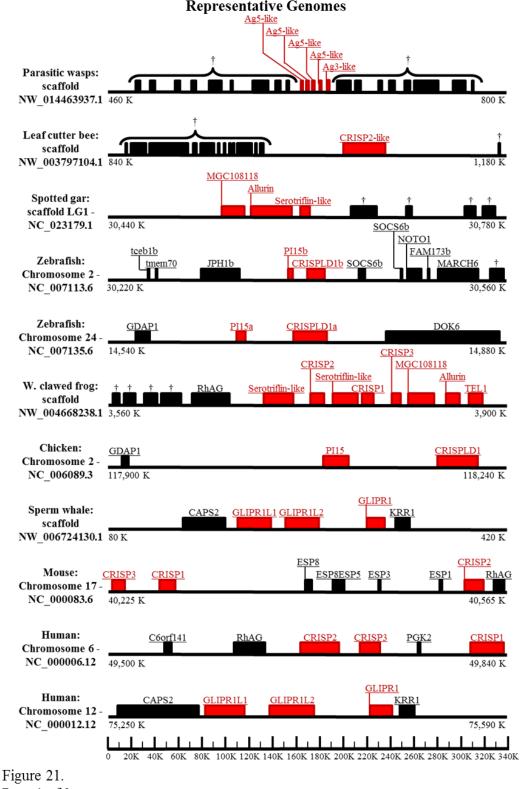


Figure 21. Genomic Clustering Data of CAP Superfamily Genes in **Representative Genomes**

Page 1 of 2

Figure 21. Genomic Clustering Data of CAP Superfamily Genes in **Representative Genomes**

Figure 21. Genomic clustering of CAP superfamily genes in representative genomes. Each entry represents gene locations within a small region of the indicated genome/chromosome. CAP genes are shown in red; non-CAP genes are shown in black. Clusters of CAP genes are common in the genomes of both vertebrates and invertebrates. Clusters of venom antigen genes are seen in insects (entry 1) and clusters of two-domain CAP proteins are seen in both insects and round worms (not shown). Clusters of CRISP genes are seen in both mammals and lower vertebrates (entries 6 and 8-11). In contrast, fish and birds do not appear to have large clusters of CAP genes (entries 3-5 and 7).

Page 2 of 2

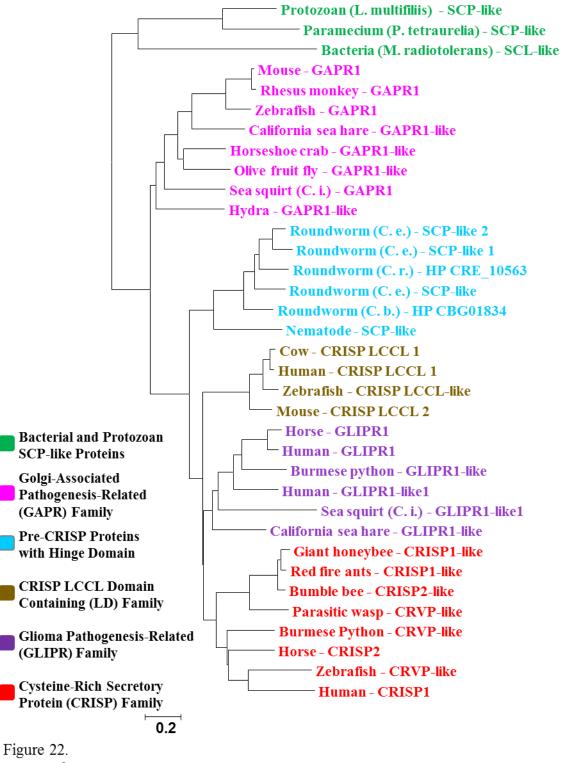


Figure 22. Evolution of Multi-Domain CAP Superfamily Proteins

Page 1 of 2

Figure 22. Evolution of Multi-Domain CAP Superfamily Proteins

Figure 22. Phylogenetic tree illustrating a proposed evolutionary pathway for vertebrate multi-domain CAP superfamily proteins. CAP/PR single domain proteins originated in bacteria and continued to be seen in protozoa, fungi, and invertebrates. These single domain proteins evolved into the PR proteins of plants (not shown), the PRY proteins of fungi (not shown) as well as the GAPR proteins of vertebrates and invertebrates (shown). Early CAP ancestral proteins giving rise to these subclasses likely had a SCP Eukaryotic type of CAP domain. The SCP Eukaryotic subclass also appears to have given rise to CAP two-domain proteins in which a Hinge domain was added to the C-terminal of a CAP/PR domain in early invertebrates including worms and insects. For simplicity, these proteins are referred to as "Pre-CRISP" proteins in this figure. This phylogenetic tree suggests that these invertebrate "Pre-CRISP" twodomain proteins gave rise to multiple subfamilies of CAP proteins characteristic of both vertebrates and higher invertebrates such as the CRISP, LD and GLIPR proteins. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 35 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 79 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 17. Page 2 of 2

CHAPTER 4: EVOLUTION OF CRISP GENE STRUCTURE

As stated previously, a full length CRISP protein contains three domains, the CAP/PR domain, the Hinge domain, and the ICR domain. The CAP/PR domain contains the four traditional CAP/SCP signature sequences and six cysteines. The Hinge domain contains four highly conserved cysteines, and the ICR domain contains six highly conserved cysteines. (Note: some literatures do combine and categorize the Hinge region and the ICR domain as a single Cysteine-Rich domain, ultimately implying a two domain protein; regardless, for the purpose of this paper, we will refer to the Hinge portion of CAP proteins as a separate domain based on the evolutionary data to be presented).

In this Chapter I summarize evidence that all CAP genes, including CRISP genes, began as single CAP/PR domain proteins, ancestors of present day bacterial, fungal and plant proteins. Subsequently, during early invertebrate evolution, specifically in roundworms and arthropods, ancestral two-domain proteins arose containing both CAP/PR and Hinge domains. Finally, ancestral three-domain CAP proteins arose, likely in the arthropods, from addition of the ICR domain to an ancestral, two-domain protein. Each of these three major steps in CRISP protein evolution will be discussed in sequence.

Evolution of the CAP/PR Domain

The CAP/PR domain was not only the earliest domain that can be recognized in bacteria but also has continued to evolve throughout invertebrate and vertebrate evolution. This continued evolution is reflected by the fact that NCBI databases (www.ncbi.nlm.nih.gov/gene/) recognize not just a single type of domain but variations of this domain that differ for almost every subfamily of CAP proteins. For example, during a BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi), CAP/PR domain amino acid sequences automatically trigger recognition of seven different types of SCP conserved domains including "bacterial", "eukaryotic", "GAPR", "GLIPR" and "CRISP". These subclassifications reflect the slow evolution of the four CAP signature sequences in the CAP/PR domain coupled to the fact that subfamilies of CAP genes, once they have diverged, evolve independently, likely due to different functional pressures. Independent evolution involved not only mutations in signature sequences, and independent physical clustering (discussed in Chapter 3), but also an increase in numbers of exons representing the CAP/PR domain as shown in Figure 16.

Evolution of the CAP/PR domain is also marked by the advent of new exon borders and new cysteines destined to become one of the six conserved cysteines of the CAP/PR domains found in modern CRISP proteins. These CAP sequence features, when traced through a series of taxonomic groups, can be used to reconstruct the changes involved in evolution of the CAP/PR domain (see the table of Figure 23B). The CAP signature sequences begin as bacterial, evolve to become eukaryotic, and then finally specialize to represent CAP subfamilies. The exon border 2/3 appears early in roundworms, 4/5 and 5/6 appear subsequently, and the 1/2 and 3/4 borders appear most recently in vertebrates. Conserved cysteines in positions 1 and 3 are found earliest in bacteria, cysteines 4 and 5 are subsequently found beginning in plants, fungi and protozoa, while a cysteine at the 6 position is not seen until vertebrates.

67

Figure 23. Evolution of CRISP Proteins is Accompanied by Placement of Introns and Cysteines at Specific Positions

А.		Exon 1/2 border		
MEI KHLL	FLVAAACLL	P M L S M K K S A	RDQFNKLVT	DLPNVQE
	_	Exon 2/3 border	·	C1
EIVNIHN	ALRRRVVPP	A S N M L K M S W S	EEAAQNARI	F S K Y <mark>C</mark> D M
	Exon 3/4 border C2			
TESNPLE	R R L P N T F C G	ENMHMTSYPV	SWSSVI GVW	YSESTSF
		Exon 4/5 border	<u>C3</u> <u>C4</u>	
KHGEWTT	ТОООІ ТТОН	Y T Q I V W A T S Y	LIGCAIASC	RQQGSPR
C5	C6, Exon 5/6 border		C7 C8 C9	C10, Exon 6/7 border
Y - LYV <mark>C</mark> H	Y C H E G N D P E	TKNEPYKTGV	P C E A C P S N C	EDKLCTN
C11	C12	C13	C14 C15	C16
P <mark>C</mark> I YYDE	Y F D C D I Q V H	YLG <mark>C</mark> N-HSTT	I L F <mark>C</mark> K A T <mark>C</mark> L	<mark>C</mark> DTEI K

В.							
ORGANISM	SCP DOMAIN TYPES	POSTULATED SCP TYPE IN CRISP AN CESTOR	N UMBERS OF EXONS CODING FOR SCP DOMAIN	EXON BORDERS PRESENT AT "CRISP" POSITIONS	CYSTEINES LOCATED AT "CRISP" POSITIONS	HINGE DOMAIN PRESENT	CRISP DOMAIN PRESENT
BACTERIA	Bacterial only	Bacterial with cysteines	One	None	1,3 4,5 occasionally	Hinge-like domain seen very rarely	Not present
PLANTS	PR only	N/A	One, rarely two	None	1,3,4,5	Not present	Not present
FUNGI	Bacterial, Eukaryotic, PRY	Bacterial, then Eukaryotic	One to three	Exon 1/2 border rarely Exon 3/4 border rarely	1,3,4,5	Hinge-like domain seen very rarely	Not present
PROTOZOA	Bacterial and Eukaryotic	Eukaryotic	One to three	Exon 2/3 border rarely Exon 4/5 border rarely Exon 5/6 border rarely	1,3,4,5	Not present	Not present
CN IDARIA	Eukaryotic	Eukaryotic	One to five	Exon 3/4 border rarely	3,5	Not present	Not present
ROUND WORMS	GAPR and Eukaryotic	Eukaryotic	Usually three to five	Exon 1/2 border rarely Exon 2/3 border Exon 4/5 border Exon 5/6 border	1,3,4,5	Hinge domain present	Not present
IN SECTS	GAPR and Eukaryotic	Eukaryotic	Usually three to five	Exon 1/2 border Exon 2/3 border	1,2,3,4,5, 6 occasionally	Hinge domain present	Present
MOLUSCS, ECHINODERMS, and UROCHORDATES	GAPR, Eukaryotic, and GLIPR	Eukaryotic	Usually four or five	Exon 2/3 border Exon 3/4 border rarely Exon 4/5 border Exon 5/6 border	1,2,3,4,5, 6 rarely	Hinge Domain Present	Variable
VERTEBRATES	GAPR, GLIPR and CRISP	Crisp	Usually five	All five exon borders are present	1,2,3,4,5,6	Hinge domain present	Present

Figure 23. Page 1 of 2

Figure 23. Evolution of CRISP Proteins is Accompanied by Placement of Introns and Cysteines at Specific Positions

Figure 23. Evolution of CRISP proteins is accompanied by placement of introns and cysteines at specific positions. Sequences of 148 CAP proteins representing the taxonomic groups indicated were aligned by web-based MUSCLE software using default parameters. Exon numbers and intron/exon borders of the corresponding genes were verified using the intron and exon prediction tracks in NCBI online databases for specific CAP genes. A. Annotated amino acid sequence of human CRISP1 specifying the location of exon borders and cysteines which, monitored over evolutionary time, suggest the sequence of events that led to vertebrate CRISP proteins. B. Table summarizing changes in the CAP gene structure that led to vertebrate CRISP proteins. SCP domain types refer to the subtypes of this conserved domain recognized by NCBI BLAST software. Exon borders were considered "CRISP-like" if they fell within the range boxed in panel A. Cysteine positions were considered "CRISP-like" if they fell within one residue of the human CRISP1 cysteines in the aligned sequences. Page 2 of 2

Addition of the Hinge Domain

The Hinge domain appears early in CAP superfamily evolution. As indicated in Figure 24, hinge-like domains can be observed infrequently within unrelated bacterial proteins and within protozoan CAP proteins. However, it should be noted that most early CAP proteins lacked a Hinge domain. It was not until the early invertebrates such as round worms (rows 3-6) and arthropods (rows 9-12) that two-domain CAP proteins containing a Hinge domain as well as a CAP/PR domain become common. It is not clear whether the domain was "borrowed" from unrelated bacterial proteins or whether it evolved in a multi-step process during CAP protein evolution. The latter possibility is suggested by the existence of two-domain CAP proteins in arthropods that have hingelike domains with two or three cysteines rather than the full set of four (rows 7 (Mosquito Ag-like) and 8 (Mosquito CRISP 3-like), Figure 24).

Subsequently, these two-domain proteins appear to have given rise to multiple subfamilies of three-domain CAP proteins. Indeed, all vertebrate CAP proteins (with the exception of GAPR proteins) carry a Hinge domain of ~26 amino acid residues. In this context this domain has been suggested to act like a swivel – that is, to allow free rotation of a third domain (e.g. the ICR domain of a CRISP protein) relative to the N-terminal CAP/PR domain [26, 90]. Interestingly, the Hinge domain, unlike the CAP/PR domain, has evolved very little since its first inclusion in CAP proteins of ancestral round worms, Apparently, its primary structure (and likely tertiary structure) has remained consistent over 1 billion years of evolution as determined by two disulfide bond linkages between four conserved cysteines (see Figure 24).

Figure 24. The Hinge Domain Was Developed Early in Evolution and Was First Seen attached to the CAP/PR Domain in Protozoa and Round Worms

								(С7	7		C	8				C	Э				(C1	0							
Bacterium - Hypoth. Prot.	Μ	ΕN	V	V	10	Q 7	TE	EE	С	Ρ	Т	С	Ρ	Е	G	τv	С	E A	S	-	D	LL	C	T	Ν	T	P	СТ	Т	T١	- ۷
Protozoan - Crisp-like	N (QS	М	Y	QS	G·	- 1	۱P	С	S	Α	С	G	Κ	Q (QS	С	N S	51	F	Т	S L	C	G	Q	Т	T /	ΑI	Ρ	G /	A T
Round Worm - SCP-Like 13	D	S D	I.	Y	QC	QG.	- E	H	С	S	Α	С	S	Е	D /	A S	С	E () D	-	S	G١	C	А	-	-	-		-		
Round Worm - SCP-Like 12	D	S D	I.	Y	QS	G	- 0) T	С	S	F	С	Ρ	S	G :	s k	C	ΕE	E A	-	S	G١	С	Ρ	-	-	-		-		
Round Worm - SCP-Like 1	GI	RP	Т	Y	KE	G	- 11	ГΤ	С	S	S	С	S	G	S (ΤK	C	D	ГΑ	-	S	G۱	С	G	-	-	-		-		
Round Worm - SCP-Like 14	GI	ИD	T	Y	K S	G	- E	ΞT	С	S	Ν	С	Ρ	D	G	ΤN	C	ЕS	5 S	-	Т	GΙ	С	V	-	-	-		-		
Mosquito - Ag-like	N (Q P	S	Y	VK	G	- H	I A	G	S	Q	С	S	Т	G		۷	S 5	5 A	Y	Ρ	G۱	C	N	-	-	-		-		
Mosquito - Ag-like	Q	Q K	Т	Y	ΤK	S F	PE	ΕP	۷	Q	Ν	С	S	L	-		Y	S A	E	Y	s	C۱	С	S	S	R	E	۶L	L	A S	S G
Mosquito - Crisp 3-like	K	ТΡ	Т	Y	AG	QG.	- E	ΕP	С	S	Q	С	Т	S	G		С	DA	K	F	D	G۱	С	Ν	Т	D	ΕI	ΡV	D	L S	s -
Mosquito - Ag5	G	ΙP	Т	Y	V S	GI	K	(P	С	S	G	С	т	т	G		С	N S	5 A	F	Ρ	G١	C	Ν	Е	S	ΕI	Ρ٧	Ρ	QI	1 1
Bumble Bee - CRISP 2	S I	RP	-	Y	VA	G	- P	(P	С	S	М	С	Κ	D	Н		С	S١	(-	-	Ν	ΚL	С	т	Ν	Α	CR	ISP	DC	MA	IN
Fire Ant - CRISP 1	G	LΡ	-	Y	VA	G	- P	(P	С	S	М	С	Κ	D	Н		С	т) -	-	D	AL	С	Т	Ν	Α	CR	ISP	DO	MA	IN
Sea Squirt - CRISP-like	ΚI	ΗP	-	Υ	S	G	- 5	S P	С	S	Α	С	Ρ	Κ	G١	(G	i C	Κ		-	N	Gι	С	S	L	S	AI) S	Ν	MI	LΚ
Amphioxus - Hypoth. Prot.	Q	(P	-	ΥI	ΙT	G	- 1	ΓР	С	Т	Q	С	Α	Ν	G١	N	C	S -		-	D	ΚL	С	G	Κ	D	-		-		
S. kowalevskii - Hypoth. Prot.	F١	V A	Κ	Y		CI	۶L	. M	С	G	۷	С	V	Ρ	EE	ΕH	C	C)	-	G	ĸ	C	s	Q	G	VI	D	G,	A T	ГС
Sea Urchin - CRISP 2-like	ΚF	R Ρ	-	Υ	ES	G	- 1	. P	С	S	s	С	Ρ	D	S E	E S	С	E ·		-	Ν	NL	С	т	М	G	-		-		
Zebra Fish - CRISP-like	T۱	/ P	Ρ	Y	S L	G	- 5	SΡ	С	Α	S	С	Ρ	Ν	N		С	E ·		-	D	NL	С	Т	Ν	Α	CR	ISP	DO	MA	IN
Python - CRVP-like	Α	ΤР	-	Y	K S	G	- F	р Р	С	G	D	С	Ρ	S	S		С	A ·		-	Ν	G۱	С	т	Ν	Ρ	CR	ISP	DC	MA	IN
Horse - CRISP 2	Ν	ΤР	-	Y	QC	G.	- 11	ΓР	С	Α	s	С	Ρ	G	Ν		С	D ·		-	Ν	G۱	С	т	Ν	S	CR	ISP	DC	MA	IN
Human, CRISP 1	N	ΕP	-	Y	K٦	G -	- 1	/ P	С	Е	A	С	Ρ	S	N		С	E -	• •	-	D	κı	C	Т	N	Ρ	CR	ISP	DO	MA	IN

Figure 24. The Hinge domain amino acid sequence of CAP proteins (blue) exhibiting a set of four highly conserved cysteines (yellow) is seen early in evolution. Hinge-like sequences are found rarely in non-CAP proteins of bacteria (row 1) and rarely in protozoa at the C-terminal of a CAP/PR domain (row 2). Round worms exhibit numerous proteins that have Hinge domains linked to CAP/PR domains (rows 3-6). Insects exhibit both Hinge domains (4 cysteines) and Hinge-like domains (1-3 cysteines) linked to CAP/PR domains (rows 7-10). In addition, insects represent the earliest example in which an ICR domain is linked to the Hinge domain (rows 11-12). Higher invertebrates also have proteins in which a Hinge domain is linked to a CAP/PR domain; however, no true ICR domains are found in these proteins (rows 13-16). In vertebrates, the Hinge domain nearly always joins a CAP/PR domain with a third domain such as ICR (rows 17-20). Sequences were aligned by MUSCLE within MEGA 7 using default parameters. Organisms and accession IDs of the amino acid sequences are available in supplemental data table 17.

Addition of the ICR Domain

The earliest appearance of the ICR domain in a three-domain CAP protein is in the Arthropoda phylum. The inclusion of the ICR in the CAP domain lineup was a separate evolutionary event that occurred subsequent to addition of the Hinge domain. This conclusion is based on extensive BLAST searches that failed to detect CAP proteins that contained ICR domains in the absence of Hinge domains. Likewise, extensive BLAST searches for ICR domains linked to Hinge domains in unrelated non-CAP proteins turned up empty handed. This evidence suggests the Hinge domain, given its separate evolutionary origin, is in fact a separate domain from the ICR domain. Thus, the designation of these two domains as one "Cysteine-Rich" or "CRISP" domain, commonly seen in the CAP superfamily literature, is erroneous and seems to stem simply from an unfounded assumption about the minimum size that a domain can have.

So where did the ICR domain come from? Indeed, BLAST searches do not reveal ICR domains in unrelated, non-CAP, multi-domain proteins. Rather, this domain bears striking resemblance to stand alone potassium channel peptide toxins of invertebrates. As shown in Figure 25A, the ICR domain amino acid sequences within a broad range vertebrate CRISP proteins (rows 1-6) share notable homology to sequences of peptide toxins found in sea anenomes (cnidaria), round worms and scorpions (arthropoda) (rows 7-12) and is highlighted by the six conserved cysteines found in each. Indeed, NMR solution and X-ray crystallographic structures (Figures 25B-E) show that the tertiary structure of the ICR domain in CRISP proteins is nearly identical to the tertiary structure of these potassium channel toxins: each is characterized by an alpha helix and a neighboring short helix-like secondary structure from which three disulfide bonds (in red) radiate, This structural similarity accounts for that fact that ICR domains, whether occurring in CRISP proteins or expressed separately in vitro, have potassium channel blocking activities analogous to those seen in these invertebrate peptide toxins (references).

Therefore, due to the remarkable sequence and structural homology of the ICR domain with these invertebrate potassium channel peptide toxins, I speculate that the ICR domain sequence of CRISP proteins originated in this toxin family and was imported into CAP proteins during invertebrate evolution.

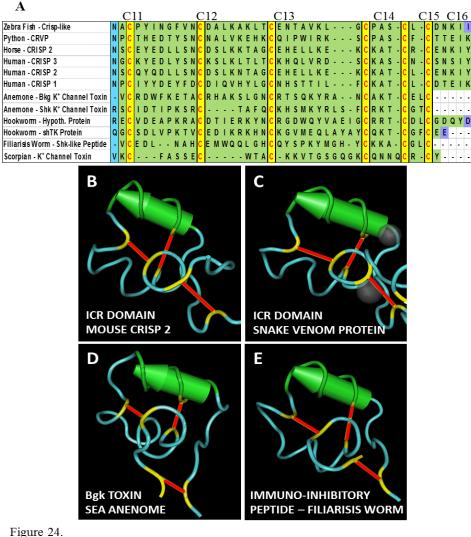




Figure 24. Page 1 of 2

Figure 25. The ICR Domain is Related to Potassium Channel Toxins

Figure 25. The ICR domain exhibits both an amino acid sequence and a tertiary structure homologous to those of potassium channel peptide toxins found in cnidarians, round worms and arthropods. A.ICR domain and toxin sequences, aligned by MUSCLE within MEGA 7 software using default parameters, each containing six conserved cysteines (yellow highlighted) that are disulfide bonded. B. NMR solution structure of the ICR domain of mouse CRISP2 [3]). C. X-ray crystal structure of the ICR domain of the cysteine-rich venom protein from the snake *Protobothrops flavoviridis* (1WVR, Shikamoto et al., 2005). D. NMR solution structure of the BgK peptide toxin from the sea anemone *Bunodosoma granuliferum* (1BGK_A, Dauplais et al. 1997). E. NMR solution structure of the homologous immuno-inhibitory toxin from *Brugia Malayi* (2MCR, Chhabra et al., 2014). The tertiary structures formed in each case are characterized by an alpha helix and a neighboring short helix-like secondary structure from which the three disulfide bonds (in red) radiate. Page 2 of 2

CHAPTER 5: CONCLUSION

The endurance and the prevalence of CAP superfamily proteins throughout organismal evolution signifies the importance of these proteins. Early Pr proteins played an important role in pathogen defense and survival of plants; however, current mammalian CAP proteins are endowed with a wide array of functions – from immunological characteristics to reproductive features. The diversity of CAP proteins has exploded, therefore it can be fairly assumed that CAP proteins have and will continue to play an important role in survival and/or reproduction.

This thesis has examined the evolution of CAP superfamily proteins as evidenced in the extensive bioinformatics data now available at the NCBI and ENSEMBL websites. These data included the sequence and annotations for genomes of over 50 organisms that span the range of evolutionary information from bacteria to mammals. As a result, I have been able to lay out a series of evolutionary steps that likely led to the diversity of CAP superfamily proteins found today.

Many of the major steps in creating this diversity are summarized in Figure 26 (which is identical to Figure 12), the earliest step (1) being the origin of the CAP/PR domain in bacteria. These one-domain proteins further diversified into other one-domain proteins including the GAPR subfamily – the earliest subfamily to diverge (step 2), the plant-specific PR proteins (step 3), and the fungi-specific PRY proteins (step 4). At later points these one-domain proteins evolved into the one-domain venom antigen proteins of round worms and insects (steps 7 and 8). The next major step was addition of the Hinge domain (step 5) to form two-domain proteins that would serve a precursors to many other

CAP subfamilies that would soon evolve. These two-domain genes/proteins were first seen in roundworms and arthropods. Surprisingly, construction of genes coding for two domains appears to have quickly led to three-domain CAP superfamily genes (step 6) as instanced in their Arthropod descendants - GLIPR1-like and CRISP-like genes that subsequently were used as templates for evolution of the GLIPR1 and CRISP subfamilies found in higher invertebrates and vertebrates. Continued diversification of the CAP superfamily occurred with spinoff of the Proteinase Inhibitor subfamily during invertebrate evolution (step 9) and the CRISP LD and C-type lectin subfamilies during early vertebrate evolution (step 10). Finally, during higher vertebrate evolution, further diversity in the CRISP, GLIPR and CLEC families is observed. CRISPs 1 and 3 split off from CRISP 2 and were first seen as distinct lineages in amphibians (step 11) while the CRISP 4 lineage did not diverge until rodents evolved (step 13). In the CLEC subfamily, CLECB was first seen in early mammals (step 12) while the CLEC18C lineage was not observed until primates evolved.

Unique to the present study is the discovery of two-domain CAP proteins and three-domain CAP proteins within invertebrates that appear to represent modern day descendants of ancestral CAP proteins that played critical roles in CAP superfamily evolution. I have been able to link these intermediate steps in evolution to what came before and what came after by using a combination of tools: BLAST to ferret out homologous proteins (often by using multiple serial BLAST searches), alignment programs such as CLUSTAL W and MUSCLE [3] to identify conserved sequence features, especially conserved cysteines and CAP signature sequences, genome browsers to identify exon borders of specific CAP genes as well as gene clusters, and phylogenetic tree generating programs such as MEGA 6 and MEGA 7 to analyze the relatedness of the CAP subfamilies.

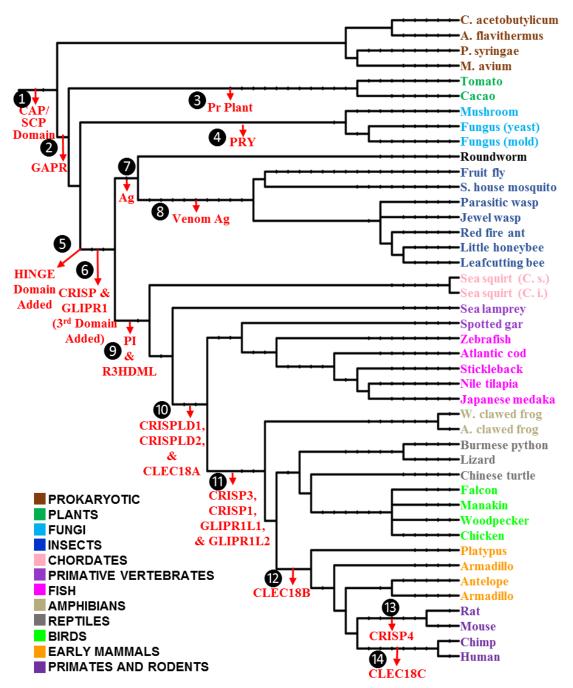


Figure 26. Point of Origin for Various CAP Superfamily Proteins

Figure 26. The figure presents the points of origin for various CAP superfamily proteins. The earliest origin of the CAP/SCP domain likely occurred in a common ancestor shared by most species. After this point of origin, we can see a burst of CAP/Pr proteins in many organism. The Hinge and the ICR domain originated in a common ancestor shared of the Animalia kingdom. A wider verity of CAP proteins start appearing in vertebrates.

Especially useful has been documentation of exon border positions within genes and clustering between genes which in this study have clearly detected periods of great change in CAP gene structure as well as periods of relatively little change. Analysis of exon border data or gene clustering data is seldom observed in the literature and here these have allowed new information about CRISP gene evolution to be leveraged from existing database entries as instanced in the following summary observations.

The earliest CRISP proteins appeared in arthropods, where it is believed they play an ion-channel regulatory roll (as seen in vertebrates). For instance, in southern house mosquito genome, we can observe eight unique CRISP proteins in two gene clusters (data not shown). Subsequently, the existence of multiple gene copies in these clusters allowed further diversification, producing in vertebrates (e.g. the human genome) three CRISP genes (CRISP1, CRISP2, and CRISP3), also present as a gene cluster (Figure 21). The arthropod and human gene products are clearly homologous as indicated by conserved cysteines, especially the six in the ICR domain that determine a tertiary structure not unlike that of potassium channel toxins also found in arthropods (Figure 25).

Thus, using CRISP amino acid sequence and exon border structure I have been able to extrapolate the origin of the CRISP protein to a shared ancestor of the phyla Arthropoda and Chordata (Figure 26). At the point of CRISP protein origin, it is likely that exon 2/exon 3 border and exon 5/exon 6 border were present in the ancestral organism (Figure 14). However, past this point of origin, both vertebrates and invertebrates have developed additional unique introns and splicing patterns despite conservation of the amino acid sequences involved. Current splice patterns for the vertebrate and invertebrate CRISP proteins can be visualized in Figures 13, 14 and 15. After the point of CRISP gene origin, Arthropoda in general developed a five exon CRISP protein while vertebrates developed a seven exon CRISP protein. These differences resulted from subsequent independent designation of exons borders in the two lineages (e.g. the exon 3/exon 4 border) and from exon borders developed in the vertebrate lineage (e.g. the exon 4/exon 5 border) but not in the invertebrate lineage.

In conclusion, new features of CAP superfamily and CRISP protein evolution in particular have been elucidated by the use of exon border detection and gene clustering data already present in genomic databases. I urge increased use of these tools in future studies of protein superfamily evolution, tools which up to this point have largely been ignored.

BIBLIOGRAPHY

- 1. Bedford, J., *Sperm/egg interaction: the specificity of human spermatozoa.* The Anatomical Record, 1977. **188**(4): p. 477-487.
- 2. Yanagimachi, R., *Mechanisms of fertilization in mammals*. Fertilization and Embryonic Development in vitro, 1981: p. 81-182.
- 3. Tamura, K., et al., *MEGA6: Molecular Evolutionary Genetics Analysis version* 6.0. Mol Biol Evol, 2013. **30**(12): p. 2725-9.
- 4. Zuckerkandl, E. and L. Pauling, *Evolutionary divergence and convergence in proteins*. Evolving genes and proteins, 1965. **97**: p. 97-166.
- 5. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Molecular biology and evolution, 1987. **4**(4): p. 406-425.
- 6. Gibbs, G.M., K. Roelants, and M.K. O'Bryan, *The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins--roles in reproduction, cancer, and immune defense.* Endocrine reviews, 2008. **29**(7): p. 865-97.
- 7. Choudhary, V., et al., *The caveolin-binding motif of the pathogen-related yeast protein Pry1, a member of the CAP protein superfamily, is required for in vivo export of cholesteryl acetate.* J Lipid Res, 2014. **55**(5): p. 883-94.
- 8. Niderman, T., et al., *Pathogenesis-related PR-1 proteins are antifungal (isolation and characterization of three 14-kilodalton proteins of tomato and of a basic PR-1 of tobacco with inhibitory activity against Phytophthora infestans.* Plant Physiology, 1995. **108**(1): p. 17-27.
- 9. Niki, T., et al., Antagonistic effect of salicylic acid and jasmonic acid on the expression of pathogenesis-related (PR) protein genes in wounded mature tobacco leaves. Plant and Cell Physiology, 1998. **39**(5): p. 500-507.
- Van Loon, L.C. and E.A. Van Strien, *The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins.* Physiological and molecular plant pathology, 1999. **55**(2): p. 85-97.

- 11. Lu, G., et al., *Sequence analysis and antigenic cross-reactivity of a venom allergen, antigen 5, from hornets, wasps, and yellow jackets.* The Journal of Immunology, 1993. **150**(7): p. 2823-2830.
- 12. Olson, J.H., et al., *Allurin, a 21-kDa sperm chemoattractant from Xenopus egg jelly, is related to mammalian sperm-binding proteins.* Proceedings of the National Academy of Sciences, 2001. **98**(20): p. 11205-10.
- 13. Aoki, N., et al., *Serotriflin, a CRISP family protein with binding affinity for small serum protein-2 in snake serum.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2008. **1784**(4): p. 621-8.
- Yamakawa, T., et al., *cDNA cloning of a novel trypsin inhibitor with similarity to pathogenesis-related proteins, and its frequent expression in human brain cancer cells.* Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression, 1998. 1395(2): p. 202-208.
- 15. Fernández, C., et al., *NMR solution structure of the pathogenesis-related protein P14a*. Journal of molecular biology, 1997. **266**(3): p. 576-593.
- 16. Bohle, B., et al., *Characterization of the human T cell response to antigen 5 from Vespula vulgaris (Ves v 5).* Clin Exp Allergy, 2005. **35**(3): p. 367-73.
- 17. Jutel, M., et al., *Bee venom immunotherapy results in decrease of IL-4 and IL-5 and increase of IFN-gamma secretion in specific allergen-stimulated T cell cultures.* The Journal of Immunology, 1995 **154**(8): p. 4187-4194.
- Mans, B.J., et al., Comparative sialomics between hard and soft ticks: implications for the evolution of blood-feeding behavior. Insect Biochem Mol Biol, 2008. 38(1): p. 42-58.
- Charlab, R., et al., *Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly Lutzomyia longipalpis.* Proceedings of the National Academy of Sciences, 1999. 96(26): p. 15115-15160.

- 20. Calvo, E., et al., *An insight into the sialome of Anopheles funestus reveals an emerging pattern in anopheline salivary protein families.* Insect Biochem Mol Biol, 2007. **37**(2): p. 164-75.
- 21. Van Vaerenbergh, M., et al., *Extending the honey bee venome with the antimicrobial peptide apidaecin and a protein resembling wasp antigen 5.* Insect Mol Biol, 2013. **22**(2): p. 199-210.
- 22. Dos Santos-Pinto, J.R., et al., Using proteomic strategies for sequencing and posttranslational modifications assignment of antigen-5, a major allergen from the venom of the social wasp Polybia paulista. J Proteome Res, 2014. **13**(2): p. 855-65.
- Kovalick, G.E., et al., *Structure and expression of the antigen 5-related gene of Drosophila melanogaster*. Insect biochemistry and molecular biology, 1998.
 28(7): p. 491-500.
- 24. Hawdon, J.M., et al., *Cloning and characterization of Ancylostoma-secreted* protein a novel protein associated with the transition to parasitism by infective hookworm larvae. Journal of Biological Chemistry, 1996. **271**(12): p. 6672-6678.
- 25. Hoffman, D.R., *Allergens in Hymenoptera venom XXIV: the amino acid sequences of imported fire ant venom allergens Sol i II, Sol i III, and Sol i IV.* Journal of allergy and clinical immunology, 1993. **91**(1): p. 71-78.
- 26. Guo, M., et al., *Crystal structure of the cysteine-rich secretory protein stecrisp reveals that the cysteine-rich domain has a K+ channel inhibitor-like fold.* J Biol Chem, 2005. **280**(13): p. 12405-12.
- 27. Yanagimachi, R., *Mammalian fertilization*. The physiology of reproduction, 1994.
 1: p. 189-317.
- 28. Da Ros, V.G., et al., *Impaired sperm fertilizing ability in mice lacking Cysteine-RIch Secretory Protein 1 (CRISP1)*. Dev Biol, 2008. **320**(1): p. 12-8.
- 29. Cohen, D.J., et al., *Evidence that human epididymal protein ARP plays a role in gamete fusion through complementary sites on the surface of the human egg.* Biology of reproduction, 2001. **65**(4): p. 1000-1005.

- 30. Legare, C., et al., *Cysteine-rich secretory protein 1 in seminal plasma: potential biomarker for the distinction between obstructive and nonobstructive azoospermia.* Fertil Steril, 2013. **100**(5): p. 1253-60.
- Maeda, T., J. Nishida, and Y. Nakanishi, *Expression pattern, subcellular localization and structure-function relationship of rat Tpx-1, a spermatogenic cell adhesion molecule responsible for association with Sertoli cells.* Development, growth & differentiation, 1999 41(6): p. 715-722.
- 32. Gibbs, G.M., et al., *The cysteine-rich secretory protein domain of Tpx-1 is related to ion channel toxins and regulates ryanodine receptor Ca2+ signaling.* J Biol Chem, 2006. **281**(7): p. 4156-63.
- Busso, D., et al., *Human testicular protein TPX1/CRISP-2: localization in spermatozoa, fate after capacitation and relevance for gamete interaction*. Mol Hum Reprod, 2005. 11(4): p. 299-305.
- 34. Busso, D., et al., *A novel function for CRISP1 in rodent fertilization: involvement in sperm-zona pellucida interaction.* Biol Reprod, 2007. **77**(5): p. 848-54.
- 35. Busso, D., et al., *Evidence for the involvement of testicular protein CRISP2 in mouse sperm-egg fusion*. Biol Reprod, 2007. **76**(4): p. 701-8.
- 36. Schwidetzky, U., B. Haendler, and W.D. Schleuning, *Isolation and characterization of the androgen-dependent mouse cysteine-rich secretory protein-3 (CRISP-3) gene.* Biochem, 1995. **309**: p. 831-836.
- Haendler, B.E.R.N.A.R.D., Krätzschmar, J., Theuring, F. R. A. N. Z., & Schleuning, W. D., *Transcripts for cysteine-rich secretory protein-1 (CRISP-1; DE/AEG) and the novel related CRISP-3 are expressed under androgen control in the mouse salivary gland*. Endocrinology, 1993 133(1): p. 192-198.
- Udby, L., A.H. Johnsen, and N. Borregaard, *Human CRISP-3 binds serum* alpha(1)B-glycoprotein across species. Biochim Biophys Acta, 2010. 1800(4): p. 481-5.
- 39. Pathak, B.R., et al., *Growth inhibition mediated by PSP94 or CRISP-3 is prostate cancer cell line specific.* Asian J Androl, 2010. **12**(5): p. 677-89.

- 40. Grupp, K., et al., *Cysteine-rich secretory protein 3 overexpression is linked to a subset of PTEN-deleted ERG fusion-positive prostate cancers with early biochemical recurrence.* Mod Pathol, 2013. **26**(5): p. 733-42.
- Henriksen, R., Lundwall, Å., Udby, L., & Fernlund, P., *The expression of β-microseminoprotein but not CRISP3 is reduced in ovarian cancer and correlates to survival*. Anticancer research, 2012. **32**(9): p. 3993-3999.
- 42. Udby, L., et al., *beta-Microseminoprotein binds CRISP-3 in human seminal plasma*. Biochem Biophys Res Commun, 2005. **333**(2): p. 555-61.
- 43. Jalkanen, J., I. Huhtaniemi, and M. Poutanen, *Mouse cysteine-rich secretory protein 4 (CRISP4): a member of the Crisp family exclusively expressed in the epididymis in an androgen-dependent manner.* Biol Reprod, 2005. **72**(5): p. 1268-74.
- 44. Nolan, M.A., et al., *Identification of rat cysteine-rich secretory protein 4 (Crisp4) as the ortholog to human CRISP1 and mouse Crisp4*. Biol Reprod, 2006. **74**(5): p. 984-91.
- 45. Xiang, X., et al., *The sperm chemoattractant "allurin" is expressed and secreted from the Xenopus oviduct in a hormone-regulated manner*. Dev Biol, 2004.
 275(2): p. 343-55.
- 46. Burnett, L.A., et al., *Mouse sperm exhibit chemotaxis to allurin, a truncated member of the cysteine-rich secretory protein family*. Developmental Biology, 2011. **360**(2): p. 318-28.
- 47. Trexler, M., L. Bányai, and L. Patthy, *The LCCL module*. European Journal of Biochemistry, 2000. **267**(18): p. 5751-5757.
- 48. Chiquet, B.T., et al., *Nonsyndromic cleft lip and palate: CRISPLD genes and the folate gene pathway connection*. Birth Defects Res A Clin Mol Teratol, 2011.
 91(1): p. 44-9.
- 49. Oyewumi, L., F. Kaplan, and N. Sweezey, *Lgl1*, a mesenchymal modulator of early lung branching morphogenesis, is a secreted glycoprotein imported by late gestation lung epithelial cells. Biochem, 2003. **376**: p. 61-69.

- 50. Kaplan, F., et al., *A novel developmentally regulated gene in lung mesenchyme: homology to a tumor-derived trypsin inhibitor*. American Journal of Physiology-Lung Cellular and Molecular Physiology, 1999. **276**(6): p. L1027-L1036.
- 51. Chiquet, B.T., et al., *CRISPLD2: a novel NSCLP candidate gene*. Hum Mol Genet, 2007. **16**(18): p. 2241-8.
- 52. Lan, J., et al., *Inflammatory cytokines, goblet cell hyperplasia and altered lung mechanics in Lgl1+/- mice.* Respir Res, 2009. **10**: p. 83.
- 53. Letra, A., et al., *CRISPLD2 variants including a C471T silent mutation may contribute to nonsyndromic cleft lip with or without cleft palate.* Cleft Palate Craniofac J, 2011. **48**(4): p. 363-70.
- 54. Girardi, A., et al., No evidence for a role of CRISPLD2 in non-syndromic cleft lip with or without cleft palate in an Italian population. Eur J Oral Sci, 2011. 119(1): p. 102-5.
- 55. Murphy, E.V., et al., *The human glioma pathogenesis-related protein is structurally related to plant pathogenesis-related proteins and its gene is expressed specifically in brain tumors.* Gene, 1995 **159**(1): p. 131-135.
- 56. Capalbo, G., et al., Endoplasmic reticulum protein GliPR1 regulates G protein signaling and the cell cycle and is overexpressed in AML. Oncol Rep, 2013.
 30(5): p. 2254-62.
- 57. Gingras, M.C. and J.F. Margolin, *Differential expression of multiple unexpected* genes during U937 cell and macrophage differentiation detected by suppressive subtractive hybridization. Experimental hematology, 2000. **28**(1): p. 65-76.
- 58. Rosenzweig, T., et al., *Related to testes-specific, vespid, and pathogenesis* protein-1 (*RTVP-1*) is overexpressed in gliomas and regulates the growth, survival, and invasion of glioma cells. Cancer Res, 2006. **66**(8): p. 4139-48.
- 59. Chilukamarri, L., et al., *Hypomethylation and Aberrant Expression of the Glioma Pathogenesis-Related 1 Gene in Wilms Tumors.* Neoplasia, 2007. **9**(11): p. 970-978.

- 60. Ren, C., et al., *RTVP-1*, a tumor suppressor inactivated by methylation in prostate cancer. Cancer research, 2004. **64**(3): p. 969-976.
- 61. Ren, C., et al., *Identification and characterization of RTVP1/GLIPR1-like genes, a novel p53 target gene cluster.* Genomics, 2006. **88**(2): p. 163-72.
- 62. Tam, M., et al., Examining Hedgehog pathway genes GLI3, SHH, and PTCH1 and the p53 target GLIPR1/GLIPR1L1/GLIPR1L2 gene cluster using fluorescence in situ hybridization uncovers GLIPR1/GLIPR1L1/GLIPR1L2 deletion in 9% of patients with multiple myeloma. J Assoc Genet Technol, 2010. 36(3): p. 111-114.
- 63. Li, L., et al., *Glioma pathogenesis-related protein 1 exerts tumor suppressor activities through proapoptotic reactive oxygen species-c-Jun-NH2 kinase signaling.* Cancer Res, 2008. **68**(2): p. 434-43.
- 64. Eberle, H.B., et al., *Identification and characterization of a novel human plant pathogenesis-related protein that localizes to lipid-enriched microdomains in the Golgi complex.* Journal of cell science, 2002 **115**(4): p. 827-838.
- 65. van Galen, J., et al., *Interaction of GAPR-1 with lipid bilayers is regulated by alternative homodimerization*. Biochim Biophys Acta, 2012. **1818**(9): p. 2175-83.
- 66. Huang, S., et al., *GLIPR-2 overexpression in HK-2 cells promotes cell EMT and migration through ERK1/2 activation.* PLoS One, 2013. **8**(3): p. e58574.
- 67. Baxter, R.M., et al., *The plant pathogenesis related protein GLIPR-2 is highly expressed in fibrotic kidney and promotes epithelial to mesenchymal transition in vitro.* Matrix Biol, 2007. **26**(1): p. 20-9.
- 68. Aalberts, M., et al., *Identification of distinct populations of prostasomes that differentially express prostate stem cell antigen, annexin A1, and GLIPR2 in humans.* Biol Reprod, 2012. **86**(3): p. 82.
- 69. Park, K.H., Kim, B. J., Kang, J., Nam, T. S., Lim, J. M., Kim, H. T., Park, J. K., Kim, Y. G., Chae, S. W., & Kim, U. H., *Ca2+ signaling tools acquired from prostasomes are required for progesterone-induced sperm motility.* Science Signaling, 2011. **4**(173): p. ra31.

- 70. Palmerini, C.A., et al., *Fusion of prostasomes to human spermatozoa stimulates the acrosome reaction.* Fertility and Sterility, 2003. **80**(5): p. 1181-1184.
- Babiker, A.A., Ronquist, G., Nilsson, U. R., & Nilsson, B. O., *Transfer of* Prostasomal CD59 to CD59-Deficient Red Blood Cells Results in Protection against Complement-Mediated Hemolysis. American Journal of Reproductive Immunology, 2002. 47(3): p. 183-192.
- 72. Miosga, T., et al., Sequence analysis of a 33 · 1 kb fragment from the left arm of Saccharomyces cerevisiae chromosome X, including putative proteins with leucine zippers, a fungal Zn (II) 2 -Cys6 binuclear cluster domain and a putative α2-SCB-α2 binding site. Yeast, 1995. 11(7): p. 681-689.
- 73. Huang, S.G., et al., *Hypoxia promotes epithelial--mesenchymal transition of hepatocellular carcinoma cells via inducing GLIPR-2 expression*. PLoS One, 2013. **8**(10): p. e77497.
- 74. Koshikawa, N., et al., *Purification and identification of a novel and four known serine proteinase inhibitors secreted by human glioblastoma cells.* Journal of biochemistry, 1996. **119**(2): p. 334-339.
- 75. Smith, D.M., et al., Cloning and expression of a novel cysteine-rich secreted protein family member expressed in thyroid and pancreatic mesoderm within the chicken embryo. Mechanisms of developmen. Mechanisms of development 2001. 102(1): p. 223-226.
- 76. Principe, S., et al., *In-depth proteomic analyses of exosomes isolated from expressed prostatic secretions in urine*. Proteomics, 2013. **13**(10-11): p. 1667-71.
- 77. Falak, S., et al., *Protease inhibitor 15, a candidate gene for abdominal aortic internal elastic lamina ruptures in the rat.* Physiol Genomics, 2014. **46**(12): p. 418-28.
- 78. Reeves, J.R., et al., *Identification, purification and characterization of a novel human blood protein with binding affinity for prostate secretory protein of 94 amino acids.* Biochem. J, 2005. **385**(1): p. 105-114.

- 79. Sheth, A.R., et al., *Characterization of a polypeptide from human seminal plasma with inhibin (inhibition of FSH secretion)-like activity.* FEBS letters, 1984.
 165(1): p. 11-15.
- 80. Chao, C.F., et al., *he porcine sperm motility inhibitor is identical to β-microseminoprotein and is a competitive inhibitor of Na+, K+-ATPase.*Biochemical and biophysical research communications, 1996. 218(2): p. 623-628.
- 81. Hirano, M., et al., *Presence of immunoglobulin binding factor on human sperm surface as sperm coating antigen*. Systems Biology in Reproductive Medicine, 1996. **37**(3): p. 163-170.
- 82. Garde, S.V., et al., *Prostate secretory protein (PSP94) suppresses the growth of androgen-independent prostate cancer cell line (PC3) and xenografts by inducing apoptosis.* The Prostate, 1999. **38**(2): p. 118-125.
- 83. Shukeir, N., et al., *Prostate secretory protein PSP-94 decreases tumor growth and hypercalcemia of malignancy in a syngenic in vivo model of prostate cancer.* Cancer research, 2003. **63**(9): p. 2072-2078.
- 84. Frost, R.J. and S. Engelhardt, *A secretion trap screen in yeast identifies protease inhibitor 16 as a novel antihypertrophic protein secreted from the heart.* Circulation, 2007. **116**(16): p. 1768-75.
- 85. Claverie, J.M., *GENE NUMBER: What If There Are Only 30,000 Human Genes?* Science, 2001. **291**(5507): p. 1255-1257.
- 86. Hartl, D.L., *Molecular melodies in high and low C.*. Nature Reviews Genetics, 2000. **1**(2): p. 145-149.
- 87. McEntyre, J. and J. Ostell, *The NCBI handbook*. 2002.
- 88. Letunic, I. and P. Bork, *Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy.* Nucleic acids research, 2011: p. gkr201.
- 89. Sneath, P.H. and R.R. Sokal, *Numerical taxonomy. The principles and practices of numerical classification.* San Fran cisco: WH Freeman, 1973.

90. Wang, J., et al., *Blocking effect and crystal structure of natrin toxin, a cysteinerich secretory protein from Naja atra venom that targets the BKCa channel.* Biochemistry, 2005. **44**(30): p. 10145-10152.

APPENDIX A

DATA COLLECTED JANUARY 2014 - DECEMBER 2015

Scientific Name:	Common Name:	Taxonomy ID:	# of CAP superfamily proteins in genome	NCBI Assigned Gene Name:	# of Protein Coding Genes	Ge nome Size (Mb)
Pseudomonas syringae	N/A	317	4	PSYR_3350, PSYR_2381, PSPTO_3579, and PSPTO_2649	5089	6.09
Anoxybacillus flavithermus	N/A	491915	2	AFLV_RS09705, and AFLV_RS08260	2,784	2.85
Pepto clostridium difficile	N/A	272563	4	Ca++-chelating serine protease, and <u>3X</u> cell wall binding protein	3,756	4.29
Solanum lycopersicum	Tomato	4081	13	PR, <u>2X</u> PR-1, PR-1A, PR-1-like, PR- 4, PR-4-like, PR-4-like, PR-6, PRB1-2- like, PRB1-3, and STS14	Incomplete	Incomplete
Theobroma cacao	Cacao	3641	11	5X CAP, Concanavalin A-like lectin protein kinase family protein, 2X PR-1, PR-1r, STS14, and STS14	Incomplete	Incomplete
Saccharomyces cerevisiae	Yeast (Fungi)	4932	3	PRY1p, PRY2p, and PRY3p	5887	12.07
Aspergillus flavus	Mold (Fungi)	5059	3	SCP-like extracellular protein, CRISP- like, and SCP domain protein PRY1	13485	36.89
Caenorhabditis elegans	Round worm	6239	34	C07A4.2, C07A4.3, C04C11.1, F09B9.5, F57B7.2, F58E2.5M, LON- 1, SCL-1, SCL-2, SCL-3, SCL-5, SCL-6, SCL-7, SCL-8, SCL-9, SCL- 10, SCL-11, SCL-12, SCL-13, SCL- 14, SCL-15, SCL-17, SCL-18, SCL- 19, SCL-20, SCL-21, SCL-22, SCL- 24, SCL-25, SCL-26, SCL-27, T05A10.6, VAP-1, and Y116F11B.13	65767	100.27
Drosophila melanogaster	Fnit fly	7227	32	Ag5r, Ag5r2, SCPr-A, SCPr-B, FLYBASE predicted and hypothetical sequence locus tag: Dmel_CG17575, Dmel_CG5106, Dmel_CG42564, Dmel_CG30488, Dmel_CG43777, Dmel_CG30486, Dmel_CG43777, Dmel_CG4270, Dmel_CG8483, Dmel_CG9822, Dmel_CG43776, Dmel_CG31296, Dmel_CG42764, Dmel_CG31296, Dmel_CG42764, Dmel_CG31482, Dmel_CG10651, Dmel_CG3640, Dmel_CG32679, Dmel_CG17974, Dmel_CG6628, Dmel_CG43775, Dmel_CG11977, Dmel_CG34002, Dmel_CG9400, Dmel_CG8072, Dmel_CG31286, and Dmel_CG34049	30277	143.73

Appendix 1A: Inventory of CAP-related Proteins in Selected Species

Appendix 1: Page 1 of 4

Ciona intestinalis	See squirt	7719	29	Cell wall protein PRY3, CiUrabin, CRISP1-like, Cysteine-rich venom protein pseudechetoxin-like, <u>2X</u> Cysteine-rich venom protein-like, HrTT- 1-like, <u>9X</u> GliPR1-like, GliPR-like2, <u>3X</u> PI16-like, proteoglycan 4, PRB1-2-like, uncharacterized LOC100181082, uncharacterized LOC100180998, uncharacterized LOC104266353, uncharacterized LOC104265473, uncharacterized LOC100179406, uncharacterized LOC10018001, and uncharacterized protein DDB G0285291-like	15438	115.23
Megachile rotundata	Leaf cutting bee	143995	3	Ag3-like, CRISP2-like, and EFG1 protein 1-like	26046	272.66
Apis florea	Little honeybee	7463	4	Ag3-like, Cysteine-rich venom protein pseudecin-like, EFG1 protein 1-like, and Uncharacterized LOC100864764	17663	230.47
S olenopsis invicta	Red fire ants	13686	9	Ag3, <u>3X</u> Ag3-like, Ag5-like, CRISP1- like, PRY1-like, U10-ctenitoxin-Pk1a- like, and uncharacterized LOC105201413	35915	396.01
Nasonia vitripennis	Jewel Wasp	7425	11	2X Ag3-like, 4X Ag5-like, allergen 5.01- like, CRISP2-like, EFG1 protein 1-like, enolase-phosphatase E1-like, and Uncharacterized LOC100680164	25548	295.79
Microplitis demolitor	Parasitic Wasp	69319	12	2X Ag3-like, 6X Ag5-like, serotriffin- like, uncharacterized LOC103580017, uncharacterized LOC103575568, and uncharacterized LOC103573858	19916	250.53
Culex quinquefasciatus	Southern house mosquito	7176	31	Salivary protein, cysteine-rich venom protein, CRSIP2, CRISP3, <u>3X</u> hypothetical protein, <u>5X</u> Ag, <u>6X</u> venom Ag, <u>2X</u> venom Ag3, <u>7X</u> venom Ag5, <u>3X</u> catrin, and salivary secreted antigen- 5 precursor Ag5-3	18883	579.04
Aedes aegypti	Yellow fever mosquito	7159	28	AAEL000793, AAEL002476, AAEL002682, AAEL002693, AAEL003053, AAEL003057, AAEL004199, AAEL004407, AAEL005531, AAEL006297, AAEL005524, AAEL008473, AAEL008479, AAEL008487, AAEL008488, AAEL009239, AAEL009695, AAEL010269, AAEL01795, AAEL01269, AAEL011795, AAEL011802, AAEL012136, AAEL011802, AAEL0124069, AAEL013406, AAEL015472, and AAEL015483	17393	1383.97

Appendix 1A: Inventory of CAP-related Proteins in Selected Species

Appendix 1: Page 2 of 4

Lepisosteus oculatus	Spotted gar	7918	11	2X Cysteine-rich venom protein ophanin- like, CRISPLD1-like, CRISPLD2-like, CLECT18A-like, 2X GliPR1-like, 2X PI15-like, Serotriffin-like, and Uncharacterized LOC102697771	15851	808.24009
Danio rerio	Zebrafish	7955	13	Cysteine-rich venom protein natrin-1- like, Cytidine monophospho-N- acetylneuraminic acid hydroxylase, CRISPLD1A, CRISPLD1B, CRISPLD2, GAPR1, GliPR1-like, GliPR1b, PI15a, PI15b, PI16-like, R3HDML, and m7150988	47848	1371.7
Oreochromis niloticus	Tilapia	8128	11	Cysteine-rich venom protein kaouthin-1- like, Cysteine-rich venom protein pseudecin-like, CRISPLD1, CRISPLD1-like, CRISPLD2-like, CLEC18A-like, GliPR1-like, GliPR1- like, PI15-like, PI16-like, and PI R3HDML-like	45431	927.69
Oryzias latipes	Meadaka (Japanese rice fish)	8090	10	Cysteine-rich venom protein pseudecin- like, CRISPLD1, CRISPLD1-like, CRISPLD2, CLECT18A-like, <u>2X</u> GliPR1-like, PI15-like, PI16-like, and GAPR1	34474	869.8
Xenopus tropicalis	Frog	8364	14	Cysteine-rich venom protein tigrin-like, CRISP1, CRISP2, CRISP2-like, CRISP3, CRISPLD1, CRISPLD2, GliPR1-like, GAPR1, GliPR1L1, MGC108118, PI15, <u>2X</u> PI16-like, R3HDML, Serotriflin-like, and Uncharacterized LOC100127722	28576	1437.51
Python bivittatus	Burmese python	176946	9	Cysteine-rich venom protein latisemin- like, CRISPLD1, CRISPLD2, GliPR1- like, <u>2X</u> GliPR1-like, PI15, PI16, and R3HDML	25736	1435.03
Falco peregrinus	Falcon	8954	10	Cysteine-rich venom protein helothermine-like, CRISPLD1, CRISPLD2, CLECT18A, <u>3X</u> GliPR1- like, PI15, PI16, and R3HDML	16079	1171.96
Picoides pubescens	Woodpecker	118200	9	CRISP2, CRISPLD1, CRISPLD2, <u>2X</u> GliPR1-like, GAPR1, PI15, PI16, and R3HDML	29768	1149.76
Gallus gallus	Chicken	9031	11	CRISP2, CRISPLD1, CRISPLD2, CLEC18A, GliPR1, GAPR1, GliPR1L, PI15, PI16, R3HDML, and Ophanin- like	32169	1046.89
Ornithorhynchus anatinus	Platypus	9258	12	CRISP2, CRISP2-like, CRISP3-like, CRISPLD2, CLEC18A-like, GliPR1, <u>2X</u> GliPR1-like, GAPR1, GliPR1L1, PI15, PI16, and R3HDML	26690	1558.51

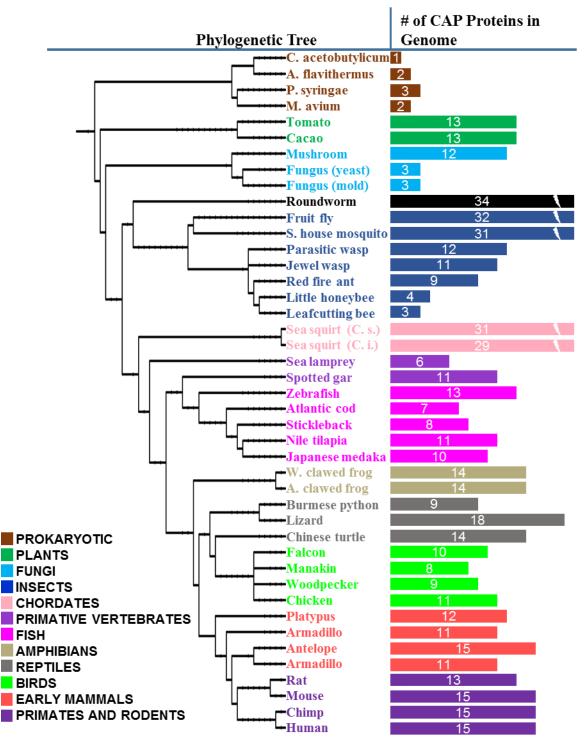
Appendix 1A: Inventory of CAP-related Proteins in Selected Species

Appendix 1: Page 3 of 4

Dasypus novemeinctus	Armadillo	9361	11	CRISP2, CRISPLD1, CRISPLD2, CLEC18A-like, GliPR1, GAPR1, GliPR1L1, GliPR1L2, PI15, PI16, and R3HDML	38189	3599.02
Pantholops hodgsonii	Antelope (chiru)	59538	15	CRISP1, CRISP2, CRISP3, CRISPLD1, CRISPLD2, CLEC18A, GliPR1, GAPR1, GliPR1L1, GliPR1L2, PI15, PI16, R3HDML, Pr protein 1C- like, and Pr protein 1C-like	32266	2696.87
Physeter catodon	Sperm whale	9755	12	CRISP1-like, CRISP2, CRISPLD1, CRISPLD2-like, CLEC18A, GliPR1, GAPR1, GliPR1L1, GliPR1L2, PI15, PI16, and R3HDML	31509	2280.71
Rattus norvegicus	Rat (Norway rat)	10116	13	CRISP1, CRISP2, CRISP4, CRISPLD1, CRISPLD2, CLEC18A, GEPR1, GAPR1, GEPR1L1, GEPR1L2, PI15, PI16, and R3HDML	50693	2870.19
Mus musculus	Mouse (house mouse)	10090	15	CRISP1, CRISP2, CRISP3, CRISP4, CRISPLD1, CRISPLD2, CLEC18A, GHPR1, GAPR1, GHPR1L1, GHPR1L2, GHPR1L3, PI15, PI16, and R3HDML	77995	2800.04
Pan troglodytes	Chimp	9598	15	CRISP1, CRISP2, CRISP3, CRISPLD1, CRISPLD2, CLEC18A, CLECT18B, CLECT18B-like, GliPR1, GAPR1, GliPR1L1, GliPR1L2, PI15, PI16, and R3HDML	56773	3309.56
Homo sapiens	Human	9606	15	CRISP1, CRISP2, CRISP3, CRISPLD1, CRISPLD2, CLEC18A, CLECT18B, CLECT18C, GliPR1, GAPR1, GliPR1L1, GliPR1L2, PI15, PI16, and R3HDML	102462	3221.5

Appendix 1A: Inventory of CAP-related Proteins in Selected Species

Appendix 1A: The following table presents an inventory of selected CAP superfamily genes in various species. CAP genes appear throughout the evolutionary tree carrying a vide array of function and diversity. Additionally, it should also be noted that the number of CAP genes present in an organisms genome is not correlated with organism complexity, quantity of protein coding gene in the genome, or the genome size. Page 4 of 4



Appendix 1B. Phylogenetic Tree vs Number of CAP Protein in Genome

Appendix 1B: The following figure presents a bar graph representation of the quantity of CAP superfamily proteins in various selected organisms. Organisms are color coded based on unique clads. Current data does not suggest any clear pattern of CAP Superfamily gene duplication within and between clads.

						-					-								-			~														
Common name - Protein name															id S				_			_		_		-										
Tomato - PR P4	-	-																					LL				M -									
Cocao Tree - CRISP-like Yeast - PRY1	-	-	-	-		• •																	LA				I - V -									
Mold - CRISP-like	-	-	-	-			-								SI		L	A		A T A T		A I					ĊI									
Roundworm - SCP-like	-	-	-	-			-		-			111	r	1 .	<u> </u>	- r	L	Ê.		нт		T.	ч А гт				c -		_			_				
Fruit Fly - CG42780	-	-	-	-			-	-	-			-	M	S 1	VF	Т	F	N		FV	. Э ГТ	s					AI					Е	xo	n 1		
Parasytic wasp - Serotriffin-like	-	-	-	-			-	-	M	RI		w	M		TI		r v	4	4	L P	A	-	U P				R -									
Jewel wasp - CRISP2-like	-	-	-	-	-			-	-			1.										-							_	_		-				
Fire Ants - CRISP1-like	-	-	-	-				-	M	RN	NL	w	T	IV	V I	V	A	L	A	LP	A	A	VP	R	RI	н	R -					E	xo	n 2		
Dwarfhoney bee - CRISP2-like		-	-						M	RI	NF		÷	i v	v ī	_				LP			V P				R -									
Leafcutter bee - CRISP2-like	-	-	-	-			-	-	M	RΥ	NL	W	ī	IV	V I	Ī		L									R -					F	xo	n 3		
Sea Squirt (C.s.) - Unchar.	-		М	A	F 1	N F	ÝY	F	Т	ΙI	ιī	L	S	v s	S I												E -						AU	u .		
Sea Squirt (C. i.) - CRISP3-like	М				MI				G	LK	ςν	Ē			IF		Ľ			IS		A I			ΝE			-								
Sea lamprey- Unchar.	-	-	-	-			-	-	-			-	1				Р			sν	A	Α.	ΑF	L	LA	A A	L -	-	Г			E	xo	n 4		
Spotted gar - Unchar.	-	-	-	-			-	-	-			-	-		- E					IΤ	V	н	VF	L	FS	L L	Ρ-	-					AU			
Zebrafish - CRISP3	-	-	-	-			-	-	-			-	-	- 1	ΜŇ	S	W	Т	0	GН	ΙT	M	FΑ	L	V I	C	F-	-								
Atlantic Cod - CRISP3	-	-	-	-			-	-	-			-	-			-	-	-	-		-	-		-		-		-				Е	xo	n 5		
Stickleback - CRISP3	-	-	-	-			-	-	-			-	-			-	-	-	-		-	M	FΤ	L	LV	C	Ι.	-								
Tilapa fish - CRISP3	-	-	-	-			-	-	-			-	-	- N	M Q) S	L	Т	G	LΕ	S	ΤÌ	мL	F	ΥТ	F	'Р-	-	_							
Japanese rice fish - CRISP3	-	-	-	-			-	-	-			-	-		- -	-	-	-	-		-	-	- P	S	LI	C	C -	-	L			Е	xo	n 6		
Western clawed frog - Allurin	-	-	-	-			-	-	-			-	-				Μ	D	Т	FN	F	Ι	I C	Ι	S A	L	F -	-								
African clawed frog - Allurin	-	-	-	-			-	-	-			-	-			-				FΝ		I	LC	Ι			Ι.			_		-			,	
Burmese python - CRISP-like	-	-	-	-			-	-	-			-	-			-	Μ	Ι	А	LΙ	V	L	LS	L	A A	A V	L -	-				E	xo	u 7		
Green anole - Cysteine-rich venom P.	-	-	-	-			-	-	-			-	-			-	м	V	L	НG	ιL	Y	I C	V	ΙI	L	L -	-								
Chinese soft-shelled turtle - Unchar.	-	-	-	-			-	-	-			-	-			-	Μ	Ι	L	LΤ	A	F	LC	L	A A	A V	L -	-			<u>_</u>	С	on	se	rve	ed
Falcon - CRISP2	-	-	-	-			-	-	-			-	-			-	м	Ι	L	ΡV	V	F	LC	L.	A A	A V	L -	-			C	0	vst	ojn	60	
Manakin - CRISP2	-	-	-	-			-	-	-			-	-			-	Μ	Ι	L	ΡV	V	F :	LC	L	ΤA	A A	L -	-				C,	150	СШ	CS	
Woodpecker - CRISP2	-	-	-	-			-	-	-			-	-			-	Μ	Ι	L	ΡL	V	F	LC	L	ΤA	A V	L -	-								
Chicken - CRISP2	-	-	-	-			-	-	-			-	-			-	Μ	Ι	L	P A	V V	L	LC	L.	A A	A V	L -	-								
Platypus - CRISP3-like	-	-	-	-			-	-	-			-	-			-	Μ	R	K	LΙ	Ι	C	LΤ	V.	A A	Ι	L -	-								
Armadillo - CRISP2	-	-	-	-			-	-	-			-	-			-	Μ	D	L	L R	L	V	VF	L	A A	Ι	L -	-								
Antelope - CRISP2	-	-	-	-			-	-								-			L	LP	V	V :	LF	L	ΤA	V V	L -	-								
Sperm whale - CRISP2	-	-	-	-		• •	-	-	-				-			-				L P		V .	LF	L	A A	A V	L -	-								
Mouse - CRISP2	-	-	-	-			-	-	-			-	-			-	М	A	W	FQ	V	M	LF	V	F A	I L	L -	-								
Rat - CRISP2	-	-	-	-			-	-	-			-	-			-				FQ		M	LF	V	F A	V V	L -	-								
Chimp - CRISP2	-	-	-	-			-		-								M		~	LP		V.	LF	L	vт	r v	L -	-								
Human - CRISP2	-	-	-	-			-	-	-			-	-			-	Μ	Α	L	LP	' -	V :	LF	L	V I	r v	L -	-								
Tomato - PR P4		-	-	-	- 1		-	-	-			-	-		V I			F			-	-		-		-		-								
Cocao Tree - CRISP-like		-	-	-			-	-	-			-	-		IV			A				Ξ.		1												
Yeast - PRY1		2	-	2	2.2			-	Ξ.			1 2			ΤV								V V			-										
Mold - CRISP-like	G	s	s	Q	P (jΝ	V E	Q	н	G I	FC	JD	Ν				_		_	CK					CG		C	G								
Roundworm - SCP-like	- <u>.</u>	-	Ē	-	2.3		Ē	1-	-	2.2			1		A G			A		FS						A		-								
Fruit Fly - CG42780	- 18	F	C	ĸ	QI		. <u>C</u>	11	L.	G .	1 1	п	I		CN SK			G		FG	G		C P	K i P	D A T B			-								
Parasytic wasp - Serotriffin-like	-	-	-	-			-	-	-			-	-	- 1	эл	. r	M	_	R	LI			х I с т	r D			LI	1								
Jewel wasp - CRISP2-like Fire Ants - CRISP1-like	-	-	-	-		- -	-	-	-	- -		-	-		S K	C P	M T		R	LI LY		D D I	o i p v	S	лл ve	. v . e	1.7	T								
Dwarf honey bee - CRISP2-like	- 1			-	1				÷.,				1		SK				R	L I T V	G	EI	n v P V	ю Г	v B V F	- T	1 1									
Leafcutter bee - CRISP2-like	- 1	-	-	2	2			2	2			-	2		NK			P	R	I V	G	D I	R V	P	VS	5 T	TR	R								
Sea Squirt (C. s.) - Unchar.	-	-	-	-				-	-			-	-		AR		-	A	K I	RA	P	GI		÷	т т т		11	v								
Sea Squirt (C. i.) - CRISP3-like		-	-	-			-	-	-			-	-		HE			R		SL	Δ	G		Δ	V N	T G	FI	ř								
Sea lamprey- Unchar.		-	-	-				-	-			-	-		LF						V D	1 1	UN	D	WR	e e	TR	r.								
Spotted gar - Unchar.														. 1	LF					NC		v	FΤ			F	IS	S								
Zebrafish - CRISP3														- 1				H			A	<u> </u>	s v	s	G.	-	v	C								
Atlantic Cod - CRISP3	-	-	- 1	.		. .	-	-	-	- .		-	- 1	- 1	- -		-	-	-		_	-		-		-	I	C								
Stickleback - CRISP3	-				- 1				- 1	- 1		1.0		- 1	LI	L J	0	0	v	H S	_	C	V D	V	SR	N	I	C								
Tilapa fish - CRISP3		-	-					-	-			-	-		ΓI								ΕQ				L	G								
Japanese rice fish - CRISP3		-	-					-	-			-	-							Υ -		-		-				C								
Western clawed frog - Allurin	-	-	-					-	-			-	-							D G		Т		-				Р								
African clawed frog - Allurin	-	-	-	-				-	-			-	-		Y S					DD				-		-	- 8	S								
Burmese python - CRISP-like	-	-	-	-				-	-			-	-	- (-	FΑ	A A	E S	S								
Green anole - Cysteine-rich venom P.	-	-	-	-				-	-			-	-							ΑE				-			LΝ									
Chinese soft-shelled turtle - Unchar.	-	-	-	-				-	-			-	-				Т	G	Q	I S	Ν	-		-	V A	A S	LS	S								
Falcon - CRISP2	-	-	-	-			-	-	-			-	-	- 1	ΡF	e s	Ι	G	E	ΕP	Е			-	LD) А	LS	S								
Manakin - CRISP2	-	-	-	-			-	-	-			-	-	- 1	ΡF	e s				ΕA				-	VВ) А	LS	S								
Woodpecker - CRISP2	-	-	-	-			-	-	-			-	-	- 1	ΡF					A P				-	FD) А	LS	S								
Chicken - CRISP2	-	-	-	-			-	-	-			-	-	- 5	SF					ΙP		Α		-	Sι	. L	LS	S								
Platypus - CRISP3-like	-	-	-					-	-			-	-	- I	ΡI		Ι					P I	r v	Р	ΥE	A	I S	S								
Armadillo - CRISP2	-	-	-	-			-	-	-			-	-	- 1	L 1			Ρ	A I	DG	ł K.		ΡG		F S	A	LI	L								
Antelope - CRISP2	-	-	-	-			-	-	-			-	-	- 1	LF	e s	F	Ρ	Т	E G	K	D	ΡS	-	FΤ	ΓA	LI	Ι								
Sperm whale - CRISP2	-	-	-	-			-	-	-			-	-	- I	LF	e s	F	Ρ	Т	E A	۱K	D	ΡS	-	FΤ	ΓA	LI	L								
Mouse - CRISP2	-	-	-	-			-	-	-			-	-	- I	LR		Р	L	Т	E G	K	D	ΡD	-	FΤ	r s	LΙ	L								
Rat - CRISP2	-	-	-	-			-	-	-			-	-		LF								ΡD				LI									
Chimp - CRISP2	-	-	-	-			-	-	-			-	-				L	Р	A	E G	ł K.	D	ΡA	-	FΤ		LΙ									
Human - CRISP2	-		-	-			-	-	-			-	-	- 1	LF	S	L	Ρ	А	EG	K	D	ΡA	-	FΤ	ΓA	LI	L								
nullai - Olubi z																																				

Appendix 2: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Sequences

Appendix 2: Page 1 of 7

Tomato - PR P4	CF	A	01	V S	P	o D	Y (LA	v		AP3		A O	V C	i V				-	G P					
Cocao Tree - CRISP-like			K											VN						ΕP					
Yeast - PRY1	Y V	E	N (ΞQ	Т	RТ	Т	ΥE	Т	LΑ	ΡA	A S	ΤА	. Т І	P T				-	S T	ΑT				
Mold - CRISP-like	VI	S	LS	SG	E.	A R	C	CC	L	GМ	ΚH	I R	S S	V 1	ΓР				-	S A	S N				
Roundworm - SCP-like	N V	Н	N C	ΓL	R	SR	Ι	ΑK	G	ΤΥ	V A		GΤ						-	A A	<u>S</u> D)	 		
Fruit Fly - CG42780	ΚL	. N	LO	βD	ΚI	N A	L	ΙK		ΗN				WA	A S	GK.	A K	IK	w	ΤА	CK]	Exor	1
Parasytic wasp - Serotriflin-like	P S		Κŀ	c v		RК		V L		ΗN	FF		тн						-	ΤА					
Jewel wasp - CRISP2-like	TS			c v		RK		V L		ΗN	FF		S R						-		A N		_	-	
Fire Ants - CRISP1-like	-	[N	SK	c v		QK		VE		ΗN	YF		TQ						-		AN			Exoi	12
Dwarf honey bee - CRISP2-like	TS		SK	C V	R	QK	. I	VE		ΗN	YL	. R	TQ						-		AN				
Leafcutter bee - CRISP2-like	TS	S S	SK	A E	V	U K I F	. 1	VE		H N H N	YF		ΤQ S-	- F					÷.		AN		1	Exor	3
Sea Squirt (C. s.) - Unchar.		. 1 T	EZ	-1 E - E	T .	TU	. 1	V R V S		ΗN	-		8 - R -		C A					D A	SD			LAUI	
Sea Squirt (C. i.) - CRISP3-like Sea lamprey- Unchar.	TV	. 1	E I	J E S V	<u> </u>		T	VE		H N	K Y D L	R	R - R N	- N	עני ר די					ΡA	GN				
Spotted gar - Unchar.	TS		6	5 V T V	Å.		1	VK		ΗN	DP		RN		L F				-	тΔ	SN]	Exor	ı 4
Zebrafish - CRISP3	TE		SS	s v	ă	ΟĒ	T	VE		ΗN	AF		R A						-	S A	SN				
Atlantic Cod - CRISP3	PT		P	Δ Τ	· ·	NO		V D		ΗN	AF		RA		-						AN		 		
Stickleback - CRISP3	PD		P	ΔT	õ	AE		ID		ΗN	AF		R A		-					та	S D]	Exoi	15
Tilapa fish - CRISP3	v T		T -	ΓĒ	ĸ	NF	Ť	v N		ΗN	AL		RN							ТА	SN				
Japanese rice fish - CRISP3	PE		Ť Å	Αv	0	A E	ī	V E		ΗN	AF		RA				2.2	1.1	1	PA	A D			-	~
Western clawed frog - Allurin	ML	D	TI	ΕТ	- 2 -	ΝŸ	L	V D		ΗN	LL	_	RS					_		TA	КD			Exoi	10
African clawed frog - Allurin	TI	E	ΤK	СТ	0	NK	I	V D		ΗN	A L	R	RS							ТА	SD				
Burmese python - CRISP-like	TS	R	RI	ΕK	Q I	ΚE	I	V D		ΗN	NL	R	RТ							ΤА	S N		1	Exoi	7
Green anole - Cysteine-rich venom P.	ΤT	R	AI	0 0	Q I	ΚE	I	V D		ΗN	NL	R	RS						1	ТА	RN			LAUI	. /
Chinese soft-shelled turtle - Unchar.	ΤD	N	AV	νò	Q I	КE	I			ΗN	A L	_	RΑ						-	ΤА	RN	T		n	
Falcon - CRISP2	ΤS		ΑI	ρġ	Q I	ΚL	I			ΗN	A L	. R	RG	VB	СР				-	ΤА	S N	T	C	Con	serve
Manakin - CRISP2	T S	R	ΑI	D Q	QI	ΚL	I	V D	R	ΗN	A L	. R	RG	VB	СР				-	ΤА	S N	T .	┙,	cyste	ines
Woodpecker - CRISP2	T S	R	ΑI	D Q	QI	ΚL	I	VΕ	κ	ΗN	ΕL	. R	RΕ	V S	S P				-	ΤА	S N	I			
Chicken - CRISP2	ΤN	R	ТΙ	D Q	QI	ΚL	I	V D		ΗN	ΑL	. R	RR	V 8	S P				-	ΡA	RN	1			
Platypus - CRISP3-like	ΤV	A	TF	εv	Q	QΕ	I	VΝ	K	ΗN	ΤL	. R	RL	VI	ΕP				-	ΡA	K N	T			
Armadillo - CRISP2	ТΤ	Q	ΤÇ	Q V	QI	ΚE	Ι	VΝ	K	ΗN	ΕL	. R	RΝ	V S	S P				-	ΡA	S N				
Antelope - CRISP2	ТТ	ΓQ.	ТС	2 V	Q	RΕ	I	V N	ΓK.	ΗN	ΕL	. R.	κs	V S	S P				-	ΡA	S N	T .			
Sperm whale - CRISP2	ТТ	ΓQ.	ТС	2 V	Q	RΕ	I	V D	νк	ΗN	ΕL	. R.	кs	V S	S P				-	ΡA	S N	T I			
						RΕ	· •	VN	K	ΗN	ΕL	. R	RS	V P					-	ΤG	S D)			
Mouse - CRISP2	ΤN		LQ	2 V	- C.																				
Rat - CRISP2	ΤN	Q		2 V 2 V	Q	RЕ	I	I A	ĸ	ΗN	ΕL		RQ						-	PG	S N				
Rat - CRISP2 Chimp - CRISP2	T N T T	Q		2 V 2 V 2 V	Q Q	R E R E	I I	I A V N	K K		E L E L	. R . R	R Q K A	. V 8	S P				2	PG PA	S N S N	T			
Rat - CRISP2	ΤN	Q			Q	RЕ	I I	I A	K K	ΗN					S P				-	PG PA PA	S N S N S N	T T			
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2	T N T T	Q Q Q Q			QQQ	RE RE RE	I I I	IA VN VN	K K K	H N H N H N	E L E L	R R	KÂ KA	. V S . V S	5 P 5 P				- - - G	PGPAPA	S N S N S N				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4	T N T T		M S	s w	Q Q Q D	R E R E R E		I A V N V N A S	K K K R	H N H N H N A Q	E L E L N Y	. R . R . A	K A K A N S	. V S . V S . R A	5 P 5 P				- - G A		S N S N S N N L				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like			M S M A	s w A W	Q Q Q D D	RE RE RE AN		I A V N V N A S A A	K K K R Y	H N H N H N A Q A Q	EL EL NY EY	R R A	K A K A N S N Q	VS VS RA RI	5 P 5 P 4 -				Α	D C	DL				
Rat - CRISP2 Chimp - CRISP2 Hum an - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1	T N T T		M S M A	s w A W	Q Q Q D D	RE RE RE AN		I A V N V N A S	K K K R Y	H N H N H N A Q	EL EL NY EY	R R A A A V V	K A K A N S	. V S . V S . R A . R 1	5 P 5 P 4 -			 	Α	D C					
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like	TN TT TT AL MK		M S M Z A I I I	S W A W P P E W	Q Q Q D D V V	RE RE AN AQ AP		I A V N V N A S A A S A A T		H N H N H N A Q A Q S N A Q	EL EL NY EY SE	R R A A O V A	K A K A N S N Q V L D S	V S V S R A R I S A	5 P 5 P 4 - 1 - 4 L		- F	ΕF	A A I N	D <mark>C</mark> SV	DL				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like	TN TT TT AL MK		M S M A A I I M B	S W A W P P E W C W	Q Q Q D D V D	RE RE AN AQ AP EG		I A V N V N A S A A S A A T		H N H N A Q A Q S N A Q A Q A Q	EL EL NY EY SE	R R A A A D V A A	K A K A N S N Q V L D S N K	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -	 G H :	- F - T	E H G H	A A IN IS	D <mark>C</mark> SV	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like	TN TT TT AL MK ML		M S M A A I I M B	S W A W P P E W C W E W	Q Q Q D D V D	RE RE AN AQ AP EG		I A V N V N A S A A S A A T A T		H N H N A Q A Q S N A Q A Q A Q	EL EL NY EY SD QL AY	R R A A A D V A A	K A K A N S N Q V L D S N K	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780	T N T T T T A L M K M L M A		M S M Z A I I M K M K M K	S W A W P P E W C W E W	Q Q Q D D V D	RE RE AN AQ AP EG		I A V N V N A S A A S A A T A T		H N H N A Q A Q S N A Q A Q A Q	EL EL NY EY SD QL AY	R R A A A D V A A	K A K A N S N Q V L D S N K	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR. P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like	T N T T T T A L M K M L M A		M S M Z A I I M K M K M K	S W A W P P E W C W E W	Q Q Q D D V D	RE RE AN AQ AP EG		I A V N V N A S A A S A A T A T		H N H N A Q A Q S N A Q A Q A Q	EL EL NY EY SD QL AY	R R A A A D V A A	K A K A N S N Q V L D S N K	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotrifin-like Jewel wasp - CRISP2-like	T N T T T T A L M K M L M L M L	V V V V V V V V V V V V V V V V V V V	M S M A A I M B M B M B M B	S W A W P P E W C W E W	Q Q Q D D V D	RE RE AN AQ AP EG		I A V N V N A S A A S A A T A T		H N H N A Q A Q S N A Q A Q A Q	EL EL NY EY SD QL AY	R R A A A D V A A	K A K A N S N Q V L D S N K	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotifiin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like	T N T T T T A L M K M L M L M L M L	P C C C C C C C C C C C C C C C C C C C	M S M A A I M B M B M B M B M B	S W A W P P E W C W E W	Q Q Q D D V D	RE RE AN AQ AP EG		I A V N V N A S A A S A A T A T		H N H N A Q A Q S N A Q A Q A Q	EL EL NY EY SD QL AY	R R A A A D V A A	K A K A N S N Q V L D S N K	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Pruit Fly - CG42780 Parasytic wasp - Serotrifin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Sea Squirt (C. s.) - Unchar.	T N T T T T A L MK M L M L M L M L M L	Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q	M A M A A I M B M B M B M B M B	S W A W P P E W C W C W C W C W C W C W	Q Q Q D D V D N H H H H	RE RE AN AQ AP EG		I A V N V N A S A A S A A T A T	K K K K K K K K K K K K K K K K K K K	H N H N A Q A Q S N A Q A Q A Q	EL EL NY EY SD QL AY	R R A A A A A A A A A A A A A A A A A A	K A K A N S N Q V L D S N K E A C R N R S N	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Hum an - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotrifiin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s.) - Unchar. Sea Squit (C. i.) - CRISP3-like	T N T T T T A L M K M L M L M L M L M L M L M L M L M L M L	P R R R R R R R R R R R R R R R R R R R	M S M A A I M B M B M B M B M B M B M B M B	S W S W F P P F W C W C W C W C W C W C W C W C W S W	Q Q Q D D V V T D V V H H H H H	R E R E A N A Q A P E G A P G A G S G A G S G P G P G		I A V N V N A S A A S A A A E K A K A K A K A K	K K K K K K K K K K K K K K K K K K K	H N H N A Q Q A Q A Q A A Q A A Q A A Q A	E L E L N Y E Y S L Q L A Y L N R V R V R V R V R V R V R V R V	R R A A A A A A A A A A A A A A A A A A	K A N S V D N K E A D R N R	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Sea Squit (C. s) - CRISP3-like Sea lamprey- Unchar.	T N T T T T A L M K M L M L M L M L M L M L M R M K K M L	P K K K K K K K K K K K K K	M S M A A H M A M H M H M H M H M H M H M H M H M H M H	S W S W F P P F W C W C W C W C W C W C W C W C W S W	Q Q Q D D D N H H H H H H	R E R E R E A N A Q A P E G A T K D S G S G S G F G N T		I A V N V N A S A A S A A A E K A K A K A K A K	K K K K K K K K K K K K K K K K K K K	H N H N A Q Q A Q A Q A A Q A A Q A A Q A	E L E L N Y E Y S L Q L A Y L N R V R V R V R V R V R V R V R V	R R A A A A A A A A A A A A A A A A A A	K A K A N S N Q V L D S N K E A C R N R S N	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR.P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Prarsytic wasp - Serotrifin-like Jewel wasp - CRISP2-like Fre Ants - CRISP2-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. s) - Unchar. Sea lamprey- Unchar. Spotted gar - Unchar.	T N T T T T A L M K M L M L M L M L M L M L M L M K K M L M K K M L M K K M K K M K K M K K M K M	P C C C C C C C C C C C C C	M S M A A I I M K M H M H M H M H M H M H M H M H M M M M	S W S W F P P F W C W C W C W C W C W C W C W C W S W	Q Q Q D D D N H H H H H H	R E R E R E A N A Q A P E G A T K D S G S G S G F G N T		I A V N V N A S A A S A A A E K A K A K A K A K	K K K K K K K K K K K K K K K K K K K	H N H N A Q Q A Q A Q A A Q A A Q A A Q A	E L E L N Y E Y S L Q L A Y L N R V R V R V R V R V R V R V R V	R R A A A A A A A A A A A A A A A A A A	K A K A N S N Q V L D S N K E A C R N R S N	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Roundworm - SCP-like Parasytic wasp - CRISP2-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Sea Squirt (C. s.) - Unchar. Sea Squirt (C. s.) - Unchar.	T N T T T T A L M K M L M L M L M L M L M L M L M L M L M L		M S M A A A A A A A A A A A A A A A A A	S W S W F P P F W C W C W C W C W C W C W C W C W S W	Q Q Q D D D N H H H H H H	R E R E R E A N A Q A P E G A T K D S G S G S G F G N T		I A V N V N A S A A S A A A E K A K A K A K A K	K K K K K K K K K K K K K K K K K K K	H N H N A Q Q A Q A Q A A Q A A Q A A Q A	E L E L N Y E Y S L Q L A Y L N R V R V R V R V R V R V R V R V	R R A A A A A A A A A A A A A A A A A A	K A K A N S N Q V L D S N K E A C R N R S N	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Fire Ants - CRISP2-like Sea Squit (C. s.) - Unchar. Sea Squit (C. s.) - Unchar. Sea Squit (C. s.) - Unchar. Spotted gar - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Matenic Cod - CRISP3	T N T I T I A L M K M L M L M L M L M L M L M L M L M L M L	V V V V V V V V V V V V V V V V V V V	M S M Z A I I M F M F M F M F M F M F M F M F M F M F	S W S W F P P F W C W C W C W C W C W C W C W C W S W	Q Q Q D D D N H H H H H H	R E R E R E A N A Q A P E G A T K D S G S G S G F G N T		I A V N V N A S A A S A A A E K A K A K A K A K	K K K K K K K K K K K K K K K K K K K	H N H N A Q Q A Q A Q A A Q A A Q A A Q A	E L E L N Y E Y S L Q L A Y L N R V R V R V R V R V R V R V R V	R R A A A A A A A A A A A A A A A A A A	K A K A N S N Q V L D S N K E A C R N R S N	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PKY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fre Arts - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s.) - Unchar. Sea lamprey: Unchar. Spotted gar - Unchar. Spotted gar - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Stickleback - CRISP3	T N T T T T A L M K M L M L M L M L M L M L M L M L M		M S M A I M A I M A I M H M H M H M H M H M H M H M M M H M M M M	S W S W F P P F W C W C W C W C W C W C W C W C W S W	Q Q Q D D D N H H H H H H	R E R E R E A N A Q A P E G A T K D S G S G S G F G N T		I A V N V N A S A A S A A A E K A K A K A K A K	K K K K K K K K K K K K K K K K K K K	H N H N A Q Q A Q A Q A A Q A A Q A A Q A	E L E L N Y E Y S L Q L A Y L N R V R V R V R V R V R V R V R V	R R A A A A A A A A A A A A A A A A A A	K A K A N S N Q V L D S N K E A C R N R S N	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotrifin-like Jewel wasp - CRISP2-like Dwarf honey bee - CRISP2-like Sea Squirt (C. 1) - CRISP3-like Sea Squirt (C. 1) - CRISP3-like Sea lamprey- Unchar. Sea Squirt (C. 1) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Tilapa fish - CRISP3	T N T T T T A II M L M L M L M L M L M L M L M L M L M L		M S M A I M A I M A I M H M H M H M H M H M H M H M M M M M M	S W S W F P P F W C W C W C W C W C W C W C W C W S W	Q Q Q D D D N H H H H H H	R E R E R E A N A Q A P E G A T K D S G S G S G F G N T		I A V N V N A S A A S A A A E K A K A K A K A K	K K K K K K K K K K K K K K K K K K K	H N H N A Q Q A Q A Q A A Q A A Q A A Q A	E L E L N Y E Y S L Q L A Y L N R V R V R V R V R V R V R V R V	R R A A A A A A A A A A A A A A A A A A	K A A SQUENCE STATES AND A SQUENCE STATES AND A SQUENCE STATES AND A SAME A SAM	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -		- F - T	E H G H	A A IN IS	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Roundworm - SCP-like Parasytic wasp - CRISP2-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - Unchar. Sea Squirt (C. i.) - Unchar. Sea Squirt (C. i.) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Stickleback - CRISP3 Tlapa fish - CRISP3 Japanese rice fish - CRISP3	T N T T T T A L M K M L M L M L M L M L M L M L M L M L M L		M 2 M 2 A 1 I 1 M k M k M k M k M k M k M k M k M k M k	S WA W P P P P E W C W E W C W C W C W	Q Q Q Q D D V D D V D D V D D V D D V D D V N H H H H H H H H S S S S	R E E A NA A A A A A A A A A A A A A A A		I A V N V N A S A A S A A A E K A K A K A K A K	K K K K K K K K K K K K K K K K K K K	H N N H N H N A Q A Q S N A Q A Q A Q A Q A Q A Q A Q A Q A Q A Q	E L E L N Y E Y S I Q L N Y C	R R A A A A A A A A A A A A A A A A A A	K A K A N S N Q V L D S N K E A C R N R S N	V S R A R C I <td>5 P 5 P 4 - 1 - 4 L 5 -</td> <td>G H : -</td> <td>- F - T D E - L - L - L - A - K - M - M - L - L - L - L</td> <td>E H G H T H V H V H S H G H I N H I K H I G H I G H</td> <td>A A N I N I N I D D I D D I D D I D D I S I S I S I S</td> <td>DC SV GA</td> <td>D L W G</td> <td></td> <td></td> <td></td> <td></td>	5 P 5 P 4 - 1 - 4 L 5 -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	E H G H T H V H V H S H G H I N H I K H I G H I G H	A A N I N I N I D D I D D I D D I D D I S I S I S I S	DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PKY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP2-like Fire Ants - CRISP2-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Spotted gar - Unchar. Spotted	T N T T T T A L M R M L M L M L M L M L M L M L M L M L M L	V V V A V A L Q K K K L K M K	M 2 M 2 A 1 M 2 M 2 M 2 M 2 M 2 M 2 M 2 M 2 M 2 M 2	S W S W A W P P P E W E W C	Q Q Q Q D D V D D V D D V D D V D D V D D V S S S S	R E E A NO A P G T D A C A A P G T D A C A C A P G T D C A C A C G G G G G T D C E E E E E E E E E E E E E E E E E E		I A V N V N A S A A A A S A A A A A A A A A A A A A	K K K K K K K K K K K K K K K K K K K	H N N H N N A Q S N A Q S N A Q A A A A	E L E L N YYES L S L N Y V V V V V V V V V V V V V V V V V V V	R R R R R R R R R R R R R R R R R R R	K A A SQUENCE STATES AND A SQUENCE STATES AND A SQUENCE STATES AND A SAME A SAM	. V S . V S . R A . R I S A C H C H	5 P 5 P 4 - 1 - 4 L 5 -	G H : -	- F - T	E H G H T H T H V H H S H G H H K H H G H H C H C H C H C H C H C H C H C H C H		DC SV GA	D L W G				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PKY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotrifin-like Jewel wasp - CRISP2-like Fre Ants - CRISP2-like Fre Ants - CRISP2-like Dwarfhonzy bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. s) - Unchar. Sea lamprey - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Hapa fish - CRISP3 Hapa fish - CRISP3 Hapa fish - CRISP3 Westem clawed frog - Allurin	T N T T T T A L M M L M L M L M L M L M L M L M L M L	V V V A V A L Q K K K L K M K K	M 2 M 2 M 2 M 2 M 2 M 2 M 2 M 2 M 2 M 2	S WA W S W C W C W	Q Q Q Q D D D D D D D D D D D N H H H H H H H H	R E E E A NO P G T D E E E G P A A O P G T D G G G G G T D E E E G P A D E E E G P O N N K T D N E E E G P		I AVN V N ASAAASAA SAAAAAAAAAAAAAAAAAAAAAAAAAAA	K K K K K K K K K K K K K K K K K K K	H N N A Q Q S N Q A Q Q A A A Q A	E LL NYYES Q LL Q LL W W YES Q LL R W W Y YES W W W A Y W Y A Y W A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	R R R R R R R R R R R R R R R R R R R	K A A SQUE A CONTRACT OF A CON	V S R A R S A C C <td>S P A - I -</td> <td>G H : -</td> <td>- F - T D E - L - L - L - A - K - M - M - L - L - L - L</td> <td>H H H H H H H H H H H H H H H H H H H</td> <td>A A N N N N N N N N N N N N N N N N N N</td> <td>D S - G A A A A A A A A A A A A A A A A A A</td> <td>DLU WG KF TG GC TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E</td> <td></td> <td></td> <td></td> <td></td>	S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GC TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - CRISP2-like Jewel wasp - CRISP2-like Jewel wasp - CRISP2-like Fire Ants - CRISP2-like Eafcutter bee - CRISP2-like Sea Squirt (C. i) - Unchar. Sea Squirt (C. j) - Unchar. Sea Squirt (C. j) - Unchar. Sea Squirt (C. j) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Tapa fish - CRISP3 Tapa fish - CRISP3 Western clawed frog - Allurin Burnese python - CRISP-like	T N T T T T T T T T T T T T T T T T T T	P K K A V I A V A L Q K K K L K M K K R	M 2 M 2 M 2 M 2 M 2 M 2 M 2 M 2 M 2 M 2	S WA W S W C W C W	Q Q Q P D D N H H H H H H H H H H H H H H H H H	R R E ANOP GITO G G G G G TO E E E G POD		I AVN V N ASAAASAA SAAAAAAAAAAAAAAAAAAAAAAAAAAA	K K K R Y S A S I L A A A A F V N N S L S T S N N N	H N N H H N A Q Q A Q Q A Q Q A A Q A Q Q A A A A	E LL NYYES Q LL Q LL W W YES Q LL R W W Y YES W W W A Y W Y A Y W A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	R R R R R R R R R R R R R R R R R R R	K A A SQUENCE STATES AND A SQUENCE STATES AND A SQUENCE STATES AND A SAME A SAM	V S R A R S A C C <td>5 P 5 P 4 - 1 - 4 L 5 -</td> <td>G H : -</td> <td>- F - T D E - L - L - L - A - K - M - M - L - L - L - L</td> <td>H H H H H H H H H H H H H H H H H H H</td> <td>A A N N N N N N N N N N N N N N N N N N</td> <td>DC SV GA</td> <td>DLU WG KF TG GG TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E</td> <td></td> <td></td> <td></td> <td></td>	5 P 5 P 4 - 1 - 4 L 5 -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	DC SV GA	DLU WG KF TG GG TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP2-like Fire Ants - CRISP2-like Leafcutter bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Spotted gar - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3 Westem clawed frog - Allurin African clawed frog - Allurin Burmese python - CRISP-like Green anole - Cysteine-rich venom P.	T N T T T T T T T T T T T T T T T T T T	P KKAVIAVALQKKKLKMKKRK	M 3 M 4 A 11 M 4 M 4 M 4 M 4 M 4 M 4 M 4 M 4 M 4 M 4	S W P P P P P P P P P P P P P P P P P P	Q Q Q D D V D D N H H H H H H D D K N S S S N S S C H D	R R E ANOP GITO G G G G G TO E E E G POD		I AVN V N ASAAASAA SAAAAAAAAAAAAAAAAAAAAAAAAAAA	K K K K K K K K K K K K K K K K K K K	H N N H H N A Q Q A Q Q A Q Q A A Q A Q Q A A A A	E LL NYYES Q LL Q LL W W YES Q LL R W W Y YES W W W A Y W Y A Y W A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	R R R R R R R R R R R R R R R R R R R	K A A SQUE A CONTRACT OF A CON		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GG TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PKY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fre Ants - CRISP2-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Deafcutter bee - CRISP2-like Sea Squit (C. s.) - Unchar. Sea Squit (C. s.) - Unchar. Spotted gar - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - Allurin Burmese python - CRISP-like Green anole - Cysteine-rich venom P. Chinese soft-shelled turte - Unchar.	T N T T T T A L M	V P K K A V I A V A L Q K K K L K M K K R K R	M 3 M 4 A 1 M 4 M 4 M 4 M 4 M 4 M 4 M 4 M 4 M 4 M 4	S W P P P E W C W C W C W C W C W C W C W C W C W	Q Q Q D D V D D N H H H H H H D D K N S S S N S S C H D S	R R E ANOP GITO G G G G G TO E E E G POD		I AVN V N ASAAASAA SAAAAAAAAAAAAAAAAAAAAAAAAAAA	KKK RYSASLAAAA FVNNSLSTSNNNN SLSTSNNNN	H N N A Q Q A A Q Q A A A Q Q A A A Q Q A A A Q Q A A A Q A A A Q A	E LL NYYES Q LL Q LL W W YES Q LL R W W Y YES W W W A Y W Y A Y W A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	R R A A A A A A A A A A A A A A A A A A	KK A SQUEKTAARR NKNTKHNTKKQKRE		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GG TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP2-like Fire Ants - CRISP2-like Leafcutter bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Spotted gar - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3 Westem clawed frog - Allurin African clawed frog - Allurin Burmese python - CRISP-like Green anole - Cysteine-rich venom P.	T N T T T T A L M	VQQ PKKAVIAVALQKKKLKMKKRKRK	M 2 A 11 M 4 A 11 M 4 M 4 M 4	S WA WP P P W S S	Q Q Q D D V D D N H H H H H H D D K N S S S N S S C H D S C	R R E NOPGTOGGGGGTDEEADEEGPDEAP		I AVN V N ASAAASAA SAAAAAAAAAAAAAAAAAAAAAAAAAAA	KKKK RYSA SAAAA FV NNSLSTSNNNNN	H N N H H N A Q Q A Q Q A Q Q A A Q A Q Q A A A A	E LL NYYES Q LL Q LL W W YES Q LL R W W Y YES W W W A Y W Y A Y W A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	R R A A A A A A A A A A A A A A A A A A	K A A SQUE A CONTRACT OF A CON		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GG TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Hum an - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s) - Unchar. Sea Squit (C. s) - CRISP3-like Sea Squit (C. s) - CRISP3- Stickleback - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - Atlurin Burmese python - CRISP3- Burmese soft-shelled turfe - Unchar. Falcon - CRISP2 Manakin - CRISP2	T N T T T T A L M	VQQQ F V PKKAVIAVALQKKKLKMKKRKRKK	M 1 M M M M M M M M M M M M M M M M M M	S WA WP P E WW S S W A WW W S S W W Y W S S W W Y W Y W Y W Y	Q Q Q D D V D D N H H H H H H D D K N S S N S S C H D S C C	R E E NOR GIO GIO DE LA PRIMERIA PRIMERIA PRIMERIA PRIMERIA PRIMERIA DE LA PRIMER		I A V N V N A A A A A A A A A A A A A A A	KKKK RYSA SAAAA FV NNSLSTSNNNNN	H N N A Q Q S N Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A A Q Q A A A C A A Q Q A A A A	E LL NYYES Q LL Q LL W W YES Q LL R W W Y YES W W W A Y W Y A Y W A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	R R A A A A A A A A A A A A A A A A A A	KK NNVDNKEEDNNSNNQNDDNDATDNNN		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GG TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PKY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP2-like Fire Ants - CRISP2-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Sotted gar - Unchar. Spotted gar - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3 Japanese rice fish - CRISP3 Mestern clawed frog - Allurin Burmese python - CRISP3 Burmese oft-shelled turtle - Unchar. Fakon - CRISP2 Woodpecker - CRISP2	T N T T T T T T T T T T T T T T T T T T	V P K K A V I A V A L Q K K K L K M K K R K K K K K K K K K K K K K K K	M 1 M M M M M M M M M M M M M M M M M M		Q Q Q D D V D D N H H H H H H D D K N S S S N S S C H D S C C S	R E E NOR GIO GIO DE LA PRIMERIA PRIMERIA PRIMERIA PRIMERIA PRIMERIA DE LA PRIMER		I A V N V N A A A A A A A A A A A A A A A	KKKK RYSA SAAAA FV NNSLSTSNNNNN	H N N A Q Q S N Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A A Q Q A A A C A A Q Q A A A A	E LL NYYES Q LL Q LL W W YES Q LL R W W Y YES W W W A Y W Y A Y W A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	R R A A A A A A A A A A A A A A A A A A	KK NNVDNKEEDNNSNNQNDDNDATDNNN		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GG TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotrifin-like Jewel wasp - CRISP2-like Fre Ants - CRISP2-like Fre Ants - CRISP2-like Dwarfhonzy bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. s) - Unchar. Sea Squirt (C. s) - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Hapa fish - CRISP3 Hamin Burmese python - CRISP-like Green anole - Cysteine-rich venom P. Chinese soft-shelled turfe - Unchar. Falcon - CRISP2 Woodpecker - CRISP2	T N T T T T T T T T T T T T T T T T T T	QQQ - VPKKAVIAVALQKKKLKMKKRKRKKKR	M 1 A H H H H H H H H H H H H H H H H H H		QQQ DDV DDN HHHHHHDDK NSSSNSSCHDSCC SS	R E E NOR GIO GIO DE LA PRIMERIA PRIMERIA PRIMERIA PRIMERIA PRIMERIA DE LA PRIMER		I A V N V N A A A A A A A A A A A A A A A	KKKK RYSA SAAAA FV NNSLSTSNNNNN	H N N A Q Q S N Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A A Q Q A A A C A A Q Q A A A A	E LL NYYES Q LL Q LL W W YES Q LL R W W Y YES W W W A Y W Y A Y W A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	R R A A A A A A A A A A A A A A A A A A	KK NNVDNKEEDNNSNNQNDDNDATDNNN		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GC TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Hum an - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PKY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Spotted gar - Unchar. Spotted gar - Unchar. Zeberafish - CRISP3 Atlantic Cod - CRISP3 Stickleback - CRISP3 Mapa fish - CRISP3 Japanese rice fish - CRISP3 Westem clawed frog - Allurin Burmese python - CRISP3 Mestern clawed frog - Allurin Burmese soft-shelled turfle - Unchar. Falcon - CRISP2 Manakin - CRISP2 Mondpecker - CRISP2 Patpuspus - CRISP2 Patpuspus - CRISP2 Patpuspus - CRISP2 Patpus - CRISP3 Patpus - CRISP2 Patpus - CRISP3 Patpus - CRISP	T N T T T T T T T T T T T T T T T T T T	QQQ PKKAVIAVALQKKKLKMKKRKRKKKKK KKKKKKKKKKKKKKKKKKKKKKK	M S M A A A A A A A A A A A A A A A A A	S WA WP P W S S S S S S S S S S S S S S S S	Q Q Q D D V D D N H H H H H H D D K N S S S N S S C H D S C S S N	R E E NQPGTDGGGGGGTDEEADEEGPDEAPR		I AV N V N AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	KKKK RYSASL AAAAAA FVNNSLSTSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	H N N A Q Q S N Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A Q Q A A A Q Q A A A C A A Q Q A A A A	E LL NYYES Q LL Q LL W W YES Q LL R W W Y YES W W W A Y W Y A Y W A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	R R AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	KK NNVDNKEEDNNSNNQNDDNDATDNNNDN		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GC TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Sotted gar - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Japanese ric Fish - CRISP3 Westem clawed frog - Allurin African clawed frog - Allurin African clawed frog - Allurin Burmese python - CRISP3 Westem clawed frog - Allurin African cRISP2 Woodpecker - CRISP2 Platypus - CRISP2 Platypus - CRISP2	T N T T T T A L M K M L M L M L M L M L M L M L M L M	QQQ PKKAVIAVALQKKKLKMKKRKRKKKKKKKKKKKKKKKKKKKKKKKKKKK	M S S S S S S S S S S S S S S S S S S S	S WA WP P W S S S S S S S S S S S S S S S S	Q Q Q D D V D D N H H H H H H D D K N S S S N S S C H D S C S S N S	R EE AAAPGTDGGGGGGTDEEAADEEEGPDEAAPRAQEE		I A V N V N A A A A A A A A A A A A A A A	KKKK RYSASLAAAAFVNNSLSTSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	H N N A Q Q A A Q Q A A A Q Q A A A Q Q A A A A Q Q A A A A Q Q A A A A A A Q Q A	E LL NYYES Q LL Q LL W W YES Q LL R W W Y YES W W W A Y W Y A Y W A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y A Y W Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	, R R A A A A A A A A A A A A A A A A A A	KK NNVDNKEEDNNSNNQNDDNDATDNNNDNE		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GC TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PKY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fre Arts - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s.) - Unchar. Sea Squit (C. s.) - CRISP3- Sickleback - CRISP3 Japanese rice fish - CRISP3 Japanese rice fish - CRISP3 Japanese rice fish - CRISP3 Japanese of - Spetien-rich venom P. Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2 Manakin - CRISP2 Platypus - CRISP3-like Amadillo - CRISP2 Antelope - CRISP2	T N T T T T T T T T T T T T T T T T T T	QQQ PKKAVIAVALQKKKLKMKKRKKKKKKKKKKKKKKKKKKKKKKKKKKKKK	M S M S M A I I M S M M M H		Q Q Q D D V D D N H H H H H H D D K N S S S N S S C H D S C S S N S S	R R R A A P G T D G G G G G T D E E A D E E E G P D E A P R A G E E E		I A A S A A A A A A A A A A A A A A A A	KKKK RYSASLAAAAFVNNSLSTSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	H N N A Q Q A A Q Q A A A Q Q A A A Q Q A A A A Q Q A A A A Q Q A A A A A A Q Q A	E L L YYUL HARRAR NASRGAARANNRK REKKNYK	, RR AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	KK NNVDNKEEDNNSNNQNDDNDATDNNNNDNEN		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - M - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GC TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Sotted gar - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Japanese ric Fish - CRISP3 Westem clawed frog - Allurin African clawed frog - Allurin African clawed frog - Allurin Burmese python - CRISP3 Westem clawed frog - Allurin African cRISP2 Woodpecker - CRISP2 Platypus - CRISP2 Platypus - CRISP2	T N T T T T T T T T T T T T T T T T T T	QQQ PKKAVIAVALQKKKLKMKKRKKKKKKKKKKKKKKKKKKKKKKKKKKKKK	M S M S M A I I M S M M M H		Q Q Q D D V D D N H H H H H H D D K N S S S N S S C H D S C S S N S S N	R R R A A P G T D G G G G G T D E E A D E E E G P D E A P R A O E E E E		I AVN AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	KKKK RYSASLAAAAFVNNSLSTSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	H N N A Q Q A A Q Q A A A Q Q A A A Q Q A A A A Q Q A A A A Q Q A A A A A A Q Q A	E L L YYUL HARRAR NASRGAARANNRK REKKNYK	, RR AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	KK NNVDNKEEDNNSNNQNDDNDATDNNNDNENN		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - M - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GC TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Human - CRISP2 Tomato - PR.P4 Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotrifin-like Jewel wasp - CRISP2-like Fre Ants - CRISP2-like Eafcutter bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. i.) - Unchar. Sea Squit (C. i.) - CRISP3-like Sea Squit (C. i.) - CRISP3-like Sea Squit (C. i.) - CRISP3-like Sea Squit (C. i.) - CRISP3 Stickleback - CRISP3 Tlapa fish - CRISP3 Westem clawed frog - Allurin Affican clawed frog - Allurin Burmese python - CRISP-like Green anole - Cysteine-rich venom P. Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2 Manakin - CRISP2 Platypus - CRISP3 Netope - CRISP2 Platypus - CRISP2 Platypus - CRISP2 Natelope - CRISP2 Natelope - CRISP2 Sperm whale - CRISP2	T N T T T T T T T T T T T T T T T T T T	QQQ PKKAVIAVALQKKKLKMKKRKRKKKKKKKKKKKKK	M : 1 I M H M H M H M H M H M H M H M H M H M	S WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW	QQQ DDV DDN HHHHH HDDKN SSSN SSCHDSCC SSN SSN S	R R R AAA EA K S A S P P N N K T D N E N E P D P N H P P P E R R R		I AVN AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	KKKK RYSASLAAAAAFVNNSLSTSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	H N N N A Q Q A A A Q Q A A A A A A A A A	E L L YYUL HARRAR NASRGAARANNRK REKKNYK	, RR AAVAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	K K NNV DNKE E DNN SNN QN DDN DATDN NNN DN ENNN		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - M - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GG TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				
Rat - CRISP2 Chimp - CRISP2 Hum an - CRISP2 Tomato - PR P4 Cocao Tree - CRISP-like Yeast - PKY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s.) - Unchar. Sea Squit (C. s.) - Unchar. Sea Squit (C. i.) - CRISP3-like Sea Squit (C. i.) - CRISP3-like Sea Squit (C. i.) - CRISP3-like Sea Squit (C. i.) - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Mestern clawed frog - Allurin African claw def frog - Allurin African cRISP3 Western clawed frog - Allurin African cRISP3 Western clawed frog - Allurin African CRISP3 Wostern CRISP2 Manakin - CRISP2 Woodpecker - CRISP2 Chicken - CRISP2 Platypus - CRISP2 Mouse - CRISP2 Sperm whale - CRISP2 Mouse - CRISP2 Mouse - CRISP2	T N T T T T T T T T T T T T T T T T T T	QQQ PKKAVIAVALQKKKLKMKKRKRKKKKKKKKKKKKKKKKKKKKKKKKKKK	M S A I I I M K M H M H M H M H M H M H M H M H M H		Q Q Q D D V D D N H H H H H H D D K N S S S N S S C H D S C S S N S S N S N S N S N S N S N S N	R R R A A A E A A S A S P P N N K T D N E N E P D P N H P P P E R R R I		I V NN SAAA EAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	KKKK RYSASLAAAAAFVNNSLSTSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	H N N H H N N A Q Q A A Q Q A A A Q Q A A A A Q Q A A A A Q Q A A A A A Q Q A	EE NESOALRRRRRNASRGAARANNRKREKKNYKKKK	, RR AAVAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	KK NNVDNKEEDNNSNNQNDDNDATDNNNDNENNNN		S P A - I -	G H : -	- F - T D E - L - L - L - A - K - M - M - L - L - L - L	H H H H H H H H H H H H H H H H H H H	A A N N N N N N N N N N N N N N N N N N	D S - G A A A A A A A A A A A A A A A A A A	DLU WG KF TG GG TG GC TG GC TG GC TG GC TG GC TG SS SS TE E E				

Appendix 2: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Sequences

Appendix 2: Page 2 of 7

							LΕ	CA	D4	1 1		- I - E														
Tomato - PR P4	ΙH	S G	-		-		A		r+ NL	AK	G	GG	D F	Т	FR		-		- /	AA	v					
Cocao Tree - CRISP-like	VН				-	- P						S D								A						
Yeast - PRY1	КТ	ΤD) S	ΤТ	Т	LΤ				SQ					ГТ	SТ	Р	- A	A A	A S 1	Т					
Mold - CRISP-like			-		-	- R	A	GQ	N L	ΥE	G	S D	P A	DI	U V I	к -	-		- 0	A (Ι					
Roundworm - SCP-like	LG	ΕN	L	ΥW	Y	WТ		ΑT	ΙT	ΝI	D	QF	GΑ	тс	βS		-				-			Fv	on 1	
Fruit Fly - CG42780	R L	SG	Q	ΝL	F	A M	G	FS	ΗA	RΙ	T	ΚТ	ΚM	IN N	1 T 1	LS	Μ	LF	ΕN	I A V	V			ĽA	011	
Parasytic wasp - Serotriflin-like	LH	ID	A		F	GS	C	GQ	ΝI	FΙ	S	ΤA	QV		VF				- 1	A	I					
Jewel wasp - CRISP2-like	LH	TD	A		F	GS	C	GQ	NI	FI	S	ΤA	RV		VF				- 1	A	v.			Ex	on 2	
Fire Ants - CRISP1-like	LY	LD	G		F	GQ	S	GQ	NI	FI	S	TG	RT		_				- 1	P .	I					
Dwarf honey bee - CRISP2-like	LY	LD	G		F	GQ	S	GQ	NI	FI	T	TR	RT		_				- 1	Р	I	_		-	_	
Leafcutter bee - CRISP2-like	LY	LD	G		F	Ģ Q	S	G Q	NI	FI	T	TG	QT	LV	VN				- 1	A .	1			Ex	on 3	
Sea Squirt (C. s.) - Unchar.	55	TN			- r	5 I e U	1	U E	NL		. Q		5 5	IN I	I K	S K	. D			» V . с т -	+					
Sea Squirt (C. i.) - CRISP3-like	DV	TT	. "			WΤ	C	GE	NI	FN	1 S	C A	PF	TV	UN N	- 1		r 1	1 1	(A)	T			F v	on A	i i
Sea lamprey- Unchar.		1 1	. 7			SE			N I N L	YN		S A T D	r r P S	TN	V IN V IN						ī			E.X	on 4	
Spotted gar - Unchar. Zebrafish - CRISP3	PM	IN	-			YE			NL	FK		TG	1 8	s v	V T		-			2 1 1	v.					
Atlantic Cod - CRISP3	RM	L N		2.2	G				NL	FY		T S	A L	DV	V A	2.2			2.4	rv.	Ť			Ex	on 5	
Stickleback - CRISP3	RM	I N				ΥĒ			NL	FY			PY	PV	v T					v	v					
Tilapa fish - CRISP3	RV	T S	_						NL	YY			Q N	TV	vs				- 1	J V I	Ť.	_		_		
Japanese rice fish - CRISP3	RM	L N	-			ΥĒ	_		NL	FΥ			PM	L. V	vк				- 3	х т -	Ť.			Ex	on 6	j –
Western clawed frog - Allurin	RQ	10	D	ΡF		γv			ΝI	ΥV			КР	DV	V A				- 7	AA	v					
African clawed frog - Allurin	RH	IK		PI	D	v v			ΝI	YN			κs	DV	V S				- 1	r v :	Ι	_	_	-	_	
Burmese python - CRISP-like	R T	VG	-		_	LQ			NL	FM			ΡY	s v	VТ		-		- /	A V C	Т			Ex	on 7	
Green anole - Cysteine-rich venom P.	R I	I N	-		G	ĸк			N L	YN		ΤА	P N	ΡV	V S				- I	A	V					
Chinese soft-shelled turtle - Unchar.	R S	Τ-	-		s	ΤG	C	GΕ	N L	ΥM	1 S	тκ	РН	S V	V S				- I	A	Ι		0	Co	nsei	rve
Falcon - CRISP2	R R	Τ-	-			ΤQ			N L	FM	1 S	S A	P F	s v	V P		-		- I	o v d	Ι		C	CV.	stein	es
Manakin - CRISP2	RR	Τ-	-			V L			ΝI	FM			P F	s v	V S		-		- I	D V I	L			- 3	, cent	
Woodpecker - CRISP2	RR	Τ-	-			V G			N L	FM			P F	ΡV	V S				- I	D A C	Ι					
Chicken - CRISP2	R E	Ι.	-			QΡ			N L	ΥM		ΤА	P S	s v	V S		-		- I	D S :	Ι					
Platypus - CRISP3-like	RQ	I K	-			ΗS			NL	L F		ТҮ	S L	S V	V S				- I	A	Ι					
Armadillo - CRISP2	RK	Τ-	-	1.1		ТК			NL	YN			PS	S V	V S				- 1	G	I					
Antelope - CRISP2	RK	T -	-			тк			NL	YN		S D	ΡM	AV	V S				- I	D A	I					
Sperm whale - CRISP2	RK	T -	-			ТК			NL	YN		S D	PT	AV	VS				- I	A	I					
Mouse - CRISP2	RK	1 -	-		N	IR			NL	YN			PT	LV	vs				- 1	ΓV.	1					
Rat - CRISP2	RK	1 -	-			IK			NL	YN		TD	PT	SV	VR				- 1	ΓV.	1					
Chimp - CRISP2	K K	+ -	-			TR		G E G E	N L N L	YN	1S	S D S D	PT	SV	V S U C					SA	1					
Human - CRISP2	K K	1 -	-		s	TR		GΕ	NL	1 10	15	5 0	r I	SV	v S	- г		C 4	- : P1	5 A (1					
Towned DD D (*** **		E D	D	_								0.1		~ ~										
Tomato - PR P4	QL	W V	S	ER	. P	D -	-	- Y	ΝΥ	A T	. N	QC	VG	GK	C M	CG	H	ΥT	QN	/ V V	W.					
Tomato - PR P4 Cocao Tree - CRISP-like	Q L K M			EK		D - Y -	2		DH			Q C T C	AS				H									
	KМ		D		V	-	- S				Ν	T C			γv	C A	Н		QI		w					
Cocao Tree - CRISP-like	KМ	W V	D	ΕK	V	Υ-	s	- Y		G S	Ν	T C	A S	GQ	γv	C A	Н	ΥT	QI	/ V V	w					
Cocao Tree - CRISP-like Yeast - PRY1	KМ	W V T P W H	D A I N	ΕK	V T K	Υ-	-	- Y Q A - Y		G S	N S N	T C A S T C	A S	G Q D S N A N 1	QV SD AI G	C A L S C G	H D H	YT FA YT	Q V S S Q V	/ V V	W L D					
Cocao Tree - CRISP-lake Yeast - PRY1 Mold - CRISP-lake Roundworm - SCP-lake Fruit Fly - CG42780	K M T S D A A A Q K	WV TP WH WE	D A IN K	EK AT EH	V T K	Ч- ТТ D-	G 1	-Y QA -Y WS	DH AA NY	GS TS DS TL KK	N S N S I	TC AS TC MS	AS SS GP	G C D S N A	QV SD AI G	C A L S C G	H D H H	YT FA YT	Q V S S Q V	/ V V 8 V I 7 S I	W L D					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like	K M T S D A A A Q K K T	WV TP WH WE	D A IN K	EK AT EH EF	V T K	Y - T T D - D Y	G 1	-Y QA -Y WS	DH AA NY SN	GS TS DS TL	N S N S T H	TC AS TC MS	AS SS GP LF	G C D S N A N T P H L S	Q V S D S Q G Q S Q	C A L S C G I G	H D H H	YT FA YT	Q V S S Q V	/ V V 8 V I 7 S I	W L D					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly- CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like	K M T S D A A A Q K K T K T	WV TP WH WE WA WY WF	A A K G L S	E K A T E H E F E E E H E E	V T Q K E R	Y - TT D - DY DY I - L -	G 1	-Y QA -Y WS	DH AA NY SN	GS TS DS TL KK GP G-	N S N S T	TC AS TC MS TP AK PD	AS SS GP LF NP ND NE	G C D S N A N T P H	Q V S D S Q G Q S Q	C A L S C G I G	H H H H H	YT FA YT AT LT		/ V V 8 V I 7 S I	W L D					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jawel wasp - CRISP2-like Fire Ants - CRISP1-like	K M T S D A A A Q K K T K T K T R M	WV TP WH WZ WA WY WF WY	D A IN G I S M	E K A T E H E F E E E H E E E Y	V T Q K E R	Y - T T D - D Y	G 1	-Y QA -Y WS	DH AA NY SN	GS TS DS TL KK GP G- GP	N S N T H S N	TC AS TC MS TP AK PD TT	AS SS GP LF NP	G C D S N Z P H L S L C I I	V SD C G C S Q S Q S C C C C C C C C C C C C C C	CALS CG IG VG VG IG	H D H H H H	YT FA YT AT LT		/ V V 8 V I 7 S I	W L D					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarfhoney bee - CRISP2-like	K M T S D A A A Q K K T K T K T R M R M	W V T P W E W A W Y W Y W Y W Y	A A A A A A A A A A A A A A A A A A A	E K A T E H E F E E E H E E E Y E Y	V T Q K E R	Y - TT D - DY DY I - L -	G 1	-Y QA -Y WS	DH AA NY SN	GS TS DS TL KK GP G- GP GD	N S N H S N D	TC AS TC MS TP AK PD TT EG	AS SS GP LF NP ND NE	G C D S N A P H L S L C I I L H	V SD CG CG V SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ	CALS CG IG VG VG IG IG	H H H H H H	YT FA YT AT LT		/ V V 8 V I 7 S I	W L D					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotrifiin-like Jawel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like	K M T S D A A A Q K K T K T K T R M	W V T P W E W A W Y W Y W Y W Y	A A A A A A A A A A A A A A A A A A A	E K A T E H E F E E E H E E E Y	V T Q K E R	Y - TT D - DY DY I - L -	G 1	-Y QA -Y WS	DH AA NY SN	G S T S D S T L K K G P G - G P G D G D	N S H S D D	TCAS MS TP AK PD TT EG KA	AS SS GP LF NP ND NE	G C D S N P F L S L C I I L H L H	V S D S D S D S D S D S D S D S D S D S	C A L S C G I G V G V G I G I G I G I G I G	H H H H H H H	YT FA YT AT LT		V V V S V I V S I M A V M A V M V V M V V V V V V V V I A V	W L D W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriftin-like Jawel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcuttre bee - CRISP2-like Sea Squirt (C. s.) - Unchar.	K M T S D A A A Q K K T K T K T R M R M	W V T P W E W A W Y W Y W Y W Y	A A A A A A A A A A A A A A A A A A A	E K A T E H E F E E E H E E E Y E Y	V T Q K E R	Y - TT D - DY DY I - L -	G 1	-Y QA -Y WS	DH AA NY SN	GSTS DSTL KKGP GP GD GD DT	N S H S D N F	TCAS MS MS TP AK PD TT EG KA	AS SS F P N P N D N E N D N E N D N D N I A	GC DS N P L S L L L L L L L L L L K	V SD C G C SQ SR SQ SR SQ SR SQ SR SQ SR SQ SR SQ SR SQ SR SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ SQ	C A L S C G I G V G V G I G I G I G I G C G	H H H H H H H H	YT FA YT AT LT		V V V S V I V S I M A V M A V M V V M V V V V V V V V I A V	W L D					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jawel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. s.) - Unchar. Sea Squirt (C. i.) - CRISP3-like	K M T S D A A A Q K K T K T K T R M R M R M R M Q A F N	WV TP WH WY WY WY WY WF WH	A A A A A A A A A A A A A A A A A A A	E K A T E F E E E E E E E Y E Y E I T E T	V T Q K E R	Y - TT D - DY DY I - L -	G 1	-Y QA -Y WS	DH AA NY SN	GS TS TL KK GP GP GD GD T ST	N S H S D N F Q	TCAS MS MS TP AK PD TT EG KA SC	AS SS GP LF NP ND NE	G C S S S S S S S S S S S S S S S S S S	V D D C C C C C C C C C C C C C	C A L S C G I G V G V G I G I G I G C G C G	H H H H H H H H H	YT FA YT AT LT		V V V S V I V S I M A V M A V M V V M V V V V V V V V I A V	W L D W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Jewel wasp - CRISP2-like Eastrutter bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. s.) - Unchar. Sea Squirt (C. s.) - CRISP3-like Sea lamprey- Unchar.	K M T S D A A A Q K K T K T R M R M R M R M Q A F N R D	W V T P W E W A W Y W Y W Y W Y W F W H W Y	A A A A A A A A A A A A A A A A A A A	E K A T E H E F E E E H E E E Y E Y	V T Q K E R	Y - TT D - DY DY I - L -	G 1	-Y QA -Y WS	DH AA NY SN	GSTS DSTL KGP GGD GGD TST D	N S H S N D N F Q G	TCAS MS MS TP AK PD TT EG KA	AS SS F P F N P N D N E N D N E N D N E N D N I A	G C S N N T P H S C I I I I I I I I I I I I I I I I I I	V C C C C C C C C C C C C C C C C C C C	C A L S C G I G V G V G I G I G I G V	H H H H H H H H H H H H H H H H	YT FA YT AT LT		V V V S V I V S I M A V M A V M V V M V V V V V V V V I A V	W L D W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Rruit Fly - CG42780 Parasytic wasp - Serotriftin-like Jewel wasp - CRISP2-like Fire Ants - CRISP2-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - Unchar. Sea Squirt (C. i.) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar.	K M T S D A A A Q K K T K T R M R M R M R M R M Q A F N R D Q S	W V T P W H W Z W Y W Y W Y W Y W Y W Y W Y W Y W Y	A A A A A A A A A A A A A A A A A A A	E K A T E F E E E E E E E Y E Y E I T E T	V T Q K E R	Y - TT D - DY DY I - L -	G 1	-Y QA -Y WS	DH AA NY SN	GS TS TL KK GP GP GD GD T ST	N S H S D N F Q	TC AS MS TP AK PD TT EG KA SC ST SV	AS SS F P F N P N D N E N D N E N D N E N D N I A	G C N A P H L S L H H K K G C	V S D S C S C S C S C S C S C S C S C S C	C A S C G I G I G V G V G V G I G I G C G V G V G I G V G I G I G I G I G I G I G I G I	H H H H H H H H H H H H H H H H H H H	YT FA YT AT LT		V V V V I	W L D W W W W W W W W W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. s.) - Unchar. Sea Squirt (C. s.) - Unchar. Sea Squirt (C. s.) - Unchar. Spotted gar - Unchar.	K M T S D A A A Q K K T K T K T R M R M R M R M R M R M R M R D Q S D A	W V T P W E W Y W Y W Y W Y W Y W Y W Y W H	A A A A A A A A A A A A A A A A A A A	E K A T E F E E E E E E E Y E Y E I T E T	V T Q K E R	Y - TT D - DY DY I - L -	G 1	-Y QA -Y WS	DH AA NY SN	GSTS DSTL KGP GGD GGD TST D	N S H S N D N F Q G	TCAS MS MS TP AK PD TT EG KA SC	AS SS F P N P N D N E N D N E N D N D N I A	G C S N P H S L L H H K G T G G C	V S D S C S C S C S C S C S C S C S C S C	C A S C G G G G G G G G G G G G G G G G G G G	H H H H H H H H H H H H H H H H H H H	YT FA YT AT LT		V V V V I	W L D W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Evel wasp - CRISP2-like Dwarfhoney bee - CRISP2-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3	K M T S D A A A Q K K T K T K T R M R M R M R M R M R M R D Q S D A T A	W V T P W H W E W A W Y W F W Y W Y W Y W Y W H W Y W Y W H W H W H	A A A A A A A A A A A A A A A A A A A	E K A T E F E E E E E E E Y E Y E I T E T	V T Q K E R	Y - TT D - DY DY I - L -	G 1	-Y QA -Y WS	DH AA NY SN	GSTS DSTL KGP GGD GGD TST D	N S T H S N D D N C F C Q C G G G	TCS MSC MSC MSC MSC MSC MSC MSC MSC MSC M	AS SS F P N P N D N E N D N E N D N D N I A	G C C C C C C C C C C C C C C C C C C C	V S D S C S C S C S C S C S C S C S C S C	C A S C G I G V G V G G V G G G C C C C C C C C C C C C C C C C C	H H H H H H H H H H H H H H H H H H H	YT FA YT AT LT		V V V V I	W L D W W W W W W W W W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Stickleback - CRISP3	K M T S D A A A Q K K T K T K T R M R M R M R M R M Q A F N R D Q S D A T A N A	W V T P W H W E W A W Y W F W Y W Y W Y W Y W H W Y W Y W H W H W H	A A A A A A A A A A A A A A A A A A A	E K A T E F E E E E E E E Y E Y E I T E T	V T Q K E R	Y - TT D - DY DY I - L -	G 1	-Y QA -Y WS	DH AA NY SN	G S S S S S S S S S S S S S S S S S S S	N S H S D D N F Q G G G G G G	TC S S S S S S S S S S S S S S S S S S S	AS SS F P N P N D N E N D N E N D N D N I A	G C C C C C C C C C C C C C C C C C C C	2 V 5 D 5 D 6 V 7 G 8 V 9 C 9 C 10 C 11 G 12 C 13 C 14 C 15 C 15 C 15 C 15 C	C A S C G C G C G C G C G C G C G C G C G C G	H H H H H H H H H H H H H H H H H H H	YT FA YT AT LT		V V V V I	W L D W W W W W W W W W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - Unchar. Sea Squirt (C. j.) - Unchar. Sea Squirt (C. j.) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Tilapa fish - CRISP3	K M T S D A A A Q K K T K T K T K T R M R M R M Q A F N R D Q S D A A S S	W V T P W H W E W A W Y W F W Y W Y W Y W Y W Y W Y W H W H W H W Y Y W Y	A A A A A A A A A A A A A A A A A A A	E KA I E H E F E E H E E H E E H E Y E Y E Y E Y E V E V E V E V E V E V E V E V	VTK QK ERKKKNHISNQAK	Y - T T T D - D Y D I I - V D - V D - V D - V D - V D - V D	G 1	- Y A - Y S - F F F F F F F F F F F F F F F F F F F	D HA A A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	G S T S T L K K G P G G G G G G G G G G G G G G G G D V G V P N G V P N G V	N S N D D C F C Q C G G G G G G G	T C S T C S T S S T S S T S S T S S T S S T S S T S S T S S T S S T S S S T S	A S S P I F P N D E N D D A N N N D A N N N D A N N N N N D N N N N N N N N N N N N N N N N N N N	G C C C C C C C C C C C C C C C C C C C	2 V 5 D 5 D 6 V 7 G 7 G 8 V 9 F 9 F 9 F 9 F 9 F 9 F 9 F 9 F 9 F 9 F 9 F 9 F	CLSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	H H H H H H H H H H H H H H H H H H H	Y T F A T T T T T T T T T T T T T T T T T T T	Q X 2 Q X 2 Q M Q M Q M Q M Q M Q M Q M Q M Q M Q M	7 V	W L D W W W W W W W W W W W W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. s.) - Unchar. Sea Squirt (C. s.) - Unchar. Sea Squirt (C. s.) - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3	K M T S D A A A Q K K T K T R M R M R M R M R M R M R M R D Q S D A T A N A S N A	W V T P W H W E W A W Y W F W Y W Y W Y W Y W Y W H W H W H W H W Y W Y W Y	A A A A A A A A A A A A A A A A A A A	E KA T E H E E E E E E E E E E E E E E E V E E V E E V E E V E E V E E V E E V E E E E	VTK QKERKKKNHISNQAKS	Y - T T T D - D Y D I I - D - V D - V D - V D - V D - V D - V D -	G 1	- Y A - Y S - F F F F F F F - F F F F F F F F F F F	D H A A N Y N	G S S S S S S S S S S S S S S S S S S S	N S N D D N C C C C C C C C C C C C C C C C	T A S A S M S M S A A A D T A A A D T A A A D T A A A A A A A A A A A A A A A A A A A	A S S S P S S P F P N N E E D D A N N N N N N N N N N N N N N N N N N N	G C C S C C C C C C C C C C C C C C C C	2 V 5 D 5 D 6 V 7 G 7 G 8 Q 1 G 1 G 2 S 2 S 2 S 3 S 4 S 5 S 5 S 5 S 5 S 5 S 5 S 5 S 5 S 5 S 5 S 5 S 5 S 5 S	C I G G I G I I G I I G I I G G I I G G I I G G G I I G G G G G G I I G	H H H H H H H H H H H H H H H H H H H	Y T F A Y T Y T Y T Y T Y T Y T Y T Y T Y T Y T		7 V	W L D W W W W W W W W W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Frie Ants - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - CRISP2-like Sea Squirt (C. i.) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3 Japanese rice fish - CRISP3	K M T S D A A A Q K K T K T R M R M R M R M R M R M R M R D Q S D A T A N A Q S N A N S	W VY T PP W H W F W YY W F W Y W Y W Y W Y W Y W Y W Y W Y W Y W Y	A A A A A A A A A A A A A A A A A A A	E KA I E H E E E E E E E E E E E E E E E V V E E V V E V V E V V V E V V V E E V V V E E E E V E E E E	VTK QKERKKKNHISNQAKSN	Y - I D Y D I - I D - D - D D - D - D D - S P N - N H - N D - H H - D D H - D 	G 1	- Y S - Y S - F F Y F F Y F F F Y F F F F F F F F F	D H A A N Y N	G S D S T L K K G P G C G C G C G C G C G C G C G C C C C V P N G V P N G V P N G V	N S S S S S S S S S S S S S S S S S S S	T A S A S M S T P A K C C T A K S C C S S S S S S S S S S S S S S S S S	A S S S P S S P I P N D E E D D A N N D E E D D A N N N N N N N N N N N N N N N N N N N	G C S S S S S S S S S S S S S S S S S S	2 V 2 V 3 D 3 D 6 V 6 V 6 V 7 C 7 C 7 C 7 C	C I G	H H H H H H H H H H H H H H H H H H H	Y T A F A T T Y T br>Y	Q X Y <td>7 V</td> <td>W L D W W W W W W W W W W W W W W W W W</td> <td></td> <td></td> <td></td> <td></td> <td></td>	7 V	W L D W W W W W W W W W W W W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriftin-like Jawel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - Unchar. Sea Squirt (C. j.) - Unchar. Sea Squirt (C. j.) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Tilapa fish - CRISP3 Tilapa fish - CRISP3 Japanese rice fish - CRISP3 Westem clawed frog - Allurin	K M T S D A A A Q K K T K T K T K T K T R M R M R M R M R M R M R M R D Q A F N R D Q A S D A S D S	W V T PP W H W F W Y W Y W Y W Y W Y W Y W Y W Y W Y W Y	DA KGLSMMM M N N N N N N N N N N N N N N N N N	E KA T E H E E E E E E E E E E E E Y E E Y E E Y E E V E E E E	V T K Q K E R K K K N H I S N Q A K S N S	Y - I D Y D I D D D D N N H H D N D -	G 1	- Y S A E F F Y F F Y F F F F F F F F F F F F F	D H A A A Y D I Y	G S S S S S S S S S S S S S S S S S S S	N S S S S S S S S S S S S S S S S S S S	TAS MSPTGAK PTTGAK SSISST SSISST SSISST NI	A S S S S S S C P L F P N D E E D D A N N E D D A N N E D D A N N E N N N N E N N N N N E N N N - N N N S - N N S - S - S	G C 2 2 N P H 2 2 I I I H K K C 2 0 G G K C K K K C K K K C K K K C K K K C K K K C K K K C K K K C K K K C K K K C K K K K C K	V V S D S D G V G V G V G V G V G V G V G V C V C V C V C V C V C V	C I G		Y T F A T T T T T T T T T T T T T T T T T T T	Q I Q I Q I Q I Q I Q I Q I I I Q I I <td>7 V</td> <td>W L D W W W W W W W W W W W W W W W W W</td> <td></td> <td></td> <td></td> <td></td> <td></td>	7 V	W L D W W W W W W W W W W W W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wap - Serotriflin-like Jewel waps - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - CRISP2-like Sea Squirt (C. i.) - CRISP3-like Sea Squirt (C. i.) - CRISP3-like Sea Imprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Adantic Cod - CRISP3 Stickleback - CRISP3 Talpa fish - CRISP3 Japanese rice fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - Allurin Burnese python - CRISP-like	K M T S D A A A Q K K T K T R M R M R M R M R M R M R M R D Q A F N R D Q S D A A N A S N S D S Q S	W V T PP W H W E W A W Y W Y W Y W Y W Y W Y W Y W Y W Y W Y	DA KGLSMM M N N N N N N N N N N N N N N N N N	E KA TI E H E F F E E E E E E E E E E E E E E E E E	V T K Q K E R K K K N H I S N Q A K S N S	Y T T D Y D I - D - D - D - P - D - P - N	G 1	- Y S A E F F Y F F Y F F F F F F F F F F F F F	D H A	G SS	H S D D T F Q C G C G C G C G C G C G C G C G C C G C	TASCAR SISSICAN SISSI	A S S S P I F P N D E E D D A P - N N N N N N N N N N N N N N N N N N N	G C C S C S C S C S C S C S C S C S C S	Q V S Q S Q G Q S <td>C A S</td> <td></td> <td>Y T F A T T Y T</td> <td>Q I<td>7 V</td><td>W L D W W W W W W W W W W W W W W W W W</td><td></td><td></td><td></td><td></td><td></td></td>	C A S		Y T F A T T Y T	Q I <td>7 V</td> <td>W L D W W W W W W W W W W W W W W W W W</td> <td></td> <td></td> <td></td> <td></td> <td></td>	7 V	W L D W W W W W W W W W W W W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Sea Squit (C. s) - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Atlantic Cod - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3 Westem clawed frog - Allurin Atlance and frog - Allurin Burmese python - CRISP-like Green anole - Cysteine-rich venom P.	K M T S D A Q K K T K T R M R M R M R M R M R M R M R M R M R M	W V T PP W H W E W A W Y W Y W Y W Y W Y W Y W Y W Y W Y W Y	A A A A A A A A A A A A A A A A A A A	E KA T E H E E E E E E E E E E E E Y E E Y E E Y E E V E E E E	V T K Q K E R K K K N H I S N Q A K S N S K H	Y T T D Y D I - D - D - D - P - D - P - N	G 1	- YA - YS 	D H A A A Y D I Y	G SS	H S D D T F Q C G C G C G C G C G C G C G C C G C	TASCAR ACCTANA AND A AND	A S S S P I F P N D E E D D A P - N N N N N N N N N N N N N N N N N N N	G C S S S S S S S S S S S S S S S S S S	Q V S Q	C A S	H H H H H H H H H H H H H H H H H H H	Y T F A T T Y T L T J T Y T		7 V	W L D W W W W W W W W W W W W W W W W W					
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wap - Serotriflin-like Jewel waps - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - CRISP2-like Sea Squirt (C. i.) - CRISP3-like Sea Squirt (C. i.) - CRISP3-like Sea Imprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Adantic Cod - CRISP3 Stickleback - CRISP3 Talpa fish - CRISP3 Japanese rice fish - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - Allurin Burnese python - CRISP-like	K M T S D A A A Q K K T K T R M R M R M R M R M R M R M R D Q A F N R D Q S D A A N A S N S D S Q S	W VY T PPW W H W E W Y W Y W Y W Y W Y W Y W Y W Y W Y W Y	A A A A A A A A A A A A A A A A A A A	E KA T E H E F E E E E E E E E E E E E E E E E E E	V T K Q K E R K K K N H I S N Q A K S N S K H E	Y T T D Y D I - I - D - D - P - P - N - N - H - D - - - - - - - - - - - - - - - - - -	G 1	- Y A E F F F Y F F F Y F F F F F F F F F F F	D H A A N Y N	G S S S S S S S S S S S S S S S S S S S	N S S S S S S S S S S S S S S S S S S S	TASC SAUT GAUCT APTEKSISSSITNI SSSPKPIT	A S S S P I P N D E E D D A P - N N N E E D D A P - N N N N N N N N N N N N N N N N N N N	G C S S S S S S S S S S S S S S S S S S	Q V S Q	Image: Constraint of the sector of		Y T A F A Y T Y T T T Y T T T Y T T Y T T Y T T Y T T Y T T Y T T								
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Rruit Fly - CG42780 Parasytic wasp - Serotriftin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcuttre bee - CRISP2-like Sea Squirt (C. i.) - Unchar. Sea Squirt (C. j.) - Unchar. Sea Squirt (C. j.) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Matantic Cod - CRISP3 Tlapa fish - CRISP3 Mestem clawed frog - Allurin Burmese python - C RISP-like Green anole - Cysteine-rich venom P. Chinese soft-shelled turtle - Unchar.	K M M T S S D A Q K K T T K T T K T T R M R M R M R M R M R M R M R M R M S S D S D S Q S Q S Q S Q S Q S Q S	W VY T PPW W H W E W Y W Y W Y W Y W Y W Y W Y W Y W Y W Y	DANKGLSMMM MMNNNNSDKNNDDSD	E KA THE A T	V T K Q K E R K K K N H I S N Q A K S N S K H E E	Y - T T T D Y D I	G 1	- Y A E F F F Y F F F F F F F F F F F F F F F	D H A A N Y	G S S S G G P C G G C T L K K K G P	N S H S N D N F Q G G G G G G G G	TASC MSPKDTGACCTVIGATAPTGAC	A S S P G P T P D D E E D D A P T N N N E E D D A P T N N N N N N N N N N N N N N N N N N N	G C S A A A A A A A A A A A A A A A A A A	V V S <td>A S G</td> <td></td> <td>Y T A F A Y T T Y T T T T</td> <td>Q I<td>7 V</td><td></td><td></td><td></td><td></td><td></td><td></td></td>	A S G		Y T A F A Y T T Y T T T T	Q I <td>7 V</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	7 V						
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly- CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - Unchar. Sea Squirt (C. i.) - Unchar. Sea Squirt (C. i.) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Manic Cod - CRISP3 Tapa fish - CRISP3 Western clawed frog - Allurin African clawed frog - Allurin Burnese python - CRISP-like Green anole - Cysteine-rich venom P. Chinese soft-shelled turtle - Unchar.	K M M T S S D A Q K K T T K T T K T T R M R M R M R M R M R M R M R M R M S S D S D S D S D S Q S Q S Q S Q S Q S	W VY T PPW W H W YY W Y W Y W Y W Y W Y W Y W Y W Y W	DANKGLSMMM MMNNNNSNSDKNNDDSDD	E KA THE E E E E E E E E E E E E E E E E E E	V T K Q K E R K K K N H I S N Q A K S N S K H E E	Y - T T T D Y D I	G 1	- Y A E F F F Y F F F F F F F F F F F F F F F	D H A A Y N S N L Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	G S S S C P P C P P C P P N N N N N N N N N N N	N S S S S S S S S S S S S S S S S S S S	TASC MSPKDTGACC TAPTGACC TSSSSSSSS SSPKPAPKK SSSSSSSS SSPKPAPKK	A S S P P N N P D D D D D D D D D D D D D D D D D	G C S A A A A A A A A A A A A A A A A A A		A S G		Y T A F A T T T T T T T T T T T T T T T T T T T	Q X <td>7 V</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	7 V						
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. s.) - Unchar. Sea Squirt (C. s.) - Unchar. Sea Squirt (C. s.) - Unchar. Sea Squirt (C. s.) - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Stickleback - CRISP3 Manin C CRISP3 Japanese rice fish - CRISP3 Mestern clawed frog - Allurin Burmese python - CRISP-like Green anole - Cysteine-rich venom P. Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2	K M T S A A A Q K K T R M R M R M R M R M R M R M R M R M R M	W VY T PPW W H W YY W Y W Y W Y W Y W Y W Y W Y W Y W	DA NKGLSMMM NNNNNSDKNNDDSDD	E K T H F F F F F F F F F F F F F F F F F F	V T K Q K E R K K K N H I S N Q A K S N S K H E E K	Y - T T T D Y D I	G 1	Y Y S E W S E F F Y F	D H A A Y N S D L Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	G S	N S S S S S S S S S S S S S S S S S S S	TASC MSPTEGACCTTVIGATAPTEGACCTTVIGATAPTEGACCTTVIGATAPTEGACCTTVIGATAPATACACCTTVIGATAACAC	A S S P P N N P D D D D D D D D D D D D D D D D D	G C S A A A A A A A A A A A A A A A A A A		C G		Y T <tht< th=""> <tht< th=""> <tht< th=""> <tht< th=""></tht<></tht<></tht<></tht<>	Q X <td>7 V</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	7 V						
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Roundworm - SCP-like Roundworm - SCP-like Parasytic wasp - CRISP2-like Jewel wasp - CRISP2-like CRISP2-like CRISP2-like CRISP2-like Sea Squirt (C. i) - CRISP3-like Sea Squirt (C. i) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Stickleback - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3 Japanese rice fish - CRISP3 Japanese rice fish - CRISP3 Westem clawed frog - Allurin African clawed frog - Allurin Burmese python - CRISP-like Green anole - Cysteine-rich venom P. Chinese soft-shelled turtle - Unchar. Fakon - CRISP2 Manakin - CRISP2	K M T S A A A Q K K T R M R M R M R M R M R M R M R M R M R M	W VY PROVIDENT OF CONTRACT	DA NKGLSMMM NNNNNSDKNNDDSDDDD	E K T H F E E E E E E E E E E E E E E E E E E	V T K Q K E R K K K N H I S N Q A K S N S K H E E K N K	Y - T T T D Y D I	G 1	Y Y S E W S E F F Y F	D H N	G S S S S S S S S S S S S S S S S S S S	N S S S S S S S S S S S S S S S S S S S	TA S C M S P K D T G A C C T Y A C S S S S S S S S S S S S S S S S S S	A S S P P P D D A P T T T A	G C S A A A A A A A A A A A A A A A A A A		A S G		Y TAATIYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	Q S V Q	7 V 1 X 1 X 1 X 1 X 1 X 1 X 1 X 1 X 1 X 1 X X 1 X X 1 X X 1 X X 1 X						
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriftin-like Jawel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcuttre bee - CRISP2-like Sea Squirt (C. i.) - Unchar. Sea Squirt (C. j.) - Unchar. Sea Squirt (C. j.) - CRISP3-like Sea lamprey- Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Stickleback - CRISP3 Tlapa fish - CRISP3 Westem clawed frog - Allurin Aftican clawed frog - Allurin Burmese python - CRISP-like Green anole - Cysteine-rich venom P. Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Woodpecker - CRISP2	K M M T S A A A A A A Q K K T R M M R K T R R M M R R M M R R M M Q S D A A T A A Q S D A A T A A Q S Q S Q S Q Q S Q Q A Q Q A	W V T P W H W Y W Y W Y W Y W Y W Y W Y W Y W Y W Y	DANKG LSMMMMNNNNSNSDKNNDDSDDDDN	E KATEEREEREEREEREEREEREEREEREEREEREEREEREER	V T K Q K E R K K K N H I S N Q A K S N S K H E E K N K K	Y - T T T D Y D I	G 1	VA-FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	D H A A Y N L S D K T K K K F T Q K K T L S A T A K E V K K Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	G S S S T L K P C P C P C P C P C P C C C C C C C C	N S I S I S I S I S I S I S I S I S I S	TA S C M S P K D T G A C C T Y A C T A P T G A C C T Y A P T T G A C C T Y A A A A A A A A A A A A A A A A A A	A S S P P P N N N E D D A P T N N N N N N N N N N N N N N N N N N	G C S A A A A A A A A A A A A A A A A A A		CLCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	H H H H H H H H H H H H H H H H H H H	Y T A T T YYYYYYYYYYYYYYYYYYYYYYYYYYYYY		7 V 1 X 1 X 1 X 1 X 1 X 1 X 1 X 1 X 1 X 1 X X 1 X X 1 X X 1 X X 1 X						
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wap - Serotrifin-like Jewel waps - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - CRISP3-like Sea Squirt (C. I.) - CRISP3 Japates - Unchar. Zebrafish - CRISP3 Japanese rice fish -	K M M T S A A A Q K K T R M R M R M R M R M R M R D Q S Q A A Q S Q S Q S Q S Q Q S Q Q S Q Q S Q Q S Q Q S	W V T P W H W Y W Y W Y W Y W Y W Y W Y W Y W Y W Y	DANKG LSMMMMNNNNSNSDKNNDDSDDDDDND	EATHEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE	V T K Q K E R K K K N H I S N Q A K S N S K H E E K N K K G	YTTDY DDY ILDDD-PP-P	G 1	YAYSEFFFYFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	D H A A Y N L S D L Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	G S S S C P D D D T V V V V V V V V V V V V V V V V	N S I S I S I S I S I S I S I S I S I S	TASCSISSISSISSISSISSAAAAAAAA	A S S P F P D E E D D A P T N N N N N N N N N N N N N N N N N N	G D N N P L 2 O L H H K H O O O H K O K K K A A A A A A A A A A A A A A A		CLCCCVICGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	H H H H H H H H H H H H H H H H H H H	Y T A T T YYYYYYYYYYYYYYYYYYYYYYYYYYYYY		7 V 1 7 V 1 1 7 V 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1						
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriffin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarfhoney bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squit (C. s.) - Unchar. Sea Squit (C. s.) - Unchar. Spotted gar - Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Atlantic Cod - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3 Westem clawed frog - Allurin Burmese python - CRISP-like Green anole - Cysteine-rich venom P. Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Woodpecker - CRISP2 Woodpecker - CRISP2 Platypus - CRISP2 like Armadillo - CRISP2	K M M T S A A A A A Q K K T T R M M R M M Q A F N D Q S D A A K T N A Q S D A A T N A Q S S Q Q S Q S	W V T P W F W Y W F W Y W Y W Y W Y W Y W Y W Y W Y W Y W Y	DANKGLSMMMMNNNSDKNNDDSDDDDNDE	EAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE	V T K Q K E R K K K N H I S N Q A K S N S K H E E K N K K G	YTD YDD I	G 1	YAYSEFFFYFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	DAAAN N N L N N N N L N	G S S S S S S S S S S S S S S S S S S S	N S I I S N D N F Q G G G G G G G G G G G G G G G G G G	TASCS MTAPTEKSISSSSPKPAPKKKTTK MTAPTEKSISSSSPKPAPKKKTTK	A S S P F P D E E D D A P T T N N N N N N N N N N N N N N N N N	G D N N P I S O I I I I I I I I I I I I I I I I I		CLCCCVICGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG		Y T A A A A A A A A A A A A A A A A A A								
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - CRISP2-like Frie Ants - CRISP2-like Fire Ants - CRISP2-like Cafcuttre bee - CRISP2-like Leafcuttre bee - CRISP2-like Sea Squirt (C. i.) - Unchar. Sea Squirt (C. j.) - Unchar. Sea Squirt (C. j.) - CRISP3-like Sea lamprey- Unchar. Zebrafish - CRISP3 Atlantic Cod - CRISP3 Stickleback - CRISP3 Tlapa fish - CRISP3 Markin and frog - Allurin African clawed frog - Allurin African clawed frog - Allurin Surmese python - CRISP-like Green anole - Cysteine-rich venom P. Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Marakin - CRISP2 Chicken - CRISP2 Chicken - CRISP2 Platypus - CRISP2	K M M T S A A A A A A A Q K K T T R M M R R M M R R M M R R M M Q S S Q S S S S S S	W VY HH HW YY YW YYY YW YY YW YYY YW YYY YW YYY YW YYY YW YYY YW YYYY YW YYYYY YW YYYYY YW YYYYYY	DANKGLSMMMNNNSNSDKNNDDSDDDDDNDEE	EAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE	V T K Q K E R K K K N H I S N Q A K S N S K H E E K N K K G H H	YTD YDD I	G 1	YAYSEFFFYFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	DAANS L YNL S L YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	G T S S S S S S S S S S S S S S S S S S	T G G G G G G G G G G G G G G G G G G G	TASCS APPTE A CCT VIGTINA PTE A CT VIGTINA PTE A CT VIGTINA PTE A CT VIGTINA PA PA A A A A A A A A A A A A A A A A	A S S P F P D E E D D A P T T N N N N N N N N N N N N N N N N N	G D S A A G G G A G G D D N A A A A A A A A A A A A A A A A A		C I C G	H H H H H H H H H H H H H H H H H H H	Y T A A T T T T T T T T T T T T T T T T								
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly- CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Sea Squirt (C. i.) - Unchar. Sea Squirt (C. i.) - Unchar. Sea Squirt (C. i.) - Unchar. Sea Squirt (C. i.) - CRISP3-like Sea lamprey- Unchar. Zebrafish - CRISP3 Mantic Cod - CRISP3 Stickleback - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - Allurin Burmese python - CRISP3 Mantim - CRISP2 Woodpecker - CRISP2 Manakin - CRISP2 Platypus - CRISP2 Platypus - CRISP2 Platypus - CRISP2 Artelope - CRISP2 Manakilo - CRISP2 Martelope - CRISP2 Sperm whale - CRISP2	K M M T S A A A A A A A Q K K T T R M M R R M M R R M M R R M M Q S S Q S S S S S S	W VY HE WW YY YW YY YY	DANKGLSMMMNNNNSNSDKNNDDSDDDDNDEEN	EAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE	V T K Q K E R K K K N H I S N Q A K S N S K H E E K N K K G H H E	Y T D Y H D D I I L D D D P P I L D D D D S N N H N D H D D D	G 1	VA VSEFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	DAAY NSDKTKKKFTQKKTLSATAKEVKKEKQIIV	G I S S S G G G G D I I V V I I V V V V V V V V I I I V S I V S A C G G G G G G G G G G G G G G G G G G	N S S S S S S S S S S S S S S S S S S S	TASCSTRATESSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	A S S F F P D D D A P T N N N N N N N N N N N N N N N N N N	G D N N P L L L H K H O O O H K O K K K A A A A A A A A A A A A A A A	VI VI SDI VI SDI VIII SDI VIII SDI VIII SDI VIII	A S G		Y YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY								
Cocao Tree - CRISP-like Yeast - PRY1 Mold - CRISP-like Roundworm - SCP-like Fruit Fly - CG42780 Parasytic wasp - Serotriflin-like Jewel wasp - CRISP2-like Fire Ants - CRISP1-like Dwarf honey bee - CRISP2-like Leafcutter bee - CRISP2-like Sea Squirt (C. i.) - CRISP3-like Sea Squirt (C. i.) - CRISP3 Stockleback - CRISP3 Atlantic Cod - CRISP3 Stockleback - CRISP3 Japanese rice fish - CRISP3 Western clawed frog - Allurin Burmese python - CRISP3 Manakin - CRISP2 Manakin - CRISP2 Woodpecker - CRISP2 Chicken - CRISP2 Patryus - CRISP2 Sperm whale - CRISP2 Mouse - CRISP2	K M T S	W VY HE WW YY YW YY YY	DANKGLSMMM NNNNSNSDKNNDDSDDDDNDEENE	EAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE	V T K Q K E R K K K N H I S N Q A K S N S K H E E K N K K G H H E E	YTDYIIIDDDDSNNHNDHD	G 1	VAYSERFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	DAAYNSDK NSDK TQKKKTLSATAKEV KKEV YN YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	G I S S S G G G G D I I V V I V N V V V V I I I V V S I V V V V V V V I I I V V S I V V S I V S A V V V V V I I I V V S I V S A V V V V V V V V V V V V V V V V V	N S S S S S S S S S S S S S S S S S S S	TASCSTRATES AND TERMINAPTER AND TERMINAPTER AND TERMINAPTER AND TERMINAPTER AND	A S S P F P D E E D D A P T N N N N N N N N N N N N N N N N N N	G D N N P I S C I I H H K H C C C H K C K K K A A A A A A A A A A A A A A A	Q V <td>C I C C I I C C I</td> <td>H D H H H H H H H H H H H H H H H H H H</td> <td>X X X X X X X X X X X X X X X X X X X</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	C I C C I I C C I	H D H H H H H H H H H H H H H H H H H H	X X X X X X X X X X X X X X X X X X X								

Appendix 2: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Sequences

Appendix 2: Page 3 of 7

Stickleback - CRISP3 N S S Y K V G C G M T L C P N G I Y F Y A Tilapa fish - CRISP3 Y R S N Q I G C A M A Y C P N S A Y K Y F Y V Japanese rice fish - CRISP3 N S S Y K V G C G V T L C S N Y K Y F Y V Japanese rice fish - CRISP3 N S S Y K V G C G V T L C S N Y K Y F Y V Western clawed fiog - Allurin A K T Y L L G C A Y N F C K E N K Y P H V S I African clawed fiog - Allurin A K S Y L L G C A Y N F C K E N K Y P H V F V Burmese python - CRISP-like Y K S R L L G C A A A K C P A S K F K Y Y V Exon 7 Green anole - Cysteine-rich venom P. Y R S Y Q V G C G F A R C P N K R Y E N Y Y V Chinese soft-shelled turtle - Unchar. Y T S Y Q V G C A I A V C L E S E F N Y Y V Y			_									ieu		~4			•						_
Casa Ters. CR19: As P = N = N = N = N = N = N = N = N = N =																							
Yaut PAUL A. E. M. S. K. K. K. A. L. H. K. D. T. A. L. S. W. D. T. A. S. Y. A. D. Y. Y. Y. Y. D. D. Y. A. D. Y. Y. Y. Y. Y. D. D. D. D. Y. Y. Y. Y. Y. Y. Y. Y. Y. D. D. D. Y.																							
Mod. C132F-Me H. H H. H K. H K. H K. M. K K.																							
Readword SCP Me KK T N K M T		A E	HI	N K	KI	ξA.	LH	K I	Т	P A	L	s w	S 1	ТС	LA	S	ΥA	QD	Y 2	A D	N	ΥD	
Fact Proc 20230 No No </td <td></td> <td>IH</td> <td>H ·</td> <td></td> <td></td> <td></td> <td></td> <td>KI</td> <td>I H</td> <td>CN</td> <td>1 A .</td> <td>LE</td> <td>V.</td> <td>A Y</td> <td>ES</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>-</td> <td>FΚ</td> <td></td>		IH	H ·					KI	I H	CN	1 A .	LE	V.	A Y	ES						-	FΚ	
Particle warp - Screeting key P T T		AK	Τ.				- N	LI	_					ЭK	DT	N	3 -	- F	NK	C V	Т	V V	Exon 1
Jeed many CRUSP2-Mac A T S · · H L V G C V X R C V X R C V V Y V Exton 2 Daarf meer (SRY2-Mac A T T · · H L V G C V X R C V X R C V V X V V			S -				- N	ΑV		~	_			A L	GΕ	I 1	R -		- 1	RΥ	N	LΑ	
Fire Am. CR18P 14a A T T H L V G C V S H C T G K G P L G K D F F M Y V Laskatterbe - CR18P 14a A S T H V G C V S H C T G S R G P L G K D F F M Y V Laskatterbe - CR18P 14a A S T H V G C G V S H C T G S R G P L G K D F F M Y V Sa Sagar(C 1) - CR18P 14a A S T H V G C G V S H C T G S R G P L G K D F F M Y V Sa Sagar(C 1) - CR18P 14a A D S H K V G C G A A C C R K O L G K V F L L V V Sa Sagar(C 1) - CR18P 14a A D S H K V G C G A A C C R K O L G K V F L L V V Sa Sagar(C 1) - CR18P 14a A D S H K V G C G A A C C R K S S S S N F F V V Zhodán - CR18P 1 S S		-	T -					RV									PR		P 1	ſF			
address c </td <td></td> <td></td> <td>-</td> <td></td> <td></td> <td></td> <td>- н</td> <td>LV</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>PR</td> <td>GΥ</td> <td>P 1</td> <td>ſF</td> <td></td> <td></td> <td>Exon 2</td>			-				- н	LV									PR	GΥ	P 1	ſF			Exon 2
Ladiant fee - CH3P2-Jake A S T + H N V G C G V S H C T C S R G P L G K D F H Y V Ses Squrt(C.) - CARS JAke A D S + H K V G C G A A F C G R A F C A R A T S T V V V Ses Squrt(C.) - CARS JAke A D S + H K V G C G A A F C G R A F C A S H S K V V V V Ses Squrt(C.) - CARS JAke A D S + H K V G C G A A F C G R Q F A S H S K V V V V Ses Squrt(C.) - CARS JAke A D S + H K V G C G A A F C G R Q F A S H S K V V V V Set Squrt C. 1 - CARS JAke A D S + H K V G C G A V N V C N F O V A K V V L V V V Set Squrt C. 1 - CARS JAke A D S		-	T -				- н	L /	/ G							. G 1	ΡL	GK	DI	FΥ			
Ses Squr(C. 1). Undur. AK T - + K I O C G A A F C R A F C R F X T V G - Y R L V S FLV S Ses Squr(C. 1). Undur. Y N S - F Q L G C A A F C R N S S O V - A F V V F Y V V FLV V Ses Squr(C. 1). Undur. Y N S S - F Q L G C A A A F C R N S S O V - A S I Y V V V FLV V Ses Squr(C. 1). Undur. Y N S S - Y S S O C A A A F C R N S S O V - A S N Y F V O C S N Y K Y F Y V FLV V F Y V Zebrafin - CRISP3 Y S S S - Y S S C G C A A A F C P S S N - - S S N Y F V O C Y A S S S N - - Y S Y F V O C Y A A A Y C P S S S N - - - Y S Y F V V FLV O C Y A A Y C P S S N - - Y S Y V Y V Y V V F Y V Y V V F Y V Y S I A A A C O V A I A Y C P S S S N - - Y F Y V Y V F V V S I A A A C O V A I A Y C F N S S N - - Y F Y V Y V V F Y V Y V V F V V S I A A A C P A S S N - - Y F Y V Y V F V V S I A A A C P A S S N - - Y F Y V Y V F V V V S I A A A C P A S S N - - Y F Y V Y V V F V V S I A A A C P A S S N - - Y F Y V Y V V F V V V V V V V V V V V V		-	T -				- н	L /	/ G	-							ΡL	GK	DI	FF	M	ΥV	
Se Sagar(C. 1) Undar. A K T			T -				- н	МV	/ G							. G 1	ΡL	G K	DI	FF	M .	ΥV	Exon 3
Set Lamprov Umbar Y N S			T -				- Y	ΚI	[G							S	ΤV	G -	YI	₹L			
Spottaging: Unchan Y R S ···· N N I G C A V A F C P N S S ···· S N Y F Y V Adarsi: Cod. CR3P3 N S S ···· V C C C A V A F C P N S S ···· S N Y F Y V Adarsi: Cod. CR3P3 N S S ···· V C C C A V A C C P N S A ···· S N Y F Y V Adarsi: Cod. CR3P3 N S S ···· V K V C C C A V A C P N S A ···· S N Y F Y V Tapa fish: CR3P3 N S S ···· V K V C C C A V A C P N S A ···· V N Y V Tapa fish: CR3P3 N S S ···· V K V C C C A V A C P N S A ···· V Y V Y Tapa fish: CR3P3 N S S ···· V V V L L G C C A V A C P N S A ···· V Y V Y Tapa fish: CR3P3 N S S ···· V V V L L G C C A V A C P N S A ··· V Y V V V Paramet in Cr3M Pish: Name N S S ···· V V V L L G C C A V A Y C P N S A ··· V V V V Creen anall - Cysteine cr3N warms N S ···· V V V G C C A A A Y C P N S R ···· V V V V Falsen odivelade varle ···· V N V V V S ···· V V V C C C A A A Y C P N S R ···· V V V V V Falsen odivelade varle ···· V N V V V V S ···· V V V C C C A A A Y C P N S R ···· V V V V V V Falsen odivelade varle ···· V N V V V V V V V V V V V V V V V	Sea Squint (C.i.) - CRISP3-like						- H	ΚV	/ G	CG	ΗA.	A F	C	QН	QF	Α	SН		- 8	SΙ	_		
Spotted are: Urebar. Y R & S Y R & S Y R & S Y R & Y	Sea lamprey- Unchar.						- F	QI	G	CA	V I	ΝΥ	C	ΡK	ΝP	G	V -		ΑI	CΥ	L	ΥV	Exon 4
Adamic Cod. CR3P3 N S S	Spotted gar - Unchar.	YR	S -				- N	ΝI	ΙG	CA	ι ν.	A F	C	ΡN	S S	-			ΥF	CΥ	F '	ΥV	
SetSetSetSet - CRISP3 N S S N Q G C M T L C P N N Q G C M T L C P N N Q G C N T L C P N N X V G C O V T L C P N N X V Y V V V V V V V V V V V V V V C C O V T L C S N	Zebrafish - CRISP3	Y S	S -				- Y	ΕV	/ G	C A	V V	ΤQ	C	3 -		-			S 1	γ	F 1	ΥG	
Tapastaria (CAISP3) Y R S N Q I Q C AM A Y C P N S A N Y Y Y Japastaria (CAISP3) Y R S A Y R Y C C C O Y L C S N N I Y L Y Y Japastaria (CAISP3) A K T Y L L G C C L A F C P G N Y Y P Y V S I Bannes python - CRISP-Me Y K S Y L L G C C A Y N C P K S K Y P Y Y Y Y Bannes python - CRISP-Me Y K S P K L L Q C A A A K C P N K R Y E N Y Y Y Crime ande - Cysine rich vrame Y R S P V Q V G C A I A V C L E S E Y N Y Y Y Crime ande - Cysine rich vrame Y R S Y R V A C C C A A X Y C P N S R Y R Y Y Y Y Chaine - CRISP2 H N S Y N V A C C C A A X Y C P N S R Y R Y Y Y Y Matkin - CRISP2 H N S Y N V A C C C A A X Y C P N S R Y R Y Y Y Y Chaine - CRISP2 H N S Y N V A C C C A A X Y C P N Q E S I K Y Y Y Y Y Chaine - CRISP2 Y N S Y N V C C C C A A Y C P N Q E S I K Y Y Y Y Chaine - CRISP2 Y N S Y N V C C C C A A Y C P N Q D N I K Y Y Y Mathematic - CRISP2 Y S S F R V C C C I A Y C P N Q D N	Atlantic Cod - CRISP3	N S	S -				- W	QV	/ G	CG	М	ΤК	C	3 -		-			SI	CΥ	F	ΥG	Exon 5
Japanes rice find. CREP3 N S	Stickleback - CRISP3	NS	S -				- Y	ΚV	/ G	CG	М	ΤL	C	ΡN		-			G	ΙY	F 1	ΥA	
Westen claved 6.02, Alknin AK K - - Y L G C N F C N Y V F Y Y Y Y Y F N Y F Y <t< td=""><td>Tilapa fish - CRISP3</td><td>YR</td><td>S -</td><td></td><td></td><td></td><td>- N</td><td>QI</td><td>ΙG</td><td>CA</td><td>M.</td><td>ΑY</td><td>C</td><td>ΡN</td><td>S A</td><td></td><td></td><td></td><td>YF</td><td>CΥ</td><td>F 1</td><td>ΥV</td><td></td></t<>	Tilapa fish - CRISP3	YR	S -				- N	QI	ΙG	CA	M.	ΑY	C	ΡN	S A				YF	CΥ	F 1	ΥV	
Westen claved 6.02, Alknin AK K - - Y L G C N F C N Y V F Y Y Y Y Y F N Y F Y <t< td=""><td>•</td><td>N S</td><td>S.</td><td></td><td></td><td></td><td>- Y</td><td>KΝ</td><td>/ G</td><td>CG</td><td>V</td><td>ΤL</td><td>C</td><td>S N</td><td></td><td>-</td><td></td><td></td><td>Ν</td><td>ΙY</td><td>L</td><td>Y G</td><td>Exon 6</td></t<>	•	N S	S.				- Y	KΝ	/ G	CG	V	ΤL	C	S N		-			Ν	ΙY	L	Y G	Exon 6
Aftican daved fog - Aturn AK is		A K	T.				- Y	LI	G			A F	C	PG	ΝY	-			Y I	ΡY	V	S I	
Barnes python - CRISP 1ke Y K S			S.		÷ .		- Y	LI	G										YI	P H	V		
Green and/s Cysteinserich verson P Y <	The second	-					- R	LI	G						S K				FB	CΥ	Y .	ΥV	Exon 7
Chinese cole added under - Urabar. Y T S Y Q V G G A I A V G L L S S E F N Y Y Y V Annakin - CRISP2 H N S Y K V A Y G P N S R F K Y F Y V Manakin - CRISP2 H N S Y K V A Y G P N S R F K Y F Y V Voodpolder - CRISP2 H N S Y L G G G V A Y G P R R F F K Y F Y V Phypa - CRISP2 Y N S Y L V G G A A N G P Y E K S F K Y F Y V Phypa - CRISP2 Y N S Y U V G G A A N G P Y E K S F K Y F Y V Phypa - CRISP2 Y S S Y U V G G G A A N G P Y E K S F K Y F Y V Phypa - CRISP2 Y S S F R V G G G I A Y G P N Q E S K Y Y Y V Arakdop - CRISP2 Y S S F R V G G G I A Y G P N Q E S		-	S.				- W	ΚV	/ G														
Fakon-CRISP2 Y Y S Y Y Y Y C C A S C S N S R Y K Y Y C C S S S N S R Y K Y Y K C C S N S R		-	S.				- Y	1 0	/ G							-			FN	V Y			
Imman CRISP2 H N S Y K V A C G N S K F K Y F Y Cystein Chikan - CRISP2 Y N S Y L L G C G A N C P K K S F K Y F Y Chikan - CRISP2 Y N S Y Q V G C A V A Y C P E R T F K Y F Y Aradope - CRISP2 F S S Y L V G C G V A Y C P N Q E Y L K Y Y Y Aradope - CRISP2 Y S S F R V G C G I A Y C P N Q E S L K Y Y Y Spem whale - CRISP2 Y S S F R V G C G I A Y C P N Q D S L K Y Y Y Aradope - CRISP2 Y S S F K V G C G I A Y C P N Q D S L K Y Y Y Mouse - CRISP2 Y S S F K V G C G I A Y C P N Q D S L K Y Y Y Human - CRISP2 Y S S F K V G C G I A Y C P N Q D S L K Y Y Y Human - CRISP2 Y S T Y O Y G C G I A Y C P N Q D S L K Y Y Y Human - CRISP2 Y S T <td< td=""><td></td><td>-</td><td>S</td><td></td><td></td><td></td><td>. Y</td><td>0</td><td>I G</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>-</td><td>YK</td><td>C Y</td><td>-</td><td></td><td>Conscived</td></td<>		-	S				. Y	0	I G									-	YK	C Y	-		Conscived
Woodpecker - CRISP2 H S I			S				v	K V	/ A										FK	c v	F	y v	cysteines
Chickan - CRISP2 Y N S			S				v	I I	G				H						Y	e v	F ·	v v	
Platypac-CRISP1-Bee S T S - - - Y Q I G G A A N G P K E K S - - F Q V I F I Arndallo - CRISP2 Y S S - - F R V G G G I A Y C P N Q E S - I K Y Y Y Mouse - CRISP2 Y S S - - F R V G G G I A Y C P N Q E S - I K Y Y Y Mouse - CRISP2 Y S S - - F K V G G G I A Y C P N Q D S - I K Y Y Y Rat-CRISP2 Y S S - - F K V G G G I A Y C P N Q D S - I K Y Y Y Rat-CRISP2 Y S S - - Y Q V G G G I A Y C P N Q D S - I K Y Y Y Hamano-CRISP2 Y S T - - Y Q V G G G I A Y C P N Q D S - - I K Y Y Y Tomato-PR 4 C N - - Y D P Y G N Y I G E N L A L G Y D Q P A A V A W D A W N N E I S Nod-CRISP-IAc - - - - - - - - - - - - - - - - - -			s				. v	0 1	G										FK	c v	F	y v	
Armidsio - CRISP2 F S S - - Y V G G G V Y V V Sperm whale - CRISP2 Y S S - - F K G G I Y P Q S - - L K Y Y Y Mouse - CRISP2 Y S - - F K G G I Y C N Q Y Y Y Manato CRISP2 Y S - - F K Y G G I Y P Q S - - L K Y Y Chara Y		-	s				v	õ	T G							S			FO	5 v	÷.	F T	
interpope - CRISP2 Y S S		-	8				v	T T								v			T R	ć v	v	v v	
Speminyhale - CRISP2 Y S			c .					DI					-		~ -	c			1 1		v ·	v v	
Monser-CRISP2 Y S Image Image C F K Image C P V P V V Ching-CRISP2 Y S Image F K G G I Y C P N Q D S Image L K Y Y V Ching-CRISP2 Y S Image Y Y V<			0.				- r												TT	~ 1 7 V	1 V ·	v v	
Part CRISP2 Y S <t< td=""><td></td><td></td><td>0</td><td></td><td></td><td></td><td>- r</td><td>V</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>TT</td><td>- 1 - V</td><td></td><td>v v</td><td></td></t<>			0				- r	V											TT	- 1 - V		v v	
			0				- r T	N I											L L T L	- 1 7 V	F.	1 V V V	
Human - CRISP2 Y S T - - Y Q V G G I A V P N Q D S - I K Y		-					- r v	0 1					-		~				TT	v v	r v	v v	
CAP2 CAP2 Tomato - PR P4 C N - Y D P V G R P </td <td></td> <td></td> <td>_</td> <td></td> <td></td> <td></td> <td>- 1 V</td> <td>Å,</td> <td></td> <td></td> <td></td> <td>A V</td> <td>E</td> <td>E IN</td> <td></td> <td></td> <td></td> <td></td> <td>TT</td> <td>~ 1 7 V</td> <td>1 V ·</td> <td>v v</td> <td></td>			_				- 1 V	Å,				A V	E	E IN					TT	~ 1 7 V	1 V ·	v v	
Tomato - PR P4 C N - Y D P V N V G E P Y -	Human - CRISF 2	1 3	1.		CAP	2	- 1	Q V				A 1	<u> </u>		U D	5			LI	1	1	1 V	-
Cocao Tree - CRISP-like C N - Y D P P G P I G E T -	Temata DP D4	CN					U C	NI N	U U	C F	D		D 1	7	_			_		-		_	
Yeast-PRY1 C S G T L F G P Y D C D A V D A V D A V D A V D A V D A V D A V D A V D A V D A V D A V D A V D A V D A V D A V D A V D A V D C N A Y D C N A Y D C N A Y D C D A D <tdd< td=""><td></td><td>C N</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>- F</td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tdd<>		C N											- F	-									
Moid - CRISP-like R I F G R I P K V V P R G V K P Q N I Y P R G N K R Q N I V R G T K R Q N I N S C T K C N C I I N			6									A T			C D	-	AV	D 4	- ·	 7 N	E I	TC	
Roundworn - SCP-like C Q Q Y K P Q N			U .	L	1 1	1 5 1		F 1		DV	L .				D C	A		UA	C		2	1 3	
Frait Fy- CG42780CNVAVTNVIOECAAGIECAAGIDNNIIOECAAGIECAAGIIONIDDRLGTFIAGNPDRIOIDDAAAPVAGNPDAAAPVAGNCIAAGNIDAAEPVAGNCIAAGNVIABCNNNNIAGNVINIAGNINNNIAGNVINNININNINNINNINNNNINNN </td <td></td> <td></td> <td>-</td> <td></td> <td>VE</td> <td></td> <td></td> <td>N X</td> <td>1.</td> <td>N O</td> <td></td> <td>v</td> <td>τ.</td> <td>1 F 7 T</td> <td>C C</td> <td>÷.</td> <td></td> <td>2 F</td> <td></td> <td></td> <td>- 11</td> <td></td> <td></td>			-		VE			N X	1.	N O		v	τ.	1 F 7 T	C C	÷.		2 F			- 11		
Parasyfic wasp - Scottriftins lake C N - Y C P A G N P D R G A G N P D R G N P D R G N K G P A G N K G L A P Y A G N K G N K G N K G N K G N K G N K G N K G N K G N V A G N V N N R U A C N G N N R N N R N		CN					2 9	IN I		TN Q	11	 	-		5 6				<u> </u>	r 5 	5	- 1 0 V	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		CN			1 /			27.3			. V		- E		A C					3 I 6 D			
Fire AntsCRISP1-likeCNNNNNNKGRLGNVAGRLGNCSMCKDCSMCKDHCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNNNNNNDNCNNN		CN																					
Dwarf honeybee - CRISP2-like C N - Y A P S G N Y K G R P Y A K C R C T M C S G N Y K G R Y K G R F C T M C Q P X A K G N Y N T Q P S G N Y N T R Q P Y A G N I N T N <t< td=""><td></td><td>C N</td><td></td><td></td><td></td><td></td><td>_</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>		C N					_																
Leafcutter be - CRISP2-tike C N - - Y A P S G N Y D R L Q P Y A G N C N C N C N C N C N C N C N C N C N C N C N C N C N C N C N C N C N N T N <th< td=""><td></td><td>CN</td><td>7</td><td></td><td>YA</td><td></td><td></td><td>NI</td><td></td><td></td><td></td><td></td><td></td><td>IV V</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></th<>		CN	7		YA			NI						IV V									
Sea Squirt (C. s.) - Unchar.CNINIAQPSRVGRPFERKGAPCSSea Squirt (C. i.) - CRISP3-ikeCNYGNKINNTRPVGRVGNKVSRCSCRNCNNTRNNNTRNNN </td <td></td> <td>C N</td> <td></td> <td></td> <td>I P</td> <td></td> <td></td> <td>IN 1</td> <td></td> <td></td> <td></td> <td></td> <td>r D</td> <td></td>		C N			I P			IN 1					r D										
Sea Squirt (C. i.) - CRISP3-like C N - - Y G P S G N Y L D W T G P X G N Y L D W T G P X G N Y L D W T N R Y G T K Y C P M G N Y V P Y N N C P N C D N C D N C D N C D N C D N N C D N N C D N N C D D N N C D N N C D N N C D N		CN			1 A			_	A I				P							U U	n		
Sea lamprey- Unchar.CHYCPMGNLNTRINRPYDLGTSCASCPNCVSpotted gar - Unchar.CHYCPLGNYQYYPYPYPYPYPYPYYRAGNYVPPPYXAGNYVGVVYNAGNYVGVVVGNNCEESickleback.CSCASCPNNCEESickleback.CRSCCASCPNNCEESickleback.CRSCCNNCEESickleback.CRSCCNNCESickleback.CRSCPNCEFNNCECASCPNCECASCPNCCDNNCESickleback.CRSCCASCPNC		CN			1 (P	5 K				ER				_	
Spotted gar - Unchar.CHYCPLGNQYEYPNAGDCPNACDZebrafish - CRISP3CHYYRAGNFRTVVP-PYXAGNFRTVP-PYXAGNNCEAdamic Cod - CRISP3CHYYRAGNFRTVPPYKKGNSCASCPNNCETilapa fish - CRISP3CHYYRAGNFRTVGNPYKKGNCASCPNACEJapanese rice fish - CRISP3CHYYRAGNFKKWP-PYKKGNCASCPNACEVWMCASCPNCCCNNCKKKKNNNNNNNNNN </td <td></td> <td>C N</td> <td></td> <td></td> <td></td> <td>_</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>-</td> <td>- K</td> <td></td> <td></td> <td></td> <td>- K</td> <td></td> <td></td> <td></td> <td>_</td> <td></td>		C N				_							-	- K				- K				_	
Zebrafish - CRISP3CHYYRAGNFRTVP-PYSLGSCASCPNCEAdamic Col - CRISP3CQYYRAGNFRTVP-PYKKGDSCPNNCEStickleback - CRISP3CHYRAGNFRTWP-PYKKGPSCASCPNCEJapanese rice fish - CRISP3CHYRAGNFKRWP-PYKKGPCASCPNCEJapanese rice fish - CRISP3CHYCPMGNINIKITYKKGPCASCPNCCESCDNCIIIYRRNNINIIINIININIININININNIN<		CH			1 C		_					NR	P		LG							_	
Adamic Cod - CRISP3 C Q - Y Y R A G N Y V G W P - P Y K E G N C E Stockleback - CRISP3 C H - Y Y R G N F R T W P - P Y K G P S C D S C D N C E Tilapa fish - CRISP3 C H - - Y R A G N F R W P - P Y K G N N N C P N C E N C D N C P Y Y R G N N N N N N C A S C P N C N C N C N N N N N N N N<		CH			1 L	_							P	I K.	AG							_	
Stickleback - CRISP3CHYYRAGNFRTWP-PYKVGPSCASCPNCETidpa fish - CRISP3CHYPPGNYFKRWP-PYKKGNYCNCEJapanese rice fish - CRISP3CHYRAGNFKRWP-PYKKGNDCNDCNNCHYPMGNNNNNNNCHYPMGNNN <t< td=""><td></td><td>CH</td><td></td><td></td><td>1 1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>_</td><td></td></t<>		CH			1 1																	_	
Tilapa fish - CRISP3CHYCPPGNQFTQ-PYKKGPSCGDCPNACEJapanese nice fish - CRISP3CHYRAGNKRWPYKAGNNCHPYKRGNNCNNINSIKTPYEASCPNDCVWestem clawed frog - AllurinCHYCPMGNMINSIKTPYEASCPNCESCDASCPNCNNINSIKTNN <td< td=""><td></td><td>C Q</td><td></td><td></td><td>1 1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>_</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>-</td><td></td></td<>		C Q			1 1								_									-	
Japanese nice fish - CRISP3CHYYRAGNFKRWP-PYKVGTPCASCPNDCVWestem clawed frog - AllurinCHYGPMGNMINSIKTPYEASCPESCESCPSCASCPESCDESVPNDINSININININNINNN<		CH			- E																		
Westem clawed fing - AllurinCHYCPMGNINSIKTPYEAGECPAGNINSIKTPYEAGECPECDASCPINSIKTPYEAGCPSCCDASCPINSIINSIINSIINSIINSIINSINSINSINSINSINSINSINSINSINSINSINSINSININNNINNNNINNNNINNNNNINNNNNINNN<																						_	
African clawed frog - AllurinCHYGPMGNMDESVPPYEGEGCASCPESCDBurmese python - CRISP-likeCHYCPAGNKIINIINIINIINIINIINNINNININNINNININNINNINN			1		_																	_	
Burmese python - CRISP-likeCHYCPAGNFINSIATPYKSGPCGDCPSSCAGreen anole - Cystemine-rich venomPCQYSPAGNKKLLNTPYAGPCGDCPKCDCPKCDCPKCDCPKCDCPKCDCPKCDCPKCDCPKCDCPKCDCPKCDCPKCDCPKCDCPKCDCPKCDCPKCDDACDDACDDACDACDDACDDACDDACDACDACDDACDACDACDACDACDACDACDACDACDAC				1																		_	
Green anole - Cysteins-rich venom P.CQ-YSPAGNRKDLNTPYAAGTPCGDCQPNCDChnese soft-shelled turde - Unchar.CHYCPAGNRKMNTPYKAGPACDCPKACDFalcon - CRISP2CQYCPMGNIRSSIPTPYKGPCQDACDMonder- CRISP2CHYCPAGNIISSIENVXKGPCADCPDACDModepeder- CRISP2CHYCPAGNIISSIETPYKGPCGDCPNACDACDACDACDACDACDCDACDACDCDACDACDAC				-																		_	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$			1																				
Falcon - CRISP2CQ-YCPMGNIRSSIPPYKEGPCQDACDManakin - CRISP2CQYCPMGNLLSSMTPYKEGPCQDACDWoodpecker - CRISP2CHYCPAGNILLSSIETPYKGNCDACEChicken - CRISP2CHYCPAGNIIGSIETPYKGQCDACDPlatyus - CRISP3-likeCQYCPAGNVNKKKGPCGDCPNCNPlatyus - CRISP2CHYCPAGNVNKKNTPYQGACDNCNLSSLNNNNNNNNNNNNNNNNNN<																							
Manakin - CRISP2CQ-YCPMGNLLSSMTPYKEGKPCQDACEWoodpecker - CRISP2CHYCPAGNILLSSIEKPYKAGNCPAGNLLSSIETPYKAGNCPAGNCPAGNDPTNN<				-																		_	
				-																		_	
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		CQ																				_	
Platypus - CRISP3-like C Q - Y C H A G N P E T N T P Y V G P C G D C N T P Y V G P C G D C N T P Y V G P C G D C N T P Y V G P C Q C P C N C N C N C N C N		CH																					
Armadillo - CRISP2 C H - Y C P A G N V N K K I P Y C A S C P N C E Armadillo - CRISP2 C Q - Y C P A G N V N K K I P Q G T P C A S C P N C A S C P N C D S C N C X S C N N S K N V S K N N N S K N N C A S C P N C D D N N N N N K K N N N N N N N				-																		_	
Antelope - CRISP2 C Q - Y C P A N V S K K N P Y Q Q A P D C D N C D S C D N C D N V S K K N P Y Q Q A P D C D D C D D C D D C D D D D D D D D D D D		CQ											P 1	ΥK	V G	P 1	P C						
Antelope - CRISP2 C Q - Y C P A G N V S K K N P Y Q Q G A C P D C D S S C N C S C D N C D N C D N C D C D C D C D C D C D C D C D C D C D C D C D C D C D C D C D N C D N C D N C D C D C D C D	Armadillo - CRISP2	CH			Y	P .	A G	NV	V V	NK	K	ΗI	P	Y Q	K G	TI	P C	A S				C E	
Sperm whale - CRISP2 C Q - Y C P A G N V S K K N P Y Q Q G T C D Mouse - CRISP2 C H - Y C P M G N V M K S T P Q Q T P C A S C P N C E Rat - CRISP2 C H - Y C P M G N V M K K S T P Q Q G T P A S C P N C E Rat - CRISP2 C H - Y C P M G N V M K K S T P Q G T P A S C P N C D D D D D	Antelope - CRISP2	CQ			Y	P.	A G	ΝN	v v	S K	K	ΝT						G S	CI	P D	N	C D	
Mouse - CRISP2 C H - Y C P M N V M K K S T P Q Q G T P N C A S C P N C E C I <thi< th=""> I I</thi<>					Y																	C D	
Rat-CRISP2 CHYCPMGNNVMKKSTPYHQGTPCASCPNNCD																						_	
		СН																				_	
		CQ			Y																		
Human-CRISP2 CQYCPAGNNMNRKNTPYQQGTPCAGCPDDC		CO			Y	P	A G	NN	M	NR	к	ΝT	P	rõ	Q G	T	PC						

Appendix 2:

Page 4 of 7

	_		_	_	_	_				_								1									-	
Tomato - PR P4		-	_	_	_	-	_	_	-	-	-		-		-	-		-		-		-	-		_	-	-	
Cocao Tree - CRISP-like	-	-	-	-	-	-	-		-	-				-		-		-					-	-			-	
Yeast - PRY1	Ν	Y	D	F	S	N	P (GF		s		TC		-		н	F 1		V V	/ w		S T	T	0	v'o	3 C	G	
Mold - CRISP-like	-	-	-	-	-	-		- -	-	-	-			-		-		-					-	-			-	
Roundworm - SCP-like	S	С	Е	А	A	T	G	LC	A	-	-			-		-		-					-	-			-	
Fruit Fly - CG42780	Y	Ρ	Ρ	L	С	A I	K	Ι-	-	-	-			-		-		-					-	-			-	Exon 1
Parasytic wasp - Serotriflin-like	V	-	K	L				S C		W	-			-		-		-		• A	D	LΨ	γA	Ν	CF	ις	2 L	
Jewel wasp - CRISP2-like	L						Ν.	A C	P	W	-							-		· A	D		γA	- H		2 Q		Exon 2
Fire Ants - CRISP1-like			Α				N.	A	Y	н	-			-						- T			V S	- H	_	A E		
Dwarf honey bee - CRISP2-like	R		K			TI		AC	H	Y	-			-						- V			V S	- H		/ E		
Leafcutter bee - CRISP2-like	R	- · ·	K			TI		AC	Y	Y				-						• T			V S				L	Exon 3
Sea Squirt (C. s.) - Unchar.	Q S		A	D		Q T		T						- N '										- T		-		
Sea Squirt (C.i.) - CRISP3-like Sea lamprey-Unchar.	N		1				NI				-		<u>_</u> A						1 4	- K	-		A		_		ΩT AM	
Spotted gar - Unchar.	D		-	÷			NI							2						. v.		ΟY		-	_			Exon 4
Zebrafish - CRISP3	D		2	ĩ			N			Ŷ				2						Ī		-	v	-			ĹĹ	
Atlantic Cod - CRISP3	D		2	L	_		NI		j.		2			-		1		-		. I	NI			- H-			L	Exon 5
Stickleback - CRISP3	D	к	-	L			NI		P	Y	-			-		-		_		·I		κу		- F			L	
Tilapa fish - CRISP3	Ν	Ν	-	L	С	ΤI	NI	PC	P	F	-			-		-		-		- S	DI	ĸΥ	N	Ν	C I	PE	L	
Japanese rice fish - CRISP3	D	ĸ	-	L	С	ΤI	NI	P	P	Y	-			-		-		-		·I	NI	R F	L	Ν	C I	P A	۱M	Exon 6
Western clawed frog - Allurin	D	ĸ	-	L			S 1		• т					-									-	-			-	
African clawed frog - Allurin	D		-	- F			WI		<u>K</u>		-			-		-		-				- -	-	-		-	-	Exen 7
Burmese python - CRISP-like	N		-				NI		_	H	-		-	-		-		-			D						L	Exon 7
Green anole - Cysteine-rich venom P.	N						NI		_	Y	-		-	-				-		. D							L	
Chinese soft-shelled turtle - Unchar.	N			L	_	-	NI		N	Y	-		-	2				-		• V	~	VF					L	C Conserved
Falcon - CRISP2 Manakin - CRISP2	N D		2				NI		K	Y	-			-						.к. .о		SF		-	_		L	C cysteines
Wanakin - CRISP2 Woodpecker - CRISP2	_	G					N I N I			Y	-			-						. Q			G					-0
Chicken - CRISP2	N			Ľ		T				Y				2						R			S	- F	_		M	
Platypus - CRISP3-like	-	R		- F			NI		P	F	2			2						. A			s	- H		LE	A	
Armadillo - CRISP2	N		2	- F	_		N		E	Ŷ	_			-						E		ΑL	S	- H-			L	
Antelope - CRISP2	N		2				N		E	Y	2			_		-		-		· E	D	LL	s				L	
Sperm whale - CRISP2	s	G	-	L		ΤI		s C	E	Y	-			-		-		-		E		LL	s	N	CI	o s	L	
Mouse - CRISP2	Ν	G	-	L	С	ΤI	N 3	s C	D	F	-			-		-		-		E	D	LL	S	Ν	CI	E S	L	
Rat - CRISP2	Ν						N		_	F	-			-				-			D			Ν			L	
Chimp - CRISP2	-	G		- F			N			Y	-			-				-		Q	D	LL			CI	D S	L	
Human - CRISP2	K	G	-	L	С	ΤI	N	s c	Q	Y	-			-		-		-		Q	D	LL	, S	Ν	CI	D S	L	
		_	_	_	_	_	_	_	_			_	_		_			_		_		_			_	_		
Tomato - PR P4	-	-	-	-	-	-			-	-	-		• •					-		• •			-	-			-	
Cocao Tree - CRISP-like Yeast - PRY1	-	ĸ	T	_		-	- ·	 V 6	- F D	- V		 T 7		- Y 1	- - P					 1 D	- Y	 <u>A</u> T				 D T	- . A	
Mold - CRISP-like		<u>к</u>	-	_			n (1			-				0 1		E		1.	<u>, 1</u>		v			. n	
Roundworm - SCP-like	-	-	-	-	-	-	-		-	-	-			-		-		-			-		-	-	_		-	
Fruit Fly - CG42780	-	-	-	-	-	-			-	-	-			-		-		-					-	-			-	
Parasytic wasp - Serotriflin-like	Y	Е	Т	W	P	N	W	LC	E	S	D	TQ	Q	G I	ΗE	R	RQ) F	CI	۲ A	T	CR	C	R	DK	ĊI	V	
Jewel wasp - CRISP2-like	R	А	Т	W	Ρ	G 1	W	LC	D	Т	D	ΤI	Q	G I	RЕ	R	RQ) F	CI	ŁΑ	Т	CR	C	D	GΚ	C I	I	
Fire Ants - CRISP1-like	V	R	Т	F	G	S 1	W	V C	E	Т	D	ΤI	ΕE	G 1	RΕ	R	RΒ	F	C	ЗA	Т	C N	C	K.	DK	C I	Y	
Dwarf honey bee - CRISP2-like	А	R	М	F	R	S 1	W	V C	E	Т	Ν	ΤS	δE	G 1	R E	R	RΚ	F	C	ЗA	Т	C N	C	E	GΚ	C I	Y	
Leafcutter bee - CRISP2-like	V	R	Т	F	R	A	W	V C	E	Т	Ν	ΤK	E	G I	RЕ	R	RK	F	C	ЗA	Т	CN	C	ĸ	GK		Y	
Sea Squirt (C. s.) - Unchar.	-	-	-	-	-	-		- -	-	-	-	- -	-	-	- -	-		-				- -	-	-	- -		-	
Sea Squirt (C. i.) - CRISP3-like						V	PI	F W	νт	Ν	v	PI	_	F			NN			εL					QN		. L	
Sea lamprey- Unchar.		A				-	-		-	-	-	- 19		N.			LK				S			Q			-	
Spotted gar - Unchar. Zebrafish - CRISP3	_	Q A		1 I	U T	-			-	-	-	-	S	N			QS	5 W . G		A PA		C N C L		TI			I	
Atlantic Cod - CRISP3		A K			G	-			-		-	Ē		N I		V				A				D 1 P 1			I	
Stickleback - CRISP3	_	T												NI			S S				S				DI		T	
Tilapa fish - CRISP3	_	Q												H						Â	_				NQ		I	
Japanese rice fish - CRISP3	_	Ā					-				-	- 7		N						A		C K		s				
Western clawed frog - Allurin								- -	-	-	-		_	-	_			_			_		_	-	_	_	-	
African dawed frog - Allurin									-					-									-				-	
Burmese python - CRISP-like	V	K	Ε	Н	K				-	-		- 🤇		I							S		C	Т	ΤI	ΕI	K	
Green anole - Cysteine-rich venom P.	к	Е	E	W	G	-	-		-	-	-	- (Ι							S						Q	
Chinese soft-shelled turtle - Unchar.	ĸ					-	-		-	-	-	- (H							S			Т			K	
Falcon - CRISP2	к					-	-		-	-	-	- (Ν							Т			Т				
Manakin - CRISP2	K					-	-		-	-	-	-		Н							Т			Т				
Woodpecker - CRISP2	K					-			-	-	-	-		Y						P A							I	
Chicken - CRISP2	_	K				-			-	-	-	- 19		H						. A		CR		L	_		Ι	
Platypus - CRISP3-like	_	K				-			-	-	-	- 19		H						-		-			_	-	_	
Armadillo - CRISP2	K K	A				-			-	-	-	- 19		H							T	_						
Antelope - CRISP2 Sperm whate CRISP2	-	K. K					-		-	-	-	- 12		H I						Δ	T						Y Y	
Sperm whale - CRISP2 Mouse - CRISP2	ĸ						-		-			1		H							T						H	
									-									A K			T						H	
	K	S	S																									
Rat - CRISP2 Chimp - CRISP2	K K						-			-	-	1																
Chimp - CRISP2 Human - CRISP2	к	Ν	Т	A	G	-	-		-	-	-		E	H H H	ΕL	L	ΚI	εĸ	C	C A	T T	C L	C	ΕI	NK	C I	Y Y	

Appendix 2: Page 5 of 7

												Г	-						1		_		1		Г						
Tomato - PR P4	-	-	-	-	-	-			-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Cocao Tree - CRISP-like	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Yeast - PRY1	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Mold - CRISP-like	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Roundworm - SCP-like	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	Exon 1
Fruit Fly - CG42780	-	-	-			-			-	-	-				-	-	-	-				-	-	-	-		-	-	-	-	L EXON 1
Parasytic wasp - Serotriflin-like	-	-	-	-	-	-		-	-	-	-	-				-	-	-					-	-	-		-	-	-	-	
Jewel wasp - CRISP2-like	-	-	-	-	-				-		-					-	-	-					-	-	-		-	-	-	-	Exon 2
Fire Ants - CRISP1-like					G											-		-						-	-		-	-	-	-	
Dwarf honey bee - CRISP2-like					-						-						-	-	-			-		-	-		-	-	-	-	
Leafcutter bee - CRISP2-like	_			_	-							-						-					-	-	-		-	-	-	-	Exon 3
Sea Squirt (C.s.) - Unchar.	_	_	-	_	_	_		_	-		-	_		_	_	-	-	-	_	_		_	-	-	-		-	-	-	-	
Sea Squirt (C.i.) - CRISP3-like	_	-	_	_	_	_	_	_	_		_	_	_	_	_	_		_	_	_	_	_	_	K.	_	LL	_	_	_	_	
Sea lamprey- Unchar.				-	-	-	_	_	-		-			_		-				_		_	-	-	_	_	-	-	-		Exon 4
Spotted gar - Unchar. Zebrafish - CRISP3	_	- T	- A	v	v	-			-	-	-				-	-	-	-				_	-	-				-		-	
Aflantic Cod - CRISP3					-			-								-			-	_		-			-			-	-	-	Exon 5
Stickleback - CRISP3		-							-		-					-	-						-	-	-		-	-	-	-	
Tilapa fish - CRISP3	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Japanese rice fish - CRISP3	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-					-	-	-	-		-	-		-	Exon 6
Western clawed frog - Allurin	-	-	-	-		-			-	-	-					-	-	-				_	-	-	-		-	-	÷		
African clawed frog - Allurin	-	-	-		-	-			-	-	-			-	-	-	-					-	-	-	-		-	-	-	-	
Burmese python - CRISP-like	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-	-	-	-			-	-	-	-		-	-	-	-	Exon 7
Green anole - Cysteine-rich venom P.	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-	- -	-	-	-	-	
Chinese soft-shelled turtle - Unchar.	-	-	-			-		_	-	-	-				-	-	-	-				-	-	-	-		-	-	-	-	Conserved
Falcon - CRISP2	-	-	-		-	-			-	-	-					-	-	-				_	-	-	-	- -	-	-	-	-	C
Manakin - CRISP2	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	cysteines
Woodpecker - CRISP2	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Chicken - CRISP2	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Platypus - CRISP3-like	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Armadillo - CRISP2	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Antelope - CRISP2	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Sperm whale - CRISP2	-	-	-		-	-		_	-	-	-			_		-	-	-					-	-	-		-	-	-	-	
Mouse - CRISP2	-	-	-			-		-	-	-	-			-		-	-	-		_		_	-	-	-		-	-	-	-	
Rat - CRISP2	-	-	-	-	-	-		-	-	-	-	-		-		-	-	-				_	-	-	-		-	-	-	-	
Chimp - CRISP2	-	-	-	-	-	-		-	-	-	-	-			-	-	-	-	-			-	-	-	-		-	-	-	-	
Human - CRISP2	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	-	-			-	-	-	-		-	-	-	-	
Tomato - PR P4				-	-		-	-	-				_	-	-	-			-	-	-	-			-	_	-	-			
Cocao Tree - CRISP-like	-	-	-		-	-		-	-	-	-	-			-				-	-	+	-			-	-	-	-			
Yeast - PRY1	-	-	-	-	-	-		-	-	-	-	-		-		-			-	-	-	-			-	-	-	-			
Mold - CRISP-like	-	-	-	-	-	-		-	-	-	-			-		-			-	-	-	-			-	-	-	-			
Roundworm - SCP-like	-		-		-	-		_	-	-	-	-		_	-	-			-	-	-	-			-	-	-	-			
Fruit Fly - CG42780	-	-	-	-	-	-			-	-	-	-		-	-	-				-	-	-			-		-				
Parasytic wasp - Serotriflin-like	-	-	-	-	-	-		-	-	-	-	-		-	-	-											-				
Jewel wasp - CRISP2-like	-	-	-	-	-	-		-	-	-	-	-		-	-	-															
Fire Ants - CRISP1-like	-	-	-	-	-	-		-	-	-	-	-		-	-	-															
Dwarf honey bee - CRISP2-like	-	-	-	-	-	-		-	-	-	-	-		-	-	-															
Leafcutter bee - CRISP2-like	-	-	-	-	-	-		-	-	-	-	-		-	-	-															
Sea Squirt (C.s.) - Unchar.	-	-	-	_	_	_		_	-		-	_		_	_	_															
Sea Squirt (C.i.) - CRISP 3-like	Q	Ν	R	Q	Q	N	RΕ	° R					R R	L	Q	G			_												
Sea lamprey- Unchar.	-	-		-	-	-		-	-	-		-		-	-	-				_	_	_					_				
Spotted gar - Unchar.	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-				_	-	_			_	_	-				
Zebrafish - CRISP3	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-					_	-			_		-	-			
Atlantic Cod - CRISP3 Stickleback - CRISP3		-							-	-	-					-			\rightarrow	-	+	-			-	-	+	-			
					-				-		-		_		-				-		-			-	-		-	-			
Tilapa fish - CRISP3 Japanese nice fish - CRISP3		-						-		-						-				-	-	-			+		+	-			
Western clawed frog - Allurin																			\rightarrow	-	+	-			+	+	+	-			
African clawed frog - Allurin																				-	+	-			+		+	-			
Burmese python - CRISP-like																			-	-	-	-			-	-	-	-			
	1.1																		\neg	+	+	-			+	+	+	-			
Green anole - Cysteine-rich venom P.	-																			+	+	-				-	1				
Green anole - Cysteine-rich venom P. Chinese soft-shelled turtle - Unchar.		-	- 1																		+						1				
	-	-		-	- 1																										
Chinese soft-shelled turtle - Unchar.	-	-	-			-	- -		-																-		-				
Chinese soft-shelled turtle - Unchar. Falcon - CRISP2	- - -	- -	-	-	-					-	-	-		-	-	-															
Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2	- - - -	- - -	- - -		-	-		-	-	-	-	-		-	-	-															
Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2 Woodpecker - CRISP2	- - - -	- - - -	- - -	- - -	- - -	- -	 	-	-	-	-	-		-	-	-					-										
Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2 Woodpecker - CRISP2 Chicken - CRISP2 Platypus - CRISP3-like Armadillo - CRISP2	- - - - -	- - - -	- - - - -	- - - -	- - - -	- - -	 		- - -	-	- -	-	 	-	-	-															
Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2 Woodpecker - CRISP2 Chicken - CRISP2 Platypus - CRISP3-like	- - - - - -	- - - - -	- - - - -	- - - - -	- - - -	- - - -	 	 	- - -	- - -	- - -	- - -	 	-	-	- - -															
Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2 Woodpecker - CRISP2 Chicken - CRISP2 Platypus - CRISP3-like Armadillo - CRISP2 Antelope - CRISP2 Spenn whale - CRISP2	- - - - - - - - - -	- - - - -	- - - - - - - -	- - - - -	- - - - -	- - - -		 	- - - -	- - -	- - - -	- - - -	 	-	- - -	- - -															
Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2 Woodpecker - CRISP2 Chicken - CRISP2 Platypus - CRISP3-like Armadilo - CRISP2 Antelope - CRISP2 Sperm whale - CRISP2 Mouse - CRISP2	- - - - - - - - - - - - -	- - - - - -	- - - - - - - -	- - - - - -	- - - - - -	- - - - -		· - · - · -	- - - - -	- - - -	- - - -	- - - -	 	-	- - - -	- - - -															
Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2 Woodpecker - CRISP2 Chicken - CRISP2 Platypus - CRISP2 Platypus - CRISP2 Antelope - CRISP2 Sperm whale - CRISP2 Mouse - CRISP2 Mate - CRISP2 Rat - CRISP2	- - - - - - - - - - - - - - - - -	- - - - - - -	- - - - - - -	- - - - - - -	- - - - - - -	- - - - - -		· - · - · - · - · -	- - - - - -	- - - -	- - - - -	- - - - - -	 		- - - - -	- - - -															
Chinese soft-shelled turtle - Unchar. Falcon - CRISP2 Manakin - CRISP2 Woodpecker - CRISP2 Chicken - CRISP2 Platypus - CRISP3-like Armadilo - CRISP2 Antelope - CRISP2 Sperm whale - CRISP2 Mouse - CRISP2	- - - - - - - - -	- - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - - - - -	- - - - - - -		· - · - · - · - · - · -	- - - - - - -	- - - -	- - - - -	- - - - - - -	 	-	- - - -	- - - -															

Appendix 2: Page 6 of 7

Appendix 2: This figure provides MEGA7 ClustalW amino acid sequence alignment for various CAP superfamily proteins. The CAP signature sequences are boxed, the highly conserved cysteine are highlighted in yellow with a red font color. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online data for specific CAP genes. Color-coding superimposed on the sequences is used to represent different exons and exon borders are indicated by a change in color. Organisms and accession IDs of the amino acids are available in the supplemental data table 15.

Page 7 of 7

Exons f	UI.	2	UI I	···				- ·			~												4				, 						
Common name - Protein name	_								_				- 1	٩n	nino	A	C10	I Se	equ	ıe n	ice			_									_
		-	-	_	-		~	-	-			~	-	-	-	-	~	-			-			-		~	_	_					
Tomato - PR P4																														Q 1			
Cocao tree - CRISP-like																														K 1			
Grape vine - PR-1																												_		Q 1			
Yeast - PRY1																														ΤV			
Mold - CRISP-like																														GΝ			
Roundworm - SCP-like	-	-	-	M	H	Т	S	Ι	L	L	L	Т	A	L	С	A	G	A	Y	A	Q	F	S	Α	Ν	G	Q	A	A	ΙI	l N	N V	H
Fruit Fly - CG42780	M																													K (
Parasytic wasp - Serotriffin-like	-	-																												S F	C F	P N	1 P
Jewelwasp - CRISP2-like	-	-																										-				- N	
Fire Ants - CRISP1-like																														S F			-
Dwarf honey bee - CRISP2-like																														S F			
Leafcutter bee - CRISP2-like	-	-	-	-	-	M	R	W	L	W	Ι	Ι	V	L	Ι	A	L	A	L	Ρ	A	A	V	Ρ	R	Κ	Т	Н	R	NF	C F	P T	P
Sea Squirt (C. s.) - Unchar.	-	-	-	М	[A	F	Ν	F	Y	F	Т	Ι	L	L	L	S	V	S	Ι	А	Q	A	F	G	Ν	D	Y	Т	S	I I	N A	A E	E
Sea Squirt (C. i.) - CRISP3-like	-	М	ΙL	A	K	Μ	Ν	F	Q	F	G	L	Κ	V	L	F	F	Ι	F	A	L	S	V	Ι	S	D	A	Ν	Ι	LÌ	N E	ΞL	, K
Human - CRISP2	-	-	-	-	-	M	A	L	L	Ρ	V	L	F	L	V	Т	V	L	L	Ρ	S	-	-	-	-	-	-	-	-			- L	, P
																																	Τ
Tomato - PR P4	D	Y	L	A	V	Н	Ν	D	A	R	A	Q	V	G	V	G	-	-	-	-	-	-	-	-	-	-	-	-	-				-
Cocao tree - CRISP-like	Е	F	L	D	A	Q	Ν	Е	А	R	A	Е	V	Ν	V	Е	-	-	-	-	-	-	-	-	-	-	-	-	-				
Grape vine - PR-1																		_	-	-	-	_	_	_	-	-	-	_	-				
Yeast - PRY1																														A S			
Mold - CRISP-like																														<mark>G</mark> ٦			
Roundworm - SCP-like																																	
Fruit Fly - CG42780																														КÌ		A T	I
Parasytic wasp - Serotriffin-like											v																						
Jewelwasp - CRISP2-like											v				-																		
Fire Ants - CRISP1-like	_																																
Dwarf honey bee - CRISP2-like																												2					I S
Leafcutter bee - CRISP2-like											L																						G
Sea Squirt (C. s.) - Unchar.																												2	Ξ.		7 т	- З Т	- 5 - E
Sea Squirt (C. i.) - CRISP3-like																												-		r v	, L	ст Ст	. E
Human - CRISP2											A																		7	1 4	ч I		
Human - CRISP2	-	A	<u>.</u>	G	r	U	P	A	r	1			P3		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-			- 4) L
Taurata DD D4	_										-															D		0					L.
Tomato - PR P4		-	-	-	-	-	-	-	-		-						2													D A			
Cocao tree - CRISP-like		-	-	-	-	-					2						2								-					D A		<u> </u>	
Grape vine - PR-1		_	-	-	_	_																			-								
Yeast - PRY1																														S I			
Mold - CRISP-like	S	G	E	A	R	С	С	С	L	G	М	K	Η	R	S															DI			
Roundworm - SCP-like	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-															D A			
Fruit Fly - CG42780																														Νŀ			
Parasytic wasp - Serotriffin-like																														H S			
Jewelwasp - CRISP2-like																														H A			
Fire Ants - CRISP1-like	K	. V	R	Q	Κ	Ι	V	D	Т	Η	Ν	Y	F	R	Т	Q	V	Κ	Ρ	Т	А	A	Ν	Μ	L	Ι	Μ	K	W	нs	s c	βL	. A
Dwarfhoney bee - CRISP2-like	K	. V	R	Q	K	Ι	V	D	Т	Η	Ν	Y	L	R	Т	Q	V	K	Ρ	Ρ	А	А	Ν	Μ	L	А	Μ	K	W	ΗI	? C	βI	, A
Leafcutter bee - CRISP2-like	K	. V	R	Q	K	Ι	V	D	Т	Η	Ν	Y	F	R	Т	Q	V	K	Ρ	Ρ	А	Α	Ν	Μ	L	V	Μ	K	W	ΗI	? C	βI	, A
Sea Squirt (C. s.) - Unchar.	Α	Е	Κ	Ι	Е	Ι	V	R	L	Н	Ν	Q	Y	R	S	R	A	Ν	-	-	А	S	D	Μ	L	А	Μ	S	W	DÌ	ΤV	ΓI	A
Sea Squirt (C. i.) - CRISP3-like	_																													וס			
Human - CRISP2																														S I			
																												-					_

Appendix 3: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Invertebrate CAP Protein Sequences

Appendix 3: Page 1 of 4

				1				1						7							1					0	AP	4
Tomato - PR P4	S	R	A	Q 1	I Y	A	N	S	R	A	-	-			- (G I		N	L	I	H S	s G	} -	-	A			LA
Cocao tree - CRISP-like	A	Υ	A	QE	ΞY	A	N	Q	R	Ι	_	_			- 1	ΑI		D	L	v	н	s e	G G	Р	Υ	G	E N	IIA
Grape vine - PR-1																												IIA
Yeast - PRY1																												LA
Mold - CRISP-like															- 1													
Roundworm - SCP-like				Q A																								LY
Fruit Fly - CG42780	_			ΙI													_	_										LF
Parasytic wasp - Serotriflin-like															N													
Jewelwasp - CRISP2-like															N									s			O N	IIF
Fire Ants - CRISP1-like															N												O N	IIF
Dwarf honey bee - CRISP2-like				Q F											N									ò		G	o N	IIF
Leafcutter bee - CRISP2-like				O F					_	•					N								G	0	s	G	O N	TT
Sea Squirt (C. s.) - Unchar.															N								' s	ī	Ť	G	τ Ε Ν	ILC
Sea Squirt (C. i) - CRISP3-like	_			к. А																				v	Ť.			ILA
Human - CRISP2				QE						1																		IL Y
11011211 - CICIOT 2	-								~														1	R	-	0		
Tomato - PR P4	ĸ	G	G	GT) F	т	G	R							_						Δ	v		T.	w	v	S F	RF
Cocao tree - CRISP-like		G																					~	_			_	KV
Grape vine - PR-1	W	~		GS											2													w T
Yeast - PRY1	T	_																										IS
Mold - CRISP-like	_																											HK
Roundworm - SCP-like	_																											FQ
	_																											
Fruit Fly - CG42780																												E K H E
Parasytic wasp - Serotriflin-like	- 1	5	1 . 	A C																								EF
Jewelwasp - CRISP2-like	1	8		А I ~ т	εv																							
Fire Ants - CRISP1-like	- 1	s	T	GI	(1		_	-																				YK
Dwarfhoney bee - CRISP2-like	-	T	Т. —	K I	(1										-													YK
Leafcutter bee - CRISP2-like	1	Т	Т	GC	21	Ľ	W								-			-	-	-	FΑ	A I	R	M	W	F 1	ΜE	YK
Sea Squirt (C. s.) - Unchar.	Т	Q	Ι	ΤS	5 S	Ν	Ι					Y																IN
Sea Squirt (C. i) - CRISP3-like	M	A		A I				S																				Τŀ
Human - CRISP2	М	S	S I	DI	р Т	S	W	/ S	-	-	-	-			-			-	-				Q	S	W	YI	DE	II
																					AP1							
Tomato - PR P4																												G
Cocao tree - CRISP-like	Y	Υ	DI	нс	3 S	Ν	T	С	A	S	G	Q	V -		-	C	A E	ΙY	Т	Q	V V	/ W	/ R	Ν	S	V	R L	G
Grape vine - PR-1																												G
Yeast - PRY1	Ν	Υ	D	F S	N	ΙP	G	F	S	S	Ν	Т			-					_			VΚ	S	Т	Т	Q V	G C
Mold - CRISP-like	D	Υ	N	ΥI) S	N	T	С	G	Ρ	Ν	A	Ι-		- (C	βE	ΙY	Т	Q	V S	š -	D	Ι	Η	Η	ΚI	H
Roundworm - SCP-like	D	Υ	G٦	W S	S	Ν	Т	L	S	М	S	LI	FΝ	I T	G	I	ΞE	ΙA	Т	Q	ΜA	N	/ A	Κ	Т	Ν	LΙ	G
Fruit Fly - CG42780												NI			V	_					LI						A V	
Parasytic wasp - Serotriflin-like	Ι	F	K .	ΥC	3 P	• н	A	K	Ν	D	L	S (Q -		- 1	v	3 H	ΙY	Т	Q	M١	7 W	Λ	Ρ	Т	н	r v	G G
Jewelwasp - CRISP2-like	\mathbf{L}	F	Τ	ΥC	÷ -	S	Ρ	D	Ν	Е	L	G]	R -		- 1	v	βE	ΙY	Т	Q	M١	7 W	/ A	Т	S	н	LV	G G
Fire Ants - CRISP1-like	D	F	K '	ΥC) P	N	Т	Т	Ν	Е	Ι	LI	Ε-		-	I	3 H	ΙY	Т	Q	V V	7 W	/ A	Т	Т	н	L V	G 🕻
Dwarf honey bee - CRISP2-like	D	Y	K '	ΥC	θE	D	Е	G	Ν	D	L	ΗI	Ε-		-	I	G H	ΙY	Т	Q	IA	ИИ	/ A	Т	Т	н	LV	G 🕻
Leafcutter bee - CRISP2-like	D	F	K '	ΥC	θE	N	K	A	Ν	D	L	ΗI	Ε-		-	I	G H	ΙY	Т	Q	IA	И	/ A	S	Т	ΗI	мv	G
Sea Squirt (C. s.) - Unchar.																												G
Sea Squirt (C. i) - CRISP3-like	D	F	Т	Y S	т	Q	Ι	С	к	Ρ	G	R	Q -		- (C	3 H	ΙY	Т	Q	v v	7 W	/ A	D	S	ΗI	кv	G
Human - CRISP2																												G
Exon 1 Exon 2			E	xon	3			ł	Exe	on 4	4			Ex	on :	5			Ex	on	6			E	xoi	1 7		

Appendix 3: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Invertebrate CAP Protein Sequences

Appendix 3: Page 2 of 4

																	Γ					CA	P2	2								Γ
Tomato - PR P4																																R
Cocao tree - CRISP-like	A	Κ	V	R	С	D	Ν	G	G	-	-	-	-	-	-	-	Т	FI	1	C	N	Y	D	Ρ	Ρ	G	Ν	Y	Ι	G	Е	Т
Grape vine - PR-1	A	R	V	Q	С	Ν	Ν	G	G	-	-	-	-	-	-	- 1	W	FΝ	7 1	C	Ν	Y	D	Ρ	Ρ	G	Ν	Y	V	G	Q	R
Yeast - PRY1	G	I	Κ	Т	С	G	G	A	W	G	-	-	-	-	-	- 1	D	γV	7 1	C	s	Y	D	Р	A	G	Ν	Y	Е	G	Е	Ϋ́
Mold - CRISP-like	М	[A	L	Е	v	A	Y	Е	S	_	_	_	_	_	-	-	-															S 1
Roundworm - SCP-like																		τv	V V	7 <mark>C</mark>	0	Y	Κ	Ρ	0	G	Ν	Y	L	N	0	N
Fruit Fly - CG42780																																v ·
Parasytic wasp - Serotriffin-like																																L
Jewel wasp - CRISP2-like																		sγ														L
Fire Ants - CRISP1-like	-			н		•												МΥ			+											L
Dwarf honey bee - CRISP2-like	_																- I.	ΜY			4											L
Leafcutter bee - CRISP2-like	-			н																												L
Sea Squirt (C. s.) - Unchar.				F																	-							_				T
Sea Squirt (C. i.) - CRISP3-like																																N '
Human - CRISP2																																K I
TIGHTER - CICIOF 2		1	A	1	-	r	14	Y		5		T.					-	1			2	1		r	А	0	11	14	141	14	K	A
Tomato - PR P4	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-
Cocao tree - CRISP-like	Y	-	-	-								-					-			-			-	-	-	-	-	-	-	-	-	-
Grape vine - PR-1	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-
Yeast - PRY1	D	Ν	V									-												-	-	-	-	-	-	-	-	-
Mold - CRISP-like	W	L	Y	Ρ	R	G	S	V	Κ	V	A	С	Ι	L	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-
Roundworm - SCP-like	Y	Т	s	G								S			s			A A					A	-	-	-	-	-	-	-	-	-
Fruit Fly - CG42780	E		_															ΥI			_		ĸ			-		-	-	-	-	-
Parasytic wasp - Serotriffin-like																							Ν	S	С	Ρ	W	Α	D	L	w	A
Jewelwasp - CRISP2-like																																A
Fire Ants - CRISP1-like	_										- H	_			- H	_						-										S I
Dwarfhoney bee - CRISP2-like	_										- H				- H							-										S I
Leafcutter bee - CRISP2-like								_								_																S I
Sea Squirt (C. s.) - Unchar.																		Q														
Sea Squirt (C. i.) - CRISP3-like																		s c														T
Human - CRISP2																																S I
	1		1	Y	Y		1	T	•	^		<u> </u>	r			<u> </u>						1	11	5	C	Y	1	Y		-		
Tomato - PR P4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		•	-	-	-	-	-	-	-	-	-	-	-	-	-
Cocao tree - CRISP-like	-	-	-	-	-	-						-					-		• •	-	-	-	-	-	-	-	-	-	-	-	-	-
Grape vine - PR-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		•	-	-	-	-	-	-	-	-	-	-	-	-	-
Yeast - PRY1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		• •	-	-	-	-	-	-	-	-	-	-	-	-	-
Mold - CRISP-like	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-
Roundworm - SCP-like	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-
Fruit Fly - CG42780	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-
Parasytic wasp - Serotriffin-like	C	R	Q	L	Y	Е	Т	W	Ρ	Ν	W	L	C	E	S	D	Т	QQ	δ	H	E	R	R	Q	F	С	R	A	Т	C	R	C
Jewelwasp - CRISP2-like	С																															C]
Fire Ants - CRISP1-like	С																															C
Dwarfhoney bee - CRISP2-like	С	-								s		- H						S E														
Leafcutter bee - CRISP2-like	C	-								A								ΚĒ									1		Т		Ν	C
Sea Squirt (C. s.) - Unchar.	-	-	-	-	-	-	-	-	-		-	-	_	-			-			-								-		-	-	-
Sea Squirt (C. i.) - CRISP3-like	Y	V	Р	F	W	_	Ν	_		_	_	F	_	I	_	_	_		-		-	-	-	-		_				N	0	V
Human - CRISP2	C	Ď					T						_	F					-	K							1	Ā		C		C

Appendix 3: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Invertebrate CAP Protein Sequences

Appendix 3: Page 3 of 4

	_	_																					-									-	_
Tomato - PR P4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cocao tree - CRISP-like	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Grape vine - PR-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Yeast - PRY1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mold - CRISP-like	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Roundworm - SCP-like	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fruit Fly - CG42780	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Parasytic wasp - Serotriflin-like	D	K	Ι	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Jewelwasp - CRISP2-like	G	K	I	Ι	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fire Ants - CRISP1-like	D	K	I	Y	Y	Η	Η	Е	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dwarf honey bee - CRISP2-like	G	łΚ	I	Y	Y	Н	н	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Leafcutter bee - CRISP2-like	G	γK	I	Y	Y	Н	Н	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sea Squirt (C. s.) - Unchar.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sea Squirt (C. i.) - CRISP3-like	N	ľL	N	G	Q	N	W	Ρ	Μ	L	Ρ	N	Ρ	Y	L	Р	Ν	S	A	A	М	Р	ĸ	w	L	L	R	A	Ι	Q	Q	N	R
Human - CRISP2	N	ΙK	I	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Appendix 3: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Invertebrate CAP Protein Sequences

Appendix 3: This figure provides MEGA7 ClustalW amino acid sequence alignment of invertebrate CRISP2-like proteins. The CAP signature sequences are boxed, the highly conserved cysteine are highlighted in yellow with a red font color. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online data for specific CAP genes. Color-coding superimposed on the sequences is used to represent different exons and exon borders are indicated by a change in color. Organisms and accession IDs of the amino acids are available in the supplemental data table 15.

Page 4 of 4

Common name - Protein name														4 n	nino	A	cid	IS	equ	ien	ice												
	-																		- 1-														
W. clawed frog - serotriflin-like	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - CRISP2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	_	-	_	-	-	-	-	-	-	-	-	-		-	-	-	-
W. clawed frog - CRISP1		-	-	-	-	-	-	-	-	-	-	-		-	-	-	-		-	-	-	-	-	-	-	-	-	-	-			_	_
W. clawed frog - serotriffin-like	-	-	-	-	-	-	-	-	-		-				-	-		-		-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - MGC 108118	-	-	-	-	-	-	-	-	-		-	-		-	-	-		-		-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - CRISP3	-	-	-	-	-	-	-	-	-		-	-		-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - allurin	-	-	-	-	-	-	-	-	-		-	-		-		-		-		-	-			-		-	-	-	-	-	-	-	-
W. clawed frog - TEL1	-	-	-	-	-	-		-	-	-		-	-		-		-				-	-		_	-	-	-	-		-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-				-	-		-			-			-	-	-	-	-	-	-	-	-	-
W. clawed frog - PI15	-	-	-	-	-		-	-	-		-								-		-	-		-		-		-	-	-	-	-	-
W. clawed frog - PI R3HDML	-	-	-	-	-	-	-	-	-	-																				-	-	-	-
W. clawed frog - CRISP LCCL1	-	-	-	-	-	-	-	-	-		-	-		-	-	-		-		-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - CRISP LCCL2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - GliPR1-like	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				A		K
W. clawed frog - GliPR2	M	I R	D	A	S	G	A	E	V	V	Е	Т	W	Y	Ν	Е	Ι	К	D	Y	S	F	G	R	Ρ	G	F	Q	S	D	Т	G	Η
	_	_						_	_																								
W. clawed frog - serotriflin-like	-	-	-	-	-	-	-	-	-		-																						
W. clawed frog - CRISP2	-	-	-	-	-	-	-	-	-	-	-		-																		Е		
V. clawed frog - CRISP1	-	-	-	-	-	-	-	-	-	-			_																	-	S		
V. clawed frog - serotriflin-like	P	R	A	Т					R		Т																				V		
W. clawed frog - MGC108118	-	-	-	-		-			-		-																						
V. clawed frog - CRISP3	-	-	-	-	-	-	-	-	-		-																						
V. clawed frog - allurin	-	-	-	-	-	-	-	-	-		-																						
W. clawed frog - TEL1	-	-	-	-	-	-					-																						
V. clawed frog - PI15	-	-	M	Ι	Е	Μ	V	S	Ι	S	A	A	F	L	L	S	L	L	С	Е	Т	С	G	L	V	L	Ρ	K	S	S	D	L	A
W. clawed frog - PI R3HDML	М	Γ	L	Ι	Н	L	Ρ	Ι	F	F	S	G	V	F	L	W	Ι	Т	Ρ	L	L	S	Α	F	L	L	D	R	A	Т	Е	L	L
W. clawed frog - CRISP LCCL1	М	[Q	V	A	A	Q	Ν	W	L	R	Α	Т	V	L	L	F	Ι	A	Q	S	V	V	Т	L	V	Ι	Ρ	Ν	A	Т	Q	L	Е
W. clawed frog - CRISP LCCL2	М	[S	S	А	Μ	N	W	Ί	L	S	L	G	L	F	F	L	L	K	Е	Q	С	Y	С	I	F	A	Ρ	Ν	S	Т	F	L	Е
W. clawed frog - GliPR1-like	Т	L	S	L	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	L	F	L	L	s	Т	Е	F	s	Т	С	s
W. clawed frog - GliPR2	Т	Q	V	V	W	K	D	S	R	Е	V	G	V	A	Κ	A	V	D	G	K	G	м	V	I	A	V	A	Q	Y	S	Ρ	A	G
W. clawed frog - serotriflin-like	-	-	`-	-	-	-	-	` _	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - CRISP2	-	-	-	-	-	-	_	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - CRISP1	-	-	_	_	-	_	_	_	_	_	2	_	2	_	2	_	2	_	2	_	2	_	2	_	2	_	-	_	2	_	2	-	-
W. clawed frog - serotriflin-like		_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	-	_	_	_	_	_	-	_	_		
W. clawed frog - MGC108118	Е	Е	L	L	Т	Т	Ν	Η	Р	L	А	R	Н	0	V	s	Т	F	Р	0	Ν	Ρ	Ν	K	R	Е	I	А	w	I	Р	Т	A
W. clawed frog - CRISP3		_	_	_	_	_	_	_	_	-	_	_	_	2	÷.	_	-	_	-	2	_	_	_	_	_	-	_	-	-	_	-	-	-
W. clawed frog - allurin		2	_	_	_	_	_	_	_	_	_	_	2	_	2	_	2	_	2	_	2	_	_	_	2	_	2	_	2	_	2	_	_
W. clawed frog - TEL1		2	2	2	2		2		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	_	2	2	2	2	2	2	2	_	_
W. clawed frog - PI15	А	А	S	N	Y	т	T	T	к	_	-		_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	-
W. clawed frog - PI R3HDML		s				2	2	2	-	_	_	_	_	_	_	_	_	_	_	_	2	_	2	_	_	_	_	_	2	_	_	_	_
W. clawed frog - CRISP LCCL1		L				_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
W. clawed frog - CRISP LCCL2	-	L			_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_		
W. clawed frog - GliPR1-like											2																						
W. clawed frog - GliPR2											-																						
				-			-	×																									
Exon 1 Exon 2]	Ex	on	3			1	Ex	on	4			1	Ex	on	5				Ex	on	6				E	x01	n 7			
Exon 8 Exon 9			E	xo	n 1	10			E	Cxo	n 1	11			F	Cxo	on :	12			F	Exc	on :	13				Ex	on	14	1		
Appendix 4:																																	

Page 1 of 7

W. clawed frog - serotriflin-like	
W. clawed frog - CRISP2 W. clawed frog - CRISP1	
W. clawed frog - serotriflin-like	
W. clawed frog - MGC108118	P F S D Y G N D T T Q H P A G G S R R K R A G G R K I T I P P E N I
W. clawed frog - CRISP3	
W. clawed frog - allurin	
W. clawed frog - TEL1	PDLSARL
W. clawed frog - PI15	
W. clawed frog - PI R3HDML	R N Q T E H M
W. clawed frog - CRISP LCCL1	
W. clawed frog - CRISP LCCL2	
W. clawed frog - GliPR1-like	
W. clawed frog - GliPR2	N V L P K G T P V
W. clawed frog - serotriflin-like	D P P F S S I S T D N S T V R Q I I I D
W. clawed frog - CRISP2	DPPFSSISTD NVTVTOIIIN
W. clawed frog - CRISP1	T L P - S S L W T T T S T V Q V I V D
W. clawed frog - serotriflin-like	I V P Y E T O S T D N A T N R O I I V D
W. clawed frog - MGC108118	EKMKNVPFSALSTD LESNRQSILN
W. clawed frog - CRISP3	V F P F S S L S T R Y A T N R Q K I V D
W. clawed frog - allurin	G T P M L D T E T O N Y L V D
W. clawed frog - TEL1	VESFDEMSTD LESNRNFIVD
W. clawed frog - PI15	DAAKVPKARRKRYI SQNDMIAIVE
W. clawed frog - PI R3HDML	F G S G I P R I R R K R Y I S P R D M S A L L D
W. clawed frog - CRISP LCCL1	GEWWTAKHRGKRAITESDMKLILD
W. clawed frog - CRISP LCCL2	T P H S R T R R A I L R T D K E E I I Q
W. clawed frog - GliPR1-like	A F P S R R G A D E N L F Q
W. clawed frog - GliPR2	S N T G T S P T A R G T S Y L G T R G T D S A L S P T A D S R E F A
	CAP3
W. clawed frog - serotriflin-like	T H N A Y R R N A S P S A R N M L K M V W N E D A A N N A A S W S A
W. clawed frog - CRISP2	A H N N Y R R N A S P S A R NML K M V W N K D A A I N A A S WA A
W. clawed frog - CRISP1	T H N G Y R R S V N P S A R NML K MMW S E A A A S N A A T W S A
W. clawed frog - serotriflin-like	V H N R W R G N V T P T A M N M L K M E W N D E A A K K A E I W A R
W. clawed frog - MGC108118	V H N E L R R N A N P P P S N M L K M V W S D L A A K S A A K W A N
W. clawed frog - CRISP3	I HNAYRRSANPTASNMLKM <mark>SWSIEAENNAKNWAT</mark>
W. clawed frog - allurin	L H N L L R R S V D P T A K D M L K M E W S P G A A L N A Q N A A A
W. clawed frog - TEL1	K H N Y Y R S W V N P P A A D M L K M H W D N Y Y L A K A K E W A L
W. clawed frog - PI15	Y H N Q V R G K V F P P A A N M E Y M V W D E N L A K L A E A W A A
W. clawed frog - PI R3HDML	Y H N Q V R S K V F P P A A N M E Y M V W D E R L A K S A E S W A N
W. clawed frog - CRISP LCCL1	L H N K L R G E V Y P P A S N M E F M I W D V E L E R S A E A W A E
0	L H N K L R G Q V H P S A S N M E Y M T W D D E L E K S A E A WA E
W. clawed frog - GliPR1-like	T Q F L E A H N K Y R K K H N V P P M R L N A E L S K S A Q T WA N
W. clawed frog - GliPR2	L E F L K A N N V Y R S R H G A K P L Q L N S K I S Q E A Q R W A E
Exon 1 Exon 2	
	Exon 10 Exon 11 Exon 12 Exon 13 Exon 14
Annondiv 1:	

Appendix 4: Page 2 of 7

			ap.	1				ne	5 11		1. 0	101					_					
			~ **					-		0	-					AP4						
W. clawed frog - serotriflin-like	GC	TG	в н	S I	? P	DK	. R	Т	I P	~			-									A S
W. clawed frog - CRISP2	TC		~	S I		D K		Т	I P	_	FO		-									A S
W. clawed frog - CRISP1	TC		ΑH										-									A S
W. clawed frog - serotriflin-like	TC		FΗ																			
W. clawed frog - MGC108118	S C		ΥH																			
W. clawed frog - CRISP3	ТС	-	ΥH												G I	ΕN	LI	M	[S	S Y	P	A S
W. clawed frog - allurin	KC	_	QН																			ΡD
W. clawed frog - TEL1	ТС		КΗ					_												_	R	H S
W. clawed frog - PI15	T C	ΙW	DH	GΙ	P S	ΥL	L	ΚI	FL	G	Q 1	V -	-		- 1	LS	VI	ιT	G	RY	ſ K	S -
W. clawed frog - PI R3HDML	QC	ΚW	DΗ	GΙ	P N	QL	Μ	R	ΥI	G	Q 1	V -	-		- 1	LS	VF	ł S	G	R Y	(R	S -
W. clawed frog - CRISP LCCL1	T C	L W	ЕН	GΙ	P A	DL	L	P '	V I	G	Q 1	V -	-		- 1	LG	Αŀ	I W	G	RY	R	Р -
W. clawed frog - CRISP LCCL2	EC	ΙW	ЕН	GΙ	Р Т	A L	L	M	S I	G	Q 1	v -	-		- 1	LA	VF	I W	G	RY	R	Q -
W. clawed frog - GliPR1-like	ΗL	LS	I N	ΚN	/1 Q	нs	G	- 2	A G	G	Εľ	V -	-	- L	Ϋ́	Y S	Y S	s s	R	GF	tΤ	LA
W. clawed frog - GliPR2	ΗL	LN	LΚ	ΝI	Σĸ	нs	D	Т	sн	G	ΕÌ	V -	-	- I	w .	ΑK	sc	ЭG	Р	S I	Т	VΤ
W. clawed frog - serotriflin-like	WΕ	ΕA	VК	ΑV	VF	DE	Ν	E	S F	Е	Y (γ£	_		-			G	Р	K S	Р	DQ
W. clawed frog - CRISP2	WE	ΕA	νк	G V	VΥ	SΕ	Y	NI	DF	0	Y (3 V	_		2			G	Р	кs	Р	GL
W. clawed frog - CRISP1	w c	EA	ΙТ	ΑV	νN	SΕ	S	0.	ΥF	õ	Y C	3 V			2			G	Р	TS	s	N O
W. clawed frog - serotriflin-like	WE	AA	νт	ΑV	VF	DE	Ι	к I	DF	D	FO	ЭK	_		_			G	Р	кт	F	G A
W. clawed frog - MGC108118		DV																				
W. clawed frog - CRISP3		EV																				
W. clawed frog - allurin		AA																				
W. clawed frog - TEL1	-	YV													_							GA
W. clawed frog - PI15		O L											P		С	NP	_	_	L	_		GP
W. clawed frog - PI R3HDML	TV	DL				DE					FI	ьн	P 1	ς - R Ε			s				T	G A
W. clawed frog - CRISP LCCL1	РТ	vн				DE					FI	ь т У V	P	ΟĒ	-	DP				RC	s	G P
W. clawed frog - CRISP LCCL2	ΡA	vн				DE					V I	o v		ΗE				C P		RC	ŝ	GP
W. clawed frog - GliPR1-like	GN	VA									v	JK								P C	ן ב ג דו	КА
W. clawed frog - GliPR2		ΕV									E S	K										КА
				CA		K L	L			11	1 1	, <u>r</u>								1		n A
W. clawed frog - serotriflin-like	v v	C U	ΥT			w v	N	0 1	Y M	1.17	C C	c s	V	S Y		РК	C					- 0
W. clawed frog - CRISP2	V V V T			~		WY			Y M					S Y		PK						~
W. clawed frog - CRISP2	VT		YT	~												n N	3		-		-	- P
W. clawed frog - serotriffin-like			YT													PN			-		-	- E
W. clawed frog - MGC108118																PI					-	- E
	Q I V I		ΥT					_				_									-	- S
W. clawed frog - CRISP3				~		w 1 W A						_							-		-	- K
W. clawed frog - allurin	MI	GH		Q										A F		PG			-		-	- 1 D
W. clawed frog - TEL1	V T			Q .		W A					_	_		A K		YG			-	 -	-	- P WR
W. clawed frog - PI15	MC	_	YT																			
W. clawed frog - PI R3HDML		TH	Y I V T	QN		W A	S	SI	NR	. 1	G	A	VI	N I		IN			W	GS	I	WR
W. clawed frog - CRISP LCCL1	VC	_	ΥT																			
W. clawed frog - CRISP LCCL2		ТН																			, 1	wΕ
W. clawed frog - GliPR1-like		GH						_													-	
W. clawed frog - GliPR2	КТ	ĠΗ	FΤ	QN	/1 V	W K	Α	S 1	ĸΕ	V	G 1	/ G	LA	AS	S	θK	G.	• -	-		-	
Exon 1 Exon 2		Exc	on 3		F	Exon	4			Ex	on 5	5		Ex	on	6 📃		E	xon	7		1
Exon 8 Exon 9		Exo	n 10		E	xon 1	11		F	Exo	n 12	2		Exc	on 1	3		Ex	on	14		1
Appendix 4:																						
D2-67																						

Page 3 of 7

	1					P°			CAI	•									-1-				-											
TTT 1 1 C 4 '0' 11	37	77	37	F	37						D		~			F	~					D	37	77		~	D	77	~		-	-		
W. clawed frog - serotriffin-like	-			F		- F												v											_	Α		_	V	
W. clawed frog - CRISP2				F														T											С	Α	- H		P	
W. clawed frog - CRISP1			_			V	_																						С	G			Ρ	
W. clawed frog - serotriffin-like	Y	R	Y	Y	Y			H		_	_	_	G	Ν	Ι	Ε	G	K	Q	F	Т	Ρ	Y	K	I	G	Ρ	Т	С	G	D	C	P :	K.
W. clawed frog - MGC108118	\mathbf{L}	E	F	Y	F			Η			_	_						-											С	Е	D	С	K	S
W. clawed frog - CRISP3				Υ														R											С	A	D	С	P	D
W. clawed frog - allurin	Υ	Ρ	Y	V	S	Ι																							С	A	S	С	Ρ	E
W. clawed frog - TEL1	Υ	Ν	Y	F	Y	V	С	Ι	Ϋ́	Y	Ρ	Т	G	Ν	R	Е	D	Κ	V	Κ	Т	Р	Y	Q	Ν	G	Т	Т	С	G	L	C	Q	ĸ
W. clawed frog - PI15	R	А	V	Y	L	V	С	N	Y	S	Ρ	K	G	Ν	W	Ι	G	-	-	Е	A	Ρ	Y	Т	Ι	G	V	Ρ	С	s	Α	С	Р	Ρ
W. clawed frog - PI R3HDML	Q	A	s	Y	L	v	С	N	Y	s	I	K	G	Ν	W	Ι	G	-	-	Е	A	Р	Y	к	L	G	R	Ρ	С	s	Α	C	Р	Ρ
W. clawed frog - CRISP LCCL1						v																							С	s	Α	С	Р	Ρ
W. clawed frog - CRISP LCCL2	N	A	v	Y	L	v	С	N	Y	s	Р	K	G	Ν	w	I	G	2	_	Е	A	Р	Y	к	Ν	G	R	Ρ	С	s	Е	C	Р	Ρ
W. clawed frog - GliPR1-like						v	_																							Р	к	-	_	_
W. clawed frog - GliPR2	_		_																													v	т	D
			-	-				~			-	~	-	- 1	-	-	- 1	-	-	-	-					-	-		Ū	Ĩ		-	-	-
W. clawed frog - serotriffin-like	Δ	C	Б	N		-		- 1	S	т	C	т	N	\mathbf{v}	C	Р	v	0															ei:	
W. clawed frog - CRISP2	A		ŧ.	N		2			G	- F	_	T	N		C	P	v	Q A																
W. clawed frog - CRISP2	-		•			-		_		- H	_	T			C		v	Q o	-	-	-	-	-	-	-	-	-	-	-	-	Γ.	2	Γ.	
0	A		_							_	_	_			C C	Р Р	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
W. clawed frog - serotriffin-like	S					-						Т	N				H					-	-	-	-	-	-	-	-	-	-	-	-	
W. clawed frog - MGC108118	S	C				-				_								Q														-	-	-
W. clawed frog - CRISP3	A		_			-					_	Т						E														-	-	
W. clawed frog - allurin	S		-			-				- F																				-	-	-	-	-
W. clawed frog - TEL1	D		_			-			Q	-		L						Y												-	-	-	-	-
W. clawed frog - PI15	-					-				- F								Ρ												-	-	-	-	-
W. clawed frog - PI R3HDML	S	Y	_			-								Ν			F											-		-	-	-	-	-
W. clawed frog - CRISP LCCL1	S	Y	G	-	-	-	-	-	G	G																						Е		
W. clawed frog - CRISP LCCL2	S	Y	G	-	-	-	-	-	G	Ν	С	Q	Ν	Ν	L	С	Y	Κ	-	-	G	D	K	Η	Y	G	R	Е	-	-	G	Ι	V	Г
W. clawed frog - GliPR1-like	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - GliPR2	D	G	G	D	Ε	D	G	F	V I	K	S	Ρ	S	S	Т	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - serotriflin-like	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - CRISP2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - CRISP1	_	_	_	_	_	_	_	_	-	_	_	2	-	-	-	-	2	2	2	_	_	_	_	2	_	2	_	_	_	_	_	_	_	
W. clawed frog - serotriffin-like		-	-	-	-	-	_	-	-	-	_	-	_	-	_	-	-	-	-	-	-	-	-	-	_	-	-	-	-	-			-	
W. clawed frog - MGC108118	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	_	-	_	-
W. clawed frog - CRISP3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	- 1	-
W. clawed frog - allurin	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
W. clawed frog - TEL1	-	-	-	-	-	-		-			-	-	-			-		-		-		-	-	-	-	-	-	_	-	_	-	_	_	_
W. clawed frog - PI15																	_								_									
W. clawed frog - PI R3HDML		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		2		
W. clawed frog - CRISP LCCL1	N	г Б	T																															
W. clawed frog - CRISP LCCL2	-																													_		_		
U							_	_																								-	_	
W. clawed frog - GliPR1-like																				- 1				-										
W. clawed frog - GliPR2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Exon 1 Exon 2			F	Cxo	n í	3			F	xo	n -	4			1	Ex	on	5			1	Ex	on	6				E	101	ı 7				
			-						-								- 14	-						-			-						1	
Exon 8 Exon 9			E	xoi	n 1	0			Ex	or	ı 1	1			E	xo	n 1	12			E	xo	n 1	13]]	Ex	on	14			I	
Appendix 4:																																		

Appendix 4: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for CAP Superfamily Genes in *X. tropicalis* Genome

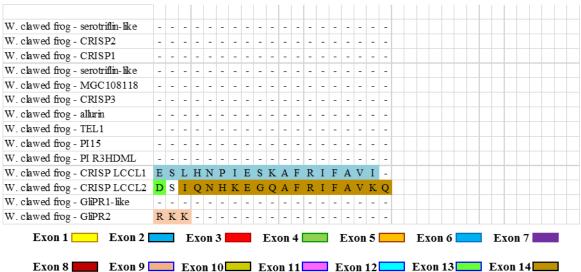
Appendix 4: Page 4 of 7

	······································
TTT 1 1 C ('0' 11	
W. clawed frog - serotriffin-like	
W. clawed frog - CRISP2	
W. clawed frog - CRISP1	
W. clawed frog - serotriffin-like	•••••••••••••••••••••••••••••••••••••••
W. clawed frog - MGC108118	
W. clawed frog - CRISP3	
W. clawed frog - allurin	
W. clawed frog - TEL1	
W. clawed frog - PI15	
W. clawed frog - PI R3HDML	· · · · · · · · · · · · · · · · · · ·
W. clawed frog - CRISP LCCL1	NNEVISTEQMSQIVS <mark>C</mark> EVRLRDQ <mark>C</mark> KGTT <mark>C</mark> NRYE <mark>C</mark>
W. clawed frog - CRISP LCCL2	T T Q T S T E N L M T Q A I K <mark>C</mark> A T K M R D S <mark>C</mark> K G S T <mark>C</mark> N R <mark>Y K C</mark>
W. clawed frog - GliPR1-like	
W. clawed frog - GliPR2	
-	
W. clawed frog - serotriffin-like	D T Y S
W. clawed frog - CRISP2	
W. clawed frog - CRISP1	DMY 8
W. clawed frog - serotrifin-like	
W. clawed frog - MGC108118	N K F S
W. clawed frog - CRISP3	
W. clawed frog - allurin	
W. clawed frog - TEL1	
W. clawed frog - PI15	G V T S
W. clawed frog - PI R3HDML	- GVKS
W. clawed frog - CRISP LCCL1	S A G C L E S N A K V I G S V H Y E M Q S S I C R A A I H S G V I D
W. clawed frog - CRISP LCCL2	
W. clawed frog - GliPR1-like	
	A V L P I P E K E L K S F R K D L L S
W. clawed frog - GliPR2	
W. clawed frog - serotriffin-like	S C S N Y T S Y C N T F P A I R D G C K A T C L C T N N Q I V -
-	
W. clawed frog - CRISP2	
W. clawed frog - CRISP1	G C S N Y I A Y C N L I P T I T N G C R G T C L C R N N E I I -
W. clawed frog - serotriffin-like	N C Q E L T T K Y S C S K Y P E L Q D D C P A H C R C T N N E I I -
W. clawed frog - MGC108118	N C D T P D T H C D T D P T A K N D C P A T C N C L K - E I H - N C D D L A S G D O C D T Y O I V K E G C P G S C L C K N N E I I -
W. clawed frog - CRISP3	
W. clawed frog - allurin	
W. clawed frog - TEL1	NCGTDKN ASLCDYSDIGCDATCKCGSEKIY -
W. clawed frog - PI15	N
W. clawed frog - PI R3HDML	N
W. clawed frog - CRISP LCCL1	
W. clawed frog - CRISP LCCL2	
W. clawed frog - GliPR1-like	
W. clawed frog - GliPR2	E H N Q Y R K L H G A G A L Q L S V A L S Q D A Q K W A D H L V G K
Exon 1 Exon 2	Exon 3 Exon 4 Exon 5 Exon 6 Exon 7
Exon 8 Exon 9	Exon 10 Exon 11 Exon 12 Exon 13 Exon 14
Appendix 4:	
Dece 5 of 7	

Page 5 of 7

	-					r				-7	_							r													
XX7 -1	_		_						_	_	_	_					_		_	_	_	_	_						_	_	+
W. clawed frog - serotriflin-like	-			-				-				_	-	-			-		_	_		-	-		_	-		-	-		-
W. clawed frog - CRISP2	-	-		-					-			_					-											-	_		-
W. clawed frog - CRISP1	-		-	-					-	_	- -	_		-			-	_	_	_	_	-			-			-	_	_	-
W. clawed frog - serotriflin-like	-	-	-	-	-	-	-	-		_		_		-		-		-	- ·	-	-	-	-	-		-	-	-	-		-
W. clawed frog - MGC108118	-	-	-	-	-	-	-	_	_	_		_	-	-		-		_		_	_	-	-	-	_	-	-	-	-		-
W. clawed frog - CRISP3	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-			-	-	-	-	-	-	-	-	-		-
W. clawed frog - allurin	-	-	-	-	-	-			-					-			-					-				-		-	-		-
W. clawed frog - TEL1	-	-	-	-	-	-	-	-	-		- -		-	-	-		-				-	-	-	-	-	-	-	-	-		-
W. clawed frog - PI15	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-			-	-	-	-	-	-	-	-	-		-
W. clawed frog - PI R3HDML	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-			-	-	-	-	-	-	-	-	-		-
W. clawed frog - CRISP LCCL1	N	S	F	V	Ι	S	к	V	S I	FO	Q A	v	Т	C	Е	Т	Т	V	ΕC	ζL	C	Ρ	F	Q	Ρ	Р	Т	Т	н	C F	þ
W. clawed frog - CRISP LCCL2												ΓI																			2
W. clawed frog - GliPR1-like	_								_		_		_				_	_		_										_	-
W. clawed frog - GliPR2	_					_		_	_	_	_	E N		-	_	_		_	_	_		-		_		_	_	v	A	E S	ć
																															T
W. clawed frog - serotriflin-like	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-			-	-	-	-	-	-	-	-	-		_
W. clawed frog - CRISP2	-	-	-	-	-			-	-	_				-			-					-	-		-			-	_		-
W. clawed frog - CRISP1	-	-	-		-			-		_		_		-			-	_	_	_	_	-			-			-	_	_	_
W. clawed frog - serotriflin-like	-	-	-	-		-			-					-								-	-			-	-	-	-	-	-
W. clawed frog - MGC108118	-	-	-			-		-	-					-			-					-		-				-	-		-
•	-	-	-	-		-	-	-				_		-						_		-		-			-	-			-
W. clawed frog - CRISP3	-		-	-	-	-	-	-		_		_		-				_						-				-	-		-
W. clawed frog - allurin	-	-	-	-					-		_	_		-										-					_	_	-
W. clawed frog - TEL1	-	-	-			-						_	-	-			-		_	_	_	-	-						_	_	-
W. clawed frog - PI15	-	-	-			-		-				_		-										-			-	-	_		-
W. clawed frog - PI R3HDML	-	-	-	-	-	-	_		-											- -				-			-				-
W. clawed frog - CRISP LCCL1	_			-								• н											_				_				
W. clawed frog - CRISP LCCL2												S Y													_					ΤA	
W. clawed frog - GliPR1-like	-	-		-		_		_	_	-	_		-		-	_	_														
W. clawed frog - GliPR2	Υ	Ν	Е	Ν	A	K.	Y	S	F.	A i	r i	, G	F	Q	S	G	S	G.	Nł	(T	Q	Μ		W	ĸ	S	s	s	Q	V C	Ĵ
																													_		_
W. clawed frog - serotriflin-like	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-			-	-	-	-	-	-	-	-	-		-
W. clawed frog - CRISP2	-	-	-	-	-	-	-	-		-			-			-		-		_				-		-	-	-	-		•
W. clawed frog - CRISP1	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-					-	-			-	-	-		-
W. clawed frog - serotriflin-like	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-			-	-	-	-	-	-	-	-	-		-
W. clawed frog - MGC108118	-	-	-	-	-	-	-	-	-	-					-	-	-	-			-	-	-	-	-	-	-	-	-		-
W. clawed frog - CRISP3	-	-	-	-	-	-	-	-	-		- -		-	-	-	-	-	-		- -	-	-	-	-	-	-	-	-	-		-
W. clawed frog - allurin	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-			-	-	-	-	-	-	-	-	-		-
W. clawed frog - TEL1	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-			-	-	-	-	-	-	-	-	-		-
W. clawed frog - PI15	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-			-	-	-	-	-	-	-	-	-		-
W. clawed frog - PI R3HDML	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-			-	-	-	-	-	-	-	-	-		-
W. clawed frog - CRISP LCCL1																															
W. clawed frog - CRISP LCCL2												ΖV																			
W. clawed frog - GliPR1-like	_	_	_	_	_	_	_	_		_			_		_	_				_		_	_	_	_	_	_	_	_	_	-
W. clawed frog - GliPR2	G	L	S	Т	D	Ν	к	G	M	Y	I A	v	G	F	Y	D	P.	A	G 1	ΙI	A	Ν	ĸ	G	Y	F	Е	D	N	VΙ	
Exon 1 Exon 2			I	Exc	on .	3			Е	xo	n 4				Ex	on	5			Е	xo	n 6				E	xoi	n 7			
		_	-			~ -		_	F			_							_	F						-			_	_	
Exon 8 Exon 9			E	XOI	n 1	.0			Ex	ton	11			E	XC	on 1	12			Ex	on	13	3]	Ex	on	14	/		
Appendix 4:																															
Daga 6 of 7																															

Page 6 of 7



Appendix 4: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for CAP Superfamily Genes in X. tropicalis Genome

Appendix 4: This figure provides MEGA7 ClustalW amino acid sequence alignment of all CAP superfamily proteins in *X. tropicalis* genome. The CAP signature sequences are boxed, the highly conserved cysteine are highlighted in yellow with a red font color. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online data for specific CAP genes. Color-coding superimposed on the sequences is used to represent different exons and exon borders are indicated by a change in color. Organisms and accession IDs of the amino acids are available in the supplemental data table 11.

Page 7 of 7

Appendix 5: Pairwise Distance of *X. tropicalis* CAP Amino Acid Sequences – Estimated Evolutionary Divergence Between Sequences

	X. topicalis protein name	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Serotriflin-like														
2	CRISP2	0.245													
3	CRISP1	0.491	0.513												
4	Serotriflin-like	0.667	0.667	0.775											
5	MGC108118	0.761	0.693	0.819	0.667										
6	CRISP3	0.570	0.547	0.629	0.733	0.605									
7	Allurin	0.999	0.981	0.963	0.963	1.035	1.092								
8	TEL1	1.092	1.054	1.054	1.073	1.174	0.897	1.092							
9	PI15	1.217	1.153	1.263	1.073	1.286	1.195	1.286	1.217						
10	PI R3HDML	1.112	1.112	1.195	1.112	1.132	1.217	1.440	1.310	0.460					
11	CRISP LCCL1	1.240	1.240	1.174	1.112	1.240	1.153	1.335	1.335	0.570	0.617				
12	CRISPLCCL2	1.217	1.174	1.217	1.054	1.174	1.240	1.217	1.310	0.535	0.680	0.419			
13	GliPR1-like	1.386	1.310	1.440	1.310	1.263	1.469	1.498	1.386	1.310	1.335	1.286	1.174		
14	GliPR2	1.558	1.440	1.386	1.558	1.527	1.590	1.498	1.527	1.657	1.558	1.558	1.590	0.913	
15	RhAG	1.888	1.888	1.846	2.079	1.933	2.028	2.028	2.251	2.028	1.888	2.028	2.134	2.079	2.316

A. Whole amino acid sequence pairwise distance

B. CAP/PR domain amino acid sequence pairwise distance

	X. topicalis protein name	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Serotriflin-like														
2	CRISP2	0.234													
3	CRISP1	0.443	0.491												
4	Serotriflin-like	0.580	0.664	0.708											
5	MGC108118	0.708	0.678	0.837	0.723										
6	CRISP3	0.567	0.594	0.635	0.664	0.664									
7	Allurin	0.891	0.855	0.820	0.872	0.986	0.891								
8	TEL1	0.966	0.986	0.928	1.006	1.069	0.855	0.966							
9	PI15	1.287	1.234	1.184	1.184	1.342	1.314	1.342	1.342						
10	PI R3HDML	1.287	1.342	1.234	1.260	1.234	1.342	1.464	1.287	0.365					
11	CRISP LCCL1	1.314	1.287	1.184	1.137	1.234	1.209	1.371	1.260	0.432	0.503				
12	CRISP LCCL2	1.209	1.184	1.184	1.160	1.260	1.314	1.371	1.314	0.443	0.567	0.333			
13	GliPR1-like	1.464	1.371	1.464	1.314	1.287	1.640	1.497	1.401	1.401	1.342	1.342	1.432		
14	GliPR2	1.602	1.497	1.531	1.497	1.497	1.640	1.762	1.602	1.853	1.679	1.762	1.679	0.787	
15	RhAG	2.125	2.065	2.190	2.701	2.190	2.333	2.500	2.595	2.065	1.953	2.333	2.413	2.595	2.413

Appendix 5: The follow tables show pairwise distance alignment of *X. tropicalis* CAP sequences. 'A' represents the pairwise alignment score for whole amino acid sequence. 'B' represents the respective protein CAP/PR amino acid alignment score. The number of amino acid substitutions per site from between sequences are shown. Analyses were conducted using the Poisson correction model. The analysis involved 15 amino acid sequences. All positions containing gaps and missing data were eliminated.

Protein name			_										An	nin	0 4	Ac	id	Se	q	ueı	nce	e										
OF FOLCA										_	_	_		_	_			_	_													
CLEC18A	-	-	-	-	-	-	-	-	-	-	-	-	-		•	• •	•		-		-		-	-	-	-	-		•	• •	-	-
CLEC18B	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-		-		-		-	-	-	-	-		•	• •	-	-
CLEC18C	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-		-		-		-	-	-	-	-		• •	• •	-	-
CRISP1	-	-	-	-	-	-	-	-	-	-	-	-	-		•	•	-		-		-		-	-	-	-	-		•	• •	-	-
CRISP3	-	-	-	-	-	-	-	-	-	-	-	-	-		•	•	-		-	-	-		-	-	-	-	-		• •	• •	-	-
CRISP2	-	-	-	-	-	-	-	-	-	-	-	-	-		•	•	-		-	-	-		-	-	-	-	-		• •	• -	-	-
PI16	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-	-	-	-	-		-	-	-	-	-		• •	• -	-	-
GLIPR1-like 1	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-	-	-	-	-		-	-	-	-	-		• •	• -	-	-
GLIPR1	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-		-	-	-		-	-	-	-	-		•	• -	-	-
GLIPR1-like 2	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-	- -	-	-	-		-	-	-	-	-		• •	-	-	-
PI15	-	-	-	-	-	-	-	-	-	-	-	-	-		•	- -	-	- -	-		- -		-	-	-	-	-				-	-
PI R3HDML	-	-	-	-	-	-	-	-	-	-	-	-	-		• •				-				-	-	-	-	-				-	-
GAPR	-	-	-	-	-	-	-	-	-	-	-	-	-				-		-				-	-	-	-	-				-	-
CRISP LCCL1	-	-	-	-	-	-	-	-	-	-	-	-	-						-				-	-	-	-	-				-	-
CRISP LCCL2	M	I S	5	7	'L	G	G															G Y										
		1		1																	1											
CLEC18A	-	-		-	-	-	-	-	-	-	- 1	-	-				- 1		-				-	-	-	-					-	-
CLEC18B	-	-			-	-	-	-	-	_	- 1	_	-					_	_	_	_		-	-	-	-	_				-	-
CLEC18D	-	-			-	-	-	-	_	_	-	_	_					_	_	_			-	-	-	-	_				-	-
CRISP1	-	-		-	-	-	-	-		_	_	_	_		+				_		_		-	-	-	_	1				-	-
CRISP3	-	-			-	-	-	-	-	-	_	-	-		+						-		Ē	-	-	-	-		+		_	Ē
CRISP2	-	-			-	-	-	-	-	-	-	-	-		-	-		-	-				-	-	-	-	-			-	-	E
PI16	-	-	-	-	-	-	-	-	-	-	-	-	-		+	-	+	-	+	-	_		-	-	-	-	-		-		-	-
GLIPR1-like 1	-	-	-	-	-	-	-	-	-	-	-	-	-	_	+	-	-	-	-	-	-		-	-	-	-	-				-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-		+	-	+	+	-		+	_	-	-	-	-	-			-	-	-
GLIPR1	-	-	-	-	-	-	-	-	-	-	-	-	-	_	_	_	_	_	-		-		-	-	-	-	-		•	• •	-	-
GLIPR1-like 2	-	-	-	-	-	-	-	-	-	-	_	-	_		-	-	-		-	-	-	_	-	-	-	-	-		• •	• •	-	-
PI15	-	-	-	-	-	-	-	-	-	-	-	-	-		•	• •	-		-		-		-	-	-	-	-		• •	• •	-	-
PI R3HDML	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-		-		-		-	-	-	-	-		• •	• •	-	-
GAPR	-	-	-	-	-	-	-	-			-							-					_		-		-		• •	• -	-	-
CRISP LCCL1	-	-																												• -	-	-
CRISP LCCL2	L	, S	S K	2	^Z Q	H	IN	E	S	H	S I	R	V	R I	R /	A i	I	P I	R	El	DI	K E	EE	I	L	M	L	H	1 k	C I	R	G
CLEC18A	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-		-	-	-		-	-	-	-	-		• •	• -	-	-
CLEC18B	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-		-	-	-		-	-	-	-	-		• •	• -	-	-
CLEC18C	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-		-		-		-	-	-	-	-		•	-	-	-
CRISP1	-	-	-	-	-	-	-	-	-	-	-	-	-		•		-		-		-		-	-	-	-	-		•	-	-	-
CRISP3	-	-	-	-	-	-	-	-	-	-	-	-	-		•	-	-		-	-	-		-	-	-	-	-			-	-	-
CRISP2	-	-	-	-	-	-	-	-	-	-	-	-	-				-		-	-	-		-	-	-	-	-				-	-
PI16	-	-	-	-	-	-	-	-	-	-	-	-	-				-		-		- 1		-	-	-	-	-				-	-
GLIPR1-like 1	-	-	-	-	-	-	-	-	-	-	-	-	-				-		-		- 1		-	-	-	-	-				-	-
GLIPR1	-	-	-	-	-	-	-	-	-	-	-	-	-				-		-				-	-	-	-	-				-	-
GLIPR1-like 2	-	-	-	-	-	-	-	-	-	-	-	-	-				-		-		_		-	-	-	-	-				-	-
PI15	-	-		-	-	-	-	-		-	-	-	_		_	_	_		_			_	-	-	-	-	_				-	-
PI R3HDML	-	-			-	-	-	-		-	-	_	_			-	-		-		-		-	-	-	_	1				-	-
GAPR	-	-		-	-	-	-	-		_	_	_	_		+		+	-	_				1	-	-		1				-	-
CRISP LCCL1	-	-	+	+	1	-	-	-	-	-	_	-	-		-		+	-	-				-	-	-	-	-		+	-	-	Ē
	-	-		- 		-	-			- P	- V?			_					-	_							- 1	_			-	
CRISP LCCL2	Q	۷	C,	(1	Q	A	. S	IN	IVI	E	11	VI	1	vv 1	71		с.	וים	C]	N i	51	A A	A	. w	A	3	Q	u .	I V	νĽ	S FI	l O
1 Exon]
8 Exon ix 6:	У				E	XO	n 1	0				E)	(0)	n 1	1				E	xo	n	12			11	ĽX	on	13	• <mark>-</mark>			E
f 8																																

Lava 2				-	- 41		~									۰.		1						- 51					-	
CRISP LCCL2					-																									
CRISP LCCL1																														
GAPR															- -															
PI R3HDML																						_								
PI15																														
GLIPR1-like 2	-	-	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-	-	- -	• -	-	-	-	- -	-	-	-	-	-
GLIPR1 GLIPR1-like 2	-	-	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-	-		• -	-	-	-	- -	-	-	-	-	-
GLIPR1-like 1	-	-	-	-	-	-		-		-	-	-	-	-		-	-	-	-			-	-	-		-	-	-	-	-
PI16	-	-	-	-	-	-	-	-		-	-	-	-	-		• -	-	-	-			-	-	-		-	-	-	-	-
CRISP2	-	-	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-	-	- -	• -	-	-	-	- -	-	-	-	-	-
CRISP3	-	-	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-	-		• -	-	-	-	- -	-	-	-	-	-
CRISP1	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• -	-	-	-	- -	-	-	-	-	-
CLEC18C	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-	- -	· -	-	-	-	- -	-	-	-	-	-
CLEC18B	-	-	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-	-		• -	-	-	-	- -	-	-	-	-	-
CLEC18A	-	-	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-	-	- -	-	-	-	-	- -	-	-	-	-	-
CRISP LCCL2	D	Y	T	Y	Ρ	Y	P	S I	EC	N	I P	W	С	P	ΕF	20	S	G	ΡI	M	C I	Ή	Y	ТC	<mark>2</mark> 1	I V	W	Α	Т	Т
CRISP LCCL1	-	-	-	-	-	-		-		-	-	-	-	-			-	-	- 1	MN	١N	IN	A	ΤJ	H I	ΙV	' H	S	G	-
GAPR	-	-	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-	-			-	-			· -	-	-	-	-
PI R3HDML	-	-	-	-	-	-	-	-		-	-	-	-	-			-	-	-			-	-	-		-	-	-	-	-
PI15	-	-	-	-	-	-	-	-		-	-	-	-	-		• -	-	-	-	- -	• -	-	-	-	- -	-	-	-	-	-
GLIPR1-like 2	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• -	-	-	-		-	-	-	-	-
GLIPR1	-	-	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-	-		-	-	-	-		-	-	-	-	-
GLIPR1-like 1	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• -	-	-	-		-	-	-	-	-
PI16	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• -	-	-	-	- -	-	-	-	-	-
CRISP2	-	-	-	-	-	-	-	-		-	-	-	-	-		• -	-	-	-		• -	-	-	-	- -	-	-	-	-	-
CRISP3	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• -	-	-	-	- -	-	-	-	-	-
CRISP1	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• -	-	-	-	- -	-	-	-	-	-
CLEC18C	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• •	-	-	-	- -	-	-	-	-	-
CLEC18B	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		-	-	-	-	- -	-	-	-	-	-
CLEC18A	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		-	-	-	-	- -	-	-	-	-	-
CLECISA	-	-	-		\square	-	_	+	_	-	-			_	+	-	-		$ \rightarrow $	_	-			+	_	-	-			_
CRISP LCCL2	Ρ	T	S	L	L	V	5	1 (σQ	1 N	1 L	G	Α	H	WC	ı R	Y	R	S	P(3 F	Η	V	9	S V			Е	V	ĸ
CRISP LCCL1	-	-	-	-	-	-	-	- -		-	-	-	-	-	- -	-		-				-	-	-	- -	-	-	-	-	-
GAPR	-	-	-	-	-	-		-		-	-	-	-	-		• •	-	-	-		• -	-	-	-		-	-	-	-	-
PI R3HDML	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• -	-	-	-	- -	-	-	-	-	-
PI15	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• -	-	-	-	- -	-	-	-	-	-
GLIPR1-like 2	-	-	-	-	-	-	-	-		-	-	-	-	-		• -	-	-	-		• •	-	-	-		• -	-	-	-	-
GLIPR1	-	-	-	-	-	-		-		-	-	-	-	-		-	-	-	-		• •	-	-	-		-	-	-	-	-
GLIPR1-like 1	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• •	-	-	-		-	-	-	-	-
PI16	-	-	-	-	-	-		-		-	-	-	-	-		-	-	-	-		• •	-	-	-		• -	-	-	-	-
CRISP2	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• •	-	-	-		• -	-	-	-	-
CRISP3	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• •	-	-	-		-	-	-	-	-
CRISP1	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• •	-	-	-		• -	-	-	-	-
CLEC18C	-	-	-	-	-	-	-	-		-	-	-	-	-		-	-	-	-		• -	-	-	-		-	-	-	-	-
CLEC18B	-	-	-	-	-	-	-	-		-	-	-	-	-	- -	-	-	-	-	- -	• -	-	-	-	- -	-	-	-	-	-
		-	L	-	-	-		- 1		-	-	-	-	-	- -	-	-	-	-		• •	-	-	- 1		-	-	-	-	-

Appe Page 2 of 8

Exons for v	CAL	5	-pe	Па		I.y	G	ene	28.1	III	un	e I		m	U S	sap	ier	15	G	eno	л	ie
CLEC18A						_																
CLEC18B						_																
CLEC18C																						
CRISP1																						
CRISP3			- <mark>M</mark>	IKQ	<u>[</u>]	LΗ	I P A	ΑL	ΕŢ	ГΤ	D	P (S	ΤC	3 F	VF	? P	A١	ИT	LΙ	FΡ	V
CRISP2											-		· -					- 1	МA	ΓI	L P	V
PI16											-		- -				· -					-
GLIPR1-like 1																		_				
GLIPR1																						
GLIPR1-like 2								- <mark>M</mark>	ΕÆ	A A	R	ΡŦ	۲A	R F	ΞW	R A	٩Q	S I	LΡ	LA	<mark>a v</mark>	' G
PI15																						
PI R3HDML				I P L																		
GAPR																						
CRISP LCCL1	NW	WG	ΗA	ΡY	ΚI	ΗG	RI	P C	S A		P	ΡS	F	GC	G G	CF	٤E	NI		YF	K E	G
CRISP LCCL2	NW																					
																-						
CLEC18A	LG				_				_		_			- 1	ГА	WA	A F	V	V P	PC	J T	0
CLEC18B	LG															WA					~	~
CLEC18D																WA					~	~
CRISP1	LG																				~	~
CRISP1 CRISP3	L F L L																					
CRISP3 CRISP2	LL															LI						
	ь-					•						_	_									
PI16				- -	- -		- -	- -	- 1	чH												
GLIPR1-like 1						• •		• •			- 1					II						
GLIPR1					1	• •		• •	1		7					SI						
GLIPR1-like 2	GV				-											LI						
PI15				- P																		
PI R3HDML				- P					-													
GAPR					1.11		1 - i -	1.1			1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -		1.1									
CRISP LCCL1																						
CRISP LCCL2	Т-	- Y	ΤP	K P	РЕ 1	ΓD) E N	ИN	ΕV	VΕ	T.	ΑI				_	IV	WI	LQ	ΡI	R V	Μ
														CA	P3							
CLEC18A	EQ																					
CLEC18B	ΕQ	A P	MA	<mark>. G</mark> A	LÌ	N R	KI	ΞS	FΙ	LL	L	S I	L H	NF	۲L	RS	s w	VQ	QΡ	P A	ΑA	D
CLEC18C	ΕQ	A P	ΜA	<mark>. G</mark> A	LÌ	N R	KI	ΞS	FΙ	LL	L	S I	L H	NF	۲L	RS	s w	VQ	Q P	P A	ΑA	D
CRISP1	DQ	FΝ	ΚL	VΊ	D	LΡ	NV	VQ	ΕI	ΕI	V1	NI	Η	N A	ΑL	RF	۲R	VV	V P	P A	A S	N
CRISP3				LI												_						
CRISP2	- A	FΤ	ΑL	LΊ	T	ζL	Q	VQ	RΗ	ΕI	V1	NK	н	ΝI	ΞL	RK	ζA	V S	SΡ	P A	A S	N
PI16	ΑT																					
GLIPR1-like 1	S S																					
GLIPR1	AN																					
GLIPR1-like 2	AR																					
PI15				RY																		
PI R3HDML	11														-							
GAPR																						
CRISP LCCL1																						NI
CRISP LCCL2	R P	ΤK	ΡK	. K 1	S I	A V	N	ΥM	1	₹V	V]	R	ם	TK	M	KI	JR	C	C G	S .	r <mark>C</mark>	N
Ener 2		•		2 -			F we	n 4			Ŧ			-				n -				
Exon 2			LXOI	13			LX0	n 4			1	LXC	013	,			-xo	ш О			Ē	Exc
Ener 0				10		1 -			_									1.7	_	_	F	
Exon 9		E	xon	10		IE	xor	111			E	XOI	112	2		E	xon	13			E	x01
x 6:																						
f 8																						

App Page 3 of 8

CLEC18C		CAT Supertaining Genes in the Homo suprens Genome
CLEC18B	CLEC19A	MDDIDWSDSIAOIAOADA
CLEC18C		
CRISP1		
CRISP3		
CRISP2		
P116		
GLIPRI-l&c 1		
GLIPRI		
GLIPRI-WC 2		
P115		
PIR3HDML GAPR CRISP LCCL1 R YEC PAGC LDSKAKVIGSVHYEMQSSICAAAIHY CRISP LCCL2 R YC PAGC LDSKAKVIGSVHYEMQSSICAAAIHY CRISP LCCL2 R YC PAGC LDSKAKVIGSVHYEMQSSICAAAIHY CRISP LCCL2 R YC PAGC LDSKAKVIGSVHYEMQSSICAAAIHY CRISP LCCL2 R YC PAGC LDSKAKVIGSVHYEMQSSICAAAIHY CRISP LCCL2 C G I P T PSLASGLWR TLQVGWNQLL PAG CLEC186 C G I P T PSLASGLWR TLQVGWNQLL PAG CRISP1 C DM TE SN PKDR - MTSLKCGENLYMSSAS CRISP2 C C NYRH SN PKDR - MTSLKCGENLYMSSAS CRISP2 C C VWGHN KERGR GENLFAITDEC GLIPRI-1ke1 Q C K FE HN DCLDKSYK QAAFEYVGEN IWLGGIK GLIPRI N CQ FSHN TRLKPPHKLHPNFTSLGEN IWLGGIK GLIPRI N CQ FSHN TRLKPPHKLHPNFTSLGEN IWLGGVPN PI15 T C IWDHGPSYLLRFLGQNLSVRTGRYRSILGUV PR3HDML Q C I WAHGPSQLMRYVGQNLSVRTGRYRSILGUV FR3HDML Q C I WAHGPSQLMRYVGQNLSVRTGRYRSICUV FR3HDML Q C I WAHGPSQLMRYVGGRYSHAAGEC ARNATCT H CRISP1 CCISP1 CCISP1 CCISP1 CCISP1 CCISP2 S SWSQAIQSWFDEYNDFDFGVGPKTPNAVVGG-HY FI16 M DV P LAME EWHHER EHYNLSAATCS S C SCWC G HY GIPR1-FR4 I TAWYNE TQ FYDFDS - LS C SRVC G HY GIPR1-FR4 I TAWYNE TQ FYDFDS - LS C SRVC G HY GIPR1-FR4 I TAWYNE TQ FYDFDS - LS C SRVC G HY GIPR1-FR4 I TAWYNE TQ FYDFDS - LS C SRVC G HY GRIPR1 GRPR C RISP1 CCISP1 CCISP1 CCISP1 CCISP1 CCISP1 CCISP1 CCISP3 S SWSQAIQSWFD I LD FYTGY C PR		
GAPR Image: Construction of the second o		MEYMVWDEN LAK SAEAWAA
CRISP LCCL1 R Y E C P A G C LD SK AK V I G S V HY EM Q S S I C RAA I HY CRISP LCCL2 R Y O C P A G C LN HK AK I F GT L F Y E S S S I C RAA I HY CLEC18A L C G T P T P S L A S G LWR T L Q V GWNM Q L L P A G I CLEC18B L C G I P T P S L A S G LWR T L Q V GWNM Q L L P A G I CLEC18C L C G I P T P S L A S G LWR T L Q V GWNM Q L L P A G I CRISP1 V C DM T E SN P LER RL P N T F C G EN I HY S S S S C RISP2 R C T LQ H SN P KD R - MT S L K G G EN L YMS S A S S C RISP2 C T LQ H SN P KD R - MT S L K G G EN L YMS S S A S C RISP2 Q C N W H	PI R3HDML	MEYMVWDKRLARAAEAWA
CRISP LCCL2 R YO C PAG C N HK AK I F GT L F Y E S S S I C A A I H Y CLEC18A L C G I P T P S L A S G LWR T L Q V GWNN Q L L PA G I CLP A G I CLEC18B L C G I P T P S L A S G LWR T L Q V GWNN Q L L PA G I CLEC18C L C G I P T P S L A S G LWR T L Q V GWNN Q L L PA G I CRISP1 W C DM T E S N P L E R L P N T F C G E N HM MT S Y P CRISP3 Q C N Y R H S N P K D R - MT S L K C G E N L YM S S A S - CRISP2 R C T LQ H S D P E D R - KT S T K C G E N L YM S S A S - PI16 Q C W G HN	GAPR	L <mark>C</mark> KNLNREAQQ <mark>YSEALA</mark> S
CLEC18A LC GTPTPSLASGLWRTLQYGWNN QLLPAGI CLEC18B LG GIPTPSLASGLWRTLQYGWNN QLLPAGI CLEC18C LC GIPTPSLASGLWRTLQYGWNN QLLPAGI CRISP1 YC DMTESNPLERRLPNTF CGENMHMTSYP- CRISP3 QCNYRHSNPLERRLPNTF CGENMHMTSYP- CRISP2 KCTLQHSNPKDR.MTSLKCGENLYMSSAS. CRISP2 KCTLQHSNPKDR.MTSLKCGENLYMSSAS. CRISP2 KCTLQHSNPKDR.MTSLKCGENLYMSSAS. CRISP2 KCTLQHSNPKDR.MTSLKCGENLYMSSAS. CRISP2 KCTLQHSNPKDR.MTSLKCGENLYMSSAS. CRISP2 KCTLQHSNPKDR.MTSLKCGENLYMSSAS. CRISP2 KCTLQHSNPKDR.MTSLKCGENLYMSSAS. CRISP2 KCTLQH.LCSNPKCYAAFEYVGENIWLGGIKS GLIPRI-Mc1 QCKFEHNDCLDKSYKCYAAFEYVGENIWLGGIKS GLIPRI-Mc2 KCLFTHNIYLQDVQMVHPKFYGIGENWVGPENI P115 TCIWDHGPSYLLRFLGQNLSYNTGRYSILQUYS GAPR TRILKHSPESSRGQCGEN CRISPLCCL1 GIIDNDGGWVDITRQGRKHYFIKSNRNGIQTIG CRISPLCCL2 GILDDKGGLVDITRQGRKHYFIKSNRNGIQTIG CRISPLCCL2 GILDDKGGLVDITRQGRKHYFIKSNRNGIQTIG CLEC18A VSFVEVVSLWFAEGQRYSHAAGECARNATCT.HX CLEC18B AS FVEVVSLWFAEGQRYSHAAGECARNATCT.HX CLEC18B AS FVEVVSLWFAEGQRYSHAAGECARNATCT.HX CRISP1 VSWSVIGWYDEILDFVYGVGPKSPNAVVG.HX CRISP3 SSWSQAIQSWFDEYNDFDFGVGPKSPNAVVG.HX CRISP2 TSWSSAIQSWYDEILDFVGVGFKSPNAVVG.HX CRISP3 SSWSQAIQSWFDEYNDFDFGVGPKSPNAVVG.HX GLIPRI-Mc1 FTPRHAITAWYNETQFYDFXRICKKVCG.HX GLIPRI-MS FTVSSSAIQSWYDEILDFVGVGPKSPNAVVG.HX GLIPRI-SVSSAIQSWYDEILDFVGVGPKSPNAVVG.HX GLIPRI-SVSSAIQSWYDEILDFVGVGPKSPNAVVG.HX GLIPRI-SVSSAIQSWYDEILDFVGVGPKSPNAVVG.HX GLIPRI-MS TASSAIQSWYDEILDFVGVGPKSPNAVVG.HX GLIPRI-MS FTVSSSAIQSWYDEILDFVGVGPKSPNAVVG.HX GLIPRI-MS FTSSAINWYDETQFYDFTSCSGCSGCS CRISPLCCL1 YQSANSFTVSKVVQQDLCYTTVAQLCPFS.HX GAPR LAWASSYDQTG	CRISP LCCL1	R Y E C P A G C L D S K A K V I G S V H Y E M Q S S I C R A A I H Y
CLEC18A LC GTPTPSLASGLWR TLQ VGWNNQLLPAGI CLEC18B LC GIPTPSLASGLWR TLQ VGWNNQLLPAGI CLEC18C LC GIPTSNPLERRLPNTFCGENMHMTSYP- CRISP1 VC DMTESNPLERRLPNTFCGENMHMTSYP- CRISP3 QC NYRHSNPLERRLPNTFCGENIYMSSAS- CRISP2 KC TLQHSDPEDR-KTSTCGENIYMSSAS- CRISP2 KC TLQHSDPEDR-KTSTCGENIYMSSAS- CRISP2 KC TLQHSDPEDR-KTSTCGENIYMSSAS- CRISP2 KC TLQHSDPEDR-KTSTCGENIYMSSAS- CRISP2 KC TLQHKERGRRGENLFAITDEC GLIPRI-Wc1 QC KFEHNDCLDKSYK YK YAAFEYVGENIWLGGIKS GLIPRI-Wc2 KC LFTHNIYLQDVQMVHPKFTSIGENNWVGPENI PI15 TC IWDHGPSYLLRFLGQNLSVRTGRYSSI CRISPLCCL1 GIIDDKGGLVDITRNGKVPFFVKSSRMGCGEN CRISPLCCL1 GIIDDKGGLVDITRNGKVPFFVKSSRMGCGEN CRISPLCCL2 GILDDKGGLVDITRNGKVPFFVKSSRMATCT-HY CLEC18A VSFVEVVSLWFAEGQRYSHAAGECARNATCT-HY CLEC18C ASFVEVVSLWFAEGQRYSHAAGECARNATCT-HY CRISP1 VSWSSVIGVWYSESTSFKHGEWTTDDDITTDHY CRISP1 VSWSSVIGVWYSESTSFKHGEWTTDDDITTDHY CRISP2 TSWSSAIQSWFDEYNDFDFS.SGCGFS-HAVYG-HY PI16 MDVPLAMEEWHHEREHYNLSAATCSPGWCG-HY GLIPRI-Wc2 FTASIAICSWYDEILDFVSGCFSSRAVGG-HY GLIPRI-Wc2 FTASIAICSWYDEILDFVSGCFSSRAVGG-HY GLIPRI-Wc2 FTASIAICSWYDESTSFKHGEWTTDDDITTDHY CRISP2 TSWSSAIQSWFDEYNDFDSSCSGCGSN-HY GLIPRI-Wc2 FTASIAIRSWHAEKKMYNFEN-GSCSRVCG-HY GLIPRI-Wc2 FTASIAIRSWHAEKKMYNFEN-GSCSRVCG-HY GLIPRI-Wc2 FTASIAIRSWHAEKKMYNFEN-GSCSSCGSDCS-NY P115 PWYDEVKDYAFPYPQDCNPRCPMRCFGPMCT-HY CRISP1 KSSAITNWYDEIQDYDFXRICKKVCG-HY GLIPRI-Wc2 FTASIAIRSWHAEKKMYNFEN-GSCSGCS-NY P115 PWYDEVKDYAFPYPQDCNPRCPMRCFGPMCT-HY CRISPLCCL1 YQSANSFTVSKVVVQATCETTVAQLC PFFK-PA CRISPLCCL2 YKPSSSFMVSKVKVVQDLDCYTTVAQLC PFFK-PA CRISPLCCL2 YKPSSFMVSKVKVVQDLDCYTTVAQLC FFFK-PA CRISPLCCL2 YKPSSFMVSKVKVVQDLDCYTTVAQLC FFFK-PA CRISPLCCL2 YKPSSFMVSKVKVVQDLDCYTTVAQLC FFFK-PA CRISPLCCL2 YKPSSFMVSKVKVVQDLDCYTTVAQLC FFFK-PA CRISPLCCL2 YKPSSFMVSKVKVVQDLDCYTTVAQLC FFFK-PA	CRISP LCCL2	R YQ <mark>C</mark> P A G <mark>C</mark> L N HK A K I F G T L F Y E S S S S I <mark>C</mark> R A A I H Y
CLEC18B LC G I P T P S L A S G LWR T L Q V GWNN Q L L PA G I CLEC18C L C G I P T P S L A S G LWR T L Q V GWNN Q L L PA G I CRISP1 Y C DMTE S N P L ER L PN T F C G ENN HMT S Y P . CRISP3 Q C N Y R H S N P K D R - MT S L K C G E N L YMS S A S . CRISP2 K C T L O H S D P E D R . K T S T R C G E N L YMS S A S P . P116 Q C V W G H N K E R G R R G E N L F A I T D E C GLIPRI-like 1 Q C K F E H N D C L D K S Y K C YA A F E Y V G E N I W L G G I K S GLIPR1 N C Q F S H N T R L K P P HK L H P N F T S L G E N I W T G S V P I GLIPR1-like 2 K C L F T H N I Y L Q D V Q M V H P K F Y G I G E N M W C G P E N I P115 T C I W D H G P S Y L L R F L G Q N L S V R T G R Y R S I L Q L V P P183HDML Q C I W A H G P S Q L M R Y V G Q N L S I H S G Q Y R S V V D L M GAPR T R I L K H S		CAP4
CLEC18C I C G I P T P S L A S G LWR TL Q V GWNN Q L P A G I CRISP1 Y C DM T E S N P L E R R L P N T F C G E N M HM T S Y P - CRISP3 Q C N Y R H S N P K D R - M T S L K C G E N L YMS S A S - CRISP2 K C T L Q H S D P H D R - K T S T R C G E N L YMS S D P - P116 Q C VWG H N K E R G R R G E N L F A I T D E C GLIPR1-lke 1 Q C K F E H N D C L D K S Y K Y A A F E Y V G E N I W L G G I K S GLIPR1 N C Q F S H N T R L K P P H K L H P N F T S L G E N I W T G S V P I GLIPR1-lke 2 K C L F T H N I Y L Q D V Q M V H P K F Y G I G E NM W V G P E N I P115 T C I WD H G P S Y L L R F L G Q N L S V R T G R Y R S I L Q V Y P1 R 3 H D M Q C I W A H G P S Q L M R Y V G Q N L S I H S G Q Y S V V D L M H GAPR T R I L K H S P E S S R G Q C G E N	CLEC18A	L <mark>C</mark> GTPTPSLASGLWRTLQVGWNMQLLPAGI
CRISP1 Y DMTE DMTE SNPLERRLPNTFC GENNMMMTSYP CRISP3 Q CNYRH SNPKDR MTSLKC GENLYMSSAS CRISP2 K CTLOHS SDPEDR KTSTC GENLYMSSAS CRISP2 K CTLOHS SDPEDR KTSTC GENLYMSSAS CRISP1 K Q VWGHN SDPEDR KTSTC GENLYMSSAS GLIPR1-% Q CKFEHNDCLDKSYKCYAAFEY GENLYMSSAS FTDEC GENNWYGPEN GLIPR1 N Q CFS NNTRLKPPHKLHPNFTSLGEN I WLGGIKS GENNWYGPEN GLIPR1-% K C LFTHN I YLQDVQMVHPKFYG I GENNWYGPEN P115 T C IWDHGPSYLLRFLGQNLSVRTGRWYSI SILQLVE P18HDML Q IWAHGPSQLMRYVGQNLSIHSGQYSVULME GANNWYGPENE GAPR TRILKHS FYENSSOG GENN SVULME GAPR TRILKHS FYENSSOG GENNWYGPENE GENNY CRISPLCCL1 GILDDKGGLVDITRNGKVPFFVKSERGQ GRANATCTHY GANNTCTHY CRISPLCCL2 GILDDKGUVSLWFAEGQRYSHAAGEC ARNATCTHY CAPI CLEC18A V	CLEC18B	LCGIPTPSLASGLWRTLQVGWNMQLLPAGI
CRISP3 Q N Y R H S N P K D R - M T S L K C G E N L YMS S A S - CRISP2 S C T L Q H S D P H D R - K T S T R C G E N L YMS S D P - P116 Q C VWG H N K E R G R R G E N L F A I T D E G G L P R - K T S V G E N I W L G G I K S G L P R - K T S L G E N I W L G G I K S G L P R - K T S L G E N I W L G G I K S G L P R - K T S L G E N I W U G G I K S G L P R - K T S L G E N I W T G S V P I G G L P R - K C L F T H N I Y L Q D V QMV H P K F Y G I G E N M W G P E N I P 15 T C U W D H G P S Y L L R F L G Q N L S V R T G R Y R S I L Q L V P I R 3 H D M L Q C I W A H G P S Q L M R Y V G Q N L S I H S G Q Y S V V D L M I G APR GAPR T R I L K H S P E S S R G Q C G E	CLEC18C	LCGIPTPSLASGLWRTLQVGWNMQLLPAGI
CRISP2 K C TLQHSDPEDR-KTSTRCGENLYMSSDP P116 Q C VWGHN	CRISP1	YCDMTE SNPLERRLPNTFCGENMHMTSYP -
CRISP2 K C TLQHSDPEDR-KTSTRCGENLYMSSDP- P116 Q C VWGHNKERGRGENLFAITDEC GLIPR1-Mc1 Q CKFEHNDCLDKSYKCYAAFEYVGENIKAFSGENLFAITDEC GLIPR1-Mc2 K CLFTHNIYLQDVQMVHPKFYGIGENIWLGGIKS GLIPR1-Mc2 K CLFTHNIYLQDVQMVHPKFYGIGENIWVGPENE P115 TC IWDHGPSYLLRFLGQNLSVRTGRYSILQUV P183HDML Q C IWAHGPSQLMRYVGQNLSIHSGQYRSVVDLMK GAPR TRILKHS - PESSRGQCGEN - CRISPLCCL1 GILDDKGGLVDITRNGKVPFFVKSERHGVQSLSK CAPI CLEC18A VSFVEVVSLWFAEGQRYSHAAGECARNATCTHY CAPI CLEC18A VSFVEVVSLWFAEGQRYSHAAGECARNATCTHY CAPI CLEC18A SSWSQAIQSWFDEYNDFDKGVGPKSPNAVVGFHY CAPI CLEC18C ASFVEVVSLWFAEGQRYSHAAGECARNATCTHY CAPI CLEC18C ASFVEVVSLWFAEGQRYSHAAGECARNATCTHY CAPI CLEC18C SSWSQAIQSWFDEYNDFDKSHAGECARNATCTHY CAPI CRISP1 VSWSSVIGVWYSESTSFKHGEWTTDDDITTDHY CRISP1 CRISP2 TSWSSAIQSWFDEYNDFDEYNDFFOSGENSPAVVGFHY GAPA GLIPR1-Mc1 MDVPLAMEEWHHEREHYNLSAATCSPGQMCGSNCGGNCGHY	CRISP3	QCNYRHSNPKDR-MTSLKCGENLYMSSAS-
P116 Q C VWG HN · · · · · · · · · · · · K E R G R R G E N L F A I TD E G GLIPR1-lke1 Q C K F E HN D C L D K S YK Y A A F E Y V G E N I W L G G I K S GLIPR1 N C Q F S HN T R L K P P HK L H P N F T S L G E N I W L G S V P I GLIPR1-lke2 K C L F T HN I Y L Q D V QM V H P K F Y G I G E N M W G P E N I P115 T C I WD H G P S Y L L R F L G Q N L S V R T G R Y R S I L Q L V I GAPR TR I L K H S · · · · · · · P E S S R G Q C G E N · · · · · · · · · · · · C CRISP LCCL1 G I D D D G G W D I T R Q G R K H Y F I K S N R N G I Q T I G R CRISP LCCL2 G I L D D K G G L V D I T R N G K V P F F V K S E R H G V Q S L S F CLEC18A V S F V E V V S L W F A E G Q R Y S H A A G E C A R N A T C T - H CLEC18B A S F V E V V S L W F A E G Q R Y S H A A G E C A R N A T C T - H CLEC18C A S F V E V V S L W F A E G Q R Y S H A A G E C A R N A T C T - H CRISP1 V S W S S V I G VWY S E S T S F K H G E W T T D D D I T T D H Y CRISP2 T S W S S A I Q S W F D E Y N D F D F G V G P K T P N A V Y G - H Y CRISP2 T S W S S A I Q S W Y D E I L D F V G V G P K S P N A V Y G - H Y P116 M D V P L A M E E W H HE R E H Y N L S A A T C S P G Q M C G - H Y GLIPR1-lke1 F T P R H A I	CRISP2	
GLIPR1 N C Q F S HN T R L K P P HK L HP N F T S L G EN I W T G S V P I GLIPR1-like 2 K C L F T HN I Y L Q D V QMV HP K F Y G I G EN M W V G P EN I PI15 T C I WD HG P S Y L L R F L G Q N L S V T G R Y R S I L Q L V I PI R3HDML Q C I WA HG P S Q LMR Y V G Q N L S I H S G Q Y R S V V D L M I GAPR T R I L K H S P E S S R G Q C G EN	PI16	
GLIPR1 N C Q F S HN T R L K P P HK L HP N F T S L G EN I W T G S V P I GLIPR1-like 2 K C L F T HN I Y L Q D V QMV HP K F Y G I G ENMWV G P EN I P I15 T C I WD HG P S Y L L R F L G Q N L S V T G R Y R S I L Q L V H P I R 3 HDML Q C I WA HG P S Q LMR Y V G Q N L S I H S G Q Y R S V V D L M H GAPR S I L Q L V H T R I L K H S P E S S R G Q C G EN	GLIPR1-like 1	OCK FEHNDCLDKSYKCYAAFEYVGENIWLGGIKS
GLIPR1-ike 2 K C L F T Y L Q Q WVGPEN Q Q WVGPEN Q Y S T Q I WDHGPSYLLRFLGQNLSVRTGRYGQVLSILSUV G Y S T Q V Y S Y Q Y S Y Q Y S Y Q Y S Y Q Y S Y Q Y S Y Q Y S Y Q Y S Y Q Y S Y Q Y S Y Q Y S Y Q Y Y Y Q Y		
P115 TC I WDHG P S Y L LR F L G QN L S VR T G R Y R S T L Q L VR PI R3HDML QC I WA HG P S Q LMR YV G QN L S I H S G Q YR S V V D LMR GAPR TR I L K H S P E S S R G Q C G E N		
PI R3HDML QC I WA HG P S Q LMR Y V G Q N L S I H S G Q Y R S V V D LMR GAPR TR I LK HS PESSRGQC GEN		
GAPRTRILKHSPESSRGQCGENCRISP LCCL1GIIDNDGGWVDITRQGRKHYFIKSNRNGIQTIGRCRISP LCCL2GILDDKGGLVDITRNGKVPFFVKSERHGVQSLSECLEC18AVSFVEVVSLWFAEGQRYSHAAGECARNATCTHYCLEC18BASFVEVVSLWFAEGQRYSHAAGECARNATCTHYCLEC18BASFVEVVSLWFAEGQRYSHAAGECARNATCTHYCLEC18CASFVEVVSLWFAEGQRYSHAAGECARNATCTHYCLEC18CASFVEVVSLWFAEGQRYSHAAGECARNATCTHYCRISP1VSWSSVIGVWYSESTSFKHGEWTTDDDITTDHYCRISP2TSWSSAIQSWYDEILDFVGVGPKSPNAVVG-HYCRISP2TSWSSAIQSWYDEILDFVGVGPKSPNAVVG-HYP116MDVPLAMEEWHHEREHYNLSAATCSPQMCG-HYGLIPR1-Ike1FTPRHAITAWYNETQFYDFDS-LSCSRVCG-HYGLIPR1FSVSSAITNWYDEIQDYDFKT-RICKVCG-HYGLIPR1FSVSSAITNWYDEIQDYDFKT-RICCHGPMCCFGPMCT-HYP115PWYDEVKDYAFPYQDCNPRCPMRCFGPMCT-HYP115PWYDEVKDYAFPYQDTG		
CRISP LCCL1 G I I DN DG GWVD I T RQ GRK H YF I K SNRN G I Q T I G F CRISP LCCL2 G I L DD K G G L V D I T RN G K V P F F V K S E R H G V Q S L S F CLEC 18A V S F V E V V S L W F A E G Q R Y S H A A G E C A R N A T C T - H CLEC 18B A S F V E V V S L W F A E G Q R Y S H A A G E C A R N A T C T - H CLEC 18B A S F V E V V S L W F A E G Q R Y S H A A G E C A R N A T C T - H CLEC 18C A S F V E V V S L W F A E G Q R Y S H A A G E C A R N A T C T - H CRISP1 V S W S S V I G V W Y S E S T S F K H G E W T T T D D D I T T D H Y CRISP3 S S W S Q A I Q S W F D E YN D F D F G V G P K T P N A V V G - H CRISP2 T S W S S A I Q S W Y D E I L D F V Y G V G P K S P N A V V G - H Y116 MD V P L A M E E W H H E R E H YN L S A A T C S P G Q M C G - H GLIPR1 F S V S S A I T N W Y D E I Q D Y D F K T - R I C K K V C G - H GLIPR1 F S V S S A I T N W Y D E I Q D Y D F K T - R I C K K V C G - H GLIPR1 F S V S S A I T N W Y D E I Q D Y D F K T - R I C K K V C G - H GLIPR1 F S V S S A I T N W Y D E I Q D Y D F K T - R I C K K V C G - H Y15 P W Y D E V K D Y A F P Y P Q D C N P R C PM R C F G P M C T - H Y R 3 HDML S W S E E K W H Y L F P A P R D C N P H C P W R C D G P T C S - H GAPR L A WA S Y D Q T G H CRISP LCCL1 Y Q S A N S F T V S K V T V Q A V T C E T T V E Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A C E X M 3		
CRISP LCCL2 G I L D D K G G L V D I T R N G K V P F F V K S E R H G V Q S L S H CAPI CLEC18A V S F V E V V S L W F A E G Q R Y S H A A G E C A R N A T C T - H CLEC18B A S F V E V V S L W F A E G Q R Y S H A A G E C A R N A T C T - H CLEC18C A S F V E V V S L W F A E G Q R Y S H A A G E C A R N A T C T - H CRISP1 V S W S S V I G V W Y S E S T S F K H G E W T T D D D I T T D H Y CRISP3 S S W S Q A I Q S W F D E Y N D F D F G V G P K T P N A V V G - H Y CRISP2 T S W S S A I Q S W Y D E I L D F V Y G V G P K S P N A V V G - H Y CRISP2 T S W S S A I Q S W Y D E I L D F V Y G V G P K S P N A V V G - H Y GLIPR1-lke 1 F T P R H A I T A W Y N E T Q F Y D F D S - L S C S R V C G - H Y GLIPR1-lke 2 F T A S I A I R S W H A E K K M Y N F E N - G S C S G D C S - N Y P115 P W Y D E V K D Y A F P Y P Q D C N P R C P M R C F G P M C T - H Y R 3 H D M L S W S E E K W H Y L F P A P R D C N P H C P W R C D G P T C S - H Y GAPR L A W A S - Y D Q T G H Y CRISP LCCL1 Y Q S A N S F T V S K V T V Q A V T C E T T V E Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F F K - P A C F S M S T V S K V K V Q D L D C Y T T V A Q L C P F F K - P A C F S M S M S F T V S K V K V Q D L D C Y T T V A Q L C P F F K - P A C S M S T S T S S F M V S K V K V Q D L D C Y T T V A Q L C P F K - P A C S M S T S M S M S F T V S K V K V Q D L D C Y T T V A Q L C P F K S S F M V S K V K V Q D L D C Y T T V A Q L C P F K S S F M V S K V K V Q D L D C Y T T V A Q L C P F K S S F M V S K V K V Q D L D C Y T T V A Q L C P F K S S F M V S K V K V Q D L D C Y T T V A Q L C P F K S S F M S K S C S M S T S T S S M S T S T S S M S T S T		
CLEC18A VSFVEVVSLWFAEGQRYSHAAGECARNATCTFH CLEC18B ASFVEVVSLWFAEGQRYSHAAGECARNATCTFH CLEC18C ASFVEVVSLWFAEGQRYSHAAGECARNATCTFH CLEC18C ASFVEVVSLWFAEGQRYSHAAGECARNATCTFH CRISP1 VSWSSVIGVWYSESTSFKHGEWTTTDDDITTDHY CRISP3 SSWSQAIQSWFDEYNDFDFGVGPKTPNAVVG-HY CRISP2 TSWSSAIQSWYDEILDFVYGVGPKSPNAVVG-HY P116 MDVPLAMEEWHHEREHYNLSAATCSPGQMCG-HY GLIPR1-ike1 FTPRHAITAWYNETQFYDFDS-LSCSRVCG-HY GLIPR1-ike2 FTASIAITNWYDEIQDYDFKT-RICKKVCG-HY P115 PWYDEVKDYAFPYPQDCNPRCPMRCFGPMCTFHY P176 MDVELWHYLFPAPRDCNPHCPWRCDGPTCS-HY GLIPR1-ike2 FTASIAIRSWHAEKKMYNFEN-GSCSGDCS-NY P115 PWYDEVKDYAFPYPQDCNPRCPMRCFGPMCT-HY P1R3HDML SWSEEKWHYLFPAPRDCNPHCPWRCDGPTCS-HY GAPR LAWAS-SSFMVSKVKVQDLDCYTVAQLCPFHK-PA CRISPLCCL1 YQSANSFTVSKVTVQAVTCETTVEQLCPFHK-PA CRISPLCCL2 YKPSSFMVSKVKVQDLDCYTVAQLCPFEK-PA Exon 2 Exon 3 Exon 4 Exon 5 Exon 6 Ex X 6:		
CLEC18A V S F V E V V S LWF A E GQ R Y S HA A G E C A R N A T C T - H Y CLEC18B A S F V E V V S LWF A E GQ R Y S HA A G E C A R N A T C T - H Y CLEC18C A S F V E V V S LWF A E GQ R Y S HA A G E C A R N A T C T - H Y CRISP1 V S W S S V I G VWY S E S T S F K H G E W T T D D D I T T D H Y CRISP3 S S W S Q A I Q S W F D E YN D F D F G V G P K S P N A V V G - H Y CRISP2 T S W S S A I Q S W Y D E I L D F V Y G V G P K S P N A V V G - H Y P116 MD V P L A M E E W H H E R E H YN L S A A T C S P G Q M C G - H Y GLIPR1-like 1 F T P R H A I T A W Y N E T Q F Y D F D S - L S C S R V C G - H Y GLIPR1 F S V S S A I T N W Y D E I Q D Y D F K T - R I C K K V C G - H Y GLIPR1 F S V S S A I T N W Y D E I Q D Y D F K T - R I C K K V C G - H Y P115 P W Y D E V K D Y A F P Y P Q D C N P R C P M R C F G P M C T - H Y P1 R 3 HDML S W S E E K W H Y L F P A P R D C N P H C P W R C D G P T C S - H GAPR L A WA S Y D Q T G H H CRISP LCCL1 Y Q S A N S F T V S K V T V Q A V T C E T T V E Q L C P F H K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F E K - P A Exon 2 Exon 3 Exon 4 Exon 5 Exon 6 Ex X 6:	CRISF LCCL2	
CLEC18B A S F V E V V S LWF A E GQ R Y S HAA G E C A R N A T C T - H Y CLEC18C A S F V E V V S LWF A E GQ R Y S HAA G E C A R N A T C T - H Y CRISP1 V S W S S V I G VWY S E S T S F K H G E W T T D D D I T T D H Y CRISP3 S S W S Q A I Q S WF D E YN D F D F G V G PK T P N A V V G - H Y CRISP2 T S W S S A I Q S WY D E I L D F V Y G V G PK S P N A V V G - H Y P116 M D V P L AME E WH H E R E H YN L S A A T C S P G Q M C G - H Y GLIPR1-like 1 F T P R HA I T AW YN E T Q F Y D F D S - L S C S R V C G - H Y GLIPR1 - like 2 F T A S I A I R S WH A E KK M YN F E N - G S C S G D C S - N Y P115 P W Y D E V K D Y A F P Y P Q D C N P R C P M R C F G P M C T - H Y P I R 3 HDML S W S E E K WH Y L F P A P R D C N P H C P W R C D G P T C S - H Y GAPR L A WA S Y D Q T G	CI EC19A	
CLEC18C A S F V E V V S LWF A E GQ R Y S HA A G E C ARN A T C T - HY CRISP1 V S W S S V I G VWY S E S T S F K H G E W T T D D D I T T D H Y CRISP3 S S W S Q A I Q S W F D E YN D F D F G V G P K T P N A V V G - H Y CRISP2 T S W S S A I Q S W Y D E I L D F V Y G V G P K S P N A V V G - H Y P116 MD V P L AME E WH H E R E H YN L S A A T C S P G Q M C G - H Y GLIPR1-like 1 F T P R HA I T A W YN E T Q F Y D F D S - L S C S R V C G - H Y GLIPR1 F S V S S A I T NWY D E I Q D Y D F K T - R I C K K V C G - H Y GLIPR1 F S V S S A I T NWY D E I Q D Y D F K T - R I C K K V C G - H Y GLIPR1 F S V S S A I T NWY D E I Q D Y D F K T - R I C K K V C G - H Y GLIPR1 S W S E E K W H Y L F P A P R D C N P R C P M R C F G P M C T - H Y P175 P W Y D E V K D Y A F P Y P Q D C N P R C P M R C F G P M C T - H Y P1 R 3 HDML S W S E E K W H Y L F P A P R D C N P H C P W R C D G P T C S - H Y GAPR L A WA S Y D Q T G H H CRISP LCCL1 Y Q S A N S F T V S K V T V Q A V T C E T T V E Q L C P F H K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F E K - P A Exon 2 Exon 3 Exon 4 Exon 5 Exon 6 Ex X 6:		- A second se
CRISP1 V S W S S V I G VWY S E S T S F K H G EW T T T D D D I T T D H Y CRISP3 S S W S Q A I Q S W F D E YN D F D F G V G P K T P N A V V G - H Y CRISP2 T S W S S A I Q S W Y D E I L D F V Y G V G P K S P N A V V G - H Y P116 MD V P L AME EWH H E R E H YN L S A A T C S P G Q M C G - H Y GLIPR1-like 1 F T P R H A I T A W YN E T Q F Y D F D S - L S C S R V C G - H Y GLIPR1 F S V S S A I T N W Y D E I Q D Y D F K T - R I C K K V C G - H Y GLIPR1-like 2 F T A S I A I R S W H A E K K M Y N F E N - G S C S G D C S - N Y P115 P W Y D E V K D Y A F P Y P Q D C N P R C P M R C F G P M C T - H Y P I R 3 HDML S W S E E K W H Y L F P A P R D C N P H C P W R C D G P T C S - H Y GAPR L A W A S Y D Q T G H H C RISP LCCL1 Y Q S A N S F T V S K V T V Q A V T C E T T V E Q L C P F H K - P Z C RISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F E K - P Z E Exon 2 E xon 10 E xon 11 E xon 13 E xo x 6: Y C 10 Y T V A Q L C P F E K - P Z Y T V A Q L C P F E K - P Z		
CRISP3 S S W S Q A I Q S W F D E YN D F D F G V G P K T P N A V V G - H Y CRISP2 T S W S S A I Q S W Y D E I L D F V Y G V G P K S P N A V V G - H Y P116 MD V P L AME E WH H E R E H YN L S A A T S P G Q M C G - H Y GLIPR1-like 1 F T P R H A I T A W YN E T Q F Y D F D S - L S S R V C G - H Y GLIPR1 F S V S S A I T N W Y D E I Q D Y D F K T - R I C K K V C G - H Y GLIPR1 F S V S S A I T N W Y D E I Q D Y D F K T - R I C K K V C G - H Y GLIPR1-like 2 F T A S I A I R S W H A E K K M Y N F E N - G S C S G D C S - N Y P115 P W Y D E V K D Y A F P Y P Q D C N P R C P M R C F G P M C T - H Y P I R 3 HDML S W S E E K W H Y L F P A P R D C N P H C P W R C D G P T C S - H Y GAPR L A W A S Y D Q T G		
CRISP2 TSWSSAIQSWYDEILDFVYGVGPKSPNAVVG-HYP16 MDVPLAMEEWHHEREHYNLSAATCSPGQMCG-HYG16 GLIPR1-like1 FTPRHAITAWYNETQFYDFDS-LSCSRVCG-HYG11PR1 GLIPR1 FSVSSAITNWYDEIQDYDFKTRICKKVCG-HYG11PR1 GLIPR1-like2 FTASIAIRSWHAEKKMYNFEN-GSCSGDCS-NY P115 PWYDEVKDYAFPYPQDCNPRCPMRCFGPMCT-HYG12 GAPR LAWASYDQTG		
P116 MD V P L AME EWHHE R EH YN L S AA T C S P G QM C G - H Y GLIPR1-like 1 F T P R HA I T AWYN E T Q F YD F D S - L S C S R V C G - H Y GLIPR1 F S V S S A I T NWYDE I Q D YD F K T - R I C K K V C G - H Y GLIPR1-like 2 F T A S I A I R S WHA E K K M YN F EN - G S C S G D C S - N Y P115 P W YD E V K D Y A F P Y P Q D C N P R C PMR C F G PM C T - H Y PI R3HDML S W S E E K WH Y L F P A P R D C N P H C PWR C D G P T C S - H Y GAPR L A WA S Y D Q T G		
GLIPR1-ike 1 F T P R HA I T AWYN E T Q F Y D F D S - L S C S R V C G - H Y GLIPR1 F S V S S A I T NWYD E I Q D Y D F K T - R I C K K V C G - H Y GLIPR1-ike 2 F T A S I A I R SWHA E K K M Y N F E N - G S C S G D C S - N Y PI15 PWYD E V K D Y A F P Y P Q D C N P R C PMR C F G P M C T - H Y PI R 3HDML S W S E E K W H Y L F P A P R D C N P H C P W R C D G P T C S - H Y GAPR L A WA S Y D Q T G		
GLIPR1 F S V S S A I T NWY D E I Q D Y D F K T R I C KK V C G - H Y GLIPR1-like 2 F T A S I A I R SWHA EKKMYN F EN - G S C S GD C S - N Y PI15 PWY D E V K D Y A F P Y P Q D C N P R C PMR C F G P M C T - H Y PIR3HDML S W S E E K WH Y L F P A P R D C N P H C PWR C D G P T C S - H Y GAPR L A WA S Y D Q T G		
GLIPR1-like 2 F T A S I A I R SWHAEKKMYN F EN - G S C S GD C S - N Y PI15 PWYD E VK D Y A F P Y PQ D C N P R C PMR C F G PM C T - H Y PI R3HDML S W S E E K WH Y L F P A P R D C N P H C PWR C D G P T C S - H Y GAPR LAWA S Y DQ T G H H CRISP LCCL1 Y Q S A N S F T V S K V T V Q A V T C E T T V E Q L C P F H K - P A CRISP LCCL2 Y K P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F E K - P A Exon 2 Exon 3 Exon 4 Exon 5 Exon 6 Exo K 6: Exon 10 Exon 11 Exon 12 Exon 13 Exo		
PI15 PWYDEVKDYAFPYPQDCNPRCPMRCFGPMCTHY PIR3HDML SWSEEKWHYLFPAPRDCNPHCPWRCDGPTCSHY GAPR LAWAS YDQTG CRISP LCCL1 YQSANSFTVSKVTVQAVTCETTVEQLCPFHK PICRISP LCCL2 YKPSSSFMVSKVKVQDLDCYTTVAQLCPFEK Exon 2 Exon 3 Exon 4 Exon 9 Exon 10 Exon 11 Exon 12 Exon 13 Exon 13 Exo x 6: Y Y Y		
PIR3HDML SWSEEKWHYLFPAPRDCNPHCPWRCDGPTCS-H GAPR LAWASYDQTG		
GAPR LAWAS YDQTG		
CRISP LCCL1 YQ SAN SFTVSKVTVQAVTCETTVEQLCPFHK-P/CRISP LCCL2 YK PSSSFMVSKVKVQDLDCYTTVAQLCPFEK-P/ Exon 2 Exon 3 Exon 2 Exon 3 Exon 4 Exon 5 Exon 9 Exon 10 Exon 11 Exon 12 Exon 13 Exo X 6:		
CRISP LCCL2 YK P S S S F M V S K V K V Q D L D C Y T T V A Q L C P F E K - P A Exon 2 Exon 3 Exon 4 Exon 5 Exon 6 Exo Exon 2 Exon 3 Exon 4 Exon 5 Exon 6 Exo Exon 9 Exon 10 Exon 11 Exon 12 Exon 13 Exo X 6: Exon 10 Exon 11 Exon 12 Exon 13 Exo		
Exon 2 Exon 3 Exon 4 Exon 5 Exon 6 Exo Exon 9 Exon 10 Exon 11 Exon 12 Exon 13 Exo X 6: Exon 10 Exon 11 Exon 12 Exon 13 Exo		
Exon 9 Exon 10 Exon 11 Exon 12 Exon 13 Exo x 6:	CRISP LCCL2	<u>YKPSSSFM</u> VSKVKVQDLD <mark>C</mark> YTTVAQL <mark>C</mark> PFE <mark>K - P</mark> A
х б:	Exon 2	Exon 3 Exon 4 Exon 5 Exon 6 Exo
		Exon 10 Exon 11 Exon 12 Exon 13 Exo
f 8	x 6:	
	f 8	

Appendix 6: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for CAP Superfamily Genes in the *Homo sapiens* Genome

	CAP1 C
CLEC18A	TQ L VWA T S S Q L G C G R H L C S A G Q A A I E A
	TQ LVWA TS SQ LGCGRHLC SAGQ T A I EA
CLEC18C	TQ L V W A T S S Q L G C G R H L C S A G Q A A I E A
	TQ I VWA TS YL I GCA I A SCRQQGS PR YL
CRISP3	TQ VVWYSSYLVG <mark>C</mark> GNAY <mark>C</mark> PNQKVLKYY
	TQ LVWYSTYQVGCGIAYCPNQDS LKYY
	TQ VVWAKTER I GCGSHFCEKLQG VEETNIEL
	TOLVWANSFYVGCAVAMCPNLGGASTAI
	TQ VVWADS YK VG <mark>C</mark> A VQ F <mark>C</mark> PK V S G F - DA L S N GAH
	IQLVWDHSYKVG <mark>C</mark> AVTP <mark>C</mark> SKIGHIIHAAI
	TOMVWATSNRIGCAIHTCONMNVWGSVWRRAVY
	TOMVWASSNRLGCAIHTCSSISVWGNTWHRAAY
	TAMVWKN TKKMG V GKA SA SDG SS F
	<mark>S H C</mark> P R V Y <mark>C</mark> P R N - <mark>C</mark> MQ A N P H Y A R V I G T R V Y S D <mark>L S</mark>
CRISP LCCL2	THCPRIHCPAH - CKDEPSYWAPVFGTNIYAD TS
	CAP2
CLEC18A	V <mark>C</mark> AYSPRGNWEVNGKT IVPYKKGAWCSLCTA
CLEC18B	V <mark>C</mark> AYSPGGNWEVNGKT I I PYKKGAWCSLCTA
CLEC18C	V <mark>C</mark> AYSPRGNWEVNGKT IVPYKKGAWCSLCTA
CRISP1	VCHYCHEGNDPETKNEPYKTGVPCEACP-
CRISP3	V <mark>C</mark> Q Y <mark>C</mark> P A G N W A N R L Y V P Y E Q G A P <mark>C</mark> A S <mark>C</mark> P -
CRISP2	V C Q Y C P A G N N M N R K N T P Y Q Q G T P C A G C P -
PI16	VCNYEP PGNVKGKR PYQEGTPCSQCPS
GLIPR1-like 1	VCNYGPAGNFANMPPYVRGESCSLCSK
GLIPR1	I CNYGPGGNYPTWPYKRGATCSACPN
GLIPR1-like 2	ICNYAPGGTLTRRPYEPGIFCTRCGR
	VCNYAPKGNWIGEAPYKVGVPCSSCPP
	VCNYAIKGNWIGESPYKMGKPCSSCPP
	VARYFPAGNVVNEGFFEENVLPPKK-
	ICRAAVHAGVVRN HGGYVDVMPVDKRK
CRISP LCCL2	I CK TAVHAGV I SNE SGGDVDVMPVDKKK
	<mark>V S GC F K AWDH A G G L C</mark> E V P R N P <mark>C</mark> R M S <mark>C</mark> Q N H G R L N
	<mark>V S GC FK AWDHAG GLC</mark> E V PRN P <mark>C</mark> RM S <mark>C</mark> QN H GR LN
CLEC18C	<mark>V S G</mark> CFK AWDHAG GLCEV PRN PCRM SCQNHGR LN
CRISP1	SN CEDK LC
CRISP3	DN
CRISP2	DD <mark>C</mark> DK G L <mark>C</mark>
PI16	YH
GLIPR1-like 1	EK
GLIPR1	DK
GLIPR1-like 2	DK
PI15	
PI R3HDML	
GAPR	
CRISP LCCL1	
CRISP LCCL2	
Exon 2	Exon 3 Exon 4 Exon 5 Exon 6 Exo
Exon 9	Exon 10 Exon 11 Exon 12 Exon 13 Exo

CLEC18A	STCHCHCPPGYTGRYCQVRCSLQCVHGRF
CLEC18B	S TCHCHCPPGYTGRYCQVRCSLQCVHGRF H
CLEC18C	STCHCHCPPGYTGRYCQVRCSLQCVHGRF
CRISP1	NPCIYYDEYFDCDIOVHYLGCNH
CRISP3	NGCKYEDLYSNCKSLKLTLTCKHC
CRISP2	NSCOYODLLSNCDSLKNTAGCEHI
PI16	PIGSPEDAODLPYLVTEAPSFRATEASDSRKMG
GLIPR1-like 1	TPQLIIPN-QNPFLKPTGRAPQQTAFNPF
	NRQRDQVKRYYSVVYPGWPIYPRNRY
	NADRDQAT - YYR FWYPK WEMPR PVVCDPL C
PI15	
PI R3HDML	
GAPR	
CRISP LCCL1	TYIASFQNGIFSESLQNPPG 0
CRISP LCCL2	TYVGSLRNGVQSESLGTPRD (
CLEC18A	E E E C S C V C D I G Y G G A Q C A T K V H F P F H T C D L R I D G
CLEC18B	E E E <mark>C</mark> S <mark>C</mark> V <mark>C</mark> D I G Y G G A Q <mark>C</mark> A T K V H F P F H T <mark>C</mark> D L R I D (
CLEC18C	E E E <mark>C</mark> S <mark>C</mark> V <mark>C</mark> D I G Y G G A Q <mark>C</mark> A T K V H F P F H T <mark>C D L R I D G</mark>
CRISP1	TTILF <mark>C</mark> KAT <mark>C</mark> L <mark>C</mark> DTEIK
CRISP3	L V R D S <mark>C</mark> K A S <mark>C</mark> N <mark>C</mark> S N S I Y
CRISP2	LLKEK <mark>C</mark> KAT <mark>C</mark> LCENKIY
PI16	P S S L A T G I P A F L V T E V S G S L A T K A L P A V E T Q A P T
GLIPR1-like 1	LGFLLLRIF
GLIPR1	SLFLIVNSVILILSVIITILVQHKYPNLVLLD -
GLIPR1-like 2	TFILLRILCFILCVITVLIVQSQFPNILLEQQM
PI15	NYLYWFK
PI R3HDML	NK F TWF
GAPR	
	K A F R V F A V V
	KAFRIFAVRO
CLEC18A	D C F M V S S E A D T Y Y R A R M K C Q R K G G V L A Q I K S Q K V
CLEC18B	DCFMVSSEADTYYRARMKCQRKGGVLAQIKSQK
CLEC18C	DC FMV S S E A D T Y YR ARMK <mark>C</mark> Q RKG G V LAQ I K S Q K
CRISP1	DC1MV55EADT1TRAKMRCQRR66VEAQTR5QR
CRISP1 CRISP3	
CRISP2	CLATKDDDCMATEADDCUTTEVDCLLAAUCLDC1
PI16	S L A TK D P P S MA T E A P P C V T T E V P S I L A A H S L P S I
GLIPR1-like 1	
GLIPR1	
GLIPR1-like 2	I F T P E E S E A GN E E E E K E E E K K E K E E MEME I MEMI
PI15	
PI R3HDML	
GAPR	
CRISP LCCL1	
CDISD I CCI 2	
CRISF LCCL2	
	Exon 3 Exon 4 Exon 5 Exon 6 Exo
Exon 2	
Exon 2	Exon 3 Exon 4 Exon 5 Exon 6 Exo Exon 10 Exon 11 Exon 12 Exon 13 Exo

Appendix 6 Page 6 of 8

CLECICA	0	F	т	T			V. 1		1 72	T	P	T	T	NT T	7 7	7.7	P	C	P	12		T		т.т		1.7	0	T	T	37	V
CLEC18A					A																										
CLEC18B					A																										
CLEC18C	~				A							_								-							-				_
CRISP1	-	-	-	-	-		_	_		-		-		_	_	_	-	-					_	_	_	-	-	-	-	-	-
CRISP3	-	-	-		-	-	_	-	-	-			-	_	-	-	-					-	_	_	-	-	-	-			
CRISP2	-			_	-	_	_			_	_		_	_	_			_		_	_	_	_	_	_					_	_
PI16	D	E	E	Ρ	V	T I	FΙ	P K	S	Т	Η	V	Р	IJ	P K	l S	А	D	Κ	V	ΤJ	DI	K 1	ΓK	L V	' P	' S	R	S	Ρ	Е
GLIPR1-like 1	-	-	-	-	-	-		• -	-	-	-	-	-		• -	-	-	-	-	-	-	-			-	-	-	-	-	-	-
GLIPR1	-	-	-	-			- -							_							_									-	-
GLIPR1-like 2	E	E	Κ	E	E	R I	ΕI	ΞE	ΕE	E	E	Т	QI	K I	ΞK	M	ſΕ	Е	Е	ΕJ	K	-	- -		-	-	-	-	-	-	-
PI15	-	-	-	-	-	-			-	-	-	-	-			-	-	-	-	-	-	-	- -		-	-	-	-	-	-	-
PI R3HDML	-	-	-	-	-	-			-	-	-	-	-			-	-	-	-	-	-				-	-	-	-	-	-	-
GAPR	-	-	-	-	-	-			-	-	-	-	-			-	-	-	-	-	-				-	-	-	-	-	-	-
CRISP LCCL1	-	-	-	-	-				-	-	-	-	-			-	-	-	-	-					-	-	-	-	-	-	-
CRISP LCCL2	-	-	-	-	-				-	-	-	-	-			-	-	-	-	-					-	-	-	-	-	-	-
														+							1										
CLEC18A	Т	A	K	D	S	FI	RV	V A	Т	G	Е	н	Q.	A I	7 1	r s	F	A	F	G	Q	P 1	וס	I F	I -	-	-	-	-	-	-
CLEC18B	-				s																					_					
CLEC18C					s																										
CRISP1		_	_	_	-	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
CRISP3		-	-	-	-	-			_												_			_			-				
CRISP2	-	-	-	-	-	-	_	_	-	-		_	_	_	_	_	-	_			_	-	_	_	-	-	-	-	-	<u> </u>	
PI16					- P:																										
GLIPR1-like 1	_				- F													~													
GLIPR1-IKE I GLIPR1		-			-		_	_	_			_		_	_	_					_	_	_	_	_	-	-				
	-	-	-	-	$ \rightarrow $	-	_	_	-	-		_	-	_		_	-	-				-	_	_	-	-	-	-	-	-	
GLIPR1-like 2	-	-	-	-	-	-			_					_							_			_			-				-
PI15	-	-	-	-	-		_	_	-	-		_			_	-	-	-				-	_	-	-	-	-	-	-	-	-
PI R3HDML	-	-	-	-	-	-			-	-		_	-			-	-				-	-	-		-	-	-		-	-	-
GAPR	-	-	-	-	-	-	-	• -	-	-		_	-		-	-	-				-	-		• -	-	-	-	-	-	-	-
CRISP LCCL1	-	-	-	-	-	-		-	-	-		_	-	- -		-	-			-	-	-	- •	• -	-	-	-	-	-	-	-
CRISP LCCL2	-	-	-	-	-	-		• •	-	-	-	-	-	- -	• •	-	-	-	-	-	-	-	- •	• •	-	-	-	-	-	-	-
CLEC18A	-	-	G	F	G	N	C 1	νE	ΕL	Q	A	S	A	AI	7 N	W	/N	D	Q	R	<mark>C</mark> I	K '	ΤF	٢N	I R	Y	ΓI	С	Q	F	A
CLEC18B	A	M	G	F	G	N	C V	νE	ΕL	Q	А	s	A	ΑI	7 N	W	'N	D	Q	R	C	K '	ΤF	١N	I R	Y	Ί	С	Q	F	A
CLEC18C					G]																										
CRISP1					-																										
CRISP3	-	-	-	-	-	- 1			-	-	-	-	-			-	-	-	-	-					-	-	-	-	-	-	-
CRISP2	-	-	-		-	-	-	-	-	-		_	-		-	-	-				-	-	-	-	-	-	-				
PI16	-				V								_	_						_	_									_	
GLIPR1-like 1	-	-	-	-	-	-	- -	. .	-	-	-	-	-	. .		-	-	-	-	-	- -		- -		-	-	-	-	-	-	-
GLIPR1	-	-	-	-	-				-	-	-	-	- 1			-	-	-	-	-	_				-	-	-	-	-	-	-
GLIPR1-like 2	-	-	-	-	-	_			-	-	-	-	-			-	-	-	-	-	_	_	_		-	-	-	-	-	-	-
PI15	-	-	-	-	-				-	-			-				-	-		-		_			-	-	-	-	-	-	-
PI R3HDML	-	-	-		-				-	-		_	-		-	-	-	-		-	-			+	-	-	1	-	-	-	-
	-	-	-	-	-	-		-	-	-		-	-		-	-	-	-				-			-	-	-	-	-	Ē	-
GAPR CRISP LCCL1	-	-	-	-	-	-			-	-			-	-		-	-	-			-	-			-	-	-	-	-	-	-
	-	-	-	-	-	-			-	-		-	-		-	-	-			-	-	-			-	-	-	-	-	-	-
CRISP LCCL2	-	-	-	-	-	-	- -		-	-	-	-	-	- -	• •	-	-	-	-	-	-	- -	- -	- -	-	-	-	-	-	-	-
Exon 2				E	xo	n 3	3			E	Cxo	n	4]	Ex	on	5				E	xo	n (5				Ex	(01
Exon 9			1	F v	~ n	1/			-	г		. 1	1		_	F				_		•	F	~ n	1	2		_	F	`v/	

Appendix 6 Page 7 of 8

CLEC18B	Q	Ε	Η	Ι	S	R	W	G	ΡQ	G.	s.	•		-	-	-	-	-	-	-	-		•		-	-	-	-	-	-	-	-
CLEC18C	Q	E	Η	Ι	S	R	W	G	P (G	s.	- -	- -	-	-	-	-	-	-	-	-	- -	•	• -	-	-	-	-	-	-	-	-
CRISP1	-	-	-	-	-	-	-	-	-	-		- -	- -	-	-	-	-	-	-	-	-	- -	•	• -	-	-	-	-	-	-	-	-
CRISP3	-	-	-	-	-	-	-	-	-	-		- -	- -	-	-	-	-	-	-	-	-	- -	•		-	-	-	-	-	-	-	-
CRISP2	-	-	-	-	-	-	-	-		-		- -	- -	-	-	-	-	-	-	-	-	- -	• •		-	-	-	-	-	-	-	-
PI16	N	Т	S	A	Т	A	N	A	T (G	GI	R /	ΑI	, A	L	Q	S	S	L	Ρ	G.	A 1	E (3 P	D	K	Ρ	s	V	V	s	G
GLIPR1-like 1	-	-	-	-	-	-	-	-	-	-		• •	- -	-	-	-	-	-	-	-	-	- -	•		-	-	-	-	-	-	-	-
GLIPR1	-	-	-	-	-	-	-	-		-				-	-	-	-	-	-	-	-				-	-	-	-	-	-	-	-
GLIPR1-like 2	-	-	-	-	-	-	-	-		-		- -		-	-	-	-	-	-	-	-		•		-	-	-	-	-	-	-	-
PI15	-	-	-	-	-	-	-	-		-		•	- -	-	-	-	-	-	-	-	-		•		-	-	-	-	-	-	-	-
PI R3HDML	-	-	-	-	-	-	-	-	-	-		•		-	-	-	-	-	-	-	-	- -	•		-	-	-	-	-	-	-	-
GAPR	-	-	-	-	-	-	-	-	-	-		• •	- -	-	-	-	-	-	-	-	-	- -	•		-	-	-	-	-	-	-	-
CRISP LCCL1	-	-	-	-	-	-	-	-		-		• •		-	-	-	-	-	-	-	-		•		-	-	-	-	-	-	-	-
CRISP LCCL2	-	-	-	-	-	-	-	-	-	•		•	• •	-	-	-	-	-	-	-	-		•	• •	-	-	-	-	-	-	-	-
CLEC18A	_	_	_	_	_	_	_	_	_	_	+				-	_	_	_	_	_	_	_			_	-					_	
CLEC18R CLEC18B	-	-	-	-	-	-	-	-							-	-	-	-	-	-	-	-	+		-	-						
CLEC18D	-	-	-	-	-	-	-	-							-	-	-	-	-	-	-				-	-	-	-	-		-	
CRISP1	-	-	-	-	-	-	-	-				-		-	-	-	-	-	-	-	-				-	-						
CRISP3	-	_	-	_	-	_	-	-	-	+		+		+	+-	-	-	-	-	_	_	_			-	-			-		-	
CRISP2		-	-	-	-	-	-		-	-		-		-	-	-	-	-	-	-	-				-	-						
PI16	L	N	s																					ΞI								
GLIPR1-like 1	-	_	-		-		-			-					-	-	-	-	-	-	_				-							
GLIPR1	-	-	-	-	-	-	-	-							-	-	-	-	-	-	-				-	1	-					
GLIPR1-like 2	-	-	-	-	-	-	-	-							-	-	-	-	-	-	-				-	-						
PI15	-	-	-	-	-	-	-	-							-	-	-	-	-	-	-				-							
PI R3HDML	-	-	-	-	-	-	-	-		-					-	-	-	-	-	-	-				-	1						
GAPR	-	-	-	-	-	-	-	-		-					-	-	-	-	-	-	-				-	t						
CRISP LCCL1	-	-	-	-	-	-	-	-		-					-	-	-	-	-	-	-				-	t						
CRISP LCCL2	-	-	-	-	-	-	-	-		-					-	-	-	-	-	-	-				-	1						

Appendix 6: This figure provides MEGA7 ClustalW amino acid sequence alignment of all CAP superfamily proteins in *Homo sapiens* genome. The CAP signature sequences are boxed, the highly conserved cysteine are highlighted in yellow with a red font color. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online data for specific CAP genes. Color-coding superimposed on the sequences is used to represent different exons and exon borders are indicated by a change in color. Organisms and accession IDs of the amino acids are available in the supplemental data table 12.

Exon 8 Exon 9 Exon 10 Exon 11 Exon 12 Exon 13 Exon 14

Page 8 of 8

Appendix 7: Pairwise Distance of *Homo sapiens* CAP Amino Acid Sequences – Estimated Evolutionary Divergence Between Sequences

	Homo sapiens	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	protein name	-	~	č	•	5	v	,	Ŭ	-	10		12	10		1.
1	CLEC18A															
2	CLEC18B	0.026														
3	CLEC18C	0.017	0.009													
4	CRISP1	1.107	1.133	1.133												
5	CRISP3	1.057	1.057	1.057	0.676											
6	CRISP2	1.133	1.133	1.133	0.660	0.293										
7	PI16	0.964	0.964	0.964	1.082	0.986	0.942									
8	GliPR1-like 1	0.986	0.986	0.986	1.082	0.879	0.942	0.964								
9	GLPR1	1.009	0.986	1.009	0.921	0.964	0.942	0.879	0.676							
10	GliPR1-like 2	1.215	1.187	1.215	1.057	1.057	1.033	1.160	0.859	0.728						
11	PI15	0.964	0.964	0.964	1.033	1.009	1.082	0.942	0.899	1.033	1.215					
12	PI R3HDML	0.986	0.986	0.986	1.187	0.986	1.160	0.986	0.986	1.057	1.160	0.340				
13	GliPR2	1.680	1.680	1.680	1.513	1.305	1.244	1.513	1.438	1.552	1.775	1.552	1.593			
14	CRISP LCCL1	2.573	2.573	2.573	2.573	2.573	2.468	2.468	2.979	2.573	2.286	2.063	2.286	2.206		
15	CRISP LCCL2	2.691	2.691	2.691	2.206	2.691	2.573	2.825	2.573	2.132	2.132	2.132	2.573	2.468	0.596	
16	RhAG	2.206	2.132	2.132	2.573	2.206	2.286	2.573	2.373	2.286	2.373	2.373	2.063	2.468	N/A	3.672

A. Whole amino acid sequence pairwise distance

B. CAP/PR domain amino acid sequence pairwise distance

	<i>Homo sapiens</i> protein name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	CLEC18A															
2	CLEC18B	0.000														
3	CLEC18C	0.000	0.000													
4	CRISP1	1.253	1.253	1.253												
5	CRISP3	1.099	1.099	1.099	0.560											
6	CRISP2	1.099	1.099	1.099	0.560	0.154										
7	PI16	0.965	0.965	0.965	1.099	0.742	0.742									
8	GLIPR1-Like 1	0.965	0.965	0.965	1.253	0.742	0.847	0.965								
9	GLPR1	1.099	1.099	1.099	0.965	0.965	0.847	0.847	0.847							
10	GliPR1-like 2	1.435	1.435	1.435	0.847	0.965	0.742	1.099	1.099	0.965						
11	PI15	0.742	0.742	0.742	1.099	0.847	0.847	1.253	0.647	1.099	1.253					
12	PI R3HDML	0.742	0.742	0.742	1.253	0.847	0.965	0.965	0.742	1.099	1.253	0.405				
13	GLPR2	1.658	1.658	1.658	1.946	1.658	1.658	2.351	1.435	1.946	1.946	1.253	1.435			
14	CRISP LCCL1	1.946	1.946	1.946	3.045	2.351	2.351	2.351	2.351	2.351	1.099	1.946	1.946	1.946		
15	CRISP LCCL2	0.847	0.847	0.847	0.847	0.742	0.742	0.647	0.847	0.847	0.847	0.742	0.847	1.435	1.658	
16	RhAG	1.946	1.946	1.946	3.045	2.351	1.946	3.045	1.946	3.045	1.946	1.946	1.946	2.351	1.253	2.351

Appendix 7: The follow tables show pairwise distance alignment of *Homo sapiens* CAP sequences. 'A' represents the pairwise alignment score for whole amino acid sequence. 'B' represents the respective protein CAP/PR amino acid alignment score. The number of amino acid substitutions per site from between sequences are shown. Analyses were conducted using the Poisson correction model. The analysis involved 16 amino acid sequences. All positions containing gaps and missing data were eliminated.

Common name - Protein name											A	mi	10	Ac	id S	See	que	enc	e				_							
Tomato - Pr-4r	-	_	_	_	-	-	-		_	_		-			-	-	-			_		-	-	-					_	
Cocoa tree - CAP-related	-	-	-				-	-	_	-	-			-		-	-			_	-	_	-	-	-	-	-	-	-	-
Yeast - PRY1	-	-	-				-				_			-		-	-			-	-		-	-	-	-	-	-	-	-
Mold - SCP-like			-			_			-	_	-			-		-					-	_	-	-	-				-	
Round worm - Unchar.			-		_	_				_						-	-				_	_	_	-	-					
		-	-	-	-	_	-		_	_	_			-		-	-			-	-	_	-	-	-				-	
Fruit fly - Unchar.			-	-		_	-		_	_	_			-	_	-	-			-	-	-	-	-	-	-	-	-	-	-
Parasitic wasp - EFG1			-						_					_					-	-	-				-	-	-	-	-	-
Jewel wasp - Unchar.															H S															
Red fire ants - PRY1-like											_			-	GΝ				_	-										
Dwarf honey bee - EFG1		-	-		-	_	-		_	_	_			-		-	-			-	-	_	-	-	-					
Leaf cutter bee - EFG1			-		_	_	_		_	_	_					-	-			_	_	_	-	-	-					
Sea squirt (C. i.) - GAPR		-	-			-	-				_			-		-	-			-	-		-	-	-				-	
Spotted gar - GAPR	-	-	-			_				_	-			-		-	-				-	_	-	-	-				-	
Zebrafish - GAPR	-	-	-	-		_	-		_	_	_			-		-	-			_	_	_	-	-	-				-	
W. clawed frog - GAPR	-	-	-				-	-				-	-	-		-	-	-	-	-			-	-	-	-	-	-	-	-
Burmese python - GAPR	-	-	-				-	-				-	-	-		-	-	-	-	-			-	-	-	-	-	-	-	-
Anole lizard - GAPR	-	-	-		- -	- -	-	-			- -	-	-	-	- -	-	-	-	-	-	- -		-	-	-	-	-	-	-	-
Chinese soft-shelled turtle - GAPR	-	-	-				-	-				-	-	-		-	-	-	-	-			-	-	-	-	-	-	-	-
Falcon - GAPR	-	-	-				-	-				-	-	-		-	-	-	-	-			-	-	-	-	-	-	-	-
Chicken - GAPR	-	-	-				-	-				-	-	-		-	-	-	-				-	-	-	-	-	-	-	-
Platypus - GAPR	-	-	-				-	-				-	-			-	-	-	-				-	-	-	-	-	-	-	-
Amadillo - GAPR	-	-	-	-		_	_		_	_	_			-		-	-			_	-	_	-	-	-				-	
Rat - GAPR	-	-	-			_				_	_			-		-	-				-	_	-	-	-	-	-	-	-	-
Mouse - GAPR	-	_	_	_		_				_	_			-		-						_	-	-	-	-	_	_	_	_
Chimp - GAPR		-	-	-	-	_	-		_	_	_			-		-	-			-	-	_	-	-	-					
Human - GAPR		-	-	_	-	_	-		_	_	_			-		-	-			_	-	_	-	-	-				-	
	-	-	-	-			-	-	-		-	-	-	-		-	-	-	-	-	-		+	-	-	-	-	-	-	-
Tomato - Pr-4r	-	-	_				-	-				-	-	-		-	-	-	-	-				_	-	-	-	-	-	-
Cocoa tree - CAP-related	-			_																			-		-					
Yeast - PRY1	-																													
Mold - SCP-like	-																													
															- -															
Round worm - Unchar.	_		-			_			_	_	_				- -	-	-				-	_	-	-	-				-	
Fruit fly - Unchar.				_																										
Parasitic wasp - EFG1																														
Jewel wasp - Unchar.	-							_							ΡI															
Red fire ants - PRY1-like	_	-	-						_	_	_	_	_	_	ΡI		_	_	_	_										_
Dwarf honey bee - EFG1				_																										
Leaf cutter bee - EFG1	-																													
Sea squirt (C. i.) - GAPR																														
Spotted gar - GAPR	-	-	-				-	-				-	-	-		-	-	-	-	-			-	-	-	-	-	-	-	-
Zebrafish - GAPR	-	-	-		- -		-	-				-	-	-	- -	-	-	-	-	-			-	-	-	-	-	-	-	-
W. clawed frog - GAPR	-	-	-				-	-				-	-	-		-	-	-	-	-			-	-	-	-	-	-	M	R
Burmese python - GAPR	-	-	-				-	-				-	-	-	- -	-	-	-	-	-			-	-	-	-	-	-	-	-
Anole lizard - GAPR	-	-	-				-	-				-	-	-		-	-	-	-				-	-	-	-	-	-	-	-
Chinese soft-shelled turtle - GAPR	-	-	-				-	-				-	-	-		-	-	-	-				-	-	-	-	-	-	-	-
Falcon - GAPR	-	-	-																											
Chicken - GAPR	-	-	-			-	-	-			_		-	-	_	-	-	-			-		-	-	-	-	-	-	-	-
Platypus - GAPR	-	_	_	_			-	-	-	-	-			-		-	-			-	-		-	-	-	-	-		_	_
Amadillo - GAPR	-	_	_	_			-	-	_	_		_	_	_	_		-	_		-	_				-	-	_	_	_	_
Rat - GAPR	-			1	+	1	-		1		-	-	-			-	-	-	-				t	-	-	F	-			
	-	-	-	-			-	-	-		_	-			_	-	-	-	-	-	-		-	-	-	-	-	-	-	-
Mouse - GAPR	-	-	-	-+-	-+-	- -	-	-	-			-		-		-	-	-	-	-	-+-		-	-	-	-	-	-	-	-
Chimp - GAPR	-	-	-	-	- -	- -	-		-		-			-		-	-		-	-		-	-	-	-		-	-	-	-
Human - GAPR	-	-	-		- -	- -	-	-		- -		-	-	-	- -	-	-	-	-	-	- -		-	-	-	-	-	-	-	-
on 1 Exon 2 Exo	Exe	n	3				ł	Exe	on	4				F	Exc	n	5				ł	Ex	on	16					E	xo
on 8 🗾 Exon 9 🗔 E ndix 8:	XO	n 1	0				E	xo	n 1	11				E	801	n 1	12				E	xo	n	13				ł	Ξx	01

Tomato - Pr-4r	_	_	_	-		_	_	-		_	_	-		_	_	-		_	_	-					-	-	-
Cocoa tree - CAP-related													-										- -		-	-	-
Yeast - PRY1	ΑI																										
Mold - SCP-like	G A																								Ι	Т	A I
Round worm - Unchar.													-												-	-	
Fruit fly - Unchar.			-	-	-			-	-	-		-	-			-	-			-	-	-		-	-	-	
Parasitic wasp - EFG1	NC	Q F	, Г	S	R	V	٧N	1 V	R	K '	ΓD	Q	R	ΤI	FS	K	G	R I	R C	G	Е	P]	LΙ	Ε	T	V	ΤF
Jewel wasp - Unchar.	NC	2 I	, L	S	R	V	٧N	1 V	R	K '	ΤD	Q	R	ΤI	FS	K	G	R I	r e	i G	E	P	LΙ	E	T	V	ΤF
Red fire ants - PRY1-like	NC	Q F	, L	, S	R	V٦	VΜ	1 V	R	Κ	ΤD	Q	R	ΤI	FS	K	G	R I	r e	i G	E	P [LΙ	Ε	T	V	ΤF
Dwarf honey bee - EFG1	NQ	Q F	۲ ر	S	R	V	VΝ	1 V	R	Κ	ΓD	Q	R	ΤI	FS	K	G	R I	R G	i G	Е	P I	LΙ	E	Т	V	ΤF
Leaf cutter bee - EFG1	NQ	Q I	۰L	S	R	٧V	νN	1 V	R	Κ	ΓD	Q	R	ΤI	FS	K	G	R I	R e	i G	Е	P I	LΙ	E	Т	V	ΤF
Sea squirt (C. i.) - GAPR		- -	-	-	-			-	-	- 1		-	-			-	-			-	-	-	- -		-	-	
Spotted gar - GAPR			-	-	-			-	-	-		-	-			-	-			-	-	-			-	-	
Zebrafish - GAPR			-	-	-			-	-			-	-			-	-			-	-				-	-	
W. clawed frog - GAPR	DA	A S	G	A	Е	٧V	VΕ	T	W	YI	N E	I	ΚI	D	Y S	F	G	R I	PG	F	Q	S I	D 1	G	H	F	ΤC
Burmese python - GAPR			-	-	-			-	-	- 1		-	-			-	-		- -	-	-				-	-	
Anole lizard - GAPR			-	-	-			-	-			-	-			-	-			-	-				-	-	
Chinese soft-shelled turtle - GAPR			-	-	-			-	-	- 1		-	-			-	-			-	-				-	-	
Falcon - GAPR			-			_			-	_			-				_	_				-				-	
Chicken - GAPR		t.	-	-		-	_	-		-	_	-	-	_	_	-			_	-		-		-	-	H	1
Platypus - GAPR		_	_					-				-	-											-	-	Ē	
Armadillo - GAPR			-	-		_	_	-		-	_	-	-	_	_	-	$ \rightarrow $		_	-			_		-	-	
Rat - GAPR		_	_	-		_	_	-		_	_		-	_	_	-			_	-				-	-	-	
		_	_			_	_	_		_	_		-	_	_	-		_	_	-		_			-	-	
Mouse - GAPR		_	_	-		_	_			_	_			_	_			_	_	-		_	_	_	_		_
Chimp - GAPR		_	_	-		_	_	-		-	_		-	_	_	-		_	_	-			_	_	-		_
Human - GAPR			-	-	-			-	-	-		-	-			-	-			-	-	-		• •	-	-	
		+	-	-		_	_	-		_	_	-		_	_	-		_	_	-		_	-	-			_
Tomato - Pr-4r		-	-	-		-	_	-		-	_	-	-		_	-		-		-		-		• -	-	-	
Cocoa tree - CAP-related													-														
Yeast - PRY1	ΤI																										
Mold - SCP-like	VI	9 F	P A	A	T	T S	s s	S	Т	Y	V P	V	Q	E I	P V	A	S	V I	E P	A	P	A	D I	/ E	Q	S	S S
Round worm - Unchar.			-	-	-			-	-	-		-	-			-	-	-		-	-	-		-	-	-	
Fruit fly - Unchar.		- -	-	-	-		- -	-	-	-		-	-			-	-	-	- -	-	-	-	- -	-	-	-	- -
Parasitic wasp - EFG1	Е 1	ΓV	ľΕ	L	F	S (G G	i R	Α	ΕI	R K	Y	Т	S I	ΕT	R	D	I	VТ	P	Κ	S			-	-	
Jewel wasp - Unchar.	ΕĨ	٢٦	/Ε	L	F	N	G G	i R	Α	ΕJ	R K	Υ	Т	S I	ΕT	R	D	Ľ	VΊ	P	K	S			-	- 1	
Redifire ants - PRY1-like	Е 1	r٦	/Ε	L	F	N	G G	βR	Α	ΕI	R K	Υ	Т	S I	ΕT	R	D	Ľ	VТ	P	K	S	S S	S M	I P	S	LÌ
Dwarf honey bee - EFG1	Е 1	ΓV	ľΕ	L	F	N	G G	βR	Α	ΕI	R K	Y	Т	S I	ΕТ	R	D	I	VТ	P	Κ	S			-	-	
Leaf cutter bee - EFG1	Εl	ΓV	ľΕ	L	F	N (G G	i R	Α	ΕI	R K	Y	Т	S I	ΕТ	R	D	I	VТ	P	Κ	S			-	-	
Sea squirt (C. i.) - GAPR			-	-	-		- -	-	-	- 1		-	-			-	-			-	-	- 1	- -		-	-	
Spotted gar - GAPR			-	-	-			-	-			-	-			-	-			-	-	-			-	-	
Zebrafish - GAPR			-	-	-			-	-	- 1		-	-			-	-			-	-	-			-	-	
W. clawed frog - GAPR	٧V	νv	VK	D	S	R I	εv	G	\mathbf{v}	AI	K A	v	D	GI	C G	ŧΜ	v	L	Ąλ	A	0	Y					
Burmese python - GAPR			-	-	-			-	-			-	-			-	-			-	-				-	-	
Anole lizard - GAPR		_	-	-		_	_	-				-	-			-				-							_
Chinese soft-shelled turtle - GAPR	_	-	-	-		-	_	-		-	_		-		_	-			_	-				_	-		-
Falcon - GAPR			-	-	-	-		-	-	-		-		_		-	-	-			-	-			-		
Chicken - GAPR				-		-		-	-	-		-	-			-	-	-		-	-	-			-	Ē	-
Platypus - GAPR		-	-	-	-	-		-	-	-		-	-			-	-	-		-	-	-		-	-	-	-
			-	-	-			-	-			-	-			-	-			-	-	-		-	-	-	
Armadillo - GAPR			-	-	-			-	-			-	-			-	-			-	-	-			-	-	
Rat - GAPR		- -	-	-	-			-	-	-		-	-			-	-			-	-	-		-	-	-	
Mouse - GAPR		- -	-	-	-			-	-	-		-	-			-	-			-	-	-		-	-	-	
Chimp - GAPR		- -	-	-	-			-	-	-		-		-		-		-	-	-		-		-	-	-	
Iuman - GAPR		- -	-	-	-	-		-	-	-		-	-			-	-	-		-	-	-	- -	-	-	-	
					_	_			_										_			_					
1 Exon 2 Exc	on (3				E	xoi	n 4	1			ł	Exe	on	5				E	(0)	n 6	5]	Ex	on
8 Exon 9 Exo	n 1	0]]	Еx	on	11	1			E	xo	n 1	2]]	Exe	on	13	3			Е	xo	n i

Page 2 of 10

Appendix 8: ClustalW Amino Acid Alignment and Color-Coded Mapping of
Exons for GAPR/GLIPR2 Sequences

Tomato - Pr-4r	-	-	-	-	-			• •		-	-	-			-	-		• •	-	-	-	-		-	-	-	-	-	-	-
Cocoa tree - CAP-related	-	-	-	-	-			- -		-	-	-		- -	· -	-		- -	-	-	-	-	- -	-	-	-	-	-	-	-
Yeast - PRY1																	ΤI								5 E				2	
Mold - SCP-like																	T /			~									~	
Round worm - Unchar.				-		_	_	_	-			-	_	_	_													L	ΤJ	K
Fruit fly - Unchar.																		_							-	-		-	-	-
Parasitic wasp - EFG1	-	-	-	-	-					-	-	-			-	_	R											Τ.	A	S
Jewel wasp - Unchar.	-	-	-	-	-	-				-	-	-				_	R													-
Red fire ants - PRY1-like	V	Η	G	T I	K	Κŀ	C S	5 7	/ G	F	V	D	QI	V S	S P	_	R													
Dwarf honey bee - EFG1	-	-	-	-	-			• •	• •	-	-	-					R													
Leaf cutter bee - EFG1	-	-	-	-	- ,	- 1	- , -	• •	• •	-	-	-	- ,-	• •			R	_		_										_
Sea squirt (C. i.) - GAPR	-	-	-	-	-			• •	• -	-	-	-				-		_			G	N	CI	, K	t S	S	K	K	E	T
Spotted gar - GAPR	-	-	-	-	-			• •		-	-	-			-	-			-	-	-	-		-	-	-	-	-	-	-
Zebrafish - GAPR	-	-	-	-	-			- -		-	-	-		- -	-	-			-	-	-		- -	-	-	-	-	-	-	-
W. clawed frog - GAPR	-	-	-	-	-					-	-	-				S	P /	4 (ĴΝ	I	T	N :	P C	ξŊ	ſF	Q	K	N	V	L
Burmese python - GAPR	-	-	-	-	-			- -		-	-	-		- -	· -	-			-	-	-	-	- -	-	-	-	-	-	-	-
Anole lizard - GAPR	-	-	-	-	-			- -		-	-	-		- -	-	-			-	-	-	-		-	-	-	-	-	-	-
Chinese soft-shelled turtle - GAPR	-														-									-			-	-	-	-
Falcon - GAPR	-	-	-	-	-					-	-	-			-	-			-	-	-	-		-	-	-	-	-	-	-
Chicken - GAPR	-	-	-	-	-				-	-	-	-			-	-			-	-	-	-		-	-	-	-	-	-	-
Platypus - GAPR	-	-	-	-	-	-			-	-	-	-			-	-		•	-	-	-	-		-	-	-	-	-	-	-
Armadillo - GAPR	-	-	-	-	-					-	-	-		- -	-	-			-	-	-		- -	-	-	-	-	-	-	-
Rat - GAPR	-	-	-	-	-					-	-	-				-			-	-	-	- 1		-	-	-	-	-	-	-
Mouse - GAPR	-	-	-	-	-					-	-	-			-	-			-	-	-	- 1		-	-	-	-	-	-	-
Chimp - GAPR	-	-	-	-	-					-	-	-				-			-	-	-	- 1		-	-	-	-	-	-	-
Human - GAPR	-	-	-	-	-					-	-	-				-			-	-	-	- 1		-	-	-	-	-	-	-
																	Т													٦
Tomato - Pr-4r	-	-	-	-	-					M	G	L	F١	NN	1 S	L	LI	LN	1 T	С	LI	M	VΙ	, A	I	F	Н	S	C !	D
Cocoa tree - CAP-related	-	-	-	-	-					-	-	- 1	M	4 V	v s	Κ	LA	ΑI	A	С	L	I	I١	7 S	F	Α	L	P	S I	L
Yeast - PRY1	L	A	Q	A	Т	Τſ	ΓS	S 1	ΓР	A	A	A	S 1	Г 1	ΓS	Т	P A	4 /	ΛT	Т	Т	Т	s ç) A	A	A	Т	S	s ,	A
Mold - SCP-like	S	A٦	W	Т	S.	A٦	N 1	ΓS	S A	W	Т	s	s v	N 1	r s	S	A A	4 () P	Т	Т	L	A S	5 1	Т	S	S	A	s (G
Round worm - Unchar.																														
Fruit fly - Unchar.	-	-	-	-	-			- -		-	-	-		- -		М	FΙ	7 1	/N	I	L	I	VΙ	, A	L	С	L	Ľ	V	L
Parasitic wasp - EFG1	Ρ	А	s	Ρ	G	ΡI	L S	s s	s s	F	Н	G	S I	FΗ	I S	S	L (G S	S D	G	Μ	s	C S	s s	i A	R	s	s	s ,	A
Jewel wasp - Unchar.	Ρ	A	s	Ρ	G	ΡI	L S	s c	3 S	L	Н	G	S I	FΗ	I G	S	L (G S	S D	G	М	s	C S	5 S	i A	R	S	S	S /	A
Red fire ants - PRY1-like	Ρ	A	s	Р	G	ΡI	L S	S N	۱S	L	Н	G	S 1	FΗ	I G	S	L	G S	S D	G	Μ	s	C S	s s	i A	R	S	S	S.	A
Dwarf honey bee - EFG1	Ρ	Α	S	P	G	ΡI	L S	S N	V S	L	Н	G	S I	FΗ	I G	S	L (3 S	S D	G	Μ	s	C S	s s	i A	. R	s	s	S /	A
Leaf cutter bee - EFG1	Ρ	Α	s	P	G	ΡI	L S	S N	v s	L	Н	G	S I	ΕH	I G	s	L (G S	5 D	G	М	s	C S	s s	A	R	s	s	s ,	A
Sea squirt (C. i.) - GAPR	Ι	Т	Т	T I	N	ΕI	Ρŀ	٢N	ΙE	Е	Т	V	D١	v 1	ГΡ	Р	Τŀ	C E	ΕE	Е	Т	v	ГЕ	E	P	Κ	G	D	G	s
Spotted gar - GAPR	-	-	-	-	-					-	-	-		- -	-	-		- -		-	-	- 1	- -	-	-	М	s	D	T /	A
Zebrafish - GAPR	-	-	-	-	-					-	-	-				-				-	-			-	-	М	G	к	s.	A
W. clawed frog - GAPR	Ρ	ĸ	G	Т	P	v :	Sľ	V I	ſG	T	S	P	ΤŻ	A F	۲G	Т	SI	ΥĪ	, G	Т	R	G	ГΙ) s	A	L	S	Ρ	T.	A
Burmese python - GAPR	-	-	-	-	-					-	-	-		- -		-				-	-	- 1	- -		-	-	-	-	-	-
Anole lizard - GAPR	-	-	-	-	-					-	-	-				-				-	-			-	-	M	G	ĸ	T.	A
Chinese soft-shelled turtle - GAPR	-	-	-	-	-					-	-	-				-				-	-	- 1			-	М	G	к	s.	A
Falcon - GAPR	-							_	_				_	_				_	_											
Chicken - GAPR	-	-	-	-	-		-																			-	-		-	
Platypus - GAPR	-	-	-	-	-				-			-	_	_	_			-	-			-	_	-	-					
Armadillo - GAPR	-	-	-	-	_		-		-			-	_	_	_			-				-	_	-	-					
Rat - GAPR	-	_	-	_	_		_	_	-			-	-	_	-			-	-			-	_	-	-					
	-	-	-	_		_	_	_	-			-	_	_	_			-	-			-	_	-	_					
Mouse - GAPR	+-	-	-	_	_				-	-		-	_	_	_			-	-			-	_	-	-					
Mouse - GAPR Chimp - GAPR				+		+	+	+	-		-	-	-	_				-	-			-	_	-	-					
Chimp - GAPR	-	_ !	_	_ !	- L	- 1										- C	- 1	11	-	17.	-	- L'	11	1.7		- 11	0	-	-	-1
	-	-	-	-	-		- -	•		-																				

Appendix 8: Page 3 of 10

Appendix 8: ClustalW Amino Acid Alignment and Color-Coded	Mapping of
Exons for GAPR/GLIPR2 Sequences	

Tomato - Pr-4r	AQN S P Q D
Cocoa tree - CAP-related	AKTLKQE
Yeast - PRY1	S S S D S D
Mold - SCP-like	ATATNAYQ ST
Round worm - Unchar.	KLDPAEYSIKDLKKW
Fruit fly - Unchar.	V I ADLOED
Parasitic wasp - EFG1	S P D S A R Y S S S O I T K T E P R K P S V R R S G P S K E F V N V
Jewel wasp - Unchar.	S P D S A R Y S T A Q I T K T D T R K P S V R R F G P D K E F I N L
Red fire ants - PRY1-like	S P D S A R Y S S S O I T K I D T R K P S V R R S G P P K E F I N V
Dwarf honey bee - EFG1	SPDSARYTSTP I TKTDTRKPSVRRLGPPKEFINI
Leaf cutter bee - EFG1	S P D S A R Y S S T P I T K T D T R K P S V R R F G P P R F V S V
Sea squirt (C. i.) - GAPR	QYAKD
Spotted gar - GAPR	SKOFAKE
Zebrafish - GAPR	SRUTARE
W. dawed frog - GAPR	DSREFALE
•	
Burmese python - GAPR Anole lizard - GAPR	SKOFADE
	SKQFADE
Chinese soft-shelled turtle - GAPR	SKQFAEE
Falcon - GAPR	SKQFAEE
Chicken - GAPR	SKQFAEE
Platypus - GAPR	SKQFADE
Armadillo - GAPR	SKQFNNE SKQFNNE
Rat - GAPR	S K Q F N N E
Mouse - GAPR	S K Q F N N E
Chimp - GAPR	SKQ F H N E
Human - GAPR	SKQFHNE
	CAP3
Tomato - Pr-4r	YLEVHNDAR AQVG - VGPMSWDADLESRAQSYANS
Cocoa tree - CAP-related	F L D A Q N E A R A E V N - V E P MAWDA Q V A A Y A Q E Y A N Q
Yeast - PRY1	V L A E HNKKR A L HK D T P A L S W S D T L A S Y A Q D Y A D N
Mold - SCP-like	VLYNHNVHR SNH S - A S SLEWDA SLEA SAQTLAAR
Round worm - Unchar.	I VHFHNKYRAHHS - SPAVTVDSNLTNLAQKWSDE
Fruit fly - Unchar.	HLNEHNRLREKHG - SPPLTLDDELTKG <mark>C</mark> EEYAK <mark>V</mark>
Parasitic wasp - EFG1	CLDTHNFYR SRHG - VPPLRLSKOLCKASOEWANV
Jewel wasp - Unchar.	CLETHNYYRNRHG - VPALRLNKQLCKTSQDWANI
Red fire ants - PRY1-like	CLDTHNFYRSRHG - VPPLRLSKQLCKTSQDWANI
Dwarf honey bee - EFG1	CLETHNFYRARHG - VPPLRLNKQLCKTSQDWANI
Leaf cutter bee - EFG1	CLETHNFYRARHG - VPPLRLSKQLCKTSQDWANI
Sea squirt (C. i.) - GAPR	MLTKHNEKRALHS - TAAMTLSVKLCEDAOKWADH
Spotted gar - GAPR	VLK SHNEYRKQHQ - VPPLKL SNKL SKEAKR YADS
Zebrafish - GAPR	ALKTHNEYRRKHQ - APPLKLSSKLCSEASRYAES
W. dawed frog - GAPR	F LKANNVYRSRHG - AKPLQLNSK I SQEAQRWAEH
Burmese python - GAPR	VLRAHNDYRKKHG - VPSLKL <mark>C</mark> KKLNRDAQQ <mark>YSEA</mark>
Anole lizard - GAPR	VLKAHNDYRKKHG-VPSLKL <mark>C</mark> KNLSREAQQYSEA
Chinese soft-shelled turtle - GAPR	
	V L K AHND Y R K K H G - V P P L K I <mark>C</mark> K K L N T G A Q Q Y A E K V L K AHND Y R K K H G - V P P L K L <mark>C</mark> K K L N R G A Q Q Y A E E
Falcon - GAPR	
Chicken - GAPR Platypus - GAPR	VLKAHNDYRKKHG - VPPLKLCKKLNRGAQQYAEE
71	VLRAHNDYRQKHG-VPSLKLCKKLNREAQQYAEA
Armadillo - GAPR	VLKAHNEYRRQHG - VPPLKLCNKLNQEAQQYSEA
Rat - GAPR	VLKAHNEYRAKHG - VPPLKLCKKLNQEAQQYSEA
Mouse - GAPR	VLKAHNEYRAQHG - VPPLKLCKKLNREAQQYSEA
Chimp - GAPR	VLKAHNEYRQKHG - VPPLKL <mark>C</mark> KKLNREAQQ <mark>YSEA</mark>
Human - GAPR	VLKAHNEYRQKHG - VPPLKLCKNLNREAQQ <mark>YSEA</mark>
	on 3 Exon 4 Exon 5 Exon 6 Exon 7
8 Exon 9 Exo	n 10 Exon 11 Exon 12 Exon 13 Exon 14

Appendix 8: Page 4 of 10

RAGDCN-LIHSGS-GENLAK-GGGD
RIADCD-LVHSGGPYGENIAW-GSDD
YDCSGT-LTHSGGPYGENLALG
CVYQHD TFSSTINGGG YGQNIGYGVSSE
MAFHKK <mark>C</mark> LVHEQPSKYGENLTSFASSKF - PSPK
LANNEK - LEHS - SSAGQN YGENL <mark>C</mark> M - RS
LATRGR - LEHRANIDYGENLY <mark>C</mark> MWSSN PKTV
LATKGR - LEHRANIEYGENLY <mark>C</mark> MWSSN PKTI
LATRGR - LEHRANIDYGENLY <mark>C</mark> MWSSN PKTI
LATRGR - LEHRANTDYGENLY <mark>C</mark> MWSSN PKTI
LAARGR - LEHRANIDYGENLY <mark>C</mark> MWSSN PKTV
I A AKNA - L E H <mark>C</mark> KDR E GA <mark>G</mark> - E N L AW - S S A E
_LASTRV-LKHSPESSRGS <mark>C</mark> GENLAW-ASYD
LASTRI-LKHSVESS <u>RGNC</u> GENLAW-ASYD
LLNLKN - LKHSDTSHGEN IWAKSGG - PSIT
LASTRI-LKHSPESSSGK <mark>C</mark> GENLAW-ASYNO
LASTRI-LKHSPESSKGK <mark>C</mark> GENLAW-ASYD
LASTRV - LQHSPESANGK <mark>C</mark> GENLAW - ASYD
LATTRV - LKHSSESANGK YGENLAW - ASYD
LAASRI - LKHSSESASGK <mark>C</mark> GENLAW - ASYD
LASTKV - LKHSPESSRGQ YGENLAW - ASYD
LASTRI-LKHSPESSRGQ <mark>C</mark> GENLAW-ASYD
_LASTRI-LKHSPESSRGQ <mark>C</mark> GENL <mark>AW-ASYD</mark>
LASTRI-LKHSPESSRGQ <mark>C</mark> GENLAW-ASYD
_LASTRI-LKHSPESSRGQ <mark>C</mark> GENL <mark>AW-ASYD</mark>
LASTRI-LKHSPESSRGQ <mark>C</mark> GENLAW-ASYD
T G R A A V E L W V S E K P N Y N Y D T N E <mark>C</mark> V S G K M
SVADAVKMWVDEKVYYDHGSNT <mark>C</mark> A SGQV
DGPAAVDAWYNE I SNYDF SNPG F S SN
I G EM I TNLMYNN EMG Y F E A L Y G E AN P SMDN F D A'
CAAAL IHGFYTEGYGFNYT <mark>RFNPG SWSK</mark>
T P L Q C V Q D W Y D E I A D Y D F E K P Q F A M S
GGEEPVNEWYAEESQHQYGKEP TTLK
NGDEPVNDWYAEESQHQYNREP TTLK
GGEEPVNDWYAEEAQHQYGKEPTTLK
HGDEPVNE <mark>WYAEEGQHQYGKEP TTLK</mark>
S G D E P V N E WYA E E A Q H Q Y G K E P T T L K
GADAAVDMWYNELKDYDFSQPGFSGS
P G K E V A D R W Y S E I K N Y N Y N C P G F S S G
T GKD V TDRWYN E VNQ YN FNQ P G F S S G
T G Q E V A D S W Y K E E K N Y N F S K P G N K A K
SGSEVADRWYSEIKNYNFQSPGFSSG
SGSEVAERWYNE INN YN FQNPG FSSG
P GNE VAE RWYNE I KNYN F QNP G F S S G
T G K D V A D R W Y S E I K N Y S F Q N P G F S S G
PGKDVADRWYSEIKNYSFQNPGFSSG
P GKEVADRWYNEIKDYNFQHPGFTSG
T G K E V A D R W Y S E I K N Y N F Q Q P G F T S G
TGKEVADRWYSEIKSYNFQQPGFTSG
TGKDVADRWYSEIKSYNFQQPGFTSG
T GKEVADRWYSEIKNYNFQQPGFTSG

Appendix 8: ClustalW Amino Acid Alignment and Color-Coded Mapping of
Exons for GAPR/GLIPR2 Sequences

Tomato - Pr-4r	CAP1 GHYTOVVWRDSVRLGCGRALCNDG
Cocoa tree - CAP-related	AHYTQVVWRNSVRLGCAKVRCDNGG
Yeast - PRY1	GHFTQVVWKSTTQVGCGIKTCGGAWG
Mold - SCP-like	GHF SQ I VWK GT THV G <mark>C</mark> A TV T <mark>C</mark> N S L GN VD S S VA V
Round worm - Unchar.	GHF TOLLWKNSRKIGVGVSVAKRGT MYH
Fruit fly - Unchar.	GHFTALVWKNAKKMGIGOAKDKKGY
Parasitic wasp - EFG1	GHFTQVVWRDSTELGVGMARNRNGE
lewel wasp - Unchar.	GHFTQIVWRSSTELGVGMARNRNGE
Red fire ants - PRY1-like	GHETOVIWEDSTELGVGMARNENGE
Dwarf honey bee - EFG1	GHFTOVVWKDSTELGVGMARNRNGE
Leaf cutter bee - EFG1	GHFTQVVWRDSTELGVGMARNRNGE
Sea squirt (C. i.) - GAPR	GHFTQVVWKASTEFGAGFAQASDGS
Spotted gar - GAPR	GHFTAMVWKSSRELGVGKATATDGS
Zebrafish - GAPR	GHFTAVVWKGSKKLGVGKAVASDGS
W. clawed frog - GAPR	GHF TQMVWKASKEVGVGLASSGKG M
Burmese python - GAPR	GHFTAMIWKNTKKMGVGKAAASDGS
Anole lizard - GAPR	GHFTAMVWKSTKKMGVGKAAASDGS
Chinese soft-shelled turtle - GAPR	GHFTAMVWKDTKKMGVGKASANDGS
Falcon - GAPR	GHFTAMVWKNTKKMGVGKASASDGS
Chicken - GAPR	GHF TAMVWK STKKMGVGKA SASDGS
Platypus - GAPR	GHF TAMVWK STTKMGVGKA STSDGS
Armadillo - GAPR	GHF TAMVWK STITKMGVGKA STSDG
Rat - GAPR	GHFTAMVWKNTKKIGVGKASASDGS
Mouse - GAPR	GHF TAMVWKNTKKI GVGKA SASDGS
Chimp - GAPR	GHFTAMVWKNTKKMGVGKASASDGS
Human - GAPR	GHFTAMVWKNTKKMGVGKASASDGS
initiani - Orit IX	CAP2
Fomato - Pr-4r	F I SCNYDPVGNWVGQRPY
Cocoa tree - CAP-related	F I TCNYDPPGNYI GE TPY
Yeast - PRY1	YVICSYDPAGNYEG EYADNVEPLA
Mold - SCP-like	FTVCNYSPAGNYAG EYADNVLR
Round worm - Unchar.	YVCIKYDPPGNMO
Fruit fly - Unchar.	WVVARVYPPVNVN - GOFFFNVLP
Parasitic wasp - EFG1	YVVANYNPPGNFLG SFAENVLPPRDASSTSS
Jewel wasp - Unchar.	YVVCNYNPAGNFLG SFAENVLPPRDASSISS YVVCNYNPAGNFLG SFMENVLPPSERSSSPS
Red fire ants - PRY1-like	VVVCNVNPAGNELG SETENVLPPVDACPAKP
Dwarf honey bee - EFG1	YVVCNYNPAGNFLG SFTENVLPLG - VKPLKK
Leaf cutter bee - EFG1	YVVCNYNPAGNFLG SFTENVLPLG - VKPLKK
	YVVGRYLPPGNMNMAGQFEENVLPLA
Sea squirt (C. i.) - GAPR Spotted gar - GAPR	FVVGRILPPGNMNMAGQFEENVLPLA
Spotted gar - GAPK Zebrafish - GAPR	FVVARYFPAGNI TNOGHFOANVLP
Zebratish - GAPR W. clawed frog - GAPR	I V V AR Y F P A GN II IN Q GH F Q AN V L P
W. clawed frog - GAPR Burmese python - GAPR	
Burmese python - GAPR Anole lizard - GAPR	F V V AR Y E P A GN I VN P GQ Y E QN V F P
Chinese soft-shelled turtle - GAPR	F V V AR Y V P A GN I VN P GQ Y E QNV L P
Falcon - GAPR	F V V AR Y D P A GN V VN P GY YE ENVL P
Chicken - GAPR	F V V AR Y D P A G N V V N P G Y Y E E N V L P
Platypus - GAPR	
Armadillo - GAPR	F V VAR Y F P A GN V VNQ GY XE ENVL P
Rat - GAPR	F V V AR Y F P A GN I VNQ GF F E E NV P P
Mouse - GAPR	F V V AR Y F P A GN I VN Q G F F E E N V P P
Chimp - GAPR	F VVAR Y F P A G N V V N E G F F E E N V L P
Human - GAPR	FVVARYFPAGNVVNEGFFEENVLP
-	
1 Exon 2 Ex	xon 3 Exon 4 Exon 5 Exon 6 Exo
8 Exon 9 Exo	on 10 Exon 11 Exon 12 Exon 13 Exon
ix 8:	

Cocoa tree - CAP-related Yeast - PRY1 Mold - SCP-like Round worm - Unchar. Fruit fly - Unchar. Parasitic wasp - EFG1 Jewel wasp - Unchar.	1000			-		-			_	-		-	_	_	-	_	_			· -	-				-	- [
Mold - SCP-like Round worm - Unchar. Fruit fly - Unchar. Parasitic wasp - EFG1	-			-			-																			
Round worm - Unchar. Fruit fly - Unchar. Parasitic wasp - EFG1							-		_	-	-	_				_	_	-	_		_					
Fruit fly - Unchar. Parasitic wasp - EFG1		_	_		_	_	_	_	_	_	_	_	_	_		_	_		_	_		_	_			-
Parasitic wasp - EFG1	1000	_	-		_	_			_		-	_	_	_		_	-		_	_		_	_		-	-
•																									-	-
Isrual man Hasher	P (Q F) -	-	- 0	ЪK	S	S 1	ΡĽ) P	V	P 1	ΚI	I M	D	ΕF	łΑ	W	QI) E	А	Γŀ	C V	Ή	NI	E)
Jewei wasp - Olicitat.			-	-	- I	. P	А	ΙI	L K	Ι	S	S٦	WΊ	ΓL	D	QS	5 A	W	QC) D	A	LÆ	٩V	Η	N I	E]
Red fire ants - PRY1-like			-	-		·I	Ν	F '	ТΕ) L	Κ	QI	ΗY	ζΙ	Ν	ΕC	Σ	W	QC) E	A	LΙ	l V	Ή	N	E 1
Dwarf honey bee - EFG1			-	-		·Τ	L	ΙI	ΕA	ΙL	Р	Y	ΥT	ΓL	D	ΕÇ	Q A	W	QQ) D	A	LΝ	ΛV	Ή	N J	E '
Leaf cutter bee - EFG1			-	-		- I	S	FΙ	LΕ) P	K	Η	ΥA	١M	[D	ΕÇ	χT	W	QQ) E	Α	LΙ	l V	Ή	N I	ΕŢ
Sea squirt (C. i.) - GAPR			-	-			-			-	-	-			-			-		-	-			-		
Spotted gar - GAPR			-	-			-			-	-	-		P	R	К.		-			-			-		
Zebrafish - GAPR			-	-											К			-						-		
W. dawed frog - GAPR	DI	ΕD	G	F	VK	c s	Р	S :	sт	' A	v	L	ΡI	I P	Е	ΚI	ΕL	К	SI	7 R	K	DI	Ĺ	S	EI	нī
Burmese python - GAPR			_													_	_	_	_	_	_		_	_	_	_
Anole lizard - GAPR																	_		_	_			_			
Chinese soft-shelled turtle - GAPR														_		_	_		_	_			_			_
Falcon - GAPR																_										
Chicken - GAPR																_	_		-	-			_			-
Platypus - GAPR																			_	_						_
Armadillo - GAPR				-			-				-			_		-										
Rat - GAPR																-	_		-	-			_		-	_
Kat - GAPK Mouse - GAPR																_										
			-	-										_	_	_										
Chimp - GAPR			-	-			-			-	-															
Human - GAPR			-	-			-			-	-	2		P	K.	к		-		-	-			-		
		_			_	_			_	_		_	_	_		_	_		_	_		_	_		_	_
Tomato - Pr-4r			-	-		• •																				
Cocoa tree - CAP-related					_																					
Yeast - PRY1			-	-		_			_				_				_		-	-			_			
Mold - SCP-like			-	-			-			-	-	-		-	-			-		-	-			-	-	
Round worm - Unchar.			-	-			-			-	-	-		-	-			-		-	-			-		
Fruit fly - Unchar.		- 0	ΰD	ΕI	NC	эQ	G	NI	LN	1 -	-	-			-			-		-	-			-	- 3	R I
Parasitic wasp - EFG1	RΗ	R R	H	R	VF	ΡE	L	R I	LN	I A	D	L	S A	۱A	A	K A	٩W	Α	C 1	ΓL	L	N 1	ΓN	K	L	II
Jewel wasp - Unchar.	RΗ	R R	Н	R	S F	P D	L	ΙI	LN	I D	E	Ľ	ΤA	۱A	A	K /	١W	A	NK	ΓL	L	N S	S N	K	L	II
Red fire ants - PRY1-like	RΗ	R K	H	R	VF	P D	L	R I	ĽΒ	S P	Е	Ľ	ΤA	۱A	. A :	K A	٩W	A	ΝΊ	ΓL	L	N 1	ΓN	K	L	II
Dwarf honey bee - EFG1	R F	R K	н	R	V S	S D	L	R I	LS	s v	Έ	Ľ	ΤS	S A	A	K A	٩w	A	N I	ΓL	L	N 1	ΓN	ĸ	L	II
Leaf cutter bee - EFG1	R F	R R	Н	R	VF	D S	L	ΤI	LS	i A	D	L	ТA	۱A	A	K A	٩w	A	NI	ΓL	L	N 1	ΓN	К	L	II
Sea squirt (C. i.) - GAPR			-	-			-			-	-	-	- -		-			-			-			-	- 1	
Spotted gar - GAPR			-	-			-			-	-	-			-			-			-			-		
Zebrafish - GAPR			-	-			-			-	-	-			-			-			-			-		
W. dawed frog - GAPR	0	YR	K	L	нс	ΞA	G	A	LC	L	S	v.	ΑI	, s	0	D A	0	K	w A	L D	H	LI	/ G	K	P	A I
Burmese python - GAPR			-	-			-			-	-	-			-			-			-			-		
Anole lizard - GAPR			-	-			-			_	-	-			-			-			-			-	- 1	
Chinese soft-shelled turtle - GAPR			-		_	_			_	-		-	_	-			_		_	-			_			-
Falcon - GAPR																										
Chicken - GAPR		_	-							-	-	-		-		-	_		_	-			-	-	-	
Platypus - GAPR			-	-		-	-	-		-	-	-		_		_	_			-	-	-	-	-	-	-
Armadillo - GAPR			-	-		-	-	-		-	-	-		-		-			-	-	-	-	-	-	-	-
			-	-		-	-			-	-	-		-		-	-	-		-	-			-		
Rat - GAPR			-	-			-			-	-	-		-				-		-	-			-	-	
Mouse - GAPR		- -	-	-		-	-			-		-		-			-			-	-		-			-
Chimp - GAPR		- -	-	-		-	-	-	- -	-				_			_	-	-	-	-		_		_	_
Human - GAPR			-	-		-	-	-	- -	-	-	-		-	-		-	-		-	-		-	-	-	- -
1 Exon 2 Ex	on	3]	Ex	on	ı 4				E	xo	n	5 🕻			F	Exc	on	6]	Ex	or

Page 7 of 10

Tomato - Pr-4r Cocoa tree - CAP-related	-	-		-		-	-		-	-	-	_	_	-	-	-	-		_	-	-	-		-	-	-	- -	-	
	-		_	_	-	-	-			-	-	_	-	-			-	_	-	-		-			-	-	_	-	-
Yeast - PRY1			_	_	_	_	-	-		-		_	_	-		-	-	_	-	_	-	-			-	-	_	-	-
Mold - SCP-like				_	_					_	_		- -	-	-	-	-		• •		-	-	-	-	-	-		-	-
Round worm - Unchar.									V					-	-	-				-	-	-	-	-	-	-		-	-
Fruit fly - Unchar.																T													
Parasitic wasp - EFG1							ŧΕ	Ν	I	Y	SI	МÇ	2 C	_		P 1	К.	L		/ S	P	R	E	V	I	S I	K۷	ΥY	S
Jewel wasp - Unchar.		S				_	ξE	Ν	Ι	Y	S T	VC	20	_	D		N	NI		7 S	A	. R	Ε	V	V.	A I	ΚV	۷Ŋ	
Red fire ants - PRY1-like																P]													
Dwarf honey bee - EFG1	Q	Т	S S	S 1	6 J	ζG	Ε	Ν	Ι	Y	SI	MQ	2	S	D	P 1	K.	V I	[]	/Т	P	R	Е	V	V	S 1	ΚV	ΥY	A
Leaf cutter bee - EFG1	Q	S	S S	S 1	6 J	ζG	Ε	Ν	Ι	Y	S 1	MQ	20	S	D	P]	ĸ	LI	[]	7 P	A	R	Ε	V	V	S]	ΚV	γY	8
Sea squirt (C. i.) - GAPR	-	-				-	-	-	-	-	-			-	-	-	-			-	-	-	-	-	-	-		-	-
Spotted gar - GAPR	-	-				-	-	-	-	-	-			-	-	-	-				-	-	-	-	-	-		-	-
Zebrafish - GAPR	-	-					-	-	-	-	-			-	-	-	-			-	-	-	-	-	-	-		-	
W. dawed frog - GAPR	0									-	-	-	-			Ν	-		-			-							
Burmese python - GAPR	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	-	_	_	_	_	_								
Anole lizard - GAPR	-		_	_		_	-	-	-	-	-	_	_	-	-		-	_	-	-	-	-			-	-	-	+	-
Chinese soft-shelled turtle - GAPR	-		-	-	-	-	-	-		-	-	-	-	-			-	-	-	-		-		-	-	-		-	F
	-					_										-													
Falcon - GAPR	-		_	_	_	_	-	-		-		_	_	-		-	-	_		_	-	-			-	-	_	-	-
Chicken - GAPR						-										-				-									
Platypus - GAPR	-				_	_										-				-								-	
Armadillo - GAPR	-	-		- -	- -	-	-	-	-	-	-			-	-	-	-		•	-	-	-	-	-	-	-	- -	-	-
Rat - GAPR	-	-				-	-	-	-	-	-			-	-	-	-			-	-	-	-	-	-	-		-	-
Mouse - GAPR	-	-			- -	- -	-	-	-	-	-			-	-	-	-			-	-	-	-	-	-	-		-	
Chimp - GAPR	-	-					-	-	-	-	-			-	-	-	-				-	-	-	-	-	-		-	
Human - GAPR	-	-					-	-	-	-	-			-	-	-	-				-	-	-	-	-	-		-	
Tomato - Pr-4r	-	-					-	-	-	-	-			-	-	-	-				-	-	-	-	-	-		-	
Cocoa tree - CAP-related	-	-					-	-	-	-	-			-	-	-	-				-	-	-	-	-	-		-	
Yeast - PRY1					-	-	-	-		-	-		-	-		-	-		-	-	-	-			-	-		-	+
Mold - SCP-like	-					_										-													
Round worm - Unchar.	-															-													
																-													
Fruit fly - Unchar.		-	_			-									÷														
Parasitic wasp - EFG1																F													
Jewel wasp - Unchar.																F													
Red fire ants - PRY1-like																F													
Dwarf honey bee - EFG1	_														+	F		- C											
Leaf cutter bee - EFG1	K	ĸ	DI	HI	ΚŦ	G	Τ	Е	P	K	V	LN	ΓI	C	Η	F	T (Q I	[]	7 W	K	N	Т	Т	El	M	G I	A	Ν
Sea squirt (C. i.) - GAPR	-	-				-	-	-	-	-	-			-	-	-	-		• •	-	-	-	-	-	-	-		-	
Spotted gar - GAPR	-	-				-	-	-	-	-	-			-	-	-	-			-	-	-	-	-	-	-		-	-
Zebrafish - GAPR	-	-					-	-	-	-	-			-	-	-	-				-	-	-	-	-	-		-	
W. dawed frog - GAPR	N	A	K	YS	S F	Ā	Т	Р	G	F	0	s c	3 S	G	N	F	T	ON	1	W	K	S	s	S	0	v	GI	10	I
Burmese python - GAPR															_	-	_	-	_	_	_	_	_	_	-	_	_	_	
Anole lizard - GAPR	-		_		_	_	-	-			-		_	-		-			_	-	-	-			-	-	_	-	-
Chinese soft-shelled turtle - GAPR	_		_	_	_	_	-	-		-	-	_	_	-		-	-	_	-	_	-	-		-	-	-	_	-	-
Falcon - GAPR	_					_								-		-				_								_	
	-	-				-	-	-	-	-	-			-	-	-	-			-	-	-	-	-	-	-		-	-
Chicken - GAPR	-	-				-	-	-	-	-	-			-	-	-	-		•	-	-	-	-	-	-	-		-	ŀ
Platypus - GAPR	-	-				-	-	-	-	-	-		• •	-	-	-	-		•	• -	-	-	-	-	-	-		-	ŀ
Armadillo - GAPR	-	-				-	-	-	-	-	-		• •	-	-	-	-		•	-	-	-	-	-	-	-		-	-
Rat - GAPR	-	-				-	-	-	-	-	-			-	-	-	-		•	-	-	-	-	-	-	-		-	
Mouse - GAPR	-	-				-	-	-	-	-	-			-	-	-	-			-	-	-	-	-	-	-		-	-
Chimp - GAPR	-	-					-	-	-	-	-			-	-	-	-				-	-	-	-	-	-		-	
Human - GAPR	-	-				-	-	-		-	-	_	_	-		-	-	_	-	-	-	-			-	-	_	-	-
1 Exon 2 Exo									t 🗖							5				E								x	
B Exon 9 Exon																													
Exon 9 Exon	10			_	Б	~		11			_			~ n	1	2			1	Ex	on	1	3				E	Ç,	or

Appendix 8: ClustalW Amino Acid Alignment and Color-Coded Mapping of
Exons for GAPR/GLIPR2 Sequences

CAP-related 1 like - Unchar Unchar PRY1-like bee - EFG1 bee - EFG1 bee - EFG1 i) - GAPR GAPR APR og - GAPR GAPR GAPR GAPR Shelled turtle - GAPR PR APR PR APR APR APR APR APR APR AP	- K K K - - - - - - - - - - - - - - - -	- - - - - - - - - - - - - - - - - - -	- - - - - - - - - - - - - - - - - - -	- - - - - - - - - - - -	- T T T T - - - - G -		 Y I	- M - V V V V - -	A A A A A A		· - · - · · · · · · · · · · · ·	- - P P P P		J N J N J N J N J N	- - I I I I	 V C V C		- - F F F	 T H T H T H		- - V V V	- L] L] L]	- K K K K	- - P P P	- - V] I]	- K K K	- S S G
like - Unchar. - Unchar. - EFG1 Unchar. - PRY1-like bee - EFG1 .i.) - GAPR GAPR APR og - GAPR GAPR GAPR GAPR GAPR Shelled turtle - GAPR PR APR PR PR PR 4r	- K K K - - - - - - - - - - - - - - - -	- R K R - - - - - - - - - - - - -	- - - - - - - - - - - - - - - - - - -	- G G G - - - - K -	- T T T T - - G C		 L I V V V V V V V V V I - - - - - - - - - - -	- M - V V V V V - - -	A A A A A		 7 Y 7 Y 7 H 7 H 7 H 7 H 7 H 7 H	- P P P P P	R C R C R C R C	J N N N N	- - I I I I	 V (0 V (0 V (0 V (0)		- F F F F	 T H T H T H		- - V V V	- - L] L] L]	- K K K K	- - P P P	- - V] I]	- K K K	- S S G
- Unchar. char. p - EFG1 Unchar. - PRY1-like bee - EFG1 i) - GAPR GAPR APR og - GAPR GAPR GAPR GAPR Shelled turtle - GAPR PR APR PR PR PR 4r	- K K K - - - - - - - - - - - - - - - -	- R K R - - - - - - - - - - - - -	- D D D - D - - - - - - - - - -	- G G G - - - - K - - - - - -	- T T T T - - - G I - - - - -		L I V V V V V V V V V I - - - - -	M - V V V V - - -	A A A A A		- 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	- P P P P P	R C R C R C R C R C	JN JN N N	- I I I I			F F F F			- V V V	- L] L] L]	K K K	- P P P	- V] I] V]	- K K	- S S G
har. p - EFG1 Unchar. - PRY1-like bee - EFG1 e - EFG1 i) - GAPR GAPR APR og - GAPR og - GAPR GAPR GAPR Shelled turtle - GAPR PR APR PR PR PR 4r	- K K K - 7 - 7 	R K R - - - - - - - - - - - - - - - - -	- D D D - - - - - - - - - -	- G G G - - - -	- T T T - - G I - - G I -		 V V V V V V V I Y I	- V V V V - -	A A A A A		7 7 7 7 7 1 7 1 7 1 7 1 7 1 7 1 7 1 7 1	P P P P P	R C R C R C R C R C	in in in in in in	I I I I			F F F F	T H T H T H T H	E N E N E N E N	V V V V		K K K	P P P	- V] I] V]	K K K	S S G
p - EFG1 Unchar. - PRY1-like bee - EFG1 e - EFG1 i) - GAPR GAPR APR og - GAPR og - GAPR GAPR GAPR Shelled turtle - GAPR PR APR PR PR PR 4r	K K K - - - - - - - - - - - - - - - - -	R K R - - - - - - - - - - - - - - - - -	D D D D - - N - - - - - - - - - -	G G G G - - -	T T T T - - G I - - - - -		V V I V V V V I Y I	V V V V - -	A A A A		7 7 7 7 1 7 1 7 1 7 1 7 1 7 1 7 1 7 1 7	P P P P P	R C R C R C R C	N N N N	I I I I I			F F F F	TH TH TH	E N E N E N E N	V V V		K K K	P P P	VI II VI	K K K	S . S G .
Unchar. - PRY1-like bee - EFG1 ee - EFG1 GAPR GAPR APR og - GAPR og - GAPR GAPR GAPR GAPR Shelled turtle - GAPR PR APR PR PR 4r	K K - - - - - - - - - - - - - - - - - -	K R - - - - - - - - - - - - -	D D D - - - N - - - - - - - - - -	G G G - - -	T T T - - G] - - -		I V V V V I - Y I	V V V - -	A A A -		7 Y 7 H 7 H 7 H 7 H	P P P P	R C R C R C R C	δΝ δΝ δΝ	I I I I	V C V C V C	iQ iQ iQ	F F F	F F F F F F	E N E N E N	V V V		K K K	P P	I] V]	K K (S G
- PRY1-like bee - EFG1 i) - GAPR GAPR APR og - GAPR og - GAPR GAPR GAPR Shelled turtle - GAPR PR APR APR PR PR PR Ar	K K - - - - - - - - - - - - - - - - - -	R R - - - - - - - - - - - - - - - - - -	D D D - - N - - - - - - - - -	G G - - K - -	T T - - G I - G I - - - - - -		V V V V Y I	V V - -	A A -		7 H 7 H 7 H	P P P	R C R C R C	SN SN	I I	V C V C	iQ iQ	F ' F '	Γ Η Γ Η	E N E N	V V	LI LI	K K	Ρ	V]	K (G
bee - EFG1 ee - EFG1 i) - GAPR GAPR APR og - GAPR og - GAPR GAPR GAPR Shelled turtle - GAPR PR APR PR PR PR 4r	K K - - - - - - - - - - - - - - - -	K R - - 2 0 - - - - - - - - - - - -	D D - - - - N - - - - - - - - - - - - -	G - - - K - - -	T T - - G] - - -		V V V I Y I	V V - -	A - -		7 H 7 H -	P P	R C R C	δN	Ι	V C	θQ	F ′	ΓĒ	ΞN	V	LI	ĸ				
ee - EFG1 i) - GAPR GAPR APR og - GAPR og - GAPR GAPR GAPR Shelled turtle - GAPR PR APR PR PR PR 4r	K - - - - - - - - - - - - - - - - - -	R - - - - - - - - - -	. D - - - - - - - - - - - - - - -	G - - - - - - -	T - - G] - - -	C - - - M	V I Y I	V - -	A - -	<mark>C</mark> 1	7 H -	P	R C											Р	T1		L.
i) - GAPR GAPR APR og - GAPR og - GAPR GAPR GAPR Shelled turtle - GAPR PR APR APR PR PR PR PR AFR	- - - - - - - - - - - - - - - - - - -	- - - - - - - - - - - - - - - - - -	- - N - - - -	- - K - -	- - G] - -	- - -	 Y I	- -	-					δN	т.					7 ЪТ	37	т.			-	Κ.	_
GAPR APR og - GAPR on - GAPR GAPR Shelled turtle - GAPR PR APR APR PR PR PR PR PR Afr	- - - - - - - - - - - - - - - - - - -	- - - - - - - - - - - - - - - - - -	- - N - - - -	- - K - -	- - G] - -	- - -	 Y I	- -	-						1	v e	θQ	F′	ΓĿ	2 T.N	•	LI	ĸ	Р	L!	ĸ	I
APR og - GAPR on - GAPR GAPR shelled turtle - GAPR PR APR APR PR PR PR PR PR Afr	T - - - - - - - - - - - -	- D - - - - - -	- - - -	- K - -	- G] -	M	Y I	-																			
og - GAPR om - GAPR GAPR shelled turtle - GAPR PR APR GAPR PR PR PR PR 4r	T - - - - - - - - - - - -	D - - - - - - -	N - - - -	- - -	G] - -	M	ΥI	-	-			-		-	-		-				-	-	-	-	-	-	-
ion - GAPR GAPR shelled turtle - GAPR PR APR GAPR PR PR PR PR 4r	- - - - - - - - - - - -	- - - -	- - -	- - -	-							-		-	-		-				-	-	-	-	-	-	-
ion - GAPR GAPR shelled turtle - GAPR PR APR GAPR PR PR PR PR 4r	- - - - - - - - - - - -	- - - -	- - -	- - -	-			A		_	-		P A	-		_	-		_		-		_	_		_	_
GAPR shelled turtle - GAPR PR APR APR GAPR PR PR PR 4r	- - - - -	- - - -	-	-	_	-	- -																				
shelled turtle - GAPR PR APR GAPR GAPR PR PR PR 4r	- - - - -	- - -	-			- 1		-	-			-		-	-		-				-	-	_	-	-	-	_
PR APR APR GAPR PR PR PR 4r	- - - - -	- - -	-			- 1		-	-			-		-	-		-	-			-	-	-	-	-	-	-
APR APR GAPR PR PR PR 4r	- - - - -	- - -	-	-		-	_			_	-		_	-		_	-	-	_	-			-			-	-
APR GAPR PR PR PR 4r	- - - -	-	-																								_
APR PR PR PR 4r	-	-			-		_			_	-		_	-		_		-	_	-			-				-
PR PR 4r	-	-																									-
PR PR 4r	-		-		_	_	_			_	_					_			_	_			-		_	_	-
PR PR 4r	-	1 -																									-
PR. 4r	_	-	-			-	_			_	_		_	-		_		_	_	_			-		_	_	-
4r	-	-	-			-	_			_	_			-		_		_	_	_			-		_	-	-
		-	-	-	-			-	-			-		-	-		-	-		-	-	-	-	-	-	-	-
	-					_																			_	_	
CAP-related	-	-	-		_	_	_			_	-		_	-		_	-	_	-	-		_	_	_	_	_	+
1	-	-	-		_	_	_			_	-		_	_		_		_	-	-			_	_	_	_	+
like	-	_	_			_	_			_	_					_		_	_	_			_	_	_	_	_
- Unchar.	-					_							_			_							_	_	_	_	
char.	-												_														
p - EFG1	Т																										
Unchar.	-																										
- PRY1-like	-																										
bee - EFG1	-																										
e - EFG1	-																										
i) - GAPR	-																									T	T
GAPR	-																										1
APR	-	1								+		Π					Π									+	1
og - GAPR	-	T							\square			Ħ	+			+	Π		+				+	1	+	+	1
0	-	1							\square	t		Ħ	+		\square	1	Π		+					1	+	+	1
GAPR	-	t	1		\dashv	+	-		\vdash	+	1	Ħ	+	1	\square	+	Ħ		+	-				+	+	+	+
	-	t	-		+	+	-		\vdash	+	-	\square	+	-	\vdash	+	Η		+				+	+	+	+	+
	-	+	-		+	+	+		\vdash	+	-	\vdash	+	+	\vdash	+	\square	+	+	-		+	+	+	+	+	╡
	-	-	-		+	+	+		\vdash	+	+	\vdash	+	+	\vdash	+	\vdash	+	+	-		+	+	+	+	+	+
	_	-	-		\vdash	+	-		\vdash	+	+	\vdash	+	+	\vdash	+	\vdash	-	+	-		\vdash	+	+	+	+	+
	_	-	-	\square	+	+	-		\vdash	+	-	\vdash	+	-	\vdash	+	\vdash		+	-		-	+	+	+	+	+
Ark	-	-	-		+	+	-		\vdash	+	-	\vdash	+	-	\vdash	+	\vdash	-	+	-		-	+	+	+	+	+
סר	-	-	-			-	-	\square	\vdash	+	-	$\left \right $	_	-	\vdash	+	\vdash		+	-			-	_	+	+	+
	-	-	-			_	-		\square	-	-		_	-	\square	-	\square		+	-			_	_	\rightarrow	+	4
PR	-	-	-			_	_		\square	_			_			_			_	_			_	_	\rightarrow	_	4
PR	-																										
						E	ion	14				E	xoi	n 5	; 🗖			E	xo	n ć	5				E	Exe	
					F																						
	10					X	on	11				Ex	on	12	2]]	Ex	on	13	3]	Ex	(0)	n
	g - GAPR on - GAPR GAPR helled turtle - GAPR R PR PR APR APR R R R R	g - GAPR - on - GAPR - GAPR - helled turtle - GAPR - PR - ON 2 Exon 3	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - PR - ON 2 Exon 3	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - ON 2 Exon 3	g - GAPR - - on - GAPR - - GAPR - - GAPR - - helled turtle - GAPR - - R - - PR - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - R -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - R - R - R - R - R - R - R - R - - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - R -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - R - R - R - R - R - R - R - PR -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - R - R - R - R - R - R - R - PR -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - R - R - R - R - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - R - R - R - R - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R <td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R<td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - - - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - R<td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - APR - - - R - R - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - <td< td=""><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - - - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - PR - PR - PR - APR - - - R - - - APR - - - R - R - R - R - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - PR - PR - APR - - - R - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td></td<></td></td></td>	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R <td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - - - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td> <td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - -</td> <td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - R<td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - APR - - - R - R - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - <td< td=""><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - - - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - PR - PR - PR - APR - - - R - - - APR - - - R - R - R - R - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - PR - PR - APR - - - R - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td></td<></td></td>	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - - - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - R <td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - APR - - - R - R - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td> <td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - <td< td=""><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - - - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - PR - PR - PR - APR - - - R - - - APR - - - R - R - R - R - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - PR - PR - APR - - - R - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td></td<></td>	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - APR - - - R - R - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - <td< td=""><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - - - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - PR - PR - PR - APR - - - R - - - APR - - - R - R - R - R - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td><td>g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - PR - PR - APR - - - R - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -</td></td<>	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - - - PR - - - APR - - - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - PR - PR - PR - APR - - - R - - - APR - - - R - R - R - R - R - R - R - R - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - R - PR - APR - - - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -	g - GAPR - on - GAPR - GAPR - GAPR - helled turtle - GAPR - R - PR - PR - APR - - - R - PR - PR - APR - - - R - R - R - R - PR - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Appendix 8: This figure provides MEGA7 ClustalW amino acid sequence alignment of invertebrate and vertebrate GAPR sequences. The CAP signature sequences are boxed, the highly conserved cysteine are highlighted in yellow with a red font color. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online data for specific CAP genes. Color-coding superimposed on the sequences is used to represent different exons and exon borders are indicated by a change in color. Organisms and accession IDs of the amino acids are available in the supplemental data table 16. Page 10 of 10

Classification	Common name - Protein name													Am	ino	A	cid	Se	equ	en	ce											
																			-													
					Γ			E	xo	n 1/	21	bor	der																	-		
	A. gephyra - SCP-like	-	Μ	G /	A A	1	G	Τč	S	δE	A	A	Т	Κ	Р	Т	A	A	Ρ	G	A	S	S S	S I	L-	-	Α	Α	ΕÌ	V N	7 A	ſ
Prokaryotes	M. stipitatus - SCP-like	s	S	QI	R S	5 1	ГТ	ΓA	۱S	6 A	, Ç	Q R	Α	Ρ	Т	Р	Α	R	Α	Α	Ρ	Ρ	ΕI	P 1	R D	F	Α	L	ЕÌ	л	7 ς	þ
	S. amylolyticus - SCP-like																															
	P. reichenowi - CRISP-like	Ν	Ι	F ·	- I	FΙ	, F	F	V	7 F	I	S	S	Υ	Ι	Υ	V	Ν	G	Q	F	C	ΚI	FΙ	N -	-	Κ	Е	F	I K	C E	5
	I. multifiliis - SCP-like	Е	R	ΜI	L C	G F	F	K	G	βN	ГI	L	Υ	L	Р	G	Q	I	Ρ	L	Q	Т	NI	Е	G-	-	Р	S	A١	V F	tΕ	ול
Protozoa	P. maninus - CRISP-like	М	Κ	ΤA	A I	I١	7 6	βF	F	· v	Ľ	, Q	S	Т	s	v	Н	G	Α	F	Е	E.	A I	E (G-	-	V	Е	G٦	N N	/ K	2
	H. hammondi - SCP-like	Е	Т	GΙ	L C	2 1	ΓF	F Q) F	εE	G	Y	R	D	Т	Т	Q	s	G	Α	Т	F	ΕI	י ס	V -	-	Т	D	E		J	
	D. discoideum - Hypoth. Prot.	М	K	FΙ	L A	A K	CΙ	J	F	L	, I	L	I	S	s	С	V	Ν	s	Q	L	Т	P I	NI	Е-	-	I	Т	r _	I Ç	χN	Ţ
	C. militaris - SCP/PRY-like	А	S	D V	V 1	ΓV	7 1	ΓΙ	T	ΓV	' A	L P	А	Ι	Ρ	S	Q	E	Ρ	S	Υ	S	ΚI	R 1	ΥV	ΓF	Т	S	Α]	VΙ	. 1	T
Transf	P. minimum - SCP-like	Ρ	S	ΡI	D 1	ΓV	/Ί	ΓV	'Ί	ΓA	A	P	S	Ι	Ρ	s	T.	Α	Ρ	Q	F	V	D	G 1	DТ	F	Т	S	Α	ΙL	. N	Ţ
Fungi	N. parvum - SCP-related	G	G	S 5	S -	- 0	βN	I N	I A	۹ G	T	N	Т	D	Т	s	S	s	G	S	G	Т	s :	s (GG	A D F A L E M V Q A - I R A V L R V L R V L R V L R V L R V L D I I K E F I K E - P S A V D I T I C I	2					
Cnidaria H. H. H. N. B. T.	B. maydis - Hypoth. Prot.	Α	V	A A	A I	? A	A E	ΗĽ	A	A A	P	R	N	V	Q	A	G	G	Α	Р	н	Q	DI	D I	ΡK	F	Ι	S	A	νN	ΛF	Ł
	H. vulgaris - PR-1b-like	Α	F	LΝ	Vŀ	CΙ	L S	δL	, W	VΕ	C	V	Α	D	L	Κ	S	Е	Α	F	Ι	Т	ΕI	N 1	D -	-	Κ	W	I.	A L	. I	5
Cuideala	H. vulgaris - GAPR1-like	Q	I	V I	V L V L L V L L V L L V L L V L L V L L V L L N N Y G S I G R A S N N C L N N C L N N C L N N C L N N C L N N C L N N C L N N C L N N C L N N C L N N C L N N C L N N N L I N	7	7																									
Cnidaria I I I	H. vulgaris - GAPR1-like	L	L	V A	A I	I S	5 V	ΓL	, L	. v	s	N	S	Υ	G	D	S	I	G	K	R	Α	S I	D	Q -	-	S	Ν	N		, K	2
	N. vectensis - Predicted Prot.	-	-		- -	- -	· -		-	· -	-	-	-	-	-	-	-	-	-	-	-	-		-	- -	-	-	-	- 1		· -	
	B. malayi - SCP-like	F	V	ΜI	FI	? A	A F	- I	T	[V	V	A	G	F	Е	С	Ρ	G	G	R	L	Т	P (Q	Q -	-	R	Ι	D	ΙV	/ N	ī
	T. spiralis - Ion-1 Protein	Ν	Q	S 1	LI	D A	V I	/ P	L	L K	i V	' K	R	S	Υ	R	F	н	Ρ	F	Υ	Κ	S I	W I	К	-	Q	S	M I	ΡL	. E)
	Nematodes (C. e.) - SCL 12	-	-	- 1	Μ	C F	F A	ΥL	, Y	ζL	, I	A	Ι	F	G	С	Α	Т	Α	Q	F	S	S 🤇	Г	A -	-	Q	Т	A	I V	/ K	2
	Nematodes (C. e.) - SCL 14	Ν	С	- - - - - - - M F P A F I T V V A G S L D A V P L K V K R - M K F A L Y L I A I Y F R L I C L I F S - - M Y F I L C L L K P T P R L Y G E R V K P T P R L Y G D R V K P T P R L Y G D R V	F	С	Е	Т	L	С	Е	F	S	ΕC	Г	G-	-	Κ	Ν	Y	ΙL	, s										
	Nematodes (C. e.) - SCL 1	-	-		- -	- N	ΛY	(F	٦ I	ΙL	C	L	L	Α	Ι	G	V	N	A	Q	F	Т	S 1	Γ	G-	-	Q	S	S	I V	ΙĽ)
	Giant honeybee - CRISP1-like	R	S	ΚI	<mark>9</mark>]	Γ	P F	ιL	Y	ζG	E	R	V	Ρ	V	Κ	L	L	R	Т	S	Ν	S I	Κľ	V -	-	R	Q	K	I V	/ D)
	Bumble bee - CRISP2-like	R	S	ΚI	P 1	Γ	PF	ιL	Y	ζG	D	R	V	Ρ	V	К	L	L	R	Т	Ν	S	G 1	R 1	V -	-	R	Q	ĸ	ΙV	/ E)
Insects	Pharaoh ant - CRISP1-like	R	S	ΚI	<mark>P</mark> 1	Γ	PF	ιL	Y	ζG	E	R	V	Ρ	V	s	L	Ι	R	Т	Т	S	S 1	Кľ	V -	-	R	Q	К	r 1	/ E)
	Red flour beetle - CRISP2-like																															
	Asian swallowtail - DIS2-like	V	R	QV	V S	5 1	N F	۲Ç	2 L	J	P	P	Ν	Κ	Ι	Ρ	D	E	Α	L	Ν	Ρ	R I	R 1	ΝV	V V	R	R	К	t V	V Q A L D A K E R V K D L I L Q N Y L N S V K L N S V K L N S V K L N S V K L L N S V K L L V Q V N Q V N Q V D T V D	
Molluscs.	Sea squirt (C. i.) - CRISP-like	-	-		- 4	A F	₹E	I R	L A	λK	R	l A	Ρ	G	D	L	L	Е	Т	V	V	L	ΤI	Εı	A -	Е	K	Ι	E	ſΝ	/ Γ	Ł
Echinoderms,	S. kowalevskii - Uhchar. Prot.	-	-			- 1	[V	7 V	' L	J	L	, S	Q	D	L	s	P	W	V	W	Ρ	F	ΤJ	ΗI	Е-	Q	K	R	A	<mark>۱ ۱</mark>	ΓE	¢.
and	Florida lancelet - Hypoth. Prot.	-	-		- -				-		-	-	-	-	-	-	-	-	Α	Т	D	L	S 1	ΗI	Е-	Q	Ι	Ν	Т	I L	. E	ţ.,
Urochordates	Purple sea urchin - CRISP2-like	-	-				• •	-	-		-	Μ	A	Е	D	L	Т	Е	Q	G	L	Т	DI	DI	Е-	-	Κ	S	Е	I I	. A	L
Urochordates	Pacific oyster - GLIPR1-like1	-	-						-		-	-		_			_	_	_	_	_	_	_	_	_			_				
	Zebrafish - CRVP-like	L	G	FΙ	Ŀ	ΗN	ΛS	5 A	۱A	V C	S	V	S	G	V	С	Т	E	L	S	S	V	-	-		-	Q	Q	E	1 1	/ Ε)
	Burmese python - CRVP-like				~				× .							_											_			r 1	/ E)
Vertebrates	E. caballus - CRISP2																															
	H. sapiens - CRISP1																											Е	E	<u>1</u>]	7 N	Ţ.
	H. sapiens - CRISP3	Α	G	LΙ	LI	2 5	ŝF	P	A	A N	E	D	Κ	D	Ρ	A	F	Т	A	L	L	Т	T (Q	ΤÇ	ξV	Q	R	E	t y	7 1	ſ
Exon 1	Exon 2 Exc	on i	3]	Ex	on	4				E	xo	n f	5				E	xo	n (5			F	Exo	n '	7			
С	Conserved cysteines					Si	ite	s 0	ofi	noi	n-a	ali	gn	ed	se	qu	ıeı	ıc	e o	m	iss	sio	n									

Appendix 9: Evolution of the Pathogenesis-Related (CAP/PR) Domain in the CRISP Protein Lineage

Appendix 9: Page 1 of 7

														Ex	on 2/3	1		
		CAP3													rder			
	A. gephyra - SCP-like	HNQA	RA	ΚA	RP	• T 1	PA	ΚA		LP	P			- I	SW	SI	P E	AA
Prokaryotes	M. stipitatus - SCP-like	HNEA	RR	ΕA	RP	• т і	РΚ	РР		LF	Р			- 1	. v w	S I	D A	A A
•	S. amylolyticus - SCP-like	HNTR	RA	ЕН				- C		A F	Р			- 1	s w	S I	ЭE	LA
	P. reichenowi - CRISP-like	HNDF	RL	КН				- N		ΑK	Р			- I	K W	SF	κ.	LE
	I. multifiliis - SCP-like	MNFL	NR	00				- S		LE	Р			- I	KL	NI	00	M S
Protozoa	P. marinus - CRISP-like	HNYF													/ E W			
	H. hammondi - SCP-like	HNKF	RΤ	ΑG		- 1	LG	νs		VF	Е			- I	V A	ΝI	D E	AR
	D. discoideum - Hypoth Prot.	HNQW	RS	ΝP				- N	- G	ΡŢ	P	V T I	PIS	GI	V Y	NB	ΞS	ΙA
	C. militaris - SCP/PRY-like	TNTY	RR	QH				- N		A S	A			- 1	L A W	N A	A Τ	LA
	P. minimum - SCP-like	TNFY	RΤ	ΕН				- N		A S	W			- 1	IRW	NI	ΞТ	LΕ
Fungi	N. parvum - SCP-related	нику	RA	КН				V D		ТF	Р			- I	ΕW	DB	E S	LA
	B. maydis - Hypoth. Prot.	нwүw	RR	LH				- C		ΑÇ	D			- I	vw	DB	ςЕ	LA
	H. vulgaris - PR-1b-like	HNKL	RS	ΜH		-		- Y		ΑI	A			- 1	LW	S 1	V Q	LE
6 H I	H. vulgaris - GAPR1-like	НИКҮ	R L	LΗ				- N		SΕ	[P]			- N	ΛTL	S 1	ΓS	LN
Cnidaria	H. vulgaris - GAPR1-like	HNRL	RΑ	КН				ΕN		ΤF	R			- I	U V L	DI	V V	LΤ
	N. vectensis - Predicted Prot.	MNK Y	RM	LН				- A		S P	Ρ			- I	RI	ΝŊ	ζD	МT
	B. malayi - SCP-like	HNEF	R I	ΗG	ΚL	, K 1	NK	DΝ	I F	ΜP	R (βK.I	ΝML	QN	<mark>1</mark> T W	S C	2 Q	LE
	T. spiralis - Ion-1 Protein	KN	- A	. R V				- N		ΜF	M			- I	KW	N S	Ϋ́	LΑ
Round Worms	Nematodes (C. e.) - SCL 12	HNDL	RΑ	LG	ΝY	D	A A	GΤ	- I	ΕP	P /	A A I	NMR	K	K W	D S	5 Т	V A
	Nematodes (C. e.) - SCL 14	ΗΝΥL	R A	LG	КΥ	V 4	A G	ΝS	- T	ΚP	S A	A S I	ΝMΝ	ſΚ Ι	JIW	DI	ΤЛ	LE
	Nematodes (C. e.) - SCL 1	HNKL	RΑ	ΚS	ΤΥ	V I	A K	GΤ	- K	ΚE	P	A Τ I	D M R	ΚN	<mark>1</mark> V V	D S	sт	V A
	Giant honeybee - CRISP1-like	ΗΝΥL	R T	QV	·					ΚP	P /	A A I	ΝML	VN	<mark>1</mark> K W	ΗI	P G	LΑ
	Bumble bee - CRISP2-like	ΗΝΥF	RΤ	QV						ΝF	ΡA	A A I	ΝML	ΑN	<mark>1</mark> K. W	н	P G	LG
Insects	Pharaoh ant - CRISP1-like	ΗΝΥF	RΤ	QV						ΚP	S A	A A I	ΝML	VN	<mark>1</mark> K W	H S	G	LΑ
	Red flour beetle - CRISP2-like	ΗΝΥF	RΤ	ΚV						VP	R A	A A I	ΝML	RN	<mark>1</mark> R W	Ηŀ	C G	A A
	Asian swallowtail - DIS2-like	HNFF	R S	ΚV						QP	ΤA	A S I	ΝML	LN	<mark>1</mark> SW	ΝI	ΡI	A A
Molluscs.	Sea squirt (C. i.) - CRISP-like	ΗNQΥ	RS	RΑ						Ν-	- 2	A S I	DΜL	AN	4 S W	DI	ТΓ	ΙA
Echinoderms.	S. kowalevskii - Uhchar. Prot.	HNML														DI	D	LΑ
and	Florida lancelet - Hypoth. Prot.	НИҮҮ														DI		LA
Urochordates	Purple sea urchin - CRISP2-like	ΗΝQΥ								<u> </u>					_	ΝI) E	ΙA
crochordates	Pacific oyster - GLIPR1-like1												ΝMV	_) G	LA
	Zebrafish - CRVP-like	H N A F	R R	. A V						QF	S A	A S I	NML		AS W	SI) A	V A
	Burmese python - CRVP-like	HNNL		тν						S P	ΤA		NML				۲D	A A
Vertebrates	E. caballus - CRISP2	HNEL		s v						S P	ΡA		NML			SI	¢Ε	ΑI
	H. sapiens - CRISP1	H N A L	R R	. R V						VF	Ρ.	AS 1	ΝML	ΚN	<mark>I</mark> SW	SI	ΕE	A A
	H. sapiens - CRISP3	HNEL	RR	. A V						S F	P .	A R	N M L	K I	<mark>a</mark> e w	NI	ΣE	A A
Exon 1	Exon 2 Exo	on 3 📃	E	xon	4 🗖		E	ion :	5 드		Ex	on 6	5]	Exon	7		I.
С	Conserved cysteines		Si	tes o	fno	n-ali	igne	d se	equei	ice	omi	ssio	n					

Appendix 9: Evolution of the Pathogenesis-Related (CAP/PR) Domain in the CRISP Protein Lineage

Appendix 9: Page 2 of 7

S. anylolyticus - SCP-like R T A Q - S W A D D L A R G - C - A F E H N R T P																									F	Exor	n 3/4	
Prokaryotes M. stipitatus - SCP-like R T A Q - S W A D L A R G - C - A F E H N P N R															Cl											bor	der	
S. myklytous - SCP-like R T A Q - S W A D D L A R G - C - A F E H N - R T P - R T P - R T P - R T P - R T P - R T P - R T P - R T P - R T P - R T P - R T P - R T P - R T P - R T P - P P - P P - P P - P P T T A D T T S CP-like E T A T T Y E A N D T A R G - C - T V S A K - R T P - R T P - P O G S T M T S P A R T A D T S CP-like K A Q - D H A N D T A L N L R N S D C - T V S A K - R T P - R T P - R T P - P O G S T M T S P A R T A D T S CP-like K A Q - D H A N D T A L N D T G K G S S E T H - R T P - R T P - R T P - R T P - P O G S T M T S P A R T A D T S CP-like D discolden - Hypoth Prot T A U N L N L N L K K R A D N C - S S E T H - R S E E A R R A D A D D C - N F S H S - R G P - C - P - C - C - C - C - C - C - C -		A. gephyra - SCP-like	K	V.	A Q	2 -	Α	Ύ	A N	I K	-		-	-	C	- I	F	Е	ΗI	N		-	-]	ΚN	R		G	Κ
Preichenwi-CRISP-like EIAT YEANE IRNSD C IVSAK	Prokaryotes	M. stipitatus - SCP-like	R	Q.	A A	L -	S	W 4	A K	ζA	-		-	-	С	- F	F	Е	ΗI	N		-	- 1	ΡN	R			G
Imakifilis - SCP-like K A A Q - D H A N D I G K G - I T G H N G S D G S T M T S Protozoa P. marinus - CRISP-like A E S Q - QW A D H L A E G - G - S L V H S G S Y S A B D. discoldeum - Hypoth Prot. T A L N - L N L R K C D S Y - S N M Y Q Y A		S. amylolyticus - SCP-like	R	Τ.	A Ç	2 -	S	W A	ΑE	D	L	A R	۲ G	÷ -	С	- A	F	Е	ΗI	N ·		-	- 1	RТ	Р			-
Protozoa P. marius - CRISP-like A E S Q - QW A D H L A E G - G - S L V H S G S Y S S A P H harmondi - SCP-like K V V L - E I V K K R A D N - C - S S E T H G S Y S S A P C. militaris - SCPPRY-like K E S S - Q W A D H L A E G - G - S L V H S G G P		P. reichenowi - CRISP-like	Е	Ι.	ΑT	-	Υ	ΕÆ	ΑP	ΙL	Ι	RN	I S	D	C	-]	V	S	A 1	K		-	-		-			-
Hammondi - SCP-like K V V L E I V K K R A D N C S S E T H S S E E A R R A Pingiano C. miliaris - SCP/PR-like S F A S S V L A A A R D A C N N V Q Y A		I multifiliis - SCP-like	K	Α.	ΑÇ	2 -	D	ΗÆ	A N	I D	Ι	GΚ	G	÷ -	Ι	ΤС	Η	N	G	S -		-	- 1	DG	S	ΤN	<u>1</u> T	S
D. discoideum - Hypoth Prot. T A L N - L N L R K C D S Y - S N M Y Q Y A	Protozoa	P. marinus - CRISP-like	Α	Е	s ç	2 -	Q	W 4	ΑI	н	L	ΑE	G	÷ -	G	- 5	L	V	н	s -		-	- (GS	Υ	S S	A	Ρ
Fungi C militaris - SCP/PRY-like S F A <td< td=""><td></td><td>H. hammondi - SCP-like</td><td>к</td><td>V '</td><td>VΙ</td><td>, -</td><td>Е</td><td>ΙV</td><td>VΚ</td><td>CΚ</td><td>R</td><td>ΑD</td><td>) N</td><td>1 -</td><td>C</td><td>- 5</td><td>S</td><td>Е</td><td>ΤI</td><td>Η</td><td></td><td>-</td><td>- 3</td><td>SΕ</td><td>Е</td><td>ΑF</td><td>ε R</td><td>Α</td></td<>		H. hammondi - SCP-like	к	V '	VΙ	, -	Е	ΙV	VΚ	CΚ	R	ΑD) N	1 -	C	- 5	S	Е	ΤI	Η		-	- 3	SΕ	Е	ΑF	ε R	Α
Fungi P. minimum - SCP-like D F A S D Y L D D M D C C D F K H S C Q F A S O Y L D D M D C C D F K H S C O F K H S C Q N N N. parvum - SCP-related Q T A T A H S Q K C V F E H S S N L S S V S P P P P T Mulgaris - Hyoth Prot. D A A R A D I E E [C T D P E H K R P G S N L S S V S P P P P T H vulgaris - GAPRI-like K A A E D R V S A I L Q N V N ME S Q L V A S K S N N V H vulgaris - GAPRI-like K F A Q E W A D Y L A N N - L F E H R - - - S D S N. vectenis - Predicted Prot. N K A E L W A D Q C T F G H S - - - S D S Nematodes (C. e) S C L 12 S S A Q Q H A N R - - C - P D D H S - - K R - - - S R - - - - - - S D S N - - - S D S N - - S D S		D. discoideum - Hypoth Prot.	Т	A	LN	τ-	L	ΝI	ĿF	ιк	С	D S	Y	7 -	S	ΝN	4 Y	Q	Y.	A ·		-	-		-			-
Fungi N. parvum - SCP-related Q T A T - A H S Q K C C · V F E H S Q N N B. maydis - Hypoth Prot D A A R - A D I E E C T D P - E H K R P G S N L S S V S P - P P D H. ulgaris - PR-1b-like K A A E - D R V S A L L Q N - Q N M E S Q L V A S K S N E N L F S L H. ulgaris - GAPR1-like K A A E - D R V S A L L Q N - Q N M E S Q L V A S K S N E N L F S L H. ulgaris - GAPR1-like K A A E - D R V S AL L Q N - Q N M E S Q L V A S K S N E N L F S L N. vectensis - Predicted Prot. N K A E - L WA V K L A K D Q S S L N V D T S S D S N. vectensis - Predicted Prot. N K A E - L WA V K L A K D Q S S L N V D T S S D S Nematodes (C. e) - SCL 12 S S A Q - Q Y A N T C C T F G H S K D R S N N R act - A A S R A N N ematodes (C. e) - SCL 14 Nematodes (C. e) - SCL 14 T T A Q - D Y S T G C C P T G H S K G T E C Nematodes (C. e) - SCL 14 T T A Q - D Y S T G C C P T G H S K G T E C Nematodes (C. e) - SCL 14 T T A Q - D Y S T G C C P T G H S K G T E C Bumble be - CNISP1-like K A A Q - RWA N R C C L G L V H D N A T G L Y L D Bumble be - CNISP1-like K A A Q - RW A D R C C L G L V H D N A T G L Y L D		C. militaris - SCP/PRY-like	S	F.	A S	-	S	ΥI	L A	ΑA	A	RΕ) A	L -	С	- 1	I F	S	Η	S -		-	- 1	RG	Ρ		-	-
N. parvin - SCP-related Q T A T - A H S Q K [C] - V F E H S Q N M B. maydis - Hypoth Prot. D A A R - A D I E E C T D P - E H K R P G S N L S S V S P - P P T H. ulgaris - PR-1b-lke K A A E - D R V S A L L Q N - Q N M E S Q L V A S K S N E N L F S I H. ulgaris - GAPR1-lke A F A Q - E W A D Y L A N N L F E H R S D S N. vectensis - Predicted Prot. N K A E - L WA V K L A K N L F K H S D S N. vectensis - Predicted Prot. N K A E - L WA V K L A K N Q S S L N V D T S S D S N. vectensis - Predicted Prot. N K A E - L WA V K L A K D Q S S L N V D T S S D S N. vectensis - Predicted Prot. N K A E - L WA V K L A K D Q S S L N V D T S S D S Nematodes (C. e) - SCL 12 S S A Q - Q Y A N T C C T F G H S K D R T Nematodes (C. e) - SCL 14 T T A Q - D Y S T G C P T G H S A S R A Nematodes (C. e) - SCL 14 A S A Q - R WA N R C P T G H S A S R A Nematodes (C. e) - SCL 14 A S A Q - R W A N R C C L G L V H D N A T G L Y L D Moluses, Giart Inneyber - CRISP1-lke K A A Q - R W A N R C C L G L V H D N A T G L Y L D Asia svalowala ⁻ DIS2-lke K A A Q - K W A N R C C L G L V H D - N A T G R Y	. .	P. minimum - SCP-like	D	F .	A S	-	D	ΥI	LΕ	D	Μ	D -	-	ļ	С	- I) F	Κ	Н	S -		-	- (GG	Ρ			-
Cnidaria H. ulgaris - PR-1b-like K. A. A. E D. R. V. S. A. L. L. Q. N Q. N. M. E. S. Q. L. V. A. S. K. S. N N -	Fungi	N. parvum - SCP-related	Q	Τ.	ΑT	- 1	А	нs	s ç	ξK	-		-	-	С	- 1	/ F	Е	Н	S -		-	-		-	- (N (Ν
Cnidaria H. wlgaris - GAPR1-like A. F. A. Q E. W. A. D. Y. L. A. N. N L. F. E. H. R E. N. L. F. S. L. F. W. L. Y. S. L. N. W. W. W. W. W. W. W. L. A. K. D Q. S. S. L. N V. D. T Q. T. F. Q. H. S V. D. T Q. T. F. Q. H. S V. D. T Q. T. F. Q. H. S V. D. T Q. T. F. Q. H. S V. D. T Q. T. F. Q. H. S V. D. T Q. T. F. Q. H. S V. D. T Q. T. F. Q. H. S V. D. T Q. P. T. G. H. S V. D. T Q. N. A. T. Q. P. Y. S. T. G C P. T. G. H. S K. D. R Q. N. M. T C P. T. G. H. S K. G. T C. P. T. G. H. S K. G. T C. P. T. G. H. S C. P. T. G. H. S		B. maydis - Hypoth. Prot.	D	Α.	A R	٤ -	А	D	ΙE	ЕΕ	С	ΤD	P	• - '	E	H K	R	. Р	G	S 1	ΝL	S	S 1	V S	Ρ	- I	• P	D
Cnidaria H. vulgaris - GAPR1-like K. S. A. Q K. Y. A. E. Y. L. A. N. N L. F. K. H		H. vulgaris - PR-1b-like	K	Α.	ΑE	- 1	D	RΥ	V S	δA	L	LΟ) N	1 -	-	QI	ΙM	ΙE	S (Q I	ιv	Α	S 1	ΚS	Ν		-	-
Mainter H utgaris - GAPR1-lake K S A Q - K Y A E Y L A N N L F K H S D S N vectensis - Predicted Prot. N K A E - L WA V K L A K D Q S S L N V D T	C 11 - 1	H. vulgaris - GAPR1-like	Α	F .	ΑÇ	<u>)</u> -	E	W A	ΑI) Y	L	ΑN	I N	1 -	-	- I	, F	Е	Η	R ·		-	-	- E	Ν	LΒ	S	L
B. malayi - SCP-like N S A Q - T W A D Q C - T F G H S P R N - Q R Q R G Round Worms Nematodes (C. e.) - SCL 12 S S A Q - Q Y A N T C - T F Q H S R K D R T Nematodes (C. e.) - SCL 12 S S A Q - Q Y A N T C - P D D H S A S R A A Q - R W A N R C L G L V H D N A T G L Y L D R A S A Q - R W A N R C L G L V H D N A T G L Y L D R A S A Q - R W A D R C L G L V H D N A T G R H I E N Y L D R A S A Q - S W A D R C I F L Q H N D P L E N T I P Y L C R A S A Q - S W A D R C I F L Q H N D P L E N T I P Y L C R A S A Q - S W A D R C I F A H T P D K E Y E A T I F C A A A Q - K W A R R C I D F A H N S R K N Q S R N N Q S R A D Q R S S I S G F V L P A S R N N Q S R R N N Q S R R N N Q S R R N N Q S R R N	Chidaria	H. vulgaris - GAPR1-like	K	ς.	A Ç	2 -	Κ	ΥÆ	A E	ΕY	L	A N	I N	1 -	-	- I	, F	Κ	н			-	-		-	- 5	5 D	S
T. spiralis - Ion-1 Protein F Q A Q - R H A N R C C - T F Q H S K D R T Round Worms Nematodes (C. e.) - SCL 12 S S A Q - Q Y A N T C C - P D D H S K D R T Nematodes (C. e.) - SCL 14 T T A Q - D Y S T G C C - P T G H S A S R A S R A Nematodes (C. e.) - SCL 14 A S A Q - N Y A N T C C - P T G H S K G T E C - P T G H S K G T E Mematodes (C. e.) - SCL 1 A S A Q - N Y A N T C C - P T G H S K G T C - P T G H S K G T Bumble be - CRISP1-like K A A Q - R W A N R C C L G L V H D N A T G L Y L D H G H S K G T Phraoh ant - CRISP1-like K A A Q - K W A N R C C L G L V H D N A T G L Y L D H G H S N A T G L Y L D H G H G D G Y L D - Molluscs, Sea squirt (C. i.) - CRISP-like R A A Q - K W A D R C C I F L Q H N D P L E N T I P Y L G G S N A G C And Sea squirt (C. i.) - CRISP-like Q F A S - N Y I S N C C D F A H S T N A Q R S S I S G F Purple sea urchin - CRISP2-like Q F A S - N Y I S N C C D F A H N S R R N N Q S S T E W Aud Urochordates S. kowalevski - Uhchar. Prot. Q L A Q - D W S E E C		N. vectensis - Predicted Prot.	N	К.	A E		L	W 4	1 1	7 K	L	A K	D) -	-	- 0) S	S	LI	N		-	- 1	V D	Т			-
Round Worms Nematodes (C. e) - SCL 12 S S A Q - Q Y A N T C P D D H S G T E C P T G H S G T E A S R A S Nematodes (C. e) - SCL 14 T T A Q - D Y S T G C P T G H S A S R A S R A Nematodes (C. e) - SCL 1 A S A Q - N Y A N T C P T G H S A S R A S R A Giant honeybee - CRISP1-like K A A Q - R W A N R C C L G L V H D N A T G L Y L D D Y L D Bumble bee - CRISP2-like K A A Q - R W A D R C C L G L V H D N A T G L Y L D D Y L D A sian swallowtail - DIS2-like K A A Q - K W S R K C C F V L T H D N I T G R H I E N Y C Asian swallowtail - DIS2-like K Q A Q - S Y A D K C C D F A H S T N A Q R S S I S G F S ca squirt (C. i) - CRISP-like Q F A S - N Y I S N C C D F A H S T N A Q R S S I S G F S kowalevski - Uhchar. Prot. Q L A Q - D W S E E C C D F A H T P D - K E Y E A T I P Purple sea urchin - CRISP2-like K M A D - R W A R R C C D F A H T P D - K E Y E A T I P Purple sea urchin - CRISP2-like K M A D - R W A R R C C D F A H S - P V N Q R T V G Purple sea urchin - CRISP2-like K M A D - R W A R R		B. malayi - SCP-like	Ν	S .	ΑÇ	2 -	T	W 4	ΑI) Q	-		-	-	С	- 1	F	G	H	S -		-	- 1	ΡR	Ν	- (Q R	Q
Nematodes (C. e.) - SCL 14 T T A Q - D Y S T G C - P T G H S A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R A S R R R A S R R R A S R R R A S R R R R A S R R R R A S R R R R A S R R R R A S R R R R R A S R R R R R A S R R R R R A S R R R R R A S R R R R R R R R R R R R R R R R R R		T. spiralis - Ion-1 Protein	F	Q.	ΑÇ) -	R	ΗA	ΑN	I R	-		-	-	С	-]	F	Q	H	S -		-	- 1	ΚD	R		-	Т
Nematodes (C. e) - SCL 1 A S A Q - N Y A N T C P T G H S K G T K G T K G T C Giant honeybee - CRISP1-like K A A Q - R W A N R C L G L V H D N A T G L Y L D Bumble bee - CRISP2-like K A A Q - R W A D R C L G L V H D N A T G L Y L D Pharaoh ant - CRISP1-like K A A Q - R W A D R C L G L V H D N A T G L Y L D Red flour beetle - CRISP2-like R A A Q - R W A D R C C I F V L T H D N I T G R H I E N Y G Asian swalowtail - DIS2-like K Q A Q - S Y A D K C C I F V L T H D N I T G R H I E N Y G Asian swalowtail - DIS2-like K Q A Q - S Y A D K C C I F A H T P D - K E Y E A T I P Sea squirt(C.i) - CRISP-like Q F A S - N Y I S N C C D F A H S T N A Q R S S I S G F S. kowalevski - Uhchar. Prot. Q L A Q - D W S E E C C D F A H N S H R G D G Y Florida lancelet - Hypoth Prot. T I A Q - G W A D G C C D F A H N S R R N N Q S Purple sea urchin - CRISP2-like K M S Q - S W T K R C C D L T L K G P S S R M L N G - Pacific oyster - GLIPR1-like! K M A D - R W A R R C C D K T H B S - N N R R R N N Q S Burmese python - CRVP-like E	ound Worms	Nematodes (C. e.) - SCL 12	s	S .	ΑÇ) -	Q	ΥĂ	٩N	Т	-		-	-	С	- F	D	D	н	s -		-	- (GТ	Е		-	-
InsectsGiart honeybee - CRISP1-likeK A A Q- R W A N R CL G L V H D N A TG L Y L DInsectsPharaoh ant - CRISP1-likeK A A Q- R W A N R CC L G L V H D N A TG R Y L DPharaoh ant - CRISP1-likeK A A Q- R W A D R CC L G L V H D N A TG L Y L DMolluscs, and UrochordatesSea squirt (C.i.) - CRISP-likeR A A Q- R W A D R CC D F A H S T N A Q R S SI S N G FMolluscs, and UrochordatesSea squirt (C.i.) - CRISP-likeQ F A S - N Y I SN CC D F A H S T N A Q R S SI S G FMolluscs, and UrochordatesSea squirt (C.i.) - CRISP-likeQ F A S - N Y I SN CC D F A H T P D K E Y E A T I FMolluscs, a HSea squirt (C.i.) - CRISP-likeQ F A S - N Y I SN CC D F A H N S H R G D G YSea squirt (C.i.) - CRISP-likeQ L A Q - D W S E E CC D F A H N S H R G D G YPurple sea urchin - CRISP2-likeK M S Q - S W T K R CC D F A H N S H R G D G YPurple sea urchin - CRISP2-likeK M A D - R W A R R CC N M T H G - P P S S R M L N GPurple sea urchin - CRISP2-likeK M A Q - K W A N K CC N M T H G - P P S S R M L N GPurple sea urchin - CRISP2-likeK N A K - R W A D K CC N M T H G - P P S S R M L N GPurple sea urchin - CRISP2A N A Q - K W A N K CC N M T H G - P P S S R M L N GPurple sea urchin - CRISP2A N A Q - K W A N		Nematodes (C. e.) - SCL 14	т	т.	A Ç) - (D	Y S	sт	G	-		_	-	С	- F	ч	G	н	s .		-	- 1	A S	R		_	Α
InsectsBumble be - CRISP2-likeK A A Q - K W A N R CCL G L V H D N A T G R Y L DInsectsPharaoh ant - CRISP1-likeK A A Q - R W A D R CC L G L V H D N A T G L Y L DRed flour beetle - CRISP2-likeR A A Q - K W S R K CC F V L T H D N I T G R H I E N Y GAsian swallowtail - DIS2-likeK Q A Q - S Y A D K CC I F L Q H N D P L E N T I P Y L GAsian swallowtail - DIS2-likeQ F A S - N Y I S N CC D F A H S T N A Q R S S I S G FSea squirt (C.i.) - CRISP-likeQ F A S - N Y I S N CC D F A H T P D - K E Y E A T I PFlorida lancelet - Hypoth Prot.T I A Q - G W A D G CC D F A H N S H R G D G YPurple sea urchin - CRISP2-likeK M S Q - S W T K R CC D F A H N S R R N N Q SPurple sea urchin - CRISP2-likeK M A D - R W A R R CC D F A H N S R R N N Q SPurple sea urchin - CRISP2-likeK M A D - R W A R R CC D F A H N S R R N N Q SPurple sea urchin - CRISP2-likeK M A D - R W A R R CC N M T H G - P P S S R M L N GPurple sea urchin - CRISP2-likeK M A D - R W A R R CC N M T H G - P P S S R M L N GPurple sea urchin - CRISP2-likeK N A K - R W A D K CC N M T H G - P P S S R M L N GPurple sea urchin - CRISP2-likeK N A K - R W A D K CC N M T H G - P P S S R M L N GPurple sea urchin - CRISP2A N A Q - K W A N K CC N M T H G - P P S S R M L N GPurple sea urchin - CRISP1Q N A R S K Y CN M T H G -		Nematodes (C. e.) - SCL 1	Α	s .	A Q) -	Ν	ΥĂ	A N	Т	2		-	-	C	- F	ч	G	н	s .		-	- 1	κG	т		_	-
InsectsPharaoh ant - CRISP1-likeK A A QR W A D RRCL G L V H DN A TG L Y L DRed flour beetle - CRISP2-likeR A A Q- K W S R KCF V L T H D N I T G R HI E N Y CAsian swallowtail - DIS2-likeK Q A QS Y A D KCI F L Q H N D P L E N TI P Y L CMolluscs, and uncholentandSea squirt (C.i.) - CRISP-likeQ F A SN Y I SNCD F A H S T N A Q R S SI S O FPerile sea urchin - CRISP2-likeQ L A QD W S E ECD F A H T P DK E Y E A T I FPurple sea urchin - CRISP2-likeQ L A QG W A D GCD F A HN S H R G D G YPurple sea urchin - CRISP2-likeK M S QS W T K RCD L T LN S R N N Q SPacific oyster - GLIPR1-like1K M A D -R W A R RCN M T H G - P P S S R M L N G-Purple sea urchin - CRISP2-likeE S A R - G W I N KCN M T H G - P P S S R M L N G-Pacific oyster - GLIPR1-like1K M A Q - K W A N KCN M T H G - P P S S R M L N G-Purple sea urchin - CRISP2A N A Q - K W A N KCN M T H G - P P S S R M L N G-Purple sea urchin - CRISP2A N A Q - K W A N K<		Giant honeybee - CRISP1-like	K	Α.	A Q) -	R	W 4	ΑN	I R	-		-	-	С	LC	÷L	V	ΗI	D -		-	N.	ΑT	G	LΥ	ΓL	D
Red flour beefle - CRISP2-like R A A Q - K W S R K C F V L T H D N I T G R H I E N Y C Asian swallowtail - DIS2-like K Q A Q - S Y A D K C I F L Q H N D P L E N T I P Y L C Molluscs, Sea squirt (C.i) - CRISP-like Q F A S - N Y I S N C D F A H S T N A Q R S S I S G F Sea squirt (C.i) - CRISP-like Q F A S - N Y I S N C D F A H S T N A Q R S S I S G F Forida lancelet - Hypoth Prot. Q L A Q - D W S E E C I F A H T P D K E Y E A T I F Purple sea urchin - CRISP2-like K M S Q - S W T K R C D L T L K G P S S T E W Pacific oyster - GLIPR1-like! K M A D - R W A R R C C D L T L K G P S S T E W Wertebrates Zebrafish - CRVP-like E S A R - G W I N K C C N M T H G - P P S S R M L N G Burmese python - CRVP-like E S A R - G W I N K C C N M T H G - P P S S R M L N G Burmese python - CRVP-like K N A Q - K W A N K C C N M T H G - P P S S A D D R K T S - T H sapiens - CRISP1 Q N A R S K Y C C N M T H G - P P S N A D D R K T S - T H sapiens - CRISP3 A N A Q - K W A N Q C N Y R - H S - N P K D R M T S - L		Bumble bee - CRISP2-like	K	Α.	A Q) -	K	w 4	A N	I R	-			_	С	LC	ιL	v	нI	D.		-	N.	ΑТ	G	RY	ΓL	D
Asian swallowtail - DIS2-like K Q A Q - S Y A D K C C I F L Q H N D P L E N T I P Y L C Molluscs, Echinoderms, and Sea squirt (C.i.) - CRISP-like Q F A S - N Y I S N C C I F A H T P D K E Y E A T I F Molluscs, Echinoderms, and Sea squirt (C.i.) - CRISP-like Q F A S - N Y I S N C C I F A H T P D K E Y E A T I F Purple sea urchin - CRISP2-like Q L A Q - D W S E E C C D F A H N S H R G D G Y Purple sea urchin - CRISP2-like K M S Q - S W T K R C C D F V H N S R R N N Q S Pacific oyster - GLIPR1-like! K M A D - RWA R R C C D N T H G - P P S S R M L N G N S R Vertebrates Zebrafish - CRVP-like E S A R - G W I N K C C N M T H G - P P S S R M L N G C Burmese python - CRVP-like K N A K - R W A D K C C N M T H G - P P S S A D D R K T S - T C H sapiens - CRISP1 Q N A R S K Y C D M T E - S - N P L E R R L P N T H sapiens - CRISP3 A N A Q - K W A N Q C N Y R - H S - N P K D R M T S - L	Insects	Pharaoh ant - CRISP1-like	ĸ	Α.	A Q) -	R	w 4	ΑI) R	2		-	-	С	L C	ιL	v	нI	D -		-	N.	ΑТ	G	LŊ	ΓL	D
Molluscs, Sea squirt (C.i) - CRISP-like Q F A S - N Y I S N C C D F A H S T N A Q R S S I S G F Schwalevskii - Uhchar. Prot. Q L A Q - D W S E E C I F A H T P D - K E Y E A T I F Forida lancelet - Hypoth Prot. T I A Q - G W A D G C D F A H N S H R G D G Y Purple sea urchin - CRISP2-like K M S Q - S W T K R C D L T L K G P S S T E W Pacific oyster - GLIPR1-likel K M A D - R W A R R C C N M T H G - P P S S R M L N Q S Zebrafish - CRVP-like E S A R - G W I N K C C N M T H G - P P S S R M L N G Burmese python - CRVP-like E S A R - G W I N K C C S F A H S - P V N Q R T V G E L E. cabalhus - CRISP2 A N A Q - K W A N K C C N M T H G - P P S S A D D R K T S - T H sapiens - CRISP1 Q N A R S K Y C D M T E - S - N P L E R R L P N T H sapiens - CRISP3 A N A Q - K W A N Q C N Y R - H S - N P K D R M T S - L		Red flour beetle - CRISP2-like	R	Α.	A C) -	K	w s	S F	ιк	2			_	С	F٦	'L	т	н	D 1	N I	т	G	RН	I	Εľ	¥Υ	G
Molniscs, S. kowalevskii - Uhchar. Prot. Q L A Q - D W S E E C I F A H T P D - K E Y E A T I F Fchinoderms, and und Forida lancelet - Hypoth Prot. T I A Q - G W A D G C C D F A H N S H R G D G Y Purple sea urchin - CRISP2-like Pacific oyster - GLIPR1-likel K M S Q - S W T K R C C D L T L K G P S S T E W Vertebrates Zebrafish - CRVP-like E S A R - G W I N K C C N M T H G - P P S S R M L N G Burmese python - CRVP-like E S A R - G W I N K C C N M T H G - P P S S R M L N G E. caballus - CRISP2 A N A Q - K W A N K C C S F A H S - P V N Q R T V G E L H. sapiens - CRISP1 Q N A R S K Y C C D M T E - S - N P L E R R L P N T H. sapiens - CRISP3 A N A Q - K W A N Q C C N Y R - H S - N P K D R M T S - L		Asian swallowtail - DIS2-like	ĸ	Q.	A Q) -	s	ΥÆ	ΑI	ЪΚ	2			_	С	IF	Ľ	Q	н	ΝI) P	L	ΕI	ΝT	Ι	ΡJ	ΓL	G
S. kowalevskii - Uhchar. Prot. QLAQ - DWSEE C IFAHTPD - KEYEAT E AT IF Fchinoderms, and Florida lancelet - Hypoth Prot. TIAQ - GWADG C C DFAH NSHRGDGY Purple sea urchin - CRISP2-like KMSQ - SWTKR C C DLTL KGPSSTEW SSTEW Pacific oyster - GLIPR1-like! KMAD - RWARR C C DLTL NSRRSTEW SSTEW Vertebrates Zebrafish - CRVP-like ESAR - GWINK C NMTHG - PPSSRML NG SRML NG Laburese python - CRVP-like ESAR - GWINK C NMTHG - PPSSRML NG G E Laburese python - CRVP-like ENAR - RWADK C NMTHG - PPSSRML NG G E H. sapiens - CRISP1 QNARSKY C DMTE - S - NPLER RLPNNT NTS - I H. sapiens - CRISP3 ANAQ - KWANQ C NYR - HS - NPKDRMT S - I		Sea squirt (C. i.) - CRISP-like	Q	F .	A S	-	Ν	ΥI	I S	5 N	-		-		С	DB	A	H	S	ΤI	ΝA	Q	R	s s	Ι	s c	F	
and UrochordatesTIAQ - GWADG CDFAH NSH RGDGYPurple sea urchin - CRISP2-like Pacific oyster - GLIPR1-like!KMSQ - SWTKR CDLTL KGP SSTEWVertebratesZebrafish - CRVP-like Burmese python - CRVP-likeESAR - GWINK CCNMTHG - PPSSRMLNGSTWVertebratesZebrafish - CRISP2 H sapiens - CRISP1A NAQ - KWANK CCNMTHG - PPSSRMLNGSTWG-VertebratesKapiens - CRISP3A NAQ - KWANQ CNMTHG - PPSR ADDRNTSTNNSTWareKapiens - CRISP3A NAQ - KWANQ CNYR - HS - NPKDRMTSINNSI	,	S. kowalevskii - Uhchar. Prot.	Q	L.	A Ç) - (D	w s	S E	Е	-		-	-	С	ΙF	A	H	ΤJ	ΡI) -	-	ΚI	ΕΥ	Е	A 1	I I	Ρ
and Purple sea urchin - CRISP2-like K M S Q - S W T K R C D L T L K G P S S T E W Pacific oyster - GLIPR1-like1 K M A D - R W A R R C C Q F V H K G P S S T E W N N Q S Zebrafish - CRVP-like E S A R - G W I N K C C N M T H G - P P S S R M L N G N Q R C N M T H G - P P S S R M L N G		Florida lancelet - Hypoth. Prot.	Т	Ι.	A Q) -	G	W 4	ΑI	G	-		-	-	С	Ι	F	Α	н			-	N	SН	R	GΙ	G	Y
Pacific oyster - GLIPR1-like1 K M A D - R W A R R C Q F V H N S R R N N Q S Zebrafish - CRVP-like E S A R - G W I N K C N M T H G - P P S S R M L N G Burmese python - CRVP-like K N A K - R W A D K C S F A H S - P V N Q R T V G E L E. caballus - CRISP2 A N A Q - K W A N K C O M T E - S - N P L E R R L P N T H. sapiens - CRISP1 Q N A R S K Y C N Y R - H S - N P K D R M T S - L			к	М	sς) -	S	W 1	ΓK	C R	-		-	-	С	Ι	L	Т	L			-	К (GΡ	S	SЗ	Έ	W
Vertebrates Burmese python - CRVP-like K N A K - R W A D K C C - S F A H S - P V N Q R T V G E L Vertebrates E. cabalhus - CRISP2 A N A Q - K W A N K C C - T L E H S - S A D D R K T S - T H. sapiens - CRISP1 Q N A R S K Y C C D M T E - S - N P L E R R L P N T H. sapiens - CRISP3 A N A Q - K W A N Q C C N Y R - H S - N P K D R M T S - L	Urochordates	Pacific oyster - GLIPR1-like1	K	М.	ΑĽ) -	R	W 4	A F	۲R	-		-	-	С	C) F	V	н			-	N	S R	R	NI	Į Q.	S
Vertebrates E. cabalhus - CRISP2 A N A Q - K W A N K C - T L E H S - S A D D R K T S - T H. sapiens - CRISP1 Q N A R S K Y C C D M T E - S - N P L E R R L P N T H. sapiens - CRISP3 A N A Q - K W A N Q C C N Y R - H S - N P K D R M T S - L		Zebrafish - CRVP-like	Е	ς.	ΑR	٤ -	G	W I	I١	ΙK	-		-		C	ΝN	1 Τ	Н	G	-]	Р Р	S	S I	RΜ	L	N (š -	-
H. sapiens - CRISP1 Q N A R S K Y C C D M T E - S - N P L E R R L P N T H. sapiens - CRISP3 A N A Q - K W A N Q C N Y R - H S - N P K D R M T S - L		Burmese python - CRVP-like	K	Ν.	ΑK	- 1	R	W 4	ΑI	ЪΚ	-		-	-	С	- 5	F	Α	H	S -	- P	V	N	QR	Т	V C	ΞE	L
H. sapiens - CRISP3 A N A Q - K W A N Q C N Y R - H S - N P K D R M T S - L	Vertebrates	E. caballus - CRISP2	A	N.	A Q) - (K	w 4	A P	łК	2			_	С	- 1	Ľ	Е	н	s -	- S	A	DI	D R	к	ΤS	5 -	т
		H. sapiens - CRISP1	0	N.	A F	t s	к	Y.			2			_	С	D N	4 T	Е	-	s -	- N	Р	L	ΕR	R	LI	Ν	Т
		•	Α	Ν.	A C) -	K	W A	ΑN	10	-		-	-	C.	NΥ	R		H	s -	- N	Р	ΚI	D R	м	ΤS	5 -	L
C Conserved cysteines Sites of non-aligned sequence omission	Exon 1		ion	3																			1 :	Exc	on 1	7		

Appendix 9: Evolution of the Pathogenesis-Related (CAP/PR) Domain in the CRISP Protein Lineage

Appendix 9: Page 3 of 7

Appendix 9: Evolution of the Pathogenesis-Related (CAP/PR) Domain in the CRISP Protein Lineage

		C2	CAP4											
	A. gephyra - SCP-like	F	GENL	- A A .	AAPP	GS	AQ	VVE	DW	VS	ΕVΑ	DY	S Y	7
Prokaryotes	M. stipitatus - SCP-like	L	GENL	- A A .	A T P W	GΤ	ΑQ	V V К	s w	A D	EAS	DY	DН	1
	S. amylolyticus - SCP-like	Y	GENL	- A A	G T S L	SΡ	QA	V V D	мw	ΗR	ERE	ERY	RΓ	H
	P. reichenowi - CRISP-like	V	DTNY	- F S 3	FFKN	ΕN	ΕA	SVD	TW	ΥE	GIN	I D Y	DF	H
	I. multifiliis - SCP-like	LGKI	GENI	- C F	GGNT	ΑV	DІ	IVQ	LI	ID	D			
Protozoa	P. marinus - CRISP-like	A	GENL	- A M (GYGA	AP	HC.	AVW	N W	ΥN	ΕΥΝ	ID F	ΜK	(
	H. hammondi - SCP-like	L	GENI	- F F .	ASDA	S C	GΥ	GVQ	MW	ΥN	ΕVΝ	ILL	ΚY	I
	D. discoideum - Hypoth. Prot.		QNSE	- W Q	T W A F	ΝL	ΤS	ILV	ΝΙ	GΕ	SSK	ΥΥ	DW	S
	C. militaris - SCP/PRY-like	Y	GENI	- A I	GYA -	ΝA	ΤA	A V A	A W	G D	ERG	ΤY	DF	C
. .	P. minimum - SCP-like	Y	GENL	- A I	G <mark>C</mark> S-	D A	ΕG	C V E	MW	GD	ERE	ЕΥ	DΥ	ł
Fungi	N. parvum - SCP-related	L	GENL	- A M (GTGL	ΤА	QQ	ТVЕ	MW	ΥN	EID) Q Y	NW	C
	B. maydis - Hypoth. Prot.	Y	E A	- W V 3	EFA-		RT.	АТН	GW	ΗE	ЕЕТ	Ϋ́ΚΥ	ΡY	Ι
	H. vulgaris - PR-1b-like		NENS	- F V	SNTH	VG	SK	VIK	YW	YK	KGE	ΕΝΥ	ΤΥ	ľ
A 11 1	H. vulgaris - GAPR1-like	R	GENI	- Y Y	S Y D P	SG	DE.	A S N	ΤW	ΥS	ЕІК	ΙNΥ	DΥ	Ι
Cnidaria	H. vulgaris - GAPR1-like	Y	GENL	- F M	LKGG	VS	- K	N I C	D W	Y S	EIK	N Y	DΥ	Ι
	N. vectensis - Predicted Prot.	Y	GQAL	- F V	SDTP	A D	- V	VNA	s w	ΥN	QIF	FY	ΝY	1
	B. malayi - SCP-like	Τ	GENL	- Y A]	YWST	A G	ΤN.	AGK	SW	WS	ELS	QL	ΥR	1
	T. spiralis - Ion-1 Protein	Τ	GENI	- W A .	APFS	DΙ	SΟ	SVK	LW	F S	EIF	NR	R C	C
Round Worms	Nematodes (C. e.) - SCL 12	Y	GENL	- Y W	S W S K	FG	VA.	A S N	IS W	ΕK	ΕFC	QΟΥ	GW	1
	Nematodes (C. e.) - SCL 14	I	GENM	YWW	ΤSΡL	LG	ΝR	SAN	ΙLW	E S	ΕFÇ	R F	G W	N
	Nematodes (C. e.) - SCL 1	Y	GENL	- Y W	S W T S	YG	ΕI	ААА	A W	ΕK	ΕFC) D F	G W	F
	Giant honeybee - CRISP1-like	FGQS	GQNI	- F I	TTRR	TL	N F	PIR	MW	$\mathbf{Y} \mathbf{M}$	ΕΥK	CD Y	ΚY	(
	Bumble bee - CRISP2-like	F G S S	GQNI	- F I	TTRR	TL	NF	AIR	MW	$\mathbf{Y} \mathbf{M}$	ΕYΚ	DF	КΥ	(
Insects	Pharaoh ant - CRISP1-like	FGQS	GQNI	- F I	STGR	TL	ΝF	PIR	MW	$\mathbf{Y} \mathbf{M}$	ΕYΚ	DF	КΥ	C
	Red flour beetle - CRISP2-like	<mark>C</mark>	GQNI	- F I .	ASDK	VP	LF.	AIK	ΤW	ΥL	ΕКΙ) N F	T F	(
	Asian swallowtail - DIS2-like	<mark>C</mark>	GQNL	- F V .	A A Q K	ТР	FF.	ALK	ΤW	F L	ΕYΚ	CNF	ΤΥ	(
Molluscs.	Sea squirt (C. i.) - CRISP-like	S - I I	GENL	GIT	SSNI	R S	D Y	ILS	V W	ΗN	ΕΙΝ	ID Y	FY	I
Echinoderms.	S. kowalevskii - Uhchar. Prot.	C D Q V	GAKT	NSI	MSYI	S N	D C	MLT	' M Y	A E	ΕΤΕ	D D C	K C	\$
and	Florida lancelet - Hypoth. Prot.	- G S V	GENI	YAD	TGRF	VT	G	ЕТЕ	NW	ΗN	EVS	DY	ΤΥ	S
Urochordates	Purple sea urchin - CRISP2-like	ACRL	GRSV	SFV	РКТК	ΗF	D	FLD	ΚW	NN	VТЕ	ΕDΥ	N F	H
Crochordates	Pacific oyster - GLIPR1-like1	FNFV	GENL	AYS	S D D R	ΚA	S	ΥVQ	MW	ΥA	EVK	CD Y	ΤF	H
	Zebrafish - CRVP-like	Y - E M	GENL	- F K .	ATGI	S S	TS	VVD	A W	ΗS	ΕVΝ	IN Y	ΚY	I
	Burmese python - CRVP-like	<mark>C</mark>	GENL	- F M	SТΗΡ	ΥS	ТА	VТQ	s w	ΥD	ERK	DF	КΥ	(
Vertebrates	E. caballus - CRISP2	<mark>C</mark>	GENI	- Y M	SSDP	ТР	SD.	ΑIQ	s w	ΥD	ΕSΙ	, D F	ΤΥ	(
	H. sapiens - CRISP1	<mark>C</mark>	GENM	- H M	ТЅҮР	V S	S S	VIG	VW	Y S	ESI	SF	КН	C
	H. sapiens - CRISP3	C	GENL	- Y M	SSAS	S S	SQ	ΑIQ	s w	F D	ΕΥΝ	ID F	D F	(
Exon 1		on 3		4 📃	Exon			Exon			Exon	7		
C Conserved cysteines Sites of non-aligned sequence omission														

Appendix 9: Page 4 of 7

												CAPI							1 .					
			_	_			_		_		_				T		_	1		_				
	4 1 000011	-										-	~ 1			xon						C3	_	
	A. gephyra - SCP-like	T	N			-	- K				K V			ΗY		- C -								ΓΑΙ
Prokaryotes	M. stipitatus - SCP-like	R				-	- T				K V					~				ΤG				Γνι
	S. amylolyticus - SCP-like	_	_					S N				Т	_	HF		<u> </u>		_		A R				r s ı
	P. reichenowi - CRISP-like	L	G	- C	IK	R		NI		-		1		EF		R				EN				ΓΑ
	I. multifiliis - SCP-like	-	-			G	V S					-		H R						K V				A A K
Protozoa	P. marinus - CRISP-like	ĸ	C			D	s w	ΥKΙ	· ·		SG	-	G I		-		ИΚ	G		D K	ΙC	-		C S G
	H. hammondi - SCP-like	G	-			E	- W		R R		GΑ	_		H F		QN				ΓG				RTI
	D. discoideum - Hypoth Prot.	1.	A	- G	CI	IN	I S	S S	5 N	-		С	_	I W	Т	N A	A N	K	S I	κs	FO	_	Αŀ	CS I
	C. militaris - SCP/PRY-like	K	Р			G	- F	EF	ΙA	-		Т		ΗF	Т	QI	L K	G	T	ГΤ	M (C	Εł	C V L
Fungi	P. minimum - SCP-like	D	Р				- F	SI		-			G I			Q I		Ν	Τ	ΤТ	V (RKI
5	N. parvum - SCP-related	K	N _			D	- V		1D			V	MI			Q 1		. К	T :	ΓК	V (ΛAΓ
	B. maydis - Hypoth. Prot.	Ν	P					DA		-		W	G 1			QN				ΓR	ΙC			VGF
	H. vulgaris - PR-1b-like								ΑP	-		-		SF		~				ΚY	V (_		- A A
Cnidaria	H. vulgaris - GAPR1-like	Е	P				GΥ	AI		-		Т		ΗF		QI				КΚ				CSY
	H. vulgaris - GAPR1-like	Е	Ρ			-	GΥ	AI		-		Т		ΗF	-	QI				КΚ	V (G 1	ΙΑΊ
	N. vectensis - Predicted Prot.	Ν					- R	. K. A	A K			Τ	_	FF	_	QI		_	_	QΕ	_	_		C A S
	B. malayi - SCP-like	Ν	P		SN	IN:	LS	PI	EV	S I	RQ	_	AI			QN	ΛG	K	ΤI	нк	IC		G 1	. V 1
	T. spiralis - Ion-1 Protein	С	Т			N	ΥF	K A	V C			С		ΗY		~	V A	E	S 1	ΕL	L (C	G I	. S 1
Round Worms	Nematodes (C. e.) - SCL 12	S	Т	ΥM	D A	LD.	LF	DS	s G			1	G 1	ΗA		_	A A	E	ΤI	N K.	1 (C	G	(K N
	Nematodes (C. e.) - SCL 14	G	Ν.	LL	ΤE	:Е	LF	N S	ŝG	-		Ι	G 1	ΗA	Т		A A	Т	ΤI	ΝK	I (C	G.	SK
	Nematodes (C. e.) - SCL 1	S	Ν.	A M	D 1	: Т.	LF	NS	s G	-		1	G	ΗA		Q 1	ΛA	. N	T :	S S	1 (C	G \	KI
	Giant honeybee - CRISP1-like	D	D			-	- E	AI		LI	ΗE	1	G 1			× .	ΙA	-		ΗL	V (_	G V	SE
	Bumble bee - CRISP2-like	N	E		ΕV	/ N I	D Y -	RH	C I	-		-	G 1			Q	ΙA			HL	V (VSH
Insects	Pharaoh ant - CRISP1-like	P .	N			-	- T	TI	V E	11	LΕ	1	G I			Q 1	/ A			ΗL	V (VSH
	Red flour beetle - CRISP2-like	S	R _			-		K I			ΜI	V		ΗY		QN				ΗE	V (LSK
	Asian swallowtail - DIS2-like	E	P -			-	- I	KI	D L		X A	_		ΗY	-	QN		-		HK	V (_		LAH
Molluscs,	Sea squirt (C. i) - CRISP-like	Т	- C	S -						_	ΧA			ΗY	Т		V A		T	YK	IC	_		A A F
Echinoderms,	S. kowalevskii - Uhchar. Prot.			ΤG	WN	1 G .				E			S I			-]		_		ΥP	M 1	_		E
and	Florida lancelet - Hypoth Prot.	S					S				4 V			ΗY	_	Q		T	S I	ĸκ	L (_	G \	VKI
Urochordates	Purple sea urchin - CRISP2-like	D					T			AI		C		RY	-	~ -	ΛA	T	T I	S F	V (C	GI	
	Pacific oyster - GLIPR1-like1						G			E		C		H Y		Q 1	7 A			EY	IO			
	Zebrafish - CRVP-like Burmese python - CRVP-like	1 V				-					QA AV	_			Т		V Y	-	S T		V (-	A	
Variation	E. caballus - CRISP2		G			-	- P	KS	· ·					H Y		~		K		K L Y R	L (V (AAK IAY
Vertebrates	H. sapiens - CRISP2	V E	U W				- P T	T		G		_		H Y J V		Q								
	•	E	w G			-	- 1 D			DI		T		HY		Q				Y L	I (V (IAS
	H. sapiens - CRISP3	v	U				- P	KI	l P	IN 1	A V	v	6	ΗY	1	Q V	vĭ	5	<u>з</u>	ΥL	V (3	GI	N A Y
Exon 1	Exon 2 Ex	on 3	3		1	Exo	n 4			E	ton	5			Е	xoi	n 6			F	Exo	n 7		
C Conserved cysteines Sites of non-aligned sequence omission																								

Appendix 9: Evolution of the Pathogenesis-Related (CAP/PR) Domain in the CRISP Protein Lineage

Conserved cysteines

Appendix 9: Page 5 of 7

Appendix 9: Evolution of the Pathogenesis-Related (CAP/PR) Domain in the
CRISP Protein Lineage

														CAP2 C5 C6 Exon 5/6										
	C4															С	5	(C6	Ex	on	5/6	1	
A. gephyra - SCP-like	C	Τŀ	[]	N S	5 F	F	G	Α	Q	FΕ	• K	Ŵ	Q	L	WY	V C	N	Υ	Т	P I	P (GΝ	F	V
M. stipitatus - SCP-like	С	Κŀ	5 1	N S	5 F	• F	G	Α	S I	FΕ	, 1	W	Q	Ľ	w v	V C	N	Υ	Т	P I	P (GΝ	F	V
S. amylolyticus - SCP-like	С	- 1	V V	ΞN	1 I) -	-	-	-		-	-	-	V	W V	V C	N	Υ	D	P 1	P (GΝ	I V	Q
P. reichenowi - CRISP-like	С	- K	C 1	ΓК	5 0	÷ -	-	-	-			-	-	Ι	L	I	K	Υ	D	NI	N 1	ΓN	K	P
I. multifiliis - SCP-like	Η	S (2 2	ΥĽ)]	C	С	V	-			L	-	D	ΥŻ	A S	Е	Υ	Т	S I	K (G N	K	С
P. maninus - CRISP-like	Ν						-	-	-			-	-	Υ	ΥV	V C	Q	Υ	G	T (GO	G S	Υ	G
H. hammondi - SCP-like	s	CI)]	N N	4 I	- ۲	-	-	-		-	-	-	Q	LI	7	v	Y	K	Ρ.	A (GΝ	W	I
D. discoideum - Hypoth. Prot.	С	- I) I	ΞI	1	- 1	-	-	-		-	-	-	V	L S	S 🕻	S	Υ	Υ	Ρ.	A (GG	F	K
C. militaris - SCP/PRY-like	C	- (7 6	VF	د د	- ť	-	-	-			W	-	F	V 4	A (E	Y	W	P :	P (GΝ	ΙV	Ç
P. minimum - SCP-like	С	- 0	3 1	ΓS	6 0	÷-	-	-	-		-	W	-	Υ	Γī	V C	E	Υ	W	P I	R (GΝ	V	I
N. parvum - SCP-related	С	- 5	6 (Q G	3 K	- 2	-	-	-		-	-	-	L	Α 1	Γ	Ν	Υ	s	Ρ.	A (GΝ	Μ	L
B. maydis - Hypoth. Prot.	С		·I	D S	5 1	V V	Ν	W	P	GF	۲	Υ	-	-		- (Y	Υ	S	F I	F (GΝ	IN	Ι
H. vulgaris - PR-1b-like	Ν	SI	I I) ç) N	۰ آ	-	-	-		-	Υ	-	V	V	Ι	Y	Υ	Е	P I	P (GΝ	IN	D
H. vulgaris - GAPR1-like	Ν		• •	βN	I K	- 1	-	-	-		-	V	-	Υ	V V	V	N	Υ	F	Ι.	A (GΝ	Μ	L
H. vulgaris - GAPR1-like	К		0	G K	2 1	I -	-	-	-		-	Т	-	V	V V	V A	Q	Υ	L	ΡI	P (GΝ	M	Е
N. vectensis - Predicted Prot.	s	A S	6 0	3 K	C 1	- 1	-	-	-		-	-	-	Υ	I	V A	Y	F	D	P I	Р 9	SΝ	Т	-
B. malayi - SCP-like	Ν	CI) (3 C	βŀ	I -	-	-	-		· T	Ľ	-	Ι	V V	V C	H	Υ	S	P.	A (GΝ	V	L
T. spiralis - Ion-1 Protein	С	ΝF	R I	Ιŀ	ł P	₹ -	-	А	P	GΗ	ΗK	F	-	I	L 1	V	H	Y	W	P	s_(GΝ	W	V
Nematodes (C. e.) - SCL 12	С	GΙ	C I	D S	5 S	; -	Μ	Ν	NI	ΜŊ	ζK	: v	-	А	V V	V C	Q	Υ	D	Q.	A (GΝ	IM	N
Nematodes (C. e.) - SCL 14	С	S S	S I	D S	5 F	-	G	Т	-		Ç	ĮΥ	-	V	٧V	V C	L	Υ	s	Ρ.	A (GΝ	Y	I
Nematodes (C. e.) - SCL 1	С	GΗ	ιI	D A	1.5	5 -	Μ	R	NI	Mľ	1 K	I I	-	A	v١	V	Q	Y	s	P 1	P	GΝ	Т	N
Giant honeybee - CRISP1-like	C	ΤC	θG	3 F	2 0	÷ -	Ρ	L	G I	ΚI) F	F	1	Μ	ΥV	V C	N	Υ	Α	P	S (GΝ	ΙY	K
Bumble bee - CRISP2-like	C	ΤC	Э С	G K	c 0	÷ -	Ρ	L	G I	ΚI) F	F	-	М	ΥV	V C	N	Υ	Α	P	S (GΝ	ΙY	K
Pharaoh ant - CRISP1-like	C	ΤC	Э С	βI	0	÷ -	Ρ	L	G I	ΚI) F	Υ	-	Μ	Υ	V C	N	Y	A	P	S (GΝ	ΙY	K
Red flour beetle - CRISP2-like	С	ΚS	5 5	5 G	3 K	C -	-	-	-	- 1	F	Y	-	Ν	Y	I	N	Υ	С	Ρ	I	GΝ	Q	P
Asian swallowtail - DIS2-like	C	ΤC	Э С	G F	, M	V -	-	-	- (GK	C F	Y	-	Ν	ΥV	V C	H	Y	С	P	G (GΝ	F	D
Sea squirt (C. i.) - CRISP-like	C	RΕ	r A	A I	r s	БТ	V	-	- (G	Y	R	L	L	V S	S C	N	Υ	G	Ρ.	A (G N	F	L
S. kowalevskii - Uhchar. Prot.	-				· I	, E	W	D	F	V A	A K	Y	С	Ρ	LN	A C	G	V	С	- 1	VI	DE	A	Y
Florida lancelet - Hypoth. Prot.	С	S D	ΓΙ	L C	3 1	C -	V	N	W	S 1	I A	N	-	L	V V	V C	N	Υ	Α	P	S (GΝ	F	V
Purple sea urchin - CRISP2-like	С	FΟ	зC	3 -			-	-	-		-	Κ	-	L	F 1	Y	M	Υ	G	P	S (GΟ	E	Т
Pacific oyster - GLIPR1-like1	C	ΑÌ	I I	LN	10	ť	-	-	-			F	-	L	V V	V C	N	Υ	Α	P.	A	GΝ	Y	P
Zebrafish - CRVP-like	С		• 0	3 S	5 1	T -	-	-	-		-	Υ	-	F	Y (G C	H	Υ	Υ	R .	A (GΝ	F	R
Burmese python - CRVP-like	C	P A	1 5	S K	ς.		-	-	-	- F	K	Y	-	Υ	ΥV	V C		- 1	С	P.	A	GΝ	F	G
E. caballus - CRISP2	С	ΡÌ	1 (ξE	Ξ-		-	-	-	S I	. K	Y	-	Υ	ΥV	V C	Q	Υ	С	P	V	GΝ	N	V
and a second	C	RC) () (÷ S	; -	-	-	-	- F	R	Y	-	L	ΥV	V C	H	Υ	C	H I	E	GΝ	D	P
H. sapiens - CRISP1	C	10.0	< >	< -																			-	_
	M. stipitatus - SCP-like S. amylolyticus - SCP-like P. reichenowi - CRISP-like P. marinus - CRISP-like H. hammondi - SCP-like D. discoideum - Hypoth. Prot C. militaris - SCP/PRY-like P. minimum - SCP-like N. parvum - SCP-like N. parvum - SCP-like H. vulgaris - PR-1b-like H. vulgaris - GAPR1-like H. vulgaris - GAPR1-like H. vulgaris - GAPR1-like M. vectensis - Predicted Prot. B. malayi - SCP-like T. spiralis - Ion-1 Protein Nematodes (C. e.) - SCL 12 Nematodes (C. e.) - SCL 12 Nematodes (C. e.) - SCL 14 Nematodes (C. e.) - SCL 1 Giant honeybee - CRISP1-like Pharaoh ant - CRISP1-like Red flour beetle - CRISP2-like Asian swallowtail - DIS2-like S. kowalevskii - Uhchar. Prot. Florida lancelet - Hypoth. Prot Purple sea urchin - CRISP2-like Pacific oyster - GLIPR1-like1 Zebrafish - CRVP-like	M. stipitatus - SCP-like C S. amylolyticus - SCP-like C P. reichenowi - CRISP-like C I. multifiliis - SCP-like H P. marinus - CRISP-like N H. hammondi - SCP-like S D. discoideum - Hypoth. Prot C C. militaris - SCP/PRY-like C P. minimum - SCP-like C N. parvum - SCP-like C B. maydis - Hypoth. Prot. C H. vulgaris - GAPR1-like N H. vulgaris - GAPR1-like N H. vulgaris - GAPR1-like N H. vulgaris - GAPR1-like N M. vectensis - Predicted Prot. S B. malayi - SCP-like N N. reatodes (C. e.) - SCL 12 Nematodes (C. e.) - SCL 12 Nematodes (C. e.) - SCL 1 Giant honeybee - CRISP1-like C Pharaoh ant - CRISP1-like C Red flour beetle - CRISP1-like C Red flour beetle - CRISP2-like C S. kowalevskii - Uhchar. Prot S. kowalevskii - Uhchar. Prot Florida lancelet - Hypoth. Prot. C Purple sea urchin - CRISP2-like C Pacific oyster - GLIPR1-like C Cabrafish - CRVP-like C	M. stipitatus - SCP-like C K K S. amylolyticus - SCP-like C - N P. reichenowi - CRISP-like C - K I. multifiliis - SCP-like K - K P. marinus - CRISP-like K - C D. discoideum - Hypoth. Prot C - I C. militaris - SCP/PRY-like C - C M. parvum - SCP-like C - C N. parvum - SCP-like K - C M. parvum - SCP-like N S P H. vulgaris - PR-1b-like N S P H. vulgaris - GAPR1-like N - S M. vectensis - Predicted Prot. S A S B. malayi - SCP-like N C I N. vectensis - Predicted Prot. S A S B. malayi - SCP-like N C I N. vectensis - Predicted Prot. S A S B. malayi - SCP-like N C I N. vectensis - Predicted Prot. S A S B. malayi - SCP-like N C I Nematodes (C. e.) - SCL 12 C G H Nematodes (C. e.) - SCL 14 C S S Nematodes (C. e.) - SCL 14 C S S Mematodes (C. e.) - SCL 14 C S S Nematodes (C. e.) - SCL 14 C S S Nematodes (C. e.) - SCL 14 C S S Saint honeybee - CRISP1-like C T C Giant honeybee - CRISP1-like C T C Saint swallowtail - DIS2-like C T C Sea squirt (C. i.) - CRISP1-like C A N S. kowalevskii - Uhchar. Prot	M. stipitatus - SCP-likeCKKS. amylolyticus - SCP-likeC-NP. reichenowi - CRISP-likeC-NP. marinus - CRISP-likeHSQP. marinus - CRISP-likeNH. hammondi - SCP-likeND. discoideum - Hypoth. ProtC-DC. militaris - SCP/PRY-likeC-GP. minimum - SCP-likeC-GS. parvum - SCP-relatedC-SB. maydis - Hypoth. Prot.C-IH. vulgaris - SCP-relatedNSNH. vulgaris - GAPR1-likeNH. vulgaris - GAPR1-likeN-CN. vectensis - Predicted Prot.SASB. malayi - SCP-likeNCDC. spiralis - Ion-1 ProteinCNRNematodes (C. e.) - SCL 12CGKNNematodes (C. e.) - SCL 11CGRRGiant honeybee - CRISP1-likeCTGCPharaoh ant - CRISP1-likeCTGGSaganit (C. i.) - CRISP1-likeCTGGSaganit (C. i.) - CRISP1-likeCTGGPurple sea urchin - CRISP2-likeCTGGSaganit (C. i.) - CRISP2-likeCTGGSaganit (C. i.) - CRISP1-likeCTGGSaganit (C. i.) - CRI	M. stipitatus - SCP-likeCKKNSS. amylolyticus - SCP-likeC-NGP. reichenowi - CRISP-likeHSQYI. multifiliis - SCP-likeHSQYP. marinus - CRISP-likeNH. hammondi - SCP-likeSCPND. discoideum - Hypoth. ProtC-DEC. militaris - SCP/RY-likeC-GTP. minimum - SCP-likeC-GTN. parvum - SCP-likeNSNDB. maydis - Hypoth. Prot.C-DDB. maydis - Hypoth. Prot.C-DDH. vulgaris - GAPR1-likeNSNDH. vulgaris - GAPR1-likeNCDGH. vulgaris - GAPR1-likeNCDGD. manatodes (C. e.) - SCL1CGKS. matodes (C. e.) - SCL1CGRMematodes (C. e.) - SCL1CGGMematodes (C. e.) - SCL1CGGGiant honeybee - CRISP1-likeCTGGPharaoh ant - CRISP2-likeCTGGS. kowalevskii - Uhchar. ProtFlorida lancelet - Hypoth. ProtCSTCPuple sea urchin - CRISP2-likeCFGGS. kowalevskii - Uhchar. Prot. <th>M. stipitatus - SCP-likeCKKNSS. amylolyticus - SCP-likeC-NGMP. reichenowi - CRISP-likeC-KTKI. multifiliis - SCP-likeHSQYYP. marinus - CRISP-likeNH. hammondi - SCP-likeND. discoideum - Hypoth. ProtC-DETC. militaris - SCP/RKY-likeC-GKKP. minimum - SCP-likeC-GGKN. parvum - SCP-relatedC-SQGB. maydis - Hypoth. Prot.C-DSGH. vulgaris - GAPR1-likeN-GNNH. vulgaris - GAPR1-likeNCDGGH. vulgaris - GAPR1-likeNCDGGB. malayi - SCP-likeNCDGGT. spiralis - Ion-1 ProteinCNRIHNematodes (C. e.) - SCL1CGRCB. malayi - SCP-likeCTGGCB. malayi - SCP-likeCCSSSSB. malayi - SCP-likeCGGKSSB. malayi - SCP-likeCGGKSSB. malayi - SCP-likeCCGGKSB</th> <th>M. stipitatus - SCP-likeCKKNSPFS. amylolyticus - SCP-likeC-NGMDP. reichenowi - CRISP-likeHSQYDICmarinus - CRISP-likeNH. hammondi - SCP-likeND. discoideum - Hypoth. ProtC-DETIC. militaris - SCP/PRY-likeC-GTSG-N. parvum - SCP-likeC-GSQGK-B. maydis - Hypoth. Prot.C-DSTVH. vulgaris - GAPR1-likeNSNQN-H. vulgaris - GAPR1-likeNCDGH-N. vectensis - Predicted Prot.SASGKNNematodes (C. e.) - SCL12CGKSS-Nematodes (C. e.) - SCL12CGKSSS-Giant honeybee - CRISP2-likeCTGGTGGR-Pharaoh ant - CRISP2-likeCTGGFGC-PES. kowalevskii - Uhchar. ProtPEFGGGK-S. audit (C. i.) - CRISP-likeCTGGG<t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGS. amylolyticus - SCP-likeC-NGMD-P. reichenowi - CRISP-likeC-KTKG-I. multifiliis - SCP-likeHSQYJCCP. marinus - CRISP-likeND. discoideum - Hypoth. ProtC-DETI-C. militaris - SCP/PRY-likeC-GTSG-N. parvum - SCP-likeC-SQGK-B. maydis - Hypoth. Prot.C-DSTVNH. vulgaris - GAPR1-likeN-GNNH. vulgaris - GAPR1-likeN-GKNH. vulgaris - GAPR1-likeNCDGHNematodes (C. e.) - SCL12CGKDN-B. malayi - SCP-likeNCDGHT. spiralis - Ion-1ProteinCNRIHNNematodes (C. e.) - SCL 12CGKGR-Giant honeybee - CRISP2-likeCTGGR-Pharaoh ant - CRISP2-likeCTGGF-<tr<tr>Red flour beetle - CRISP2-li</tr<tr></th><th>M. stipitatus - SCP-likeCKKNSPFGAS. amylolyticus - SCP-likeC-NGMDP. reichenowi - CRISP-likeC-KTKGI. multifiliis - SCP-likeHSQYJICCVP. marinus - CRISP-likeND. discoideum - Hypoth. ProtC-DETI<t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGASS. amylolyticus - SCP-likeC-NGMDP. reichenowi - CRISP-likeHSQYDICCV-P. marinus - CRISP-likeNP. marinus - CRISP-likeND. discoideum - Hypoth. ProtC-DETI</th><th>M. stipitatus - SCP-likeCKKNSPFGASFFGASCC</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTS. amylolyticus - SCP-likeC-NGM<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWS. amylolyticus - SCP-likeC-NGM<</th><th>M. stipitaus - SCP-likeCKKNSPFGASFPTWQS. amylolyticus - SCP-likeC-NGM<!--</th--><th>M. stipitaus - SCP-likeCKKNSPFGASFPTWQLS. amylolyticus - SCP-likeC-NGMDVP. reichenowi - CRISP-likeHSQYDICCVVP. marinus - CRISP-likeNVC-DETIVC. discoideum - Hypoth. ProtC-DETIVC. militaris - SCP/PRY-likeC-GTSGVVN. parvum - SCP-relatedC-SQGKVVVN. parvum - SCP-relatedNSNDQNVVVVVVYNSN<d< td="">QNVVVVVYNSN<d< td="">QNVVVYNSN<d< td="">QNVYVVYNNNNNN<n< td="">N<t< th=""><th>M. stipitaus - SCP-likeCKKNSPFGASFTWQLWWS. amylolyticus - SCP-likeC-NGMDVWWP. reichenowi - CRISP-likeHSQYDICCVLDYYP. marinus - CRISP-likeNVYYD. discoideum - Hypoth. ProtC-DETIVYYC. militaris - SCP/PRY-likeC-GTSGVYYY<</th><th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWVCS. amylolyticus - SCP-likeC-NGMDVWVCP. reichenowi - CRISP-likeHSQYDICCVDYASP. marinus - CRISP-likeNVVVCD. discoideum - Hypoth. ProtC-DETIVVSCC. militaris - SCP/PRY-likeC-GTSGVVICD. discoideum - Hypoth. Prot.C-DSTNNVVICD. markins - SCP-likeC-GTSQGKVVVCB. markis - Hypoth. Prot.C-SNNNNVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWCNS. amylolyticus - SCP-likeC-NGMDVWVCNP. reichenowi - CRISP-likeHSQYDICCVVVQVVQVVVQVVQVVVVCVVVQVVVCVVV<t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGASFTWQLWCNNS. amylolyticus - SCP-likeC-NGMDILICKYYCQYP. reichenowi - CRISP-likeHSQYDICVILICKYYCQYP. mainius - CRISP-likeNYYCQYD. discoideum - Hypoth. Prot.C-DETIVXSSYCVYYCQYYNYCQYYNYCQYYCQYYCQYYCQYYCQYYCQYYCQYYYCQYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWUWVNTTS. amylolyticus - SCP-likeC-NGMIIIIKYDP. reichenowi - CRISP-likeHSQYICCVYVQQYDICCVVVCQYCQYGQYGQYGQYGQYCQYGNNNNNNNNNNNNNNNNNNN<t< th=""><th>M. stipitatus - SCP-like C K K N S P T W V<!--</th--><th>M. stipitatus - SCP-likeCKKKNSPFGASFTWQLWCNYTPPPS. amylolyticus - SCP-likeC-NGMDIIICKXNNDPPPPPP. narinus - CRISP-likeHSQYICCVVVQYGGGP. marinus - CRISP-likeSCPNNVVSCYPPPD. discoideum - Hypoth ProtC-DETIVVSCSYPP<th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWCNYTPPGNS. amylolyticus - SCP-likeC-NGMDIIICKKGNN<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWVNTPPGNFS. anylolyticus - SCP-likeCNGMDILICNYDPPGNVP. reichenowi - CRISP-likeC-KTKGNKKGNKKGNKGNKKGNKKGNNTNNTNKKGNKKGNNTNNTNNN<t< th=""></t<></th></th></th></t<></th></t<></th></t<></n<></d<></d<></d<></th></th></t<></th></t<></th>	M. stipitatus - SCP-likeCKKNSS. amylolyticus - SCP-likeC-NGMP. reichenowi - CRISP-likeC-KTKI. multifiliis - SCP-likeHSQYYP. marinus - CRISP-likeNH. hammondi - SCP-likeND. discoideum - Hypoth. ProtC-DETC. militaris - SCP/RKY-likeC-GKKP. minimum - SCP-likeC-GGKN. parvum - SCP-relatedC-SQGB. maydis - Hypoth. Prot.C-DSGH. vulgaris - GAPR1-likeN-GNNH. vulgaris - GAPR1-likeNCDGGH. vulgaris - GAPR1-likeNCDGGB. malayi - SCP-likeNCDGGT. spiralis - Ion-1 ProteinCNRIHNematodes (C. e.) - SCL1CGRCB. malayi - SCP-likeCTGGCB. malayi - SCP-likeCCSSSSB. malayi - SCP-likeCGGKSSB. malayi - SCP-likeCGGKSSB. malayi - SCP-likeCCGGKSB	M. stipitatus - SCP-likeCKKNSPFS. amylolyticus - SCP-likeC-NGMDP. reichenowi - CRISP-likeHSQYDICmarinus - CRISP-likeNH. hammondi - SCP-likeND. discoideum - Hypoth. ProtC-DETIC. militaris - SCP/PRY-likeC-GTSG-N. parvum - SCP-likeC-GSQGK-B. maydis - Hypoth. Prot.C-DSTVH. vulgaris - GAPR1-likeNSNQN-H. vulgaris - GAPR1-likeNCDGH-N. vectensis - Predicted Prot.SASGKNNematodes (C. e.) - SCL12CGKSS-Nematodes (C. e.) - SCL12CGKSSS-Giant honeybee - CRISP2-likeCTGGTGGR-Pharaoh ant - CRISP2-likeCTGGFGC-PES. kowalevskii - Uhchar. ProtPEFGGGK-S. audit (C. i.) - CRISP-likeCTGGG <t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGS. amylolyticus - SCP-likeC-NGMD-P. reichenowi - CRISP-likeC-KTKG-I. multifiliis - SCP-likeHSQYJCCP. marinus - CRISP-likeND. discoideum - Hypoth. ProtC-DETI-C. militaris - SCP/PRY-likeC-GTSG-N. parvum - SCP-likeC-SQGK-B. maydis - Hypoth. Prot.C-DSTVNH. vulgaris - GAPR1-likeN-GNNH. vulgaris - GAPR1-likeN-GKNH. vulgaris - GAPR1-likeNCDGHNematodes (C. e.) - SCL12CGKDN-B. malayi - SCP-likeNCDGHT. spiralis - Ion-1ProteinCNRIHNNematodes (C. e.) - SCL 12CGKGR-Giant honeybee - CRISP2-likeCTGGR-Pharaoh ant - CRISP2-likeCTGGF-<tr<tr>Red flour beetle - CRISP2-li</tr<tr></th><th>M. stipitatus - SCP-likeCKKNSPFGAS. amylolyticus - SCP-likeC-NGMDP. reichenowi - CRISP-likeC-KTKGI. multifiliis - SCP-likeHSQYJICCVP. marinus - CRISP-likeND. discoideum - Hypoth. ProtC-DETI<t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGASS. amylolyticus - SCP-likeC-NGMDP. reichenowi - CRISP-likeHSQYDICCV-P. marinus - CRISP-likeNP. marinus - CRISP-likeND. discoideum - Hypoth. ProtC-DETI</th><th>M. stipitatus - SCP-likeCKKNSPFGASFFGASCC</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTS. amylolyticus - SCP-likeC-NGM<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWS. amylolyticus - SCP-likeC-NGM<</th><th>M. stipitaus - SCP-likeCKKNSPFGASFPTWQS. amylolyticus - SCP-likeC-NGM<!--</th--><th>M. stipitaus - SCP-likeCKKNSPFGASFPTWQLS. amylolyticus - SCP-likeC-NGMDVP. reichenowi - CRISP-likeHSQYDICCVVP. marinus - CRISP-likeNVC-DETIVC. discoideum - Hypoth. ProtC-DETIVC. militaris - SCP/PRY-likeC-GTSGVVN. parvum - SCP-relatedC-SQGKVVVN. parvum - SCP-relatedNSNDQNVVVVVVYNSN<d< td="">QNVVVVVYNSN<d< td="">QNVVVYNSN<d< td="">QNVYVVYNNNNNN<n< td="">N<t< th=""><th>M. stipitaus - SCP-likeCKKNSPFGASFTWQLWWS. amylolyticus - SCP-likeC-NGMDVWWP. reichenowi - CRISP-likeHSQYDICCVLDYYP. marinus - CRISP-likeNVYYD. discoideum - Hypoth. ProtC-DETIVYYC. militaris - SCP/PRY-likeC-GTSGVYYY<</th><th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWVCS. amylolyticus - SCP-likeC-NGMDVWVCP. reichenowi - CRISP-likeHSQYDICCVDYASP. marinus - CRISP-likeNVVVCD. discoideum - Hypoth. ProtC-DETIVVSCC. militaris - SCP/PRY-likeC-GTSGVVICD. discoideum - Hypoth. Prot.C-DSTNNVVICD. markins - SCP-likeC-GTSQGKVVVCB. markis - Hypoth. Prot.C-SNNNNVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWCNS. amylolyticus - SCP-likeC-NGMDVWVCNP. reichenowi - CRISP-likeHSQYDICCVVVQVVQVVVQVVQVVVVCVVVQVVVCVVV<t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGASFTWQLWCNNS. amylolyticus - SCP-likeC-NGMDILICKYYCQYP. reichenowi - CRISP-likeHSQYDICVILICKYYCQYP. mainius - CRISP-likeNYYCQYD. discoideum - Hypoth. Prot.C-DETIVXSSYCVYYCQYYNYCQYYNYCQYYCQYYCQYYCQYYCQYYCQYYCQYYYCQYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWUWVNTTS. amylolyticus - SCP-likeC-NGMIIIIKYDP. reichenowi - CRISP-likeHSQYICCVYVQQYDICCVVVCQYCQYGQYGQYGQYGQYCQYGNNNNNNNNNNNNNNNNNNN<t< th=""><th>M. stipitatus - SCP-like C K K N S P T W V<!--</th--><th>M. stipitatus - SCP-likeCKKKNSPFGASFTWQLWCNYTPPPS. amylolyticus - SCP-likeC-NGMDIIICKXNNDPPPPPP. narinus - CRISP-likeHSQYICCVVVQYGGGP. marinus - CRISP-likeSCPNNVVSCYPPPD. discoideum - Hypoth ProtC-DETIVVSCSYPP<th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWCNYTPPGNS. amylolyticus - SCP-likeC-NGMDIIICKKGNN<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWVNTPPGNFS. anylolyticus - SCP-likeCNGMDILICNYDPPGNVP. reichenowi - CRISP-likeC-KTKGNKKGNKKGNKGNKKGNKKGNNTNNTNKKGNKKGNNTNNTNNN<t< th=""></t<></th></th></th></t<></th></t<></th></t<></n<></d<></d<></d<></th></th></t<></th></t<>	M. stipitatus - SCP-likeCKKNSPFGS. amylolyticus - SCP-likeC-NGMD-P. reichenowi - CRISP-likeC-KTKG-I. multifiliis - SCP-likeHSQYJCCP. marinus - CRISP-likeND. discoideum - Hypoth. ProtC-DETI-C. militaris - SCP/PRY-likeC-GTSG-N. parvum - SCP-likeC-SQGK-B. maydis - Hypoth. Prot.C-DSTVNH. vulgaris - GAPR1-likeN-GNNH. vulgaris - GAPR1-likeN-GKNH. vulgaris - GAPR1-likeNCDGHNematodes (C. e.) - SCL12CGKDN-B. malayi - SCP-likeNCDGHT. spiralis - Ion-1ProteinCNRIHNNematodes (C. e.) - SCL 12CGKGR-Giant honeybee - CRISP2-likeCTGGR-Pharaoh ant - CRISP2-likeCTGGF- <tr<tr>Red flour beetle - CRISP2-li</tr<tr>	M. stipitatus - SCP-likeCKKNSPFGAS. amylolyticus - SCP-likeC-NGMDP. reichenowi - CRISP-likeC-KTKGI. multifiliis - SCP-likeHSQYJICCVP. marinus - CRISP-likeND. discoideum - Hypoth. ProtC-DETI <t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGASS. amylolyticus - SCP-likeC-NGMDP. reichenowi - CRISP-likeHSQYDICCV-P. marinus - CRISP-likeNP. marinus - CRISP-likeND. discoideum - Hypoth. ProtC-DETI</th><th>M. stipitatus - SCP-likeCKKNSPFGASFFGASCC</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTS. amylolyticus - SCP-likeC-NGM<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWS. amylolyticus - SCP-likeC-NGM<</th><th>M. stipitaus - SCP-likeCKKNSPFGASFPTWQS. amylolyticus - SCP-likeC-NGM<!--</th--><th>M. stipitaus - SCP-likeCKKNSPFGASFPTWQLS. amylolyticus - SCP-likeC-NGMDVP. reichenowi - CRISP-likeHSQYDICCVVP. marinus - CRISP-likeNVC-DETIVC. discoideum - Hypoth. ProtC-DETIVC. militaris - SCP/PRY-likeC-GTSGVVN. parvum - SCP-relatedC-SQGKVVVN. parvum - SCP-relatedNSNDQNVVVVVVYNSN<d< td="">QNVVVVVYNSN<d< td="">QNVVVYNSN<d< td="">QNVYVVYNNNNNN<n< td="">N<t< th=""><th>M. stipitaus - SCP-likeCKKNSPFGASFTWQLWWS. amylolyticus - SCP-likeC-NGMDVWWP. reichenowi - CRISP-likeHSQYDICCVLDYYP. marinus - CRISP-likeNVYYD. discoideum - Hypoth. ProtC-DETIVYYC. militaris - SCP/PRY-likeC-GTSGVYYY<</th><th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWVCS. amylolyticus - SCP-likeC-NGMDVWVCP. reichenowi - CRISP-likeHSQYDICCVDYASP. marinus - CRISP-likeNVVVCD. discoideum - Hypoth. ProtC-DETIVVSCC. militaris - SCP/PRY-likeC-GTSGVVICD. discoideum - Hypoth. Prot.C-DSTNNVVICD. markins - SCP-likeC-GTSQGKVVVCB. markis - Hypoth. Prot.C-SNNNNVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWCNS. amylolyticus - SCP-likeC-NGMDVWVCNP. reichenowi - CRISP-likeHSQYDICCVVVQVVQVVVQVVQVVVVCVVVQVVVCVVV<t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGASFTWQLWCNNS. amylolyticus - SCP-likeC-NGMDILICKYYCQYP. reichenowi - CRISP-likeHSQYDICVILICKYYCQYP. mainius - CRISP-likeNYYCQYD. discoideum - Hypoth. Prot.C-DETIVXSSYCVYYCQYYNYCQYYNYCQYYCQYYCQYYCQYYCQYYCQYYCQYYYCQYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWUWVNTTS. amylolyticus - SCP-likeC-NGMIIIIKYDP. reichenowi - CRISP-likeHSQYICCVYVQQYDICCVVVCQYCQYGQYGQYGQYGQYCQYGNNNNNNNNNNNNNNNNNNN<t< th=""><th>M. stipitatus - SCP-like C K K N S P T W V<!--</th--><th>M. stipitatus - SCP-likeCKKKNSPFGASFTWQLWCNYTPPPS. amylolyticus - SCP-likeC-NGMDIIICKXNNDPPPPPP. narinus - CRISP-likeHSQYICCVVVQYGGGP. marinus - CRISP-likeSCPNNVVSCYPPPD. discoideum - Hypoth ProtC-DETIVVSCSYPP<th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWCNYTPPGNS. amylolyticus - SCP-likeC-NGMDIIICKKGNN<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWVNTPPGNFS. anylolyticus - SCP-likeCNGMDILICNYDPPGNVP. reichenowi - CRISP-likeC-KTKGNKKGNKKGNKGNKKGNKKGNNTNNTNKKGNKKGNNTNNTNNN<t< th=""></t<></th></th></th></t<></th></t<></th></t<></n<></d<></d<></d<></th></th></t<>	M. stipitatus - SCP-likeCKKNSPFGASS. amylolyticus - SCP-likeC-NGMDP. reichenowi - CRISP-likeHSQYDICCV-P. marinus - CRISP-likeNP. marinus - CRISP-likeND. discoideum - Hypoth. ProtC-DETI	M. stipitatus - SCP-likeCKKNSPFGASFFGASCC	M. stipitatus - SCP-likeCKKNSPFGASFPTS. amylolyticus - SCP-likeC-NGM<	M. stipitatus - SCP-likeCKKNSPFGASFPTWS. amylolyticus - SCP-likeC-NGM<	M. stipitaus - SCP-likeCKKNSPFGASFPTWQS. amylolyticus - SCP-likeC-NGM </th <th>M. stipitaus - SCP-likeCKKNSPFGASFPTWQLS. amylolyticus - SCP-likeC-NGMDVP. reichenowi - CRISP-likeHSQYDICCVVP. marinus - CRISP-likeNVC-DETIVC. discoideum - Hypoth. ProtC-DETIVC. militaris - SCP/PRY-likeC-GTSGVVN. parvum - SCP-relatedC-SQGKVVVN. parvum - SCP-relatedNSNDQNVVVVVVYNSN<d< td="">QNVVVVVYNSN<d< td="">QNVVVYNSN<d< td="">QNVYVVYNNNNNN<n< td="">N<t< th=""><th>M. stipitaus - SCP-likeCKKNSPFGASFTWQLWWS. amylolyticus - SCP-likeC-NGMDVWWP. reichenowi - CRISP-likeHSQYDICCVLDYYP. marinus - CRISP-likeNVYYD. discoideum - Hypoth. ProtC-DETIVYYC. militaris - SCP/PRY-likeC-GTSGVYYY<</th><th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWVCS. amylolyticus - SCP-likeC-NGMDVWVCP. reichenowi - CRISP-likeHSQYDICCVDYASP. marinus - CRISP-likeNVVVCD. discoideum - Hypoth. ProtC-DETIVVSCC. militaris - SCP/PRY-likeC-GTSGVVICD. discoideum - Hypoth. Prot.C-DSTNNVVICD. markins - SCP-likeC-GTSQGKVVVCB. markis - Hypoth. Prot.C-SNNNNVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWCNS. amylolyticus - SCP-likeC-NGMDVWVCNP. reichenowi - CRISP-likeHSQYDICCVVVQVVQVVVQVVQVVVVCVVVQVVVCVVV<t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGASFTWQLWCNNS. amylolyticus - SCP-likeC-NGMDILICKYYCQYP. reichenowi - CRISP-likeHSQYDICVILICKYYCQYP. mainius - CRISP-likeNYYCQYD. discoideum - Hypoth. Prot.C-DETIVXSSYCVYYCQYYNYCQYYNYCQYYCQYYCQYYCQYYCQYYCQYYCQYYYCQYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWUWVNTTS. amylolyticus - SCP-likeC-NGMIIIIKYDP. reichenowi - CRISP-likeHSQYICCVYVQQYDICCVVVCQYCQYGQYGQYGQYGQYCQYGNNNNNNNNNNNNNNNNNNN<t< th=""><th>M. stipitatus - SCP-like C K K N S P T W V<!--</th--><th>M. stipitatus - SCP-likeCKKKNSPFGASFTWQLWCNYTPPPS. amylolyticus - SCP-likeC-NGMDIIICKXNNDPPPPPP. narinus - CRISP-likeHSQYICCVVVQYGGGP. marinus - CRISP-likeSCPNNVVSCYPPPD. discoideum - Hypoth ProtC-DETIVVSCSYPP<th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWCNYTPPGNS. amylolyticus - SCP-likeC-NGMDIIICKKGNN<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWVNTPPGNFS. anylolyticus - SCP-likeCNGMDILICNYDPPGNVP. reichenowi - CRISP-likeC-KTKGNKKGNKKGNKGNKKGNKKGNNTNNTNKKGNKKGNNTNNTNNN<t< th=""></t<></th></th></th></t<></th></t<></th></t<></n<></d<></d<></d<></th>	M. stipitaus - SCP-likeCKKNSPFGASFPTWQLS. amylolyticus - SCP-likeC-NGMDVP. reichenowi - CRISP-likeHSQYDICCVVP. marinus - CRISP-likeNVC-DETIVC. discoideum - Hypoth. ProtC-DETIVC. militaris - SCP/PRY-likeC-GTSGVVN. parvum - SCP-relatedC-SQGKVVVN. parvum - SCP-relatedNSNDQNVVVVVVYNSN <d< td="">QNVVVVVYNSN<d< td="">QNVVVYNSN<d< td="">QNVYVVYNNNNNN<n< td="">N<t< th=""><th>M. stipitaus - SCP-likeCKKNSPFGASFTWQLWWS. amylolyticus - SCP-likeC-NGMDVWWP. reichenowi - CRISP-likeHSQYDICCVLDYYP. marinus - CRISP-likeNVYYD. discoideum - Hypoth. ProtC-DETIVYYC. militaris - SCP/PRY-likeC-GTSGVYYY<</th><th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWVCS. amylolyticus - SCP-likeC-NGMDVWVCP. reichenowi - CRISP-likeHSQYDICCVDYASP. marinus - CRISP-likeNVVVCD. discoideum - Hypoth. ProtC-DETIVVSCC. militaris - SCP/PRY-likeC-GTSGVVICD. discoideum - Hypoth. Prot.C-DSTNNVVICD. markins - SCP-likeC-GTSQGKVVVCB. markis - Hypoth. Prot.C-SNNNNVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWCNS. amylolyticus - SCP-likeC-NGMDVWVCNP. reichenowi - CRISP-likeHSQYDICCVVVQVVQVVVQVVQVVVVCVVVQVVVCVVV<t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGASFTWQLWCNNS. amylolyticus - SCP-likeC-NGMDILICKYYCQYP. reichenowi - CRISP-likeHSQYDICVILICKYYCQYP. mainius - CRISP-likeNYYCQYD. discoideum - Hypoth. Prot.C-DETIVXSSYCVYYCQYYNYCQYYNYCQYYCQYYCQYYCQYYCQYYCQYYCQYYYCQYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWUWVNTTS. amylolyticus - SCP-likeC-NGMIIIIKYDP. reichenowi - CRISP-likeHSQYICCVYVQQYDICCVVVCQYCQYGQYGQYGQYGQYCQYGNNNNNNNNNNNNNNNNNNN<t< th=""><th>M. stipitatus - SCP-like C K K N S P T W V<!--</th--><th>M. stipitatus - SCP-likeCKKKNSPFGASFTWQLWCNYTPPPS. amylolyticus - SCP-likeC-NGMDIIICKXNNDPPPPPP. narinus - CRISP-likeHSQYICCVVVQYGGGP. marinus - CRISP-likeSCPNNVVSCYPPPD. discoideum - Hypoth ProtC-DETIVVSCSYPP<th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWCNYTPPGNS. amylolyticus - SCP-likeC-NGMDIIICKKGNN<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWVNTPPGNFS. anylolyticus - SCP-likeCNGMDILICNYDPPGNVP. reichenowi - CRISP-likeC-KTKGNKKGNKKGNKGNKKGNKKGNNTNNTNKKGNKKGNNTNNTNNN<t< th=""></t<></th></th></th></t<></th></t<></th></t<></n<></d<></d<></d<>	M. stipitaus - SCP-likeCKKNSPFGASFTWQLWWS. amylolyticus - SCP-likeC-NGMDVWWP. reichenowi - CRISP-likeHSQYDICCVLDYYP. marinus - CRISP-likeNVYYD. discoideum - Hypoth. ProtC-DETIVYYC. militaris - SCP/PRY-likeC-GTSGVYYY<	M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWVCS. amylolyticus - SCP-likeC-NGMDVWVCP. reichenowi - CRISP-likeHSQYDICCVDYASP. marinus - CRISP-likeNVVVCD. discoideum - Hypoth. ProtC-DETIVVSCC. militaris - SCP/PRY-likeC-GTSGVVICD. discoideum - Hypoth. Prot.C-DSTNNVVICD. markins - SCP-likeC-GTSQGKVVVCB. markis - Hypoth. Prot.C-SNNNNVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV<	M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWCNS. amylolyticus - SCP-likeC-NGMDVWVCNP. reichenowi - CRISP-likeHSQYDICCVVVQVVQVVVQVVQVVVVCVVVQVVVCVVV <t< th=""><th>M. stipitatus - SCP-likeCKKNSPFGASFTWQLWCNNS. amylolyticus - SCP-likeC-NGMDILICKYYCQYP. reichenowi - CRISP-likeHSQYDICVILICKYYCQYP. mainius - CRISP-likeNYYCQYD. discoideum - Hypoth. Prot.C-DETIVXSSYCVYYCQYYNYCQYYNYCQYYCQYYCQYYCQYYCQYYCQYYCQYYYCQYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWUWVNTTS. amylolyticus - SCP-likeC-NGMIIIIKYDP. reichenowi - CRISP-likeHSQYICCVYVQQYDICCVVVCQYCQYGQYGQYGQYGQYCQYGNNNNNNNNNNNNNNNNNNN<t< th=""><th>M. stipitatus - SCP-like C K K N S P T W V<!--</th--><th>M. stipitatus - SCP-likeCKKKNSPFGASFTWQLWCNYTPPPS. amylolyticus - SCP-likeC-NGMDIIICKXNNDPPPPPP. narinus - CRISP-likeHSQYICCVVVQYGGGP. marinus - CRISP-likeSCPNNVVSCYPPPD. discoideum - Hypoth ProtC-DETIVVSCSYPP<th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWCNYTPPGNS. amylolyticus - SCP-likeC-NGMDIIICKKGNN<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWVNTPPGNFS. anylolyticus - SCP-likeCNGMDILICNYDPPGNVP. reichenowi - CRISP-likeC-KTKGNKKGNKKGNKGNKKGNKKGNNTNNTNKKGNKKGNNTNNTNNN<t< th=""></t<></th></th></th></t<></th></t<>	M. stipitatus - SCP-likeCKKNSPFGASFTWQLWCNNS. amylolyticus - SCP-likeC-NGMDILICKYYCQYP. reichenowi - CRISP-likeHSQYDICVILICKYYCQYP. mainius - CRISP-likeNYYCQYD. discoideum - Hypoth. Prot.C-DETIVXSSYCVYYCQYYNYCQYYNYCQYYCQYYCQYYCQYYCQYYCQYYCQYYYCQYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	M. stipitatus - SCP-likeCKKNSPFGASFPTWUWVNTTS. amylolyticus - SCP-likeC-NGMIIIIKYDP. reichenowi - CRISP-likeHSQYICCVYVQQYDICCVVVCQYCQYGQYGQYGQYGQYCQYGNNNNNNNNNNNNNNNNNNN <t< th=""><th>M. stipitatus - SCP-like C K K N S P T W V<!--</th--><th>M. stipitatus - SCP-likeCKKKNSPFGASFTWQLWCNYTPPPS. amylolyticus - SCP-likeC-NGMDIIICKXNNDPPPPPP. narinus - CRISP-likeHSQYICCVVVQYGGGP. marinus - CRISP-likeSCPNNVVSCYPPPD. discoideum - Hypoth ProtC-DETIVVSCSYPP<th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWCNYTPPGNS. amylolyticus - SCP-likeC-NGMDIIICKKGNN<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWVNTPPGNFS. anylolyticus - SCP-likeCNGMDILICNYDPPGNVP. reichenowi - CRISP-likeC-KTKGNKKGNKKGNKGNKKGNKKGNNTNNTNKKGNKKGNNTNNTNNN<t< th=""></t<></th></th></th></t<>	M. stipitatus - SCP-like C K K N S P T W V </th <th>M. stipitatus - SCP-likeCKKKNSPFGASFTWQLWCNYTPPPS. amylolyticus - SCP-likeC-NGMDIIICKXNNDPPPPPP. narinus - CRISP-likeHSQYICCVVVQYGGGP. marinus - CRISP-likeSCPNNVVSCYPPPD. discoideum - Hypoth ProtC-DETIVVSCSYPP<th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWCNYTPPGNS. amylolyticus - SCP-likeC-NGMDIIICKKGNN<</th><th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWVNTPPGNFS. anylolyticus - SCP-likeCNGMDILICNYDPPGNVP. reichenowi - CRISP-likeC-KTKGNKKGNKKGNKGNKKGNKKGNNTNNTNKKGNKKGNNTNNTNNN<t< th=""></t<></th></th>	M. stipitatus - SCP-likeCKKKNSPFGASFTWQLWCNYTPPPS. amylolyticus - SCP-likeC-NGMDIIICKXNNDPPPPPP. narinus - CRISP-likeHSQYICCVVVQYGGGP. marinus - CRISP-likeSCPNNVVSCYPPPD. discoideum - Hypoth ProtC-DETIVVSCSYPP <th>M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWCNYTPPGNS. amylolyticus - SCP-likeC-NGMDIIICKKGNN<</th> <th>M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWVNTPPGNFS. anylolyticus - SCP-likeCNGMDILICNYDPPGNVP. reichenowi - CRISP-likeC-KTKGNKKGNKKGNKGNKKGNKKGNNTNNTNKKGNKKGNNTNNTNNN<t< th=""></t<></th>	M. stipitatus - SCP-likeCKKKNSPFGASFPTWQLWCNYTPPGNS. amylolyticus - SCP-likeC-NGMDIIICKKGNN<	M. stipitatus - SCP-likeCKKNSPFGASFPTWQLWVNTPPGNFS. anylolyticus - SCP-likeCNGMDILICNYDPPGNVP. reichenowi - CRISP-likeC-KTKGNKKGNKKGNKGNKKGNKKGNNTNNTNKKGNKKGNNTNNTNNN <t< th=""></t<>

Appendix 9: Page 6 of 7

Appendix 9: Evolution of the Pathogenesis-Related (CAP/PR) Domain in the CRISP Protein Lineage

Appendix 9: This figure provides MEGA7 ClustalW amino acid sequence alignment of invertebrate and vertebrate CRISP and CRISP-like sequences. The CAP signature sequences are boxed, the highly conserved cysteine are highlighted in yellow with a red font color. Exon numbers and intron/exon borders were verified using the intron and exon prediction tracks in NCBI online data for specific CAP genes. Color-coding superimposed on the sequences is used to represent different exons and exon borders are indicated by a change in color. Organisms and accession IDs of the amino acids are available in the supplemental data table 17. Page 7 of 7

APPENDIX B

DATA COLLECTED JANUARY 2014 - DECEMBER 2015

Scientific Name:	Common Name:	Taxonomy ID:	Assigned Protein Name:	Protein Accession ID:
Peptoclostridium difficile 630	N/A	272563	Ca++ chelating serine	YP_001087944.1
Clostridium acetobutylicum	N/A	1488	SCP-like Transporter	WP_034580941.1
Bacillus koreensis	N/A	284581	SCP-like	WP_053403069.1
Pseudomonas mendocina	N/A	300	SCP-like	WP_012019283.1
Pseudomonas syringae pv. syringae B728a	N/A	205918	Allergen V5/Tpx-1 related	YP_236420.1
Vitis vinifera	Grape Vine	29760	Pr	XP_002273416.1
Theobroma cacao	Cocao	3641	CRISP-like CAP	XP_007041421.1
Sesamum indicum	Sesame	4182	Pr	XP_011092473.1
Solanum tuberosum	Potato	4113	Pr 1b	NP_001275095.1
Solanum lycopersicum	Tomato	4081	Pr P4	AAA03615.1
Leptosphaeria maculans JN3	N/A	985895	SCP-PRY-like	XP_003833499.1
Cordyceps militaris CM01	Mold	983644	PRY1	XP_006673225.1
Candida orthopsilosis Co 90-	Yeast (C. orthopsilosis)	1136231	PRY1P	XP_003866625.1
Candida glabrata CBS 138	Yeast (C. glabrata)	284593	PRY-like	XP_446695.1
Saccharomyces cerevisiae S288c	Yeast (S. cerevisiae)	559292	PRY1P	NP_012456.1
Homo sapiens	Human	9606	GAPR1	NP_001273939.1
Pan troglodytes	Chimpanzee	9598	GAPR1	XP_009454888.1
Rattus norvegicus	Rat (Norway rat)	10116	GAPR1	XP_002726558.2
Gallus gallus	Chicken	9031	GAPR1	XP_419085.1
Python biv ittatus	Burmese python	176946	GAPR1	XP_007442486.1
Aedes aegypti	Yellow fever mosquito	7159	Ag-like	 XP_001662120.1

Supplemental Data Table 1: Subfamilies of the CAP Superfamily

Supplementary data table 1: Page 1 of 2

Culex	Southern house			
quinquefasciatus	mosquito	7176	Ag5	XP_001845617.1
Microplitis demolitor	Parasitic wasp	69319	Ag3	XP_008545885.1
Solenopsis invicta	Red fire ant	13686	Ag3	NP_001291520.1
Nasonia vitripennis	Jewel wasp	7425	Ag5-like	XP_001603551.3
Rattus norvegicus	Rat (Norway rat)	10116	CRISP LCCL 2	NP_612527.2
Homo sapiens	Human	9606	CRISP LCCL 2	NP_113664.1
Xenopus (Silurana) tropicalis	W. clawed frog	8364	CRISP LCCL 2	NP_001027499.1
Rattus norvegicus	Rat (Norway rat)	10116	CRISP LCCL 1	NP_001128435.1
Xenopus (Silurana) tropicalis	W. clawed frog	8364	CRISP LCCL 1	XP_012820697.1
Homo sapiens	Human	9606	GliPR1	NP_006842.2
Pan troglodytes	Chimpanzee	9598	GliPR1	XP_009424141.1
Rattus norvegicus	Rat (Norway rat)	10116	GliPR1	NP_001011987.1
Python bivittatus	Burmese python	176946	GliPR1	XP_007445182.1
Homo sapiens	Human	9606	CRISP1	NP_001122.2
Xenopus (Silurana) tropicalis	W. clawed frog	8364	CRISP1	XP_002933627.1
Xenopus (Silurana) tropicalis	W. clawed frog	8364	CRISP2	NP_001008204.1
Python bivittatus	Burmese python	176946	CRISP-like	XP_007436999.1
Rattus norvegicus	Rat (Norway rat)	10116	CRISP2	XP_006244705.1
Homo sapiens	Human	9606	CRISP3	NP_001177915.1
Homo sapiens	Human	9606	CRISP2	NP_001135879.1
Pan troglodytes	Chimpanzee	9598	CRISP2	XP 009449701.1

Supplemental Data Table 1: Subfamilies of the CAP Superfamily

Supplementary data table 1: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 1 Subfamilies of the CAP Superfamily.

Page 2 of 2

Scientific Name:	Common Name:	Taxonomy	Assigned Protein	Protein Accession
Scientific Name:	Common Name:	ID:	Name:	ID:
Solanum	Tomato	4081	Pr P4	AAA03615.1
lycopersicum	Tomato	4001	F 1 F 4	AAA03013.1
Culex	Southern house	7176	Ag5	XP 001868686.1
quinquefasciatus	mosquito	/1/0	Ag5	AF_001000000.1
Vananus laguis	African clawed	8355	Allurin (CRISP2-	AAL12003.2
Xenopus laevis	frog	0333	like)	AAL12003.2
Homo sapiens	Human	9606	CRISP1	NP_001122.2
Homo sapiens	Human	9606	CRISP2	NP_003287.1
Homo sapiens	Human	9606	CRISP3	NP_001177915.1
Homo sapiens	Human	9606	CRISP4	NP_084309.1
Homo sapiens	Human	9606	GliPR1	NP_006842.2
Homo sapiens	Human	9606	GliPR1L1	NP_689992.1
Homo sapiens	Human	9606	GliPR1L2	NP_001257325.1
Homo sapiens	Human	9606	CRISPLD1	NP_113649.1
Homo sapiens	Human	9606	CRISPLD2	NP_113664.1
Homo sapiens	Human	9606	PI15	NP_056970.1
Homo sapiens	Human	9606	PI16	NP_001186088.1
Homo sapiens	Human	9606	R3HDML	NP_848586.1
Homo sapiens	Human	9606	CLEC18A	NP_001129686.1
Homo sapiens	Human	9606	CLEC18B	NP_001011880.2
Homo sapiens	Human	9606	CLEC18C	NP_775890.2
Homo sapiens	Human	9606	GAPR1	NP_001273939.1

Supplemental Data Table 2: Domain Structure of CAP Proteins

Supplementary data table 2: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 2. Domain Structure of CAP Proteins

Scientific Name:	Common Name:	Taxonomy ID:	Assigned Protein Name:	Prote in Accession ID:
Clostridium acetobutylicum	N/A	272562	SCP-like	NP_348847.1
Pseudomonas syringae	N/A	205918	SCP-Allergen-like	YP_236420.1
Pseudomonas mendocina	N/A	399739	SCP-like	WP_012019283.1
Anoxybacillus flavithermus	N/A	491915	Membrane Protein	WP_012575415.1
Mycobacterium avium	N/A	262316	SCP-like	WP_003874141.1
Peptoclostridium difficile	N/A	272563	Ca++ chelating serine	YP_001087944.1
Bacillus koreensis	N/A	284581	SCP-like	WP_053400013.1

Supplemental Data Table 3: Accession ID and Organism Information for ClustalW Amino Acid Alignment of CAP/SCP Proteins in Bacteria

Supplementary data table 3: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 3 ClustalW Amino Acid Alignment of CAP/SCP in Bacteria

Supplemental Data Table 4: Accession ID and Organism Information for
ClustalW Amino Acid Alignment of CAP/SCP Proteins in Plants

Scientific Name:	Common Name:	Taxonomy	Assigned Protein	Protein Accession
Scientific Name.	Common Name.	ID:	Name:	ID:
Solanum lycopersicum	Tomato	4081	PR P4	AAA03615.1
Solanum tuberosum	Potato	4113	PR 1b	NP_001275095.1
Theobroma cacao	Cocao	3641	CRISP-like	XP_007041421.1
Vitis vinifera	Grape (wine grape)	29760	PR-1	XP_002273416.1
Sesamum indicum	Sesame	4182	PR related	XP_011092473.1
Prunus persica	Peach	3760	PR related	XP_007201285.1

Supplementary data table 4: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 4 ClustalW Amino Acid Alignment of CAP/SCP in Plants

Scientific Name:	Common Name:	Taxonomy	Assigned Protein	Protein Accession
Sekiline Nume.	common rume.	ID:	Name:	ID:
Moniliophthora	Mushroom (M.	1381753	Pr-1	XP 007856948.1
roreri	roreri)	1301/33	F1-1	AF_007830948.1
Saccharomyces	Yeast (S.	559292	PRY1	NP 012456.1
cerevisiae	cerevisiae)	555252	FKII	NF_012430.1
Caudida alabuata	Yeast (C.	284593	PRY-like	XP_446695.1
Candida glabrata	glabrata)			
Candida	Yeast (C.	1136231	PRY1	XP 003866625.1
orthopsilosis	orthopsilosis)	1136231	FKII	AF_003800023.1
Leptosphaeria	Yeast (L.	985895	SCP-PRY-like	XP_003833499.1
maculans	maculans)		SCI-IKI-IKC	
Cordyceps militaris	Mold (C.	082644	PRY1	XP_006673225.1
	militaris)	983644		

Supplemental Data Table 5: Accession ID and Organism Information for ClustalW Amino Acid Alignment of CAP/SCP Proteins in Fungus

Supplementary data table 5: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 5 ClustalW Amino Acid Alignment of CAP/SCP in Fungus.

Scientific Name:	Common Name:	Taxonomy ID:	Assigned Protein Name:	Protein Accession ID:
Microplitis demolitor	Parasitic wasp	69319	Ag3	XP_008545885.1
Microplitis demolitor	Parasitic wasp	69319	Ag5	XP_008545751.1
Nasonia vitripennis	Jewel wasp	7425	Ag5-like	XP_001603551.3
Nasonia vitripennis	Jewel wasp	7425	Ag3-like	XP_008214984.1
Solenopsis invicta	Red fire ants	13686	Ag3	NP_001291520.1
Megachile rotundata	Leafcutter bee	143995	Ag3-like	XP_003708569.1

Supplemental Data Table 6: Accession ID and Organism Information for ClustalW Amino Acid Alignment of Insect Antigen-Related Proteins

Supplementary data table 6: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 7 ClustalW Amino Acid Alignment of Insect Antigen-Related Proteins.

Supplemental Data Table 7: Accession ID and Organism Information for ClustalW Amino Acid Alignment of Vertebrate CRISP2 and CRISP2-like Proteins

	Proteins					
Scientific Name:	Common Name:	Taxonomy	Assigned Protein	Protein Accession		
		ID:	Name:	ID:		
Petromyzon	S ea lamprey	7757	Uncharacterized	S4RMJ8_PETMA		
marinus	5 eu lamprey	1151	protein	SHICKING_I LINEY		
Lepisosteus	Spotted gar	7918	Uncharacterized	W5NHS1 LEPOC		
oculatus	Spotted gar	/ 210	protein			
			CRISP3 (cysteine-			
Danio rerio	Zebrafish	7955	rich venom natrin-	XP_003200319.1		
			like)			
Gadus morhua	Atlantic cod	8049	CRISP3	ENSGMOT000000		
Oudus mornud	Auanue cou	0049		07884		
Gasterosteus	Stickleback					
aculeatus	(three-spined	69293	CRISP3	G3PI03_GASAC		
acuteatus	stickleback)					
Oreochromis	Tilapia (nile	8128	CRISP3	I3JFJ5_ORENI		
niloticus	tilapia)	8128	CRISPS			
Oryzias latipes	Japanese rice fish	8000	CRISP3	H2MPL3_ORYLA		
Oryzius iuripes	Japanese nee lish	8090				
Xenopus (Silurana)	western clawed	8364	CRISP2	NP 001008204.1		
tropicalis	frog	0304	CRISF2	NF_001008204.1		
Xenopus laevis	African clawed	8355	CRISP2	NP_001082594.1		
Aenopus idevis	frog	0333				
			CRISP-like (cysteine-			
Python bivittatus	Burmese python	176946	rich venom latisemin-	XP_007436999.1		
			like)			
Anolis carolinensis	Green anole	28377	Cysteine-rich venom	XP 003215243.1		
Anous curounensis	Oreen anote	20377	helothermine-like	AI_003213243.1		
Pelodiscus sinensis	Chinese soft-	13735	Uncharacterized	VD 006127110 1		
r elouiscus sinensis	shelled turtle	13/33	LOC102443713	XP_006137110.1		
Falco peregrinus	Peregrine falcon	8954	CRISP2	XP_005239276.1		
	Golden-collared	270015	CDISD2	VD 0080107671		
Manacus vitellinus	manakin	328815	CRISP2	XP_008919767.1		
Picoides pubescens	Downy	110000	CRISP2	VD 0000025261		
	woodpecker	118200		XP_009903526.1		
Gallus gallus	Chicken	9031	CRISP2	XP_420051.3		
Ornithorhynchus	Diatrona		CDICD2 Has	VD 001512420.2		
anatinus	Platypus	9258	CRISP3-like	XP_001512430.2		
	•		•	•		

Supplementary data table 7:

Page 1 of 2

Supplemental Data Table 7: Accession ID and Organism Information for ClustalW Amino Acid Alignment of Vertebrate CRISP2 and CRISP2-like Proteins

Dasypus novemcinctus	Armadillo (nine- banded armadillo)	9361	CRISP2	XP_004471246.1
Mus musculus	House mouse	10090	CRISP2	NP_001191000.1
Rattus norvegicus	Norway rat	10116	CRISP2	XP_006244705.1
Pan troglodytes	Chimpanzee	9598	CRISP2	XP_009449701.1
Homo sapiens	Human	9606	CRISP2	NP_001135879.1
Xenopus (Silurana) tropicalis	western clawed frog	8364	Allurin (CRISP2- like)	NP_001188271.1
Xenopus laevis	African clawed frog	8355	Allurin (CRISP3- like)	AAL12003.2
Taeniopygia guttata	Zebra finch	59729	CRISP2	XP_012428365.1

Supplementary data table 7: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 8 - ClustalW Amino Acid Alignment Vertebrate CRISP2 and CRISP2-like Proteins; Figure 13 - Genomic Exon Border Structure In Selected Vertebrate CRISP2 and CRISP2-like Genes; and Figure 14 - ClustalW Amino Acid Alignment for Exon Borders in Selected CRISP and CRISP-like Genes.

Page 2 of 2

Supplemental Data Table 8: Accession ID and Organism Information for ClustalW Amino Acid Alignment and Secondary Structure of Chinese Cobra CRISP Protein

Scientific Name:	Common Name:	Taxonomy	Assigned Protein	Protein Accession
Scientific Name.	Common Name.	ID:	Name:	ID:
Naja atra	Chinese cobra	8656	Cysteine-rich venom protein natrin-1	Q7T1K6.1
Petromyzon marinus	Sea lamprey	7757	Uncharacterized protein	S4RMJ8_PETMA
Oreochromis niloticus	Tilapia (nile tilapia)	8128	CRISP3	I3JFJ5_ORENI
Python bivittatus	Burmese python	176946	CRISP-like (cysteine- rich venom latisemin- like)	XP_007436999.1
Falco peregrinus	Peregrine falcon	8954	CRISP2	XP_005239276.1
Homo sapiens	Human	9606	CRISP2	NP_001135879.1
Xenopus laevis	African clawed frog	8355	Allurin (CRISP3- like)	AAL12003.2

Supplementary data table 8: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 10 ClustalW Amino Acid Alignment and Secondary Structure of Chinese Cobra CRISP Protein.

Scientific Name:	Common Name:	Taxonomy	Assigned Protein	Protein Accession
Sciencific Ivanie.	common reame.	ID:	Name:	ID:
Caenorhabditis	Roundworm	6239	SCP-Like	ND 502506 1
elegans	Kouldwollii	0239	extracellular protein	NP_502506.1
Drosophila	Fruit fly	7227	CG42780	NP 001188554.1
melanogaster		1221	042/80	NP_001100334.1
Microplitis	Demositie ween	69319	Serotriflin	VD 009552510 1
demolitor	Parasitic wasp	09319	Serouniiin	XP_008553510.1
Nasonia vitripennis	Jewel wasp	7425	CRISP2-like	XP_008207728.1
Solenopsis invicta	Red fire ants	13686	CRISP1-like	XP_011162333.1
Apis florea	Little honeybee	7463	CRISP2-like	XP_003690882.1
Megachile	Leafcutter bee	143995	CRISP2-like	VD 003701024 1
rotundata	Lealeutter Dee	143993	CRISF 2-like	XP_003701024.1
Conus textile	Snail (cloth-of-	6494	Cysteine-rich venom	Q7YT83.1
	gold cone)	0494	protein	Q/1103.1
Ciona savignyi	Sea squirt (C. s.)	51511	Uncharacterized	H2Z9A1_CIOSA
			protein	
Ciona intestinalis	Sea squirt (C. i.)	7719	CRISP3-like	XP_002121335.1

Supplemental Data Table 9: Accession ID and Organism Information for ClustalW Amino Acid Alignment of Invertebrate CRISP2-like Proteins

Supplementary data table 9: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 11 ClustalW Amino Acid Alignment Invertebrate CRISP2-like Proteins

Supplemental Data Table 10: Accession ID and Organism Information for Genomic Exon Border Structure In Selected Invertebrate CRISP and CRISP-like Amino Acid Sequences

Scientific Name:	Common Name:	Taxonomy ID:	Assigned Protein Name:	Protein Accession ID:
Solanum lycopersicum	Tomato	4081	PR P4	AAA03615.1
Theobroma cacao	Cocao	3641	CRISP-like	XP_007041421.1
Vitis vinifera	Grape (wine grape)	29760	PR-1	XP_002273416.1
Saccharomyces cerevisiae	Yeast (S. cerevisiae)	559292	PRY1	NP_012456.1
Aspergillus flavus	Mold (A. f.)	332952	CRISP-like	XP_002372142.1
Caenorhabditis elegans	Roundworm	6239	SCP-Like extracellular protein	NP_502506.1
Drosophila melanogaster	Fruit fly	7227	CG42780	NP_001188554.1
Solenopsis invicta	Red fire ants	13686	CRISP1-like	XP_011162333.1
Megachile rotundata	Leafcutter bee	143995	CRISP2-like	XP_003701024.1
Apis florea	Dwarf honey bee	7463	CRISP2-like	XP_003690882.1
Microplitis demolitor	Parasitic wasp	69319	Serotriflin	XP_008553510.1
Nasonia vitripennis	Jewel wasp	7425	CRISP2-like	XP_008207728.1
Ciona savignyi	Sea squirt (C. s.)	51511	Uncharacterized protein	H2Z9A1_CIOSA
Ciona intestinalis	Sea squirt (C. i.)	7719	CRISP3-like	XP_002121335.1
Homo sapiens	Human	9606	CRISP2	NP_001135879.1

Supplementary data table 10: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 15 Genomic Exon Border Structure In Selected Invertebrate CRISP and CRISP-like Amino Acid Sequence.

Supplemental Data Table 11: Accession ID and Organism Information for
Phylogenetic Relationships of All CAP Superfamily Proteins In X. tropicalis
(Western Clawed Frog) Genome

Assigned Protein	Protein Accession
Name:	ID:
Serotriflin-like	XP_002933626.1
CRISP2	NP_001008204.1
CRISP1	XP_002933627.1
Serotriflin-like	XP_004915026.1
MGC108118	NP_001025526.1
CRISP3	NP_989314.1
Allurin	NP_001188271.1
TEL1	XP_012819576.1
PI15	XP_002937120.1
PI R3HDML	XP_012808160.1
CRISP LCCL1	XP_012820697.1
CRISP LCCL2	NP_001027499.1
GliPR1-like	XP_002940783.2
GliPR2	NP_001107504.1
RhAG	XP_002933645.2

Supplementary data table 11: The following table presents the assigned protein name and the protein accession ID number for all the amino acid sequences used in Figure 16 - Phylogenetic Relationships of All CAP Superfamily Proteins In X. tropicalis (Western Clawed Frog) Genome; Figure 17 - Genomic Exon Border Structure In CAP Superfamily Proteins in the *X. tropicalis* (Western Clawed Frog) Genome. All sequences belong to *Xenopus (Silurana) tropicalis* (Western clawed frog) – Taxonomy ID: 8355).

(IIuman) Genome					
Assigned Protein	Protein Accession				
Name:	ID:				
CLEC18A	NP_001129686.1				
CLEC18B	NP_001011880.2				
CLEC18C	NP_775890.2				
CRISP1	NP_001122.2				
CRISP3	NP_001177915.1				
CRISP2	NP_001135879.1				
PI16	NP_001186088.1				
GliPR1-like 1	NP_001291893.1				
GliPR1	NP_006842.2				
GliPR1-like 2	NP_001257325.1				
PI15	NP_056970.1				
PI R3HDML	NP_848586.1				
GliPR2	NP_001273939.1				
CRISP LCCL1	NP_001273706.1				
CRISP LCCL2	NP_113664.1				
RhAG	NP_000315.2				

Supplemental Data Table 12: Accession ID and Organism Information for Phylogenetic Relationships Of All CAP Superfamily Proteins In the *Homo sapiens* (Human) Genome

Supplementary data table 12: The following table presents the assigned protein name and the protein accession ID number for all the amino acid sequences used in Figure 18 - Phylogenetic Relationships Of All CAP Superfamily Proteins In Homo sapiens (Human) Genome. All sequences belong to *Homo Sapiens* (human) – Taxonomy ID: 9606).

Scientific Name:	Common Name:	Taxonomy	Assigned Protein	Protein Accession
Scientific Ivanie.	Common Name.	ID:	Name:	ID:
Homo sapiens	Human	9606	CRISP2	NP_001122.2
Danio rerio	Zebrafish	7955	CRVP-like	XP_003200319.1
Python bivittatus	Burmese Python	176946	CRVP-like	XP_007436999.1
Equus caballus	Horse	9796	CRISP2	NP_001075401.1
Microplitis demolitor	Parasytic wasp	69319	CRVP-like	XP_014298016.1
Solenopsis invicta	Red fire ants	13686	CRISP1-like	XP_011162337.1
Apis dorsata	Giant honeybee	7462	CRISP1-like	XP_006608542.1
Bombus impatiens	Bumble bee	132113	CRISP2-like	XP_012250250.1
Ciona intestinalis	Sea squirt (C.i.)	7719	GliPR1-like 1	XP_004227533.1
Aplysia californica	California sea hare	6500	GliPR1-like	XP_012943149.1
Python bivittatus	Burmese Python	176946	GliPR1-like	XP_007445182.1
Homo sapiens	Human	9606	GliPR1-like 1	NP_001291893.1
Homo sapiens	Human	9606	GliPR1	NP_006842.2
Equus caballus	Horse	9796	GliPR1	XP_001914748.2
Mus musculus	Mouse	10090	CRISP LCCL 2	EDL11624.1
Danio rerio	Zebrafish	7955	CRISP LCCL-like	XP_009296842.1
Homo sapiens	Human	9606	CRISP LCCL 1	NP_113649.1
Bos taurus	Cow	9913	CRISP LCCL 1	NP_001094635.1
Brugia malayi	Nematode	6279	SCP-like	XP_001894274.1
Caenorhabditis remanei	Roundworm (C. r.)	31234	HP CRE_10563	XP_003095745.1
Caenorhabditis briggsae	Roundworm (C. b.)	6238	HP CBG01834	XP_002634263.1
Caenorhabditis elegans	Roundworm (C. e.)	6239	SCP-like	NP_502502.1
Caenorhabditis elegans	Roundworm (C. e.)	6239	SCP-like 1	NP_504055.1
Caenorhabditis elegans	Roundworm (C. e.)	6239	SCP-like 2	NP_504056.1
Bactrocera oleae	Olive fruit fly	104688	GAPR1-like	XP_014099442.1
Hydra vulgaris	Hydra	6087	GAPR1-like	XP_012559751.1

Supplemental Data Table 13: Accession ID and Organism Information for Evolution of Multi-Domain CAP Superfamily Proteins

Supplementary data table 13:

Page 1 of 2

Supplemental Data Table 13: Accession ID and Organism Information for Evolution of Multi-Domain CAP Superfamily Proteins.

Ciona intestinalis	Sea squirt (C. i.)	7719	GAPR1	XP_002128906.1
Limulus polyphemus	Horseshoe crab	6850	GAPR1-like	XP_013778105.1
Aplysia californica	California sea hare	6500	GAPR1-like	XP_012946763.1
Danio rerio	Zebrafish	7955	GAPR1	XP_009290605.1
Macaca mulatta	Rhesus monkey		GAPR1	NP_001253658.1
Mus musculus	Mouse	10090	GAPR1	NP_081726.1
Methylobacterium radiotolerans	Bacteria (M. radiotolerans)	31998	SCP-like	WP_058607187.1
Paramecium tetraurelia	Paramecium (P. tetraurelia)	5888	SCP-like	XP_001462314.1
Ichthyophthirius multifiliis	Protozoan (L. multifiliis)	5932	SCP-like	XP_004035428.1

Supplementary data table 13: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Figure 21 Evolution of Multi-Domain CAP Superfamily Proteins.

Page 2 of 2

Scientific Name:	Common Name:	Taxonomy	Assigned Protein	Protein Accession
~ 1		ID:	Name:	ID:
Solanum lycopersicum	Tomato	4081	PR P4	AAA03615.1
Theobroma cacao	Cocao	3641	CRISP-like	XP_007041421.1
Saccharomyces cerevisiae	Yeast (S. cerevisiae)	559292	PRY1	NP_012456.1
Aspergillus flavus	Mold (A. f.)	332952	CRISP-like	XP 002372142.1
Caenorhabditis elegans	Roundworm	6239	SCP-Like extracellular protein	NP_502506.1
Drosophila melanogaster	Fruit fly	7227	CG42780	NP_001188554.1
Microplitis demolitor	Parasitic wasp	69319	S erotriflin	XP_008553510.1
Nasonia vitripennis	Jewel wasp	7425	CRISP2-like	XP_008207728.1
Solenopsis invicta	Red fire ants	13686	CRISP1-like	XP_011162333.1
Apis florea	Dwarf honey bee	7463	CRISP2-like	XP_003690882.1
Megachile rotundata	Leafcutter bee	143995	CRISP2-like	XP_003701024.1
Ciona savignyi	Sea squirt (C. s.)	51511	Uncharacterized protein	H2Z9A1_CIOSA
Ciona intestinalis	Sea squirt (C. i.)	7719	CRISP3-like	XP 002121335.1
Petromyzon marinus	Sea lamprey	7757	Uncharacterized protein	S4RMJ8_PETMA
Lepisosteus oculatus	Spotted gar	7918	Uncharacterized protein	W5NHS1_LEPOC
Danio rerio	Zebrafish	7955	CRISP3 (cysteine- rich venom natrin- like)	XP_003200319.1
Gadus morhua	Atlantic cod	8049	CRISP3	ENSGMOT000000 07884
Gasterosteus aculeatus	Stickleback (three-spined stickleback)	69293	CRISP3	G3PI03_GASAC
Oreochromis niloticus	Tilapia (nile tilapia)	8128	CRISP3	I3JFJ5_ORENI

Supplemental Data Table 14: Accession ID and Organism Information for ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Protein Sequences (Appendix 2)

Supplementary data table 14:

Page 1 of 2

Selected Sequences (Appendix 2)					
Oryzias latipes	Japanese rice fish	8090	CRISP3	H2MPL3_ORYLA	
Xenopus (Silurana) tropicalis	Western clawed frog	8364	Allurin (CRISP2- like)	NP_001188271.1	
Xenopus laevis	African clawed frog	8355	Alhrin (CRISP3- like)	AAL12003.2	
Python biv ittatus	Burmese python	176946	CRISP-like (cysteine- rich venom latisemin- like)		
Anolis carolinensis	Green anole	28377	Cysteine-rich venom helothermine-like	XP_003215243.1	
Pelodiscus sinensis	Chinese soft- shelled turtle	13735	Uncharacterized LOC102443713	XP_006137110.1	
Falco peregrinus	Peregrine falcon	8954	CRISP2	XP_005239276.1	
Manacus vitellinus	Golden-collared manakin	328815	CRISP2	XP_008919767.1	
Picoides pubescens	Downy woodpecker	118200	CRISP2	XP_009903526.1	
Gallus gallus	Chicken	9031	CRISP2	XP_420051.3	
Ornithorhynchus	Platypus	9258	CRISP3-like	XP_001512430.2	

anatinus

Dasypus

novemcinctus

Physeter catodon

Rattus norvegicus

Pan troglodytes

Homo sapiens

Mus musculus

Pantholops

hodgsonii

Supplemental Data Table 14: Accession ID and Organism Information for ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Sequences (Appendix 2)

Supplementary data table 14: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Appendix 2 ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Sequence. Page 2 of 2

9361

59538

9755

10090

10116

9598

9606

CRISP2

CRISP2

CRISP2

CRISP2

CRISP2

CRISP2

CRISP2

XP 004471246.1

XP 005960701.1

XP 007101973.1

NP 001191000.1

XP 006244705.1

XP 009449701.1

NP 001135879.1

Armadillo (nine-

Chiru (Antelope)

Sperm whale

House mouse

Norway rat

Chimpanzee

Human

banded

armadillo)

Scientific Name:	Common Name:	Taxonomy ID:	Assigned Protein Name:	Protein Accession ID:
Solanum lycopersicum	Tomato	4081	PR P4	AAA03615.1
Theobroma cacao	Cocao	3641	CRISP-like	XP_007041421.1
Vitis vinifera	Grape (wine grape)	29760	PR-1	XP_002273416.1
Saccharomyces cerevisiae	Yeast (S. cerevisiae)	559292	PRY1	NP_012456.1
Aspergillus flavus	Mold (A. f.)	332952	CRISP-like	XP_002372142.1
Caenorhabditis elegans	Roundworm	6239	SCP-Like extracellular protein	NP_502506.1
Drosophila melanogaster	Fruit fly	7227	CG42780	NP_001188554.1
Microplitis demolitor	Parasitic wasp	69319	Serotriffin	XP_008553510.1
Nasonia vitripennis	Jewel wasp	7425	CRISP2-like	XP_008207728.1
Solenopsis invicta	Red fire ants	13686	CRISP1-like	XP_011162333.1
Apis florea	Dwarf honey bee	7463	CRISP2-like	XP_003690882.1
Megachile rotundata	Leafcutter bee	143995	CRISP2-like	XP_003701024.1
Ciona savignyi	Sea squirt (C. s.)	51511	Uncharacterized protein	H2Z9A1_CIOSA
Ciona intestinalis	Sea squirt (C. i.)	7719	CRISP3-like	XP_002121335.1
Homo sapiens	Human	9606	CRISP2	NP_001135879.1

Supplemental Data Table 15: Accession ID and Organism Information for ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Invertebrate Sequence (Appendix 3)

Supplementary data table 15: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Appendix 3: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for Selected Invertebrate Sequence.

Scientific Name:	Common Name:	Taxonomy ID:	Assigned Protein Name:	Protein Accession ID:
Solanum	Tomato	4081	PR P4	AAA03615.1
lycopersicum	10111110	4001	rKr4	AAA03013.1
Theobroma cacao	Cocao	3641	CRISP-like	XP_007041421.1
Saccharomyces	Yeast (S.	559292	PRY1	NP 012456.1
cerevisiae	cerevisiae)	557272		
Aspergillus flavus	Mold (A. f.)	332952	CRISP-like	XP_002372142.1
Caenorhabditis elegans	Roundworm	6239	Uncharacterized protein	NP_509707.2
Drosophila melanogaster	Fruit fly	7227	Uncharacterized protein	NP_731097.1
Microplitis demolitor	Parasitic wasp	69319	EFG1	XP_008214060.1
Nasonia vitripennis	Jewel wasp	7425	Uncharacterized protein	XP_014296097.1
Solenopsis invicta	Red fire ants	13686	PRY1-like	XP_011157510.1
Apis florea	Dwarfhoney bee	7463	EFG1	XP_012340180.1
Megachile rotundata	Leafcutter bee	143995	EFG1	XP_012138605.1
Ciona intestinalis	Sea squirt (C. i.)	7719	GAPR	XP_002128928.1
Lepisosteus oculatus	Spotted gar	7918	GAPR	XP_006635797.1
Danio rerio	Zebrafish	7955	GAPR	NP_001005978.1
Xenopus (Silurana) tropicalis	Western clawed frog	8364	GAPR	NP_001107504.1
Python bivittatus	Burmese python	176946	GAPR	XP 007442486.1
Anolis carolinensis	Green anole	28377	GAPR	XP_003222287.1
Pelodiscus sinensis	Chinese soft- shelled turtle	13735	GAPR	XP_006113099.1
Falco peregrinus	Peregrine falcon	8954	GAPR	XP_005240380.2
Gallus gallus	Chicken	9031	GAPR	XP_419085.1
Ornithorhynchus anatinus	Platypus	9258	GAPR	XP_001511959.1

Supplemental Data Table 16: Accession ID and Organism Information for ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for GAPR/GLIPR2 Sequences (Appendix 8)

Supplementary data table 16:

Page 1 of 2

Supplemental Data Table 16: Accession ID and Organism Information for ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for GAPR/GLIPR2 Sequences (Appendix 8)

Dasypus novemcinctus	Armadillo (nine- banded armadillo)	9361	GAPR	XP_012380143.1
Mus musculus	House mouse	10090	GAPR	XP_002726558.2
Rattus norvegicus	Norway rat	10116	GAPR	NP_081726.1
Pan troglodytes	Chimpanzee	9598	GAPR	XP_009454888.1
Homo sapiens	Human	9606	GAPR	NP_001273939.1

Supplementary data table 16: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Appendix 8: ClustalW Amino Acid Alignment and Color-Coded Mapping of Exons for GAPR/GLIPR2 Sequences

Supplemental Data Table 17: Accession ID and Organism Information for Evolution of the Pathogenesis-Related (PR/SCP) Domain in the CRISP Protein Lineage (Appendix 9)

Scientific Name:	Common Name:	Taxonomy ID:	Assigned Protein Name:	Protein Accession ID:
Archangium gephyra	A. gephyra	48	SCP-like HP	WP_015353234.1
Myxococcus stipitatus	M. stipitatus	83455	SCP-like family protein	WP_015353234.1
Sandaracinus amylolyticus	S. amylolyticus	927083	SCP-like HP	WP_053237857.1
Plasmodium reichenowi	P. reichenowi	5854	CRISP-like	XP_012762159.1
Ichthyophthirius multifiliis	I. multifiliis	5932	SCP-like	XP_004035428.1
Perkinsus marinus	P. marinus	423536	CRISP-like	XP_002768599.1
Hammondia hammondi	H. hammondi	99158	SCP-like	XP_008887770.1
Dictyostelium discoideum	D. discoideum	352472	HP DDB_G0288221	XP_636852.1
Cordyceps militaris	C. militaris	983644	Extracellular SCP domain protein PRY1	XP_006673225.1
Phaeoacremonium minimum	P. minimum	1286976	Putative scp-like extracellular protein	XP_007915157.1
Neofusicoccum parvum	N. parvum	1287680	Putative allergen v5 tpx-1- related protein	XP_007589546.1
Bipolaris maydis	B. maydis	665024	HP COCC4DRAFT_36619	XP_014082918.1
Hydra vulgaris	H. vulgaris	6087	PR-1b-like	XP_012559646.1
Hydra vulgaris	H. vulgaris	6087	GAPR1-like	XP_002158876.3
Hydra vulgaris	H. vulgaris	6087	GAPR1-like	XP_012559751.1
Nematostella vectensis	N. vectensis	45351	Predicted Protein	XP_001631455.1
Brugia malayi	B. malayi	6279	SCP-like	XP_001894274.1
Trichinella spiralis	T. spiralis	6334	ION-1 Protein	XP_003377286.1
Caenorhabditis elegans	Nematodes (C. e.)	6239	SCP-like 12	NP_504056.1
Caenorhabditis elegans	Nematodes (C. e.)	6239	SCP-like 14	NP_502498.1
Caenorhabditis elegans	Nematodes (C. e.)	6239	SCP-like 1	NP_502502.1

Supplementary data table 17:

Page 1 of 2

Supplemental Data Table 17: Accession ID and Organism Information for Evolution of the Pathogenesis-Related (PR/SCP) Domain in the CRISP Protein Lineage (Appendix 9)

Apis dorsata	Giant honeybee	7462	CRISP1-like	XP_006608542.1
Bombus impatiens	Eastern bumble bee	132113	CRISP2-like	XP_012250250.1
Monomorium pharaonis	Pharaoh ant	307658	CRISP1-like	XP_012529279.1
Tribolium castaneum	Red flour beetle	7070	CRISP2-like	XP_973363.1
Papilio xuthus	Asian swallowtail	66420	CRISP DIS2-like	XP_013168180.1
Ciona savignyi	Pacific transparent sea squirt	51511	CRISP-like	H2Z9A1_CIOSA
Saccoglossus kowalevskii	S. kowalevskii	10224	Uncharacterized protein LOC102806400	XP_006826071.1
Branchiostoma floridae	Florida lancelet	7739	Hypothetical protein BRAFLDRAFT_220379	XP_002604410.1
Strongylocentrotus purpuratus	Purple sea urchin	7668	CRISP2-like	XP_011662194.1
Crassostrea gigas	Pacific oyster	29159	GLIPR1-like1	XP_011424284.1
Danio rerio	Zebrafish	7955	CRVP natrin-1-like	XP_003200319.1
Python bivittatus	Burmese python	176946	CRVP latisemin-like	XP_007436999.1
Equus caballus	Horse	9796	CRISP2	NP_001075342.1
Homo sapiens	Human	9606	CRISP1	NP_001122.2
Homo sapiens	Human	9606	CRISP3	NP_006052.2

Supplementary data table 17: The following table presents the scientific name, common name, taxonomy ID number, assigned protein name, and the protein accession ID number for all the amino acid sequences used in Appendix 9: Evolution of the Pathogenesis-Related (PR/SCP) Domain in the CRISP Protein Lineage Page 2 of 2