

5-1-2015

A Novel Family of Major Histocompatibility Complex Class I Genes in Marsupials and Monotremes

Katina Krasnec

Follow this and additional works at: https://digitalrepository.unm.edu/biol_etds

Recommended Citation

Krasnec, Katina. "A Novel Family of Major Histocompatibility Complex Class I Genes in Marsupials and Monotremes." (2015). https://digitalrepository.unm.edu/biol_etds/66

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Biology ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Katina Krasnec

Candidate

Biology

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Rob Miller, Chairperson

Irene Salinas

Tom Turner

Michael Criscitiello

**A NOVEL FAMILY OF MAJOR HISTOCOMPATIBILITY
COMPLEX CLASS I GENES IN MARSUPIALS AND
MONOTREMES**

by

KATINA KRASNEC

B.A., Biology, Bryn Mawr College, 2007

M.S., Anthropology, University of New Mexico, 2011

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

**Doctor of Philosophy
Biology**

The University of New Mexico
Albuquerque, New Mexico

May, 2015

ACKNOWLEDGMENTS

I would like to thank the members of my committee, including Rob Miller for his guidance and patience, Irene Salinas for her incredible support and advice, and Tom Turner and Mike Criscitiello for their help and advice.

Thank you to my lab members Tori Hansen, Bethaney Fehrenkamp, Ben Wheaton, and Gaby Chacon for not only help, but also encouragement and camaraderie.

Thank you to my parents, Joseph and Maria Krasnec, for their guidance and support and for pushing me to challenge my intellectual boundaries. Thanks also to Alex Krasnec, for being the big brother we all need and providing unconditional encouragement.

And my greatest thanks to Sam Golbuff for his unlimited support and love. Sappy platitudes aside, I would have never finished without you.

A Novel Family of Major Histocompatibility Complex Class I Genes in Marsupials and Monotremes

By

Katina Krasnec

B.A., Biology, Bryn Mawr College, 2007

M.S., Anthropology, University of New Mexico, 2011

Ph.D., Biology, University of New Mexico, 2015

ABSTRACT

The Major Histocompatibility Complex (MHC) class I family of genes encode for molecules that have well-conserved structures, but have evolved to perform a diverse functions. The availability of an opossum genome from the grey, short-tailed opossum, *Monodelphis domestica*, has allowed for analysis of MHC class I genes in a marsupial. Traditional methods for gene discovery uncovered 13 MHC class I genes in the opossum. Utilization of a novel method to search for MHC domain structures discovered a family of 17 novel MHC class I genes. These genes, named *ModoUT1-17*, were located in a cluster on chromosome 1, unlinked to the MHC. *UT* homologues are only found in marsupial and monotreme genomes, consistent with being ancient in mammals yet lost in eutherians.

Twelve of the *ModoUT* loci are transcribed in the opossum thymus. The majority of *UT* transcription is in the thymus or skin, with limited expression in other tissues. Full-

length sequencing of eleven transcribed *ModoUT* genes revealed between five and eight exons, with typical class I gene structure and few alternative splice variants. A survey of *ModoUT* polymorphism in different *M. domestica* populations found low levels of polymorphism. Limited positive selection occurs in any of the *ModoUT* genes, suggesting they may not be under pathogen-mediated pressure.

Also uncovered in *M. domestica* genome search was the presence of two additional loci of the *ModoUA* gene, now designated *ModoUA3* and *ModoUA4*. The *ModoUA* gene is thought to be the class I molecule involved in peptide presentation. These new genes were uncovered in a region of the genome that was expanded and more complete than in earlier genome assemblies. The occurrence of five to six alleles in individual *M. domestica* indicates three loci being transcribed. The *ModoUA1* and *ModoUA3* genes are highly similar and alleles cannot be distinguished, while *ModoUA4* is easily identifiable, although less common and also relatively non-polymorphic and not under positive selection. The use of later assemblies and novel search methods confirmed the existence of three related MHC class I genes in the opossum, making opossums more typical of mammals by having multiple classical MHC class I loci.

TABLE OF CONTENTS

LIST OF FIGURES.....	viii
LIST OF TABLES.....	ix
CHAPTER 1: Introduction to MHC Class I gene evolution and marsupial immunology.....	1
Evolution of Structure-Function Relationships.....	2
MHC Class I Family as a research model.....	3
Theories on origin of MHC class I structure.....	15
Theories on origin and generation of MHC class I genes.....	17
Searching genomes for more MHC class I molecules.....	18
Marsupials as a model research organisms.....	19
Early immunological research on marsupials.....	21
The unique marsupial immune system.....	22
MHC research in marsupials.....	23
References.....	25
CHAPTER 2: Marsupials and monotremes possess a novel family of MHC class I genes that is lost from the eutherian lineage.....	32
Abstract.....	33
Introduction.....	34
Materials and Methods.....	38
Results.....	45
Sensitive peptide searches for MHC class I proteins.....	45
Sensitive genome searches for MHC class I genes.....	47
Phylogenetic analysis.....	49
Chromosomal location.....	52
Sequencing and gene expression.....	52
Homology mapping.....	55
Discussion.....	58
Conclusion.....	62
References.....	63
Supplementary Figures and Tables.....	68

CHAPTER 3: The <i>UT</i> family of MHC class I loci unique to non-eutherian mammals have limited polymorphism and tissue specific patterns of expression in the opossum.....	80
Abstract.....	81
Introduction.....	82
Materials and Methods.....	85
Results.....	92
Analysis of <i>ModoUT</i> transcription.....	92
<i>ModoUT</i> genes organization.....	94
Polymorphism and evidence of positive selection.....	96
Expression of <i>ModoUT8</i> in thymic but not peripheral $\alpha\beta$ T cells.....	101
Discussion.....	103
References.....	107
Supplementary Figures and Tables.....	110
CHAPTER 4: The MHC genomic region of the opossum <i>Monodelphis domestica</i> reexamined.....	112
Abstract.....	113
Introduction.....	114
Materials and Methods.....	118
Results & Discussion.....	123
References.....	131
APPENDIX 1: Additional publication.....	134
The pathology and pathogenicity of a novel <i>Haemoproteus</i> spp. infection in wild Little Penguins (<i>Eudyptula minor</i>)	
APPENDIX 2: Large Supplemental Tables from Chapter 2.....	145
Supplemental Table 2.....	145
Supplemental Table 3.....	158
Supplemental Table 4.....	211

LIST OF FIGURES

CHAPTER 1

Figure 1.....	4
Figure 2.....	14
Figure 3.....	16

CHAPTER 2

Figure 1.....	40
Figure 2.....	50
Figure 3.....	53
Figure 4.....	54
Figure 5.....	56
Supplemental Figure 1.....	68
Supplemental Figure 2.....	69
Supplemental Figure 3.....	70
Supplemental Figure 4.....	71
Supplemental Figure 5.....	72

CHAPTER 3

Figure 1.....	93
Figure 2.....	95
Figure 3.....	97
Figure 4.....	100
Figure 5.....	102
Supplemental Figure 1 & 2.....	111

CHAPTER 4

Figure 1.....	124
Figure 2.....	126
Figure 3.....	127
Figure 4.....	128

LIST OF TABLES

CHAPTER 1

Table 1.....	5
--------------	---

CHAPTER 2

Table 1.....	46
Supplemental Table 1.....	72
Supplemental Table 2.....	73
Supplemental Table 3.....	74
Supplemental Table 4.....	75
Supplemental Table 5.....	76
Supplemental Table 6.....	77
Supplemental Table 7.....	78

CHAPTER 3

Table 1.....	89
Supplemental Table 1.....	110

CHAPTER 4

Table 1.....	117
--------------	-----

CHAPTER 1

INTRODUCTION TO MHC CLASS I GENE EVOLUTION AND MARSUPIAL IMMUNOLOGY

By Katina V. Krasnec and Robert D. Miller

Center for Evolutionary & Theoretical Immunology, Department of Biology, University
of New Mexico, Albuquerque, NM, 87131, USA

KVK and RDM drafted manuscript. KVK generated figures

THE EVOLUTION OF STRUCTURE-FUNCTION RELATIONSHIPS

The relationship between the structure of a molecule and its function is a fundamental question in both structural and evolutionary biology. Model systems can be used to thoroughly examine how the structure of a molecule affects function. An ideal model system in which to examine this relationship is the Major Histocompatibility Complex class I family. The MHC class I molecules are best defined by the complex structure they all share. However, a striking difference can be seen in how these molecules have evolved to perform a wide variety of functions, indicating the plasticity of molecular structures (Kasahara, 1997, Kelley et al. 2005, Klein et al. 2007).

Examples of other molecules with conserved structures and diverse functions include serpins and lipocalins. Serpins, short for serine proteinase inhibitors, have been found across a vast number of organisms, from humans down to viruses (Silverman et al. 2001). The majority of serpins are comprised of a conserved structure comprised of three β sheets, a minimum of 7 α -helices, and a reactive site loop (RSL) that contains approximately 17 flexible residues that tether two of the β sheets together (Silverman et al. 2001). Researchers have identified serpins that inhibit proteinases, including caspase and cysteine proteinase, while others have been shown to have functions outside inhibition, and serve important roles in hormone transport and blood pressure regulation. Another molecule, lipocalin, shares conserved regions of sequence motifs that result in conserved protein folds resulting in an antiparallel β barrel. Unlike serpins, lipocalin has been found primarily in vertebrates. Lipocalins are involved in diverse functions including ligand binding, binding of receptors for retinol and bilin, along with formation of complexes with macromolecules (Flower, 1996).

THE MHC CLASS I FAMILY AS A RESEARCH MODEL

The MHC class I molecules are encoded by a family of genes that are found in all jawed vertebrates. A number of MHC Class I genes and other important immune and non-immune genes can be found in a cluster together in a region known as the Major Histocompatibility Complex (MHC) region. The MHC class I genes are best defined not by their function, but by their conserved structure. The conserved structure across MHC Class I molecules make them useful for examining the relationship between structure and function of a molecule.

The conserved structure of the MHC Class I molecule is defined by an α chain, comprised of three extracellular domains: $\alpha 1$, $\alpha 2$, and $\alpha 3$ domains, paired with $\beta 2$ -microglobulin (Figure 1). A transmembrane domain and a cytoplasmic tail extend from the $\alpha 3$ domain of the MHC molecule and anchor the molecule to the cell (Yeager and Hughes, 1999). While the $\beta 2$ -microglobulin is not encoded within the MHC region, it is necessary for the proper cell surface expression of most MHC class I molecules (Zijstra et al. 1990). Together, the $\alpha 1$ and $\alpha 2$ domains fold and combine to form a groove comprised of β -sheets with two α helices (Adams and Luoma, 2012). Overall, MHC class I molecules, while having similar structures, have evolved to perform a variety of functions (Table 1). Examples of known functions performed by MHC class I molecules are:

I. Presentation of Lipids to T cells

The groove formed by the $\alpha 1$ and $\alpha 2$ domains in MHC Class I molecules may serve to bind many different antigenic molecules recognized by the immune system. CD1 is an example of a molecule that binds a large array of hydrophobic lipids. CD1 has a

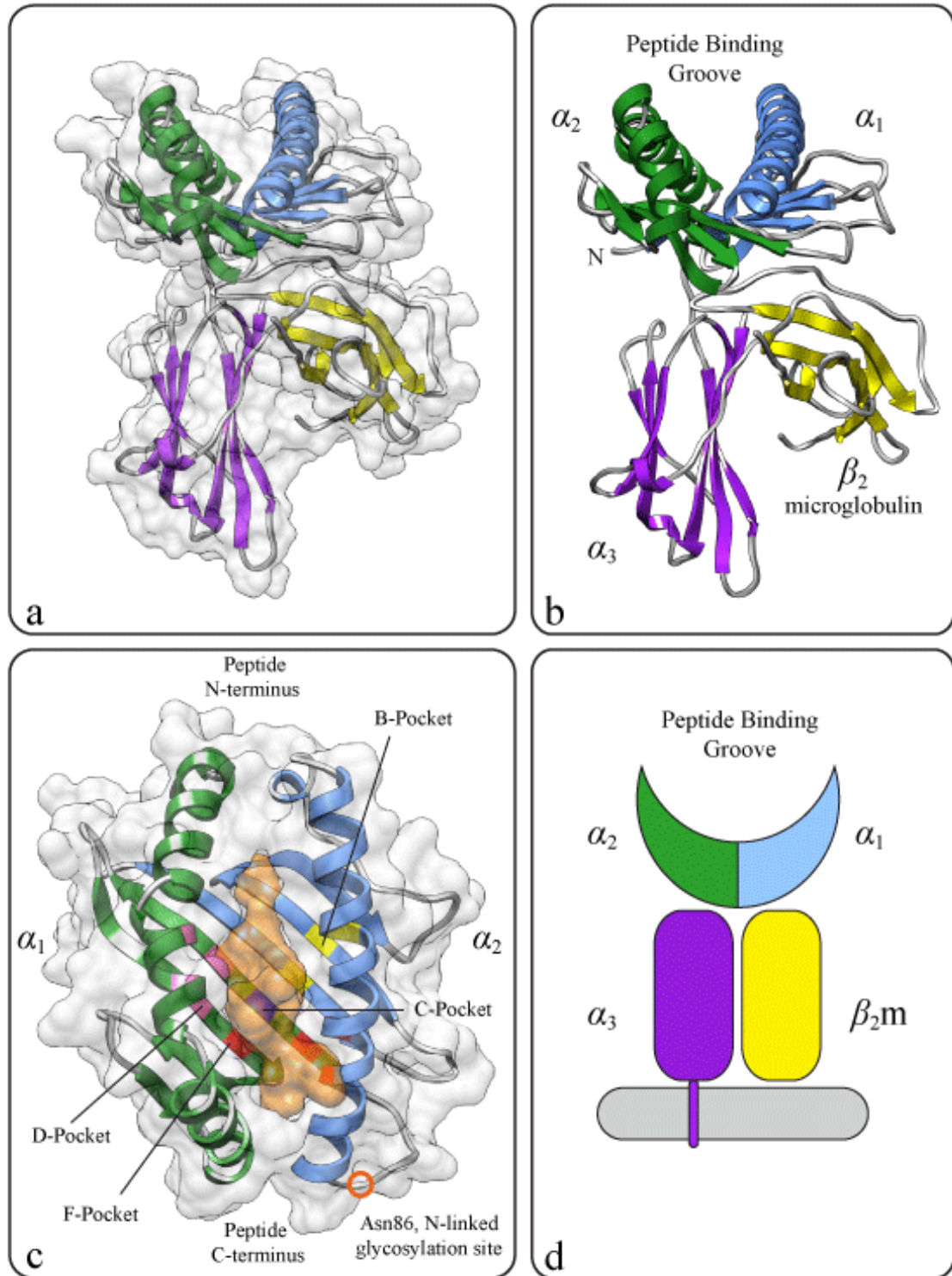


Figure 1: Conserved structure of MHC Class I molecules. Ribbon structure shows the β sheets and the α helices. The groove formed by the α_1 and α_2 domains can also be seen.

Table 1: List of molecules considered to be MHC Class I with functionality, examples and characteristics. Except for M10, these MHC Class I molecules can be found in different species within the vertebrate lineage.

Roles MHC Class I Molecules Have Evolved	Immune/Non-Immune	Examples	Polymorphic	Tissue Specificity
Present antigenic lipids to α/β T cells	Immune	CD1	No	Group 1: cortical thymocytes, Langerhans cells, dendritic cells, B cell subset Group 2: GI epithelium, Bone marrow derived cells
Receptor for transferring maternal IgG to fetus	Immune	FcRn	No	Neonatal epithelium, lung, liver, placenta
Regulation of NK Cells	Immune	All MHC Class I	N/A	Ubiquitously expressed
Binding of Transferrin Receptor	Non-immune	HFE	No	Liver, Gut
Chaperone for Receptor of Pheromones in Rodents	Non-immune	M10	No	Brain
Lipid Catabolization and Fat Storage Reduction	Non-immune	ZAP	No	Adipocytes
Present antigenic peptides to CD8+ T cells	Immune	HLA-A, H2-K	Highly Polymorphic	Ubiquitously expressed

compressed and hydrophobic binding groove, and functions as an antigen presenting molecule by binding self and bacterial lipids for presentation to T cells (Lawton and Kronenberg 2004, Rodgers and Cook 2005, Shawar et al. 1994). In mammals, CD1 is encoded outside the MHC region, while in avian species, CD1 is encoded within the MHC region.

In eutherian mammals, multiple functional CD1 genes exist. An example of this can be seen in humans, who have five CD1 gene isoforms, whereas mice only have one CD1 gene (Brigl and Brenner 2004, Shawar et al. 1994). The presence and/or absence of CD1 genes in non-eutherian and non-mammalian vertebrates, along with being non-orthologous to eutherian isoforms suggests the duplications that generated multiple CD1 groups occurred 170-180 million years ago, after divergence of the eutherian-marsupial lineage (Baker and Miller 2007). CD1 is a pseudogene in the marsupial *M. domestica*, and is encoded outside the MHC region. However, CD1 is functional in other marsupial species (Baker and Miller 2007).

No other molecules that specialize in chaperoning or in loading lipids have been found outside of the CD1 system. This potentially indicates that the antigenic lipid presentation by CD1 molecules may have diverged early from the MHC system in vertebrates, and remains a more primitive system for presentation of antigenic lipids (Hughes 1991, Lawton and Kronenberg 2004). Regardless of the evolutionary divergence, CD1 molecules play an important role in immunity via effector functions for CD1d restricted T cells (NKT cells), pathogenic *Mycobacterium* responses, and presentation of atypical and self lipids (Brigl and Brenner 2004, Rodgers and Cook 2005, Shinkai and Locksley 2000).

II. Transport of Immunoglobulins

Neonatal Fc receptors (FcRn) are a group of MHC class I molecules that do not bind any lipids or peptides, but instead serve as an Fc receptor for IgG. A key difference between FcRn and the other MHC class I molecules structure is the groove is occluded in FcRn and thus cannot bind any molecules (Ghetie and Ward 2000). The occlusion of this binding groove however, does not limit the function of the molecule. FcRn was first discovered as the receptor for transference of maternal IgG to neonate (Story et al. 1994). Subsequent research has shown FcRn is also responsible for maintaining appropriate serum levels of IgG, an indicator that the FcRn molecule is able to transport IgG within and across cells (Ghetie and Ward 2000). Despite the closed binding groove, FcRn binds the IgG on the external surface of the molecule. The FcRn gene is found outside of the MHC region across the mammalian lineage.

IgY is the avian equivalent of IgG, and in avian species, including the chicken, a yolk sac receptor called FcRY plays a similar role to that of FcRn in transferring IgY (West Jr. et al. 2004). However, the chicken FcRY is not homologous in structure to MHC Class I molecules. Instead, the FcRY is homologous to mammalian phospholipase A₂ receptors. This suggests that the use of the MHC Class I molecule FcRn for IgG transport is a recent evolutionary adaptation in mammals, as it is not seen in more distant vertebrates like birds (Roopenian and Akilesh 2007, West Jr. et al. 2004).

III. Regulation of NK Cells

An important mechanism used in innate immunity employs the use of MHC class I molecules and Natural Killer cells (NK cells). NK cells are similar to CD8⁺ cytotoxic T cells, but instead of detecting specific antigenic peptides, they respond rapidly to virally

infected cells or tumor cells that do not express MHC class I on the cell surface (Kärre, 2008). If a cell is detected as lacking MHC class I molecules, cell destruction is initiated via apoptosis by NK cells (Grommé and Neefjes 2002, Terunuma et al. 2008).

IV. Binding of Transferrin Receptor and the Transfer of Iron

While a number of MHC class I molecules are involved in various immune functions, other MHC class I molecules have evolved roles in non-immune functions. One of these non-immune roles was discovered in individuals with hereditary haemochromatosis (HH), a disease that causes an overload of iron in the blood (Feder et al. 1996). These individuals had defective copies of the HFE gene, a molecule homologous to the MHC class I molecules, that has a narrowed binding groove that cannot bind peptides or lipids (Adams and Luoma 2012, Braud et al. 1999). The role of the HFE molecule is to bind the transferrin receptor, which then reduces the receptor's affinity to load transferrin molecules bound with iron (Lebrón et al. 1998). Without the HFE molecule reducing the receptor's affinity for iron rich molecules, large deposits of iron can form in organs, leading to organ failure (Braud et al. 1999, Feder et al. 1996). The gene encoding for the HFE molecule is located within the MHC region, despite its non-immune function.

The most common mutation resulting in HH prevents the association of β 2-microglobulin with the HFE α chain, resulting in the molecule not being presented on the cell surface. Defects resulting in the loss of β 2-microglobulin association with the α chain often result in the absence of surface MHC molecules (Braud et al. 1999). The linking of an MHC Class I gene deficiency to a common hereditary disease has brought more focus to the different non-immune roles the MHC class I genes may play.

V. Chaperone for Receptor of Pheromones in Rodents

In rodents, a highly unique MHC class I molecule has been discovered in the vomeronasal organ. The vomeronasal organ serves as the peripheral sensory organ of the accessory olfactory system, and is involved in chemical and pheromone communication (Keverne 1999). The molecule, called M10, is homologous to MHC class I structure and associates with β 2-microglobulin. M10 serves to bind a specialized olfactory receptor V2R in the vomeronasal organ (Adams and Luoma 2012). The M10 molecule and V2R form a multi-molecular complex located at the ends of the vomeronasal neurons, which is the region where pheromone detection occurs (Dulac and Torello 2003). Based on the location of M10 in the vomeronasal organ and the interaction with V2R, it is suspected that it may be a chaperone or co-receptor for the detection of pheromones or receptor localization (Dulac and Torello 2003). The M10 genes are not found in species like humans that lack vomeronasal organs, which is an indicator that they play a highly specific role in pheromone detection (Kumánovics et al. 2003).

VI. Lipid Catabolization and Fat Storage Reduction

Another MHC class I molecule with non-immune functions is the zinc α 2 glycoprotein (ZAG) molecule. ZAG has a more open binding groove, and binds small hydrophobic molecules that resemble fatty acids. This is unlike many of the other MHC class I molecules not involved in peptide binding that have closed or minimal binding grooves (Adams and Luoma 2012, Sánchez 1999). Interestingly, the ZAG molecule functions in the absence of β 2-microglobulin. The expression of the ZAG molecule serves to catabolize lipids and reduce fat stores by breaking down lipids in adipocytes (Sánchez 1999). Because of the role ZAG molecules play in fat depletion, increased

expression of the molecules can be found in patients suffering from cachexia (Adams and Luoma 2012).

VII. Presentation of Peptides Derived from Self & Foreign Antigens to T cells.

A. Discovery, Function, and Evolution

The role of peptide presentation for MHC class I molecules has been left for last because it is the most well studied function. Peptide presentation is the most ubiquitous function and therefore likely the ancestral role of MHC class I molecules. Because peptide presentation was the first function described for MHC class I molecules, they are often referred to in the literature as classical MHC class I.

Peter Gorer discovered the MHC while investigating the genetic basis for tumor rejection and/or survival in mice (Klein 2001). Gorer's (1938) work suggested there were two dominant genes that influence tumor growth or rejection, one of which was linked to a blood group antigen, called antigen-2. Gorer and others began to refer to such loci as histocompatibility loci (H) and the one linked to blood group antigen-2 became known as the H-2 locus. Another researcher, George Snell, began to experiment with mice that differed only by a single H loci. He made the discovery that the H locus was likely a complex of a minimum of two loci, but one locus (H-2) was more critical in the rapid tissue rejection response or for acceptance of transplanted tumors (Snell 1951). Other H loci did not generate tissue rejection so rapidly. The more rapidly responding locus was named the Major Histocompatibility Gene. The region where these loci were uncovered was termed the Major Histocompatibility Complex or the MHC.

The MHC class I genes are best known for their role in encoding for molecules that present antigenic peptides to CD8⁺ T cells (Lawlor et al. 1990). Presentation of self-derived peptides to T cells allows for cells in the body to be frequently monitored for pathogenic infection. If the peptide presented to the CD8⁺ T cell is recognized as non-self, apoptosis of the cell presenting the foreign peptide will be initiated. Examples of antigenic peptide presenting MHC class I molecules include HLA-A in humans, H2-k in mice, and MdoUA in the opossum.

A key feature of MHC class I molecules like human HLA-A is that they can be found ubiquitously expressed on all nucleated cells. Another family of MHC molecules exists (MHC class II) that only are expressed on antigen presenting cells, including macrophages and dendritic cells (Rodgers and Cook 2005, Shawar et al. 1994). When intra-cellularly derived non-self peptides are presented to T-cells from MHC class I molecules, an adaptive immune response is generated. This response includes recruitment of macrophages to the area, B-cell activation, and killing of infected cells (Yeager and Hughes 1999). In order for molecules like HLA-A to bind a wide variety of self or non-self peptides, high levels of polymorphism are present in the residues making up the peptide-binding region. Pathogen-mediated selection can act on the genes and stimulate nucleotide changes at sites where peptide binding occurs, resulting in high levels of polymorphism in MHC class I genes (Hughes and Nei 1989, Messaoudi et al. 2002).

The MHC class I genes that encode for peptide-presenting class I molecules are found at a specific chromosomal region where MHC class II genes and antigen processing molecules are also located, known as the MHC region. Many other MHC class I genes, including CD1 and FcRn, can be found entirely outside of this region, and

even on different chromosomes (Shawar et al. 1994). Peptide-presenting MHC class I genes are typically organized with five exons, encode only one mRNA form, and do not contain any unpaired cysteines. Other MHC class I genes that encode for molecules in non-peptide presenting roles may have more atypical gene organization and exon composition that results in features like unpaired cysteines and multiple alternative mRNA forms (Baker et al. 2009, Fujii et al. 1994).

B. Organization of the MHC Region

The genomic organization of the MHC can be highly variable between vertebrates. In eutherians, such as humans and mice, the peptide presenting MHC class I and class II genes are separated by MHC class III with antigen processing and transporter genes found in the MHC class II region (Edwards and Hedrick 1998, Flajnik and Kasahara 2001, Figure 2). The non-eutherian opossum has an MHC region that rivals the length and complexity of eutherian MHC, but is organized more similarly to the MHC of non-mammals. In the opossum, MHC class I and class II genes are interspersed in the genome without separation by MHC class III genes seen in eutherians (Belov et al. 2006, Figure 2). Another feature of non-eutherian MHC organization is that the antigen processing and transporter genes can be found near MHC Class I genes, a trait also seen in frogs and chickens, suggesting the linkage of these genes is the ancestral organization (Baker et al. 2009, Siddle et al. 2009).

C. Evolution and the MHC Class I Genes

Flajnik and Kasahara (2001) postulated that the proximity of the MHC class I genes to antigen-processing or transporter genes is an indicator that these genes evolved their

functions together. However, the co-evolution of these genes could be problematic as it may restrict diversity. If both MHC class I and antigen-processing and transporter genes are in close proximity, Flajnik and Kasahara (2001) predict these genes would then have lower numbers of duplications or deletions, resulting in less diversity within them. Species with unlinked MHC class I and antigen processing/transporter genes (e.g. antigen processing/transporter genes in the MHC class II region) would then have greater levels of duplications and deletions. Because of the separation of these genes in eutherian mammals, one would expect greater level of diversification of the MHC class I and antigen processing/transport genes that have occurred as the result of duplication, deletion, and divergence.

However, the genomes of marsupials, such as the opossum and tammar wallaby, *Macropus eugenii*, reveal that this is not the case. The close proximity of antigen processing/transporter genes to MHC class I genes or the displacement of the antigen processing or transporter genes in these species to outside the MHC region does not hinder the ability of the genes to undergo deletions and duplications necessary for selection (Baker et al. 2009, Belov et al. 2006, Siddle et al. 2009, Siddle 2011). The tammar wallaby has an increased number of transporter and antigen processing genes, some of which are located outside of their core MHC region, while opossums have these genes in proximity to MHC Class I genes (Belov et al. 2006, Siddle et al. 2011). Based upon the overall patterns of MHC gene organization, it is presumed the current form of eutherian MHC organization (Figure 2) occurred after divergence with marsupials, as non-eutherian MHC organization more closely resembles that of non-mammalian species (Miller 2010).

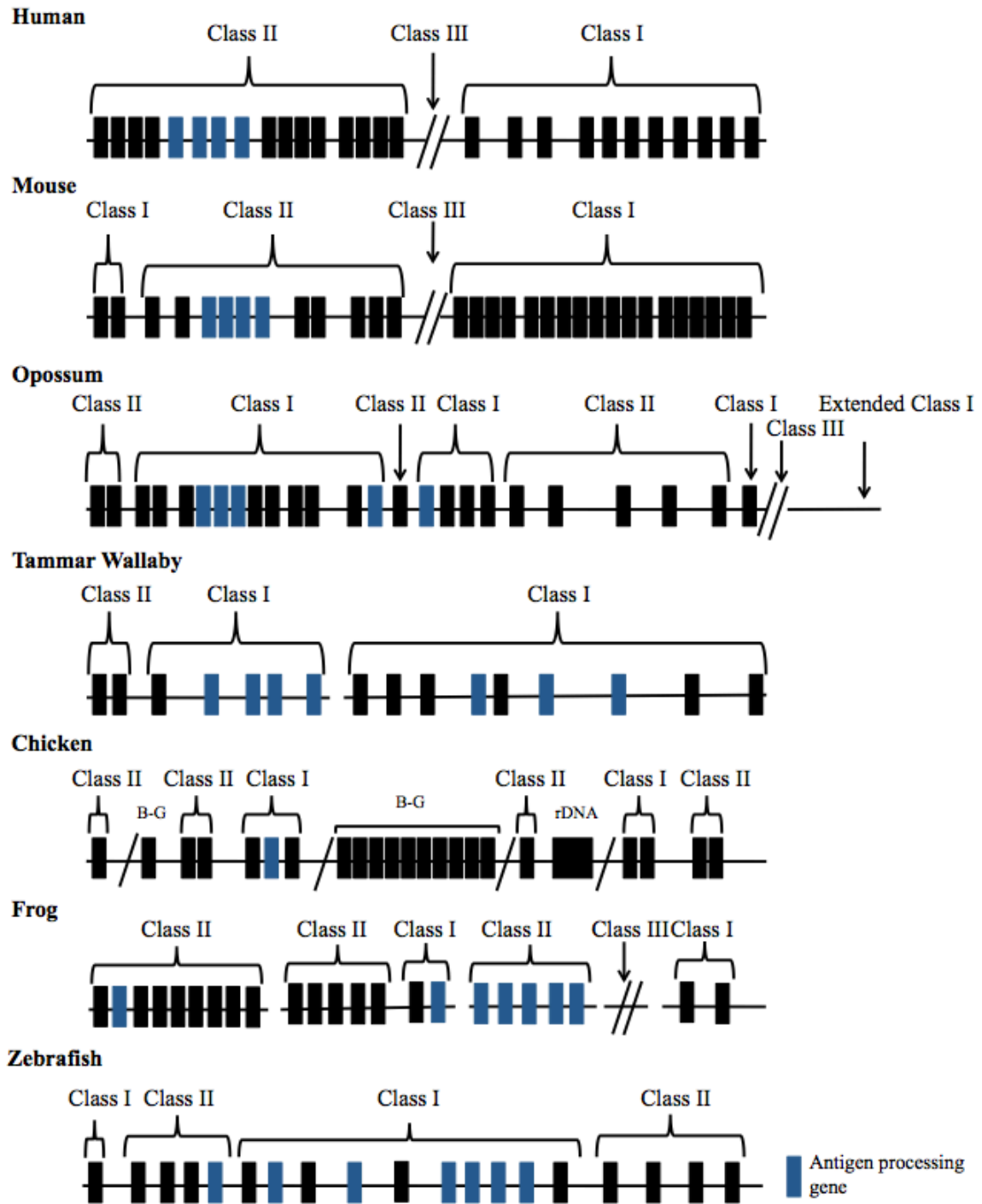


Figure 2: Organization of MHC genes across different species including eutherians, non-eutherians, avian, amphibian, and bony fish. Adapted from Belov et al. 2006, Edwards and Hendrick, 1998, Flajnik and Kasahara, 2001, Lukacs et al. 2007, Ohta et al. 2006, and Siddle et al. 2011.

THEORIES ON THE ORIGIN OF THE MHC CLASS I DOMAIN STRUCTURE.

Most of the theories regarding the origin of the unique MHC class I structure are based on its peptide-presenting role for several reasons. The first reason is that peptide presentation is seen as the classical role of MHC molecules, the second being peptide presentation is the most ubiquitous role and therefore believed to be its primordial function.

The evolutionary path and development of MHC Class I and II are debated, but two main hypotheses have been put forward, both involving exon shuffling (Figure 3). The first hypothesis, proposed by Flajnik et al. (1991), posits an ancestral immunoglobulin domain combined with a peptide-binding domain derived from heat-shock proteins to generate an ancestral MHC class I molecule. The ancestral class I molecules comprised of a single α chain containing an immunoglobulin domain and peptide binding domain, and were associated with a β 2-microglobulin domain as well. Duplication resulted in multiple ancestral class I and β 2-microglobulin genes. An ancestral class I molecule then likely donated an α 1 domain onto the β 2-microglobulin gene, forming a β chain gene. The remaining α chain from the ancestral class I would readily associate with the newly formed β chain, generating ancestral class II molecules.

The second hypothesis on the origin of MHC class I and II genes was formulated by Klein and O'hUigin (1993) and supports an ancestral immunoglobulin-like domain combined with membrane-anchoring domains and peptide-binding domains to make an ancestral class II molecule. Two instances of whole genome duplication resulted in four ancestral class II molecules. Deletion of the immunoglobulin-like domain and membrane-

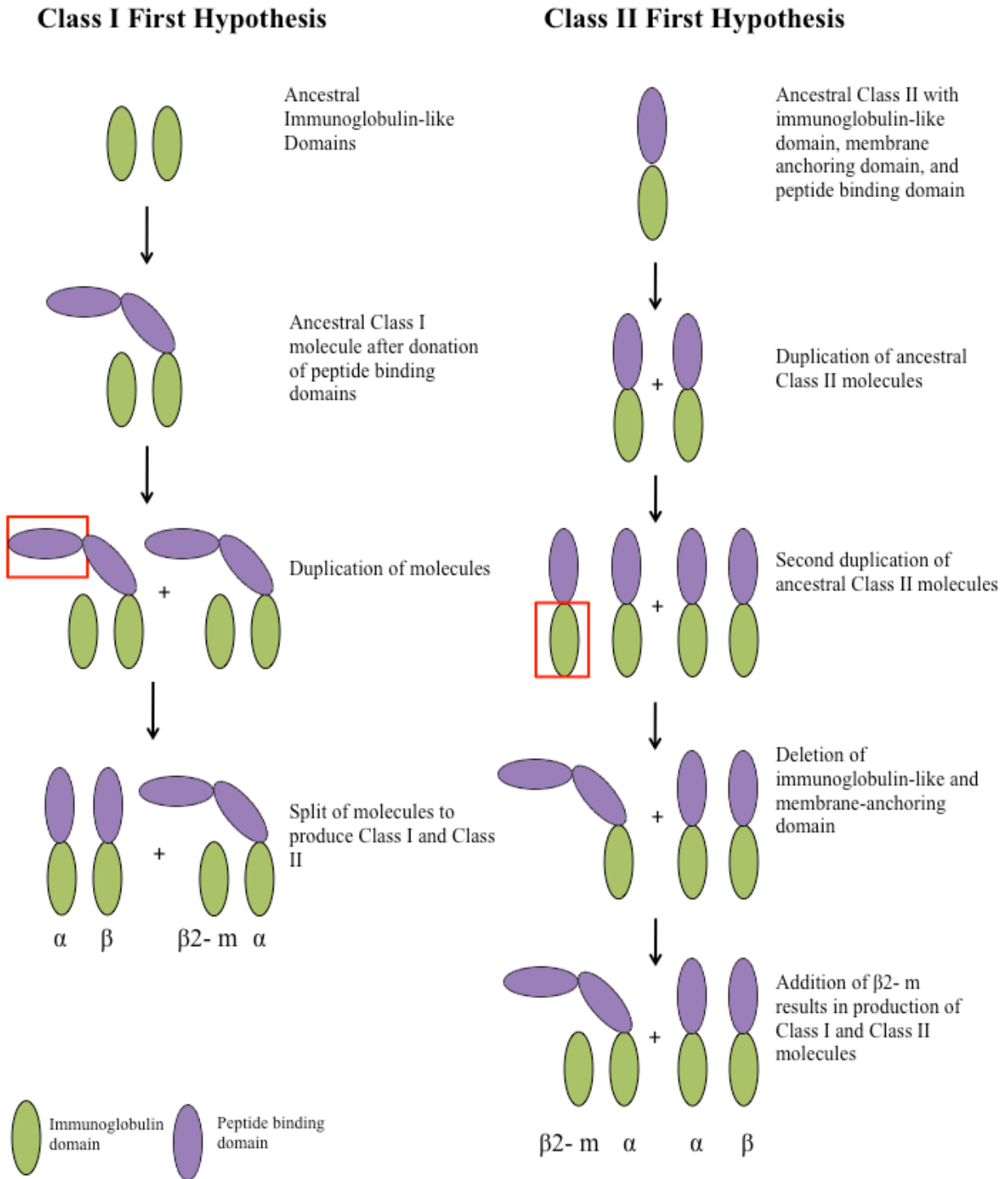


Figure 3: Diagram of the hypothetical evolutionary pattern of MHC. Adapted from Flajnik et al. 1991 (Class I First) and Klein and O’hUigin, 1993 (Class II First)

anchoring domain in one of the class II molecules, along with combining with another ancestral class II molecule resulted in the formation of class I molecules.

While the two presented theories both use exon shuffling for evolution of the genes encoding the MHC class I molecule, the development of MHC region occurred via genome wide duplications that also resulted in the emergence of genes like -classical MHC Class I genes being found in the same region of the genome (Nei and Rooney 2005). However, in 1997, Masatoshi Nei and colleagues developed an alternative hypothesis to explain the diversity generated within the MHC gene family. Phylogenetic analysis of MHC and immunoglobulin genes families resulted in the development of a new evolutionary pattern called “birth-and-death” evolution (Nei et al. 1997, Nei and Rooney, 2005).

THEORIES ON THE ORIGIN AND GENERATION OF MHC CLASS I GENES

Until the early 1990's, it was hypothesized that the diversity of MHC genes across species was the result of concerted evolution. Concerted evolution can be explained as homologous genes being more similar to their paralogs within a species than of a locus in another species. The genes generated from concerted evolution could be modified or changed if gene conversion occurs. Gene conversion occurs as the result of the transfer of DNA from homologous genes, often by unequal cross-over during meiosis. The conversion of genes could generate polymorphism and differences within these genes (Nei et al. 1997). This view was problematic, because it could not explain how gene conversion could occur between two previously differentiated loci, nor could it resolve the existence of classical and non-classical MHC Class I genes being found in the same region of the genome (Nei and Rooney 2005). However, in 1997, Masatoshi Nei and

colleagues developed an alternative hypothesis to explain the diversity generated within the MHC gene family. The expansion and contraction of these important immunological genes prompted an in-depth phylogenetic analysis of MHC and immunoglobulin gene families. This resulted in the development of a new evolutionary pattern called “birth-and-death” evolution, initially referred to as the accordion model (Klein et al. 1993, Nei et al. 1997, Nei and Rooney, 2005).

In the birth and death model, new genes are generated via gene duplications. Sometimes these duplicated genes may be a functional gene (birth), while others become pseudogenes (death) (Nei and Rooney 2005). Overall, different families or orders of mammals do not have wholly orthologous genes, and instead have different genes or loci due to their birth and death process in individual species (Nei et al. 1997, Nei and Rooney 2005). It is important to note that many MHC genes have survived between species for hundreds of millions of years (e.g. *MICA* and *HFE*), while others (*HLA-A*, *B*, and *C*) have diverged more recently between humans and great apes (Adams and Parham 2002, Nei et al. 1997). The development of high levels of diversity in the MHC region along with the visible differences between MHC across species has been the result of generation, duplication, and loss of MHC genes (Nei and Rooney 2005).

SEARCHING GENOMES FOR MORE MHC CLASS I MOLECULES

Traditionally, novel genes in genomes have been found using *in silico* search methods that rely on sequence identity and similarity to identify new genes. This methodology relies heavily on conserved nucleotide identity or similarity for identification. MHC class I genes offer an ideal model for the new search methods that search using structure rather than sequence. Because nearly all MHC class I genes can be

grouped via their conserved structure consisting of an α chain comprised of three domains and an associated β 2-microglobulin, we can utilize new search methodologies that look for the defined protein domain structures of Class I molecules. This method uses a search algorithm that looks for the conserved structure of the MHC Class I molecules rather than sequence homology. Using this novel technique, available full genomes can be searched to see if the algorithm could uncover previously identified MHC class I genes found using traditional search methods or any new Class I genes.

MARSUPIALS AS A MODEL RESEARCH ORGANISM

Because of this ability to search and possibly uncover novel families of MHC class I genes, the genomes of non-traditional organisms must be examined. This includes marsupials and monotremes, including the marsupial species the grey-short tailed opossum, *Monodelphis domestica*.

I. Origins and History of Marsupials

Marsupials are a unique part of the animal kingdom, as they represent the most evolutionarily distant living species of viviparous mammals to eutherians (Cifelli and Davis 2003). Marsupials, or Marsupialia, are a distinct infraclass of mammals that contains nearly 300 different species. While the majority of the known marsupial species are found in Australia and Oceania, many can also be found in the Americas (Wilson and Reeder eds. 2005). South and Central America contain the bulk of the Western Hemisphere's approximately 80 marsupial species. One species can be found in North America, the Virginia opossum, *Didelphis virginiana* (Graves and Westerman 2002). Currently, there are no marsupial species found in Africa, Asia, or Europe. However, a

fossil specimen of a marsupial (*Sinodelphys szalayi*) has been found in China dating to 125 million years ago, along with the earliest placental mammal fossil, indicating that Asia may have been the location where diversification between early metatherians and eutherians occurred (Luo et al. 2003).

II. Unique Birth and Development in Marsupials

Marsupials differ from eutherian mammals in several ways, with the most noticeable difference being the birth of highly altricial or relatively underdeveloped young. The altricial birth state is due to a short gestational period and comparatively less invasive placenta (Graves and Westerman 2002). As the result of these early developmental differences, the overall development of the immune system of marsupials differs from that of eutherians. Marsupial young are born both physically and immunologically underdeveloped and develop some of their lymphoid cells post-natally (Baker et al. 1999, Old and Deane 2000, Wang et al. 2012). However, the pouch environment in which marsupial young enter frequently encounter pathogenic bacteria, and therefore defense mechanisms must be in place for the young to fight microorganisms. As a result, some of the protective mechanisms marsupials have developed to protect altricial young include transfer of immunoglobulins and immune compounds in milk and antimicrobial peptides secreted by the mother in the pouch (Edwards et al. 2012). These characteristics have made marsupials a valuable tool for the study of early immunological development in mammals.

III. Use of the Gray Short-tailed Opossum, *Monodelphis domestica*, in Research

Beginning in 1978, *M. domestica* were imported into the United States from four

sites in Brazil and one Bolivian site for use as a captive laboratory bred marsupial. Since then, *M. domestica* has been used in biomedical research as an animal model for melanoma, corneal cancer, and hypercholesterolaemia (Samollow 2006, VandeBerg and Robinson 1997). Other research, lead by Fry and Saunders (2000), has used *M. domestica* as a model to examine neural fiber regeneration following spinal cord injury. The uniquely altricial development of *M. domestica* allows the species to be used as a model for early neurological and immunological development, while also providing insight about the evolutionary lineage by sharing the last common viviparous ancestor to eutherian mammals. Additionally, *M. domestica* was the first marsupial to have its entire genome sequenced (Mikkelsen et al. 2007), a critical resource for searching genomes.

EARLY IMMUNOLOGICAL RESEARCH ON MARSUPIALS

Early research on the immune systems of marsupials focused on the thymus. It was noted by Johnstone (1898) that some species of marsupials had only a thoracic thymus, as is the case in eutherians, while others had both a cervical and a thoracic thymus. Thymectomy studies done with the quokka, *Setonix brachyuras*, where either the cervical or thoracic thymus was removed suggested that the roles of both of the thymuses were the same, despite their differing locations and development patterns (Stanley et al., 1972). Thymectomy studies on opossum pups a week after birth found a reduction in lymphocytes in lymphatic tissue with an increase of myeloid tissue, confirming that one or both thymuses in marsupials have a critical role in the maturation of T cells and other lymphoid tissues (Miller et al. 1965). Mixed lymphocyte reactions (MLR) of peripheral blood leukocytes (PBL) from two different populations of Virginia opossum failed to

respond and proliferate, but did respond to mitogens (Fox et al. 1976). It was inferred from this that the opossums might lack antigen reactive lymphocytes or may not have MHC, as they could not reject non-self leukocytes from the same species. The application of molecular genetic technologies, however, has confirmed the presence of MHC genes in marsupials.

THE UNIQUE MARSUPIAL IMMUNE SYSTEM

The work done in marsupial immunology has shown that components of immunity have been lost in non-eutherians and found in eutherians and vice versa. In comparisons of the immune systems of eutherians and non-eutherians, the research allows us to see the differences that exist between close relatives. Components of the adaptive immune system, such as immunoglobulins, have been shown to differ between eutherian and non-eutherian mammals. While IgM, IgG, IgE, and IgA have been found in all mammalian species, the marsupial *M. domestica* lacks IgD, which is postulated to be an ancient immunoglobulin in jawed vertebrates (Ohta and Flajnik 2006, Sun et al. 2011, Wang et al. 2009). It is likely the lack of IgD represents a significant gene loss in *M. domestica*, given that IgD can also be found in eutherians, non-eutherians, and other jawed vertebrates (Flajnik 2002, Miller and Belov 2000, Ohta and Flajnik 2006).

All jawed vertebrates have the four T cell receptors (TCR), TCR- α , TCR- β , TCR- δ and TCR- γ . While marsupials and monotremes have all four traditional TCRs, they also have a fifth atypical T cell receptor, TCR- μ (Parra et al. 2007, 2008). Even between marsupials and monotremes TCR- μ differs. In marsupials, pre-joined V genes encode the TCR isoform, whereas the platypus has doubly-rearranging V domains (Miller 2010,

Parra et al. 2007). While the function of TCR- μ and the presumed partner chain remain unresolved, TCR- μ is related to TCR- δ , but remains divergent in both sequence and structure from traditional mammalian T- cell receptors (Parra et al. 2008, 2012). The loss of the immunoglobulin IgD in *M. domestica* and the gain of TCR- μ in marsupials and monotremes have shed new light onto the evolutionary trajectory and development of mammalian immunity.

MHC RESEARCH IN MARSUPIALS

MHC class I genes have been characterized the most in mammalian species such mice and humans, but also have been examined in a more distant group of non-eutherian mammals, the marsupials. Early work on MHC in koalas generated numerous MHC class I gene sequences, and defined MHC class I in the species as a large, multi-gene family (Houlden et al. 1996). In contrast, low levels of polymorphism were found in the MHC class II β chain of tammar wallabies in comparison to eutherians (McKenzie and Cooper 1994). Currently, research on the MHC in marsupials focuses mostly on MHC class I genes.

The model organism for the investigation of marsupial MHC Class I genes is the gray short-tailed opossum, *Monodelphis domestica* (Baker et al. 2006, Baker and Miller 2007, Baker et al. 2009, Belov et al. 2006, Gouin et al. 2006, Miska et al. 2004, Miska and Miller 1999). Using traditional *in silico* search methods, eleven genes encoding the MHC Class I α chain have been uncovered in the *M. domestica* genome, along with three MHC Class I pseudogenes (Baker et al. 2009, Belov et al. 2006, Gouin et al. 2006).

The function and features of MHC class I genes have been fairly well defined in the opossum. Work by Gouin et al. (2006) revealed that a non-classical MHC class I gene, *ModoUG*, can be expressed in three different alternatively spliced mRNA forms. In these forms, a short cytoplasmic tail has been found that does not have traditional phosphorylation sites. In Baker et al. (2009), *M. domestica* MHC class I genes including *ModoUE*, *UI*, *UJ*, *UK*, and *UM* were characterized. These *M. domestica* class I genes had atypical genomic characteristics including alternative mRNA splicing, low polymorphism, and unpaired cysteine residues that could form homodimers. This is indicative that these MHC genes likely do not produce molecules involved in traditional antigenic peptide presentation.

References

- Adams, E.J., and Luoma, A.M. (2012) The Adaptable Major Histocompatibility Complex (MHC) Fold: Structure and Function of Nonclassical and MHC Class I-Like Molecules. *Annu. Rev. Immunol.* **31**(1):529-561.
- Adams, E., and Parham, P. (2002) Specific-specific evolution of MHC class I genes in the high primates. *Immunological Reviews.* **183**(1):41-64.
- Baker, M.L., Gemmel, E. and Gemmel, R.T. (1999) Ontogeny of the immune system of the brushtail possum, *Trichosurus vulpecula*. *The Anatomical Record.* **256**(4):354-365.
- Baker, M.L. and Miller, R.D. (2007) Evolution of mammalian CD1: marsupial CD1 is not orthologous to eutherian isoforms and is a pseudogene in the opossum *Monodelphis domestica*. *Immunology.* **121**(1):113-121.
- Baker, M.L., Melman, S.D., Huntley, J. and Miller, R.D. (2009) Evolution of the opossum MHC: Evidence for diverse alternative splice patterns and low polymorphism among class I genes. *Immunology.* **128**:e418-e431.
- Belov, K., Deakin, J.E., Papenfuss, A.T., Baker, M.L., Melman, S.D., Siddle, H.V., Gouin, N., Goode, D.L., Sargeant, T.J., Robinson, M.D., Wakefield, M.J., Mahony, S., Cross, J.G.R., Benos, P.V., Samollow, P.B., Speed, T.P., Graves, J.A.M. and Miller, R.D. (2006) Reconstructing an ancestral mammalian immune supercomplex from a marsupial MHC. *PLoS Biology.* **4**(3):e46.
- Braud, V., Allan, D., and McMichael, A. (1999) Functions of nonclassical MHC and non-MHC-encoded class I molecules. *Current Opinion in Immunology.* **11**(1):100-108.
- Brigl, M. and Brenner, M.B. (2004) CD1: Antigen Presentation and T Cell Function. *Annu. Rev. Immunol.* **22**:817-890.
- Cifelli, R.L and Davis, B.M. (2003) Marsupial Origins. *Science.* **302**(5652):1899-1900.
- Deakin, J.E., Waters, P.D., Marshall Graves, J.A. eds. (2010) Marsupial Genetics and Genomics. Springer Life Sciences Press.
- Dulac, C. and Torello, A. (2003) Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nat. Rev. Neurosci.* **4**(7):551-562.
- Edwards, M.J., Hinds, L.A., Deane, E.M. and Deakin, J.E. (2012) A review of the complementary mechanisms which protect the developing marsupial pouch young. *Development and Comparative Immunology.* **37**:213-220.
- Edwards, S.V. and Hedrick, P.W. (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology and Evolution.* **13**(8):305-311.

- Feder J.N. et al. (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature Genetics*. **13**(4):399-408.
- Flajnik, M.F. (2002) Comparative analyses of immunoglobulin genes: surprises and portents. *Nat. Rev. Immunology*. **2**(9):688-98.
- Flajnik, M.F., Canel, C., Kramer, J., and Kasahara, M. (1991) Which came first, MHC class I or class II? *Immunogenetics*. **33**(5-6):295-300.
- Flajnik, M.F. and Kasahara, M. (2001) Comparative genomics of the MHC: Glimpses into the evolution of the adaptive immune system. *Immunity*. **15**:351-362.
- Flower, D.R. (1996) The Lipocalin protein family: structure and function. *Biochem J*. **318**(1):1-14.
- Fox, D.H., Rowlands, D.T.J., and Wilson, D.B. (1976) Proliferative Reactivity of Opossum Peripheral Blood Leukocytes to Allogeneic Cells, Mitogens, and Specific Antigens. *Transplantation* **21**(2):164-168.
- Fry, E.J. and Saunders, N.R. (2000) Spinal Repair in the Immature Animals: A Novel Approach using the South American Opossum *Mondelphis domestica*. *Clinical and Experimental Pharmacology and Physiology*. **27**:542-547.
- Ghetie, V., and Ward, E.S. (2000) Multiple roles for the major histocompatibility complex class I-related receptor FcRn. *Annual Review of Immunology*. **18**(1):739-766.
- Gorer, P.A. (1938) The antigenic basis of tumour transplantation. *J. Pathol. Bacteriol.* **47**(2):231-252.
- Gouin, N., Wright, A.M., Miska, K.B., Parra, Z.E., Samollow, P.B., Baker, M.L. and Miller, R.D. (2006) Modo-UG, a marsupial nonclassical MHC class I locus. *Immunogenetics*. **58**:396-406.
- Graves, J.A.M. and Westerman, M. (2002) Marsupial genetics and genomics. *Trends in Genetics* **18**(10):517-521.
- Grommé, M. and Neefjes, J. (2002) Antigen degradation or presentation by MHC Class I molecules via classical and non-classical pathways. *Molecular Immunology*. **39**(3-4): 181-202.
- Houlden, B.A., Greville, W.D., and Sherwin, W.B. (1996) Evolution of MHC Class I Loci in Marsupials: Characterization of Sequences from Koala (*Phascolarctos cinereus*). *Mol. Biol. Evol.* **13**(8): 1119-1127.

- Hughes, A.L. (1991) Evolutionary Origins and Diversification of the Mammalian CD1 Antigen Genes. *Mol. Biol. Evol.* **8**(2):185-201.
- Hughes, A.L. and Nei, M. (1988) Pattern of nucleotide substitutions at major Histocompatibility complex class I loci reveals overdominant selection. *Nature*. **335**:167-170.
- Hughes, A.L. and Nei, M. (1989) Evolution of the major Histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Molecular Biology and Evolution*. **6**(6):559-579.
- Johnstone, J. (1898) The Thymus in the Marsupials. *Journal of the Linnean Society of London, Zoology*. **26**(171): 537-557.
- Kärre, K. (2008) Natural kill cell recognition of missing self. *Nature Immunology*. **9**:477-480.
- Kasahara M. (1997) New insights into the genomic organization and origin of the major histocompatibility complex: Role of chromosomal (genome) duplication in the emergence of the adaptive immune system. *Hereditas*. **127**:59-65.
- Kelley, J., Walter, L., and Trowsdale, J. (2005) Comparative genomics of the major Histocompatibility complexes. *Immunogenetics*. **56**:683-695.
- Keverne, E.B. (1999) The Vomeronasal Organ. *Science*. **286**(5440):716-720.
- Klein, J. (2001). George Snell's First Foray Into the Unexplored Territory of the Major Histocompatibility Complex. *Genetics*. **159**(2):435-439.
- Klein, J. and O'hUigin C. (1993). Composite origin of major histocompatibility complex genes. *Curr Opin Genet Dev*. **3**(6):923-930.
- Klein, J., Ono, H., Klein, D., and O'hUigin C. (1993) The Accordion Model of *Mhc* Evolution. *Progress in Immunology Vol. VIII*. 137-143.
- Klein, J., Sato, A., and Nikolaidis, N. (2007) MHC, TSP, and the Origin of Species: From Immunogenetics to Evolutionary Genetics. *Annu. Rev. Genet.* **41**:281-304.
- Klein, J., Satta, Y., O'hUigin C., and Takahata, N. (1993) The molecular descent of the major histocompatibility complex. *Annu. Rev. Immunol.* **11**(1):269-295.
- Kumánovics, A., Takada, T., and Lindahl, K.F. (2003) Genomic organization of the mammalian MHC. *Annu. Rev. Immunol.* **21**(1):629-657.

- Lawlor, D.A., Zemmour, J., Ennis, P.D., and Parham, P. (1990) Evolution of class-I MHC genes and proteins: from natural selection to thymic selection. *Annu. Rev. Immunol.* **8**(1):23-63
- Lawton, A.P. and Kronenberg, M. (2004) The Third Way: Progress on pathways of antigen processing and presentation by CD1. *Immunology and Cell Biology.* **82**(3):295-306.
- Lebrón, J.A., Bennett, M.J., Vaughn, D.E., Chirino, A.J., Snow, P.M. Mintier, G.A., Feder, J.N., and Bjorkman, P.J. (1998) Crystal Structure of the Haemochromatosis Protein HFE and Characterization of Its Interaction with Transferrin Receptor. *Cell.* **93**(1):111-123.
- Lukacs, M.F., Harstad, H., Grimholt, U., Beetz-Sargent, M., Cooper, G.A., Reid, L., Bakke, H.G., Phillips, R.B., Miller, K.M., Davidson, W.S., and Koop, B.F. (2007) Genomic organization of duplicated major histocompatibility complex class I regions in the Atlantic salmon (*Salmo salar*). *BMC Genomics.* **8**(1):251-267.
- Luo, Z.X., Qiang, J., Wible, J.R. and Yuan, C.X. (2003) An Early Cretaceous Tribosphenic Mammal and Methaterian Evolution. *Science.* **302**(5652):1934-1940.
- McKenzie, L.M. and Cooper, D.W. (1994) Low MHC Class II variability in a Marsupial. *Reprod. Fertil. Dev.* **6**:721-726.
- Messaoudi, I., Guevara Patiño, J.A., Dyall, R., LeMaout J., and Nikolich-Zuglich, J. (2002). Direct Link Between *mhc* Polymorphism, T Cell Avidity, and Diversity in Immune Defense. *Science.* **298**(5599):1797-1800
- Miller, J.F.A.P., Block, M., Rowlands, D.T., and Kind, P. (1965) Effect of thymectomy on haematopoietic organs of the opossum "embryo". *Proceedings of the Society for Experimental Biology and Medical Science* **118**:916-921.
- Miller, R.D. (2010) Those other mammals: The immunoglobulin and T cell receptors of marsupials and monotremes. *Seminars in Immunology.* **22**:3-9.
- Mikkelsen et al. (2007). Genome of the marsupial *Monodelphis domestica* reveals innovations in non-coding sequences. *Nature.* **447**(7141):167-177.
- Miska, K.B. and Miller, R.D. (1999). Marsupial MHC class I: classical sequences from the opossum, *Monodelphis domestica*. *Immunogenetics.* **50**:89-93.
- Miska, K.B., Wright, A.M., Lundgren, R., Sasaka-McClees, R., Osterman, A., Gale, J.M. and Miller R.D. (2004). Analysis of a marsupial MHC region containing two recently duplicated class I loci. *Mamm. Genome.* **15**:851-864.

- Nei, M., Gu, X., and Sitnikova, T. (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA*. **94**(15):7799-7806.
- Nei, M. and Rooney, A.P. (2005) Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **39**:121-152.
- O'Callaghan, C.A. and Bell, J.I. (1998) Structure and function of the human MHC class Ib molecules HLA-E, HLA-F and HLA-G. *Immunological Reviews*. **163**:129-138.
- Ohta, Y. and Flajnik, M.F. (2006) IgD, like IgM, is a primordial immunoglobulin perpetuated in most jawed vertebrates. *Proc Natl Acad Sci USA*. **103**(28):10723-10728.
- Ohta, Y., Goetz, W., Hossain, M., Nonaka, M., and Flajnik, M.F. (2006) Ancestral organization of the MHC revealed in the amphibian *Xenopus*. *Journal of Immunology*. **176**(6):3675-3685.
- Old, J. and Deane, E. (2000) Development of the immune system and immunological protection in marsupial pouch young. *Dev Comp Immunol*. **24**:445-454.
- Parra, Z.E., Baker, M.L., Schwartz, R.S., Deakin, J.E., Lindblad-Toh, K., and Miller, R.D. (2007) A unique T cell receptor discovered in marsupials. *Proc. Natl. Acad. Sci. USA*. **104**:9776-9781.
- Parra, Z.E., Baker, M.L., Hathaway, J., Lopez, A.M., Trujillo, J., Sharp, A., and Miller, R.D. (2008) Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica*. *BMC Genomics*. **9**:111.
- Parra, Z.E., Mitchell, K., Dalloul, R.A. and Miller, R.D. (2012). A Second TCR δ Locus in Galliformes Uses Antibody-like V Domains: Insight in the evolution of TCR δ and TCR μ Genes in Tetrapods. *Journal of Immunology*. **188**:3912-3919.
- Rodgers, J.R. and Cook, R.G. (2005) MHC Class Ib Molecules Bridge Innate and Acquired Immunity. *Nat. Rev. Immunol.* **5**(6):459-471.
- Roopenian, D.C., and Akilesh, S. (2007) FcRn: the neonatal Fc receptor comes of age. *Nature Reviews Immunology*. **7**(9):715-725.
- Sánchez, L.M, Chirino, A.J., and Bjorkman, P.J. (1999) Crystal Structure of Human ZAG, A Fat-Depleting Factor Related to MHC Molecules. *Science*. **283**(5409):1914-1919.
- Shawar, S.M., Vyas, J.M. Rodgers, J.R. and Rich, R.R. (1994) Antigen Presentation by Major Histocompatibility Complex Class I-B Molecules. *Annual Review of Immunology*. **12**(1):839-880.

- Shinkai, K. and Locksley, R.M. (2000) CD1, Tuberculosis, and the Evolution of Major Histocompatibility Complex Molecules. *Journal of Experimental Medicine*. **191**(6):907-914.
- Siddle, H.V., Deakin, J.E., Coggill P., Hart, E., Cheng Y., Wong, E. S.W., Harrow, J., Beck, S. and Belov, K. (2009). MHC-lined and un-linked class I genes in the wallaby. *BMC Genomics*. **10**:310.
- Siddle, H.V. Deakin, J.E., Coggill, P., Whilming, L. Harrow, J., Kaufman, J, Beck, S., and Belov, K. (2011) The tammar wallaby major Histocompatibility complex shows evidence of past genomic instability. *BMC Genomics*. **12**(1):421-436.
- Silverman, G.A. et al. (2001) The Serpins Are an Expanding Superfamily of Structurally Similar but Functionally Diverse Proteins: Evolution, Mechanisms of Inhibition, Novel Functions, and Revised Nomenclature. *J. Biol. Chem*. **276**:33293-33296.
- Snell, G. D. (1951) A fifth allele at the histocompatibility-2 locus of the mouse as determined by tumor transplantation. *J. Natl. Cancer Inst*. **11**: 1299–1305.
- Spurgin, L.G. and Richardson, D.S. (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceeding of the Royal Society B: Biological Sciences*. **277**(1684):978-988.
- Stanley, N.F., Yadav, M., Waring, H., and Eadie, M. (1972) The effect of thymectomy on response to various antigens of a marsupial *Setonix brachyurus* (Quokka). *Australian Journal of Experimental Biology & Medical Science* **50**(6): 689-702.
- Story, C.M., Mikulska, J.E., and Simister, N.E. (1994) A Major Histocompatibility Complex Class I-like Fc Receptor Cloned from Human Placenta: Possible Role in Transfer of Immunoglobulin G from Mother to Fetus. *J. Exp. Med*. **180**: 2377-2381.
- Sun Y., Wei, Z., Hammarstrom, L. and Zhao, Y. (2011) The immunoglobulin delta gene in jawed vertebrates: A comparative overview. *Developmental and Comparative Immunology*. **35**(9):975-981.
- Terunuma, H., Deng, X., Dewan, Z., Fujimoto, S., and Yamamoto, N. (2008) Potential Role of NK Cells in the Induction of Immune Responses: Implications for NK Cell-Based Immunotherapy for Cancer and Viral Infections. *International Reviews of Immunology*. **27**:93-110.
- Tyndale-Biscoe, H. and Renfree, M. eds. (1987) Monographs on Marsupial Biology: Reproductive physiology of marsupials. Cambridge: Cambridge University Press.
- VandeBerg, J.L. and Robinson, E.S. (1997) The Laboratory Opossum (*Monodelphis domestica*) in Biomedical Research In Saunders, N.R. and Hinds, L.A. (Eds.)

Marsupial Biology: Recent Research, New Perspectives (Ch 14, pp 238-253)
University of New South Wales Press.

Wang, X., Olp, J. J., and Miller R. D. (2009) On the genomics of immunoglobulins in the gray, short-tailed opossum *Monodelphis domestica*. *Immunogenetics* **61**:581–596.

Wang, X., Sharp, A.R., and Miller, R.D. (2012) Early Postnatal B Cell Ontogeny and Antibody Repertoire Maturation in the Opossum, *Monodelphis domestica*. *PLoS ONE* **7**(9):e45931

West Jr., A.P, Herr, A.B., and Bjorkman, P.J. (2004) The Chicken Yolk Sac IgY Receptor, a Functional Equivalent of Mammalian MHC-Related Fc Receptor, Is a Phospholipase A₂ Receptor Homolog. *Immunity* **20**(5):601-610.

Wilson, D. E.; Reeder, D. M. eds. (2005). *Mammal Species of the World* (3rd ed.). Baltimore: Johns Hopkins University Press.

Yang, Z. (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**(8):1586-1591.

Yeager, M. and Hughes, A.L. (1999) Evolution of mammalian MHC: natural selection, recombination, and convergent evolution. *Immunol. Rev.* **167**:45-58

Zijlstra, M., Bix, M., Simister, N.E., Loring, J.M., Raulet, D.H., and Jaenisch, R. (1990) β 2-Microglobulin deficient mice lack CD4-8⁺ cytolytic T cell. *Nature*. **344**:742-746.

CHAPTER 2

MARSUPIALS AND MONOTREMES POSSESS A NOVEL FAMILY OF MHC CLASS I GENES THAT IS LOST FROM THE EUTHERIAN LINEAGE

By Anthony T Papenfuss^{1,2,3,4,*}, Zhi-Ping Feng^{1,2}, Katina Krasnec⁵, Janine E Deakin^{6,7},
Michelle L Baker^{5,8}, Robert D Miller^{5,*}

¹Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, 1G
Royal Pde, Parkville 3052, Australia

²Department of Medical Biology, University of Melbourne, Australia

³Peter MacCallum Cancer Centre, East Melbourne 3002, Australia

⁴Sir Peter MacCallum Department of Oncology, University of Melbourne, Australia

⁵Center for Evolutionary and Theoretical Immunology, Department of Biology,
University of New Mexico, Albuquerque, New Mexico, United States of America

⁶Research School of Biology, Australian National University, Canberra, Australia

⁷Institute for Applied Ecology, University of Canberra, Canberra, Australia

⁸Australian Animal Health Laboratory, CSIRO, Geelong, Australia

ATP and RDM drafted manuscript, ATP, ZF, KK, JED, MLB collected data and
performed data analysis and generated figures.

Manuscript submitted to: *Genome Biology*

ABSTRACT

Major histocompatibility complex (MHC) class I genes are found in the genomes of all jawed vertebrates. This gene family has evolved to perform immune-related and non-immune functions, including presenting endogenously-derived peptide antigens to cytotoxic T-cells.

Sensitive sequence searches of representative vertebrate genomes using hidden Markov models revealed an extensive, novel sub-family of divergent MHC class I genes, denoted as *UT*, which have not previously been characterized. These genes are found in both American and Australian marsupials and in monotremes, but are absent from non-mammalian genomes and lost from the eutherian lineage. We show that *UT* genes are expressed in immune tissues of the gray short-tailed opossum and several Australian marsupials. Structural homology modeling predicts that *UT* family members have an open, though short, antigen-binding groove. The function of this family is as yet unknown, however their predicted structure may be consistent with being able to present antigens to T-cells.

INTRODUCTION

The major histocompatibility complex (MHC) is a region unique to the genomes of jawed vertebrates and contains genes that are critical to the generation of immune responses. It is the most gene dense and polymorphic region in the genome (reviewed in ¹). The MHC is named for its role in recognition of ‘self’ and ‘non-self’ and was first identified in connection with tumour transplant rejection². Genes in the MHC are also associated with resistance to infectious diseases, autoimmunity, reproductive success, inflammatory response and innate immunity (reviewed in ^{3,4}).

The genes of the MHC are sub-divided into class I, II and III. The MHC class I genes are particularly noteworthy for having undergone gene duplication and divergence, resulting in an extended gene family whose members perform a broad range of functions. The classical role of class I molecules is to present endogenously-derived peptides to CD8⁺ T cells to stimulate cytotoxic responses against virus infected or tumour cells. The class I molecules performing this role are sometimes referred to as classical MHC class I. Examples of classical class I genes include *HLA-A*, *-B* and *-C* in humans and *H2-K*, *H2-D* and *H2-L* in mouse. Classical MHC class I genes are generally broadly expressed in nucleated cells and highly polymorphic. Class I molecules performing other functions, collectively known as non-classical MHC class I, generally have low polymorphism, may have tissue-specific expression and in some cases have evolved functions other than antigen-presentation, including immuno-regulatory and non-immune roles. Examples of non-classical class I genes include *HLA-E*, *-F* and *-G* in human, *B1* and *Qa1* in mouse, as

well as *MIC*. The function of non-classical molecules is not limited to the immune system. The *HFE* gene, for example, serves as part of the transferrin complex involved in iron storage (reviewed in ⁵). Others, such as the neonatal Fc receptor, *FcRN*, that transports maternal IgG to fetal or neonatal mammals, has a role in the immune system that is distinctly different from conventional class I (reviewed in ⁶). Typically, classical and some non-classical genes are located in the MHC, although many of the non-classical are located elsewhere in the genome⁷.

In humans, the MHC is located on chromosome 6p¹. Additionally, there are three regions of the genome that are paralogues of the MHC, indicative of the two rounds of whole genome duplication thought to have occurred in early vertebrate evolution⁸. These paralogous regions are located on chromosomes 1q, 9q, and 19p. They contain additional non-classical class I genes, including the *CDI* gene family, *MRI* and *FCGRT*. Other non-classical class I genes are found on chromosome 20 (*PROCR*), chromosome 7 (*AZGP1*) and chromosome 6q (*ULBP* and *RAET* families), suggesting that duplication and translocation have acted to further distribute MHC class I genes throughout the genome.

In some species, similar processes have acted to spread class I genes from the MHC. Two tightly linked, classical class I-like genes (*UB* and *UC*) in the opossum, *Monodelphis domestica*, for example, were translocated outside the MHC although they remain syntenic to the MHC on chromosome 2^{9,10}. In a more extreme example, in the tammar wallaby, *Macropus eugenii*, the classical class I-like genes have been completely translocated out of the MHC and are distributed across multiple chromosomes¹¹.

Both classical and non-classical class I molecules have a conserved and distinctive protein domain structure. MHC class I genes typically have 5-9 exons encoding proteins with well-defined domain organization (Figure 1A and 1B). The first exon encodes a signal peptide. Exons 2 and 3 encode the α_1 and α_2 domains, which together make up the antigen-presenting domain (APD). An immunoglobulin domain (Ig or α_3) is encoded by exon 4. Additional exons may encode one or more transmembrane domains and the final exon contains a conserved cytoplasmic domain at the C-terminus of some MHC class I genes. The α_1 , α_2 and Ig domains are the hallmark of MHC class I genes. However, different isoforms of some MHC class I genes exist. These may splice out some of these domains to produce other membrane bound versions of the protein or secreted forms. Additionally, the UL16-binding protein (*ULBP*) and retinoic acid early transcript (*RAET*) families, known in eutherians, are MHC class I-related genes that lack immunoglobulin domains and may utilize a GPI-anchor rather than a transmembrane domain¹²⁻¹⁵.

To better understand the evolution of MHC class I genes, particularly in mammals, we undertook to catalogue the class I genes. Here, we describe a sensitive comparative genomic analysis of MHC class I genes spanning the vertebrates. This was achieved using a novel approach based upon combining profile hidden Markov models (HMMs), which represent the separate domains characteristic of MHC class I genes. Our results reveal a new sub-family of MHC class I genes in marsupials and monotremes, which are not found in non-mammals and have been lost from the eutherian lineage. We show that these genes are transcribed in immune tissues in the gray short-tailed opossum, tammar

wallaby, brushtail possum and Tasmanian devil. Structural homology mapping is used to begin to investigate the function of these genes.

Materials and Methods

Collection of annotated protein sequences

Predicted MHC class I proteins were identified and extracted from the Ensembl genebuilds (Release 75) of a selection of species spanning the gnathostomes, a jawless vertebrate, 2 invertebrate species, and a fungus. Protein sequences from human³³, mouse³⁴, dog³⁵, cow³⁶, opossum³⁷, wallaby³⁸, Tasmanian devil¹⁹, platypus³⁹, chicken⁴⁰, zebra finch⁴¹, turkey (The Turkey Genome Consortium), green anole lizard⁴², *Xenopus tropicalis*⁴³, zebrafish⁴⁴, pufferfish⁴⁵, lamprey⁴⁶, sea squirt⁴³, fruitfly⁴⁷ and yeast⁴⁸ were searched using profile hidden Markov models (HMMs) representing the MHC class I APD (PFAM:PF00129 and SUPFAM:0045513), C1-type Ig domain (PFAM:PF07654) and MHC class II β domain (PFAM:PF00969) using HMMer version 2 (fs and ls) and HMMer version 3 (PFAM models only). The separate domain searches were integrated and MHC class I proteins predicted using a simple heuristic: proteins were annotated as predicted MHC class I proteins if they had a significant match to the MHC class I APD (E-value $<10^{-5}$) or a weak match to the APD (score >0) and a significant match to the Ig domain model (E-value $<10^{-5}$) in the correct order, with the additional requirement that the MHC class I APD model score is higher than the MHC class II β domain model score. Where a gene had multiple isoforms, the longest protein was selected as representative. The most sensitive approach (based on the number of proteins matched) used the SUPFAM MHC class I APD model, the PFAM Ig domain and HMMer2 in fs-mode. ULBPs and RAETs, which do not possess immunoglobulin domains, were identified by searching with the MHC class I APD HMM only.

Sensitive genome search

The predicted MHC class I proteins were used to construct custom profile HMM models in HMMer2. The 6-frame translations of the human (hg19), mouse (mm9), dog, cow, opossum (mondom5), wallaby (Meug_1.0), Tasmanian devil (assembly 7), platypus (OANA5), chicken, zebra finch, turkey, green anole, *Xenopus tropicalis*, zebrafish, tetraodon, lamprey, sea squirt, fruitfly and yeast genome sequences (Ensembl Release 75) were searched using profile HMMs representing the MHC class I APD (PFAM:PF00129, SUPFAM:0045513 and a custom model), C1-type Ig (PFAM:PF07654 and the custom model), C-terminal (PFAM:PF06623) and MHC class II β (PFAM:PF00969) domains with HMMer (version 2) with an E-value threshold of 10. Local alignment models (fs) were used.

The coordinates of predicted domains in the 6-frame translation were then transformed back to genomic coordinates. Genomic regions matching the first half of the MHC class I domain model were annotated as α_1 domains, while features matching the second half were annotated as α_2 domains. Regions also matching MHC class II β domains were removed if the class II match scores were greater than the class I match scores.

Genomic regions containing matches to the α_1 , α_2 , Ig and C-terminal domains with the correct orientation and order and intron-like separation were identified by aligning a model representing the canonical domain architecture of class I genes to the predicted domains (Figure 1). The alignment algorithm was implemented using dynamic programming. It used weighted HMMer scores as match scores. The weights were

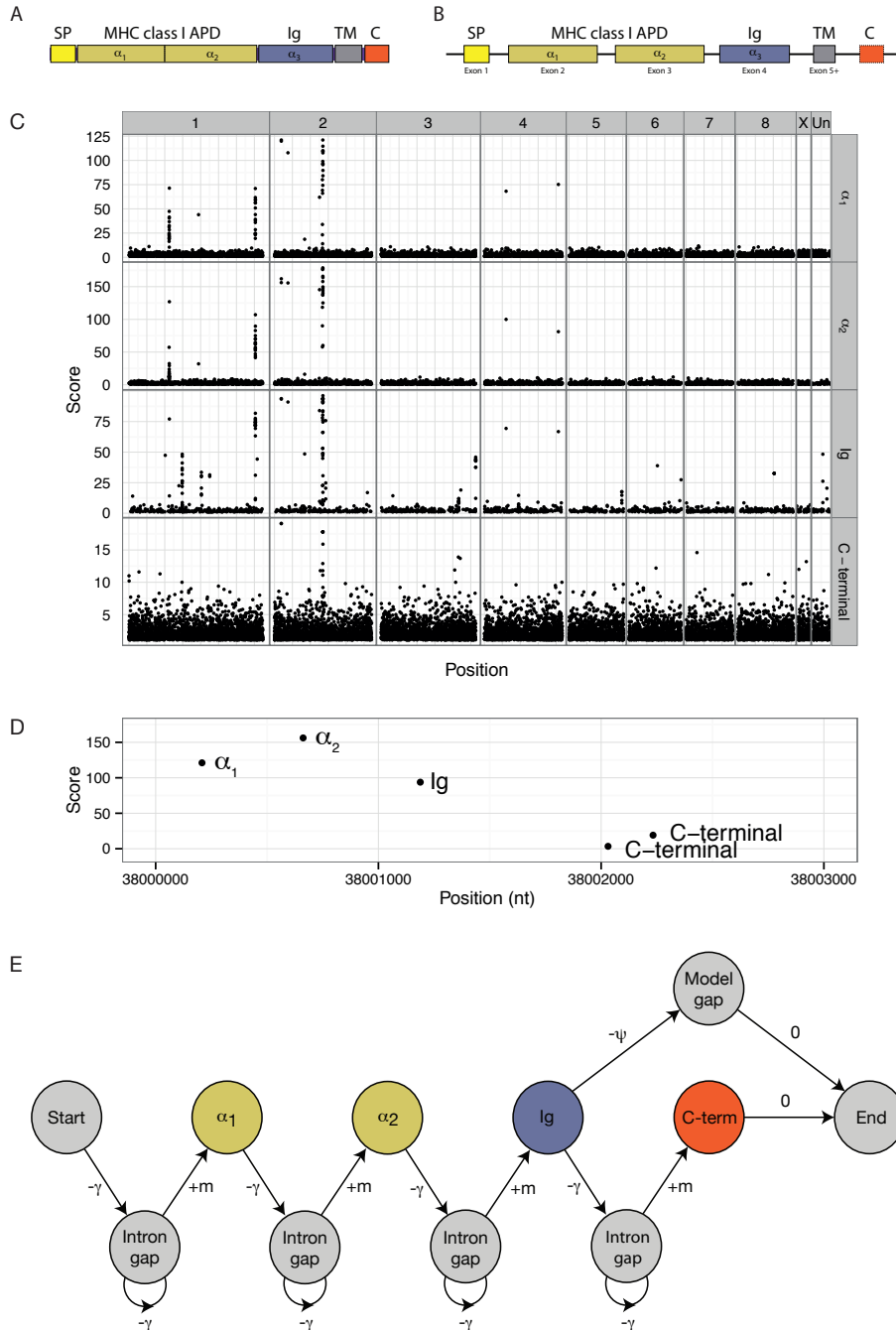


Figure 1. Sensitive pan-genome search for MHC class I genes. The canonical domain structure of MHC class I (A) proteins and (B) genes. (C) The location in the opossum genome and score of matches to profile hidden Markov models representing the antigen-presenting domain (split into α_1 and α_2 regions), C-type immunoglobulin domain and C-terminal domain. (D) Example of a high-scoring run of α_1 , α_2 , Ig and C-terminal domains in the opossum genome. (E) Finite state automata of the alignment algorithm to search for runs of α_1 , α_2 , Ig and C-terminal domains, taking domain score and distance between domains into account. The nodes (circles) show match states. Symbols on edges show scores/penalties: $+m$ is the match score, which is based on the HMM match score; $-\gamma$ is a distance-dependent affine gap penalty, which models introns and allows the alignment to skip over matches that interrupt a run of domains; $-\psi$ is a constant penalty for dropping the C-terminal domain.

selected to approximately normalise the contributions from each domain (weights were α_1 : 1, α_2 : 1, Ig: 2, C-terminal: 20). An affine gap penalty was used to model introns with gaps shorter than 5000nt penalty-free and calibrated so that a 20,000nt gap gets a penalty of 300. Mismatches are effectively disallowed by applying a very large mismatch score (-20,000), but the affine gap function can skip over mismatching domains. Parameters were selected to have maximum sensitivity on the well-annotated human MHC class I genes and then tested on the mouse (positive control) and lamprey and sea squirt genomes (negative controls).

The workflow is summarized in Supplementary Figure 5 and code is provided in the Supplementary Materials and Methods.

Phylogenetic analysis

Multiple sequence alignments of predicted peptide sequences were generated using Clustal Omega^{49,50} and edited in jalview⁵¹.

The phylogeny of the 449 predicted MHC class I genes identified in the representative jawed vertebrates was inferred using the JTT model⁵² in BEAST2⁵³. A discrete Gamma distribution with 4 categories was used to model evolutionary rate differences among sites. Four Markov chains were run for 3,000,000 steps each starting from random trees. Trees were output every 1000 steps. The consensus tree was estimated from the last 500,000 steps of the 4 chains.

To infer the evolutionary history of 30 selected human, mouse, marsupial and monotreme MHC class I genes, and the gene tree of the 46 *UT* family members, the best phylogenetic model was first selected using PROTTEST3⁵⁴. In both cases, the best model based on AIC was the JTT method⁵² with invariant sites, gamma rate distribution, and empirical amino acid frequency. Phylogenetic trees were estimated using the maximum likelihood method with MEGA5⁵⁴. The bootstrap consensus tree inferred from 500 replicates was taken to represent the evolutionary history of the genes analyzed. A discrete Gamma distribution was used to model evolutionary rate differences among sites with 4 categories. The rate variation model allowed for some sites to be evolutionarily invariable. The *UT* gene tree and the species tree were reconciled using NOTUNG⁵⁵ to identify gene duplication and loss events.

BAC library screening

Overgo probes representing each of the wallaby and platypus novel class I genes were designed from genomic sequence using the Overgo Maker program. The specificity of the resulting overgos was judged by using the 40bp probe sequence to BLASTN search the tammar wallaby or platypus genomes. All overgoes used to screen the BAC libraries are listed in Supplementary Table 6. Overgos were radioactively labelled, pooled and hybridised to tammar wallaby (Me_KBa; Arizona Genomics Institute) or platypus (Oa_Bb, Clemson University Genomics Institute) BAC library filters as previously described¹¹. Positive BACs from this initial screening were spotted onto Hybond N+ and subjected to a further round of screening with individual probes as previously described⁵⁶.

Fluorescence in situ hybridisation

BAC DNA from each positive BAC clone was directly labelled with either SpectrumOrange dUTP or SpectrumGreen dUTP (Abbott Molecular Inc., Des Plaines, IL, USA). Labelled BACs were hybridised to male metaphase chromosomes spreads, visualised and imaged as previously described⁵⁶.

RT-PCR of predicted transcripts in opossum

Coding sequences of *mdUT* genes were amplified by targeted PCR with primers designed based on predicted exon 2 and 3 gene sequence, using a cDNA library constructed from opossum thymus mRNA (Supplementary Table 7). The PCR was done using Advantage HF 2 PCR kit (Clontech, Mountain View, CA), with the following parameters for all primers: 94°C for 1 minute, 35 cycles of 94°C for 30 seconds and 61 to 65.1°C gradient for 4 minutes, and 68°C for 5 minutes. The amplified DNA was then ligated into the pCR4-TOPO TA vector, transformed into One Shot Chemically Competent TOP10 *E. coli*, and incubated with 250µL LB medium at 37° while shaking for 1 hr (Invitrogen, Carlsbad, CA). A total of 120 µL of the transformed cells were then plated on ampicillin agar plates and incubated between 12-18 hours at 37°C. A minimum of 8 clones per plate were chosen, and plasmid DNA were generated using the boiling lysis method. Both the forward and reverse strands were sequenced with BigDye Terminator v3.1 Cycle Sequencing Kit (Invitrogen, Carlsbad, CA) Analysis of the sequences was done using Sequencher 5.0 (Gene Codes, Ann Arbor, MI).

Searching marsupial immune tissue transcriptome data

To find support for the expression of UTs in several marsupials, sequencing data from the following immune tissue cDNA or EST libraries were downloaded: Roche 454 sequencing data from tammar wallaby, *Macropus eugenii*, thoracic and cervical thymus cDNA libraries (NCBI SRA accessions: SRX019250 and SRX019249)⁵⁷; Roche 454 sequencing data from Tasmanian devil, *Sarcophilus harrisii*, spleen and lymph node cDNA libraries (NCBI SRA accession: TBA); Roche 454 sequencing data from the opossum, *Monodelphis domestica*, thymus cDNA libraries (Katina Krasnec and Robert Miller, unpublished data); 17,818 ESTs from brushtail possum, *Trichosurus vulpecula*, spleen, lymph node and stimulated splenocytes (NCBI EST: LIBEST_019237); and a small set of 1319 ESTs from a northern brown bandicoot, *Isodon macrourus*, thymus ESTs (GenBank accession: EE743888-EE745206) *ENREF_41*⁵⁸.

Reads from each library were aligned to predicted tammar wallaby UTs, or Tasmanian devil UTs, in the case of the devil spleen and lymph libraries, using BLASTN. An E-value threshold of 10^{-5} was used and only a single best hit was recorded.

Structural homology modelling

Structure prediction used the I-TASSER method²⁰. Structural similarity or divergence was evaluated by a pairwise root mean square deviation (RMSD) value upon superposition of the backbone C α trace from the two groups of structurally equivalent atoms in MHC class I α_1 and α_2 domains. Structure visualization and the RMSD calculation are using Pymol (<http://pymol.sourceforge.net/>).

Results

Sensitive peptide searches for MHC class I proteins

We first set out to identify all annotated MHC class I proteins in 15 representative species sampled from across vertebrate life. The selected species comprised human, mouse, dog, cow, three species of marsupials with sequenced genomes, platypus, three avian species, a lizard, a frog, and two fish species. Additionally, we selected 4 eukaryotic species known to lack MHC class I genes as negative controls (lamprey, sea squirt, fruitfly and yeast). Predicted protein sequences from these species were obtained from Ensembl and searched using profile HMMs representing the MHC class I APD and the C1-type Ig domain, which are characteristic of MHC class I genes, and the MHC class II β domain, with HMMer. The separate domain searches were integrated and MHC class I proteins predicted using a simple heuristic: proteins were annotated as predicted MHC class I proteins if they had a significant match to the MHC class I APD or a weak match to the APD and a significant match to the Ig domain model in the correct order, with the additional requirement that the MHC class I APD model matched with higher score than the class II β domain model. MHC class I genes frequently encode multiple isoforms; in these cases, we selected the longest protein as the representative protein. A variety of HMMs were tested (e.g. PFAM, SUPFAM and iteratively constructed custom models; see Methods for details) and the most sensitive combination was adopted.

Our search identified 348 MHC class I proteins across the 15 jawed vertebrate species searched (summarized in Table 1). This included all 24 known human and 41 mouse

Table 1. Summary of the number of MHC class I genes across species. The number MHC class I genes identified in each species by searching annotated proteins using customized models and by sensitive genome search.

Species	Number of predicted MHC class I genes		
	Protein search	Genome search	Merged total
Human	24	26	33
Mouse	41	49	55
Dog	19	21	24
Cow	47	39	55
Opossum	28	40	47
Tammar wallaby	17	35	41
Tasmanian devil	22	23	25
Platypus	19	10	21
Chicken	24	21	26
Zebrafinch	11	3	11
Turkey	7	3	7
Green anole	25	19	26
Frog	26	31	32
Zebrafish	28	31	33
Tetraodon	10	10	13
Lamprey	0	0	0
Sea squirt	0	0	0
Fruitfly	0	0	0
Yeast	0	0	0
	348	361	449

MHC class I proteins with no false positives. Searches of several negative controls—lamprey, sea squirt, fruitfly and yeast—did not identify any MHC class I proteins. Aligning all PFAM-A domain models to the set of predicted MHC class I proteins using hmmpfam showed that for each protein the strongest matches consisted only of the MHC class I APD, Ig and in some cases the conserved MHC C-terminal domains, with no other unexpected high quality matches. MHC class II genes were never misidentified as class I genes in the searches of any jawed vertebrate protein databases. Taken together these observations indicate the approach has high sensitivity and specificity.

Sensitive genome searches for MHC class I genes

Next, we set out to identify any unannotated MHC class I genes in these genomes using a highly sensitive search method designed to take advantage of the conserved exon/domain organisation of MHC class I genes (Figure 1A). Profile HMMs representing the MHC class I APD, C1-type Ig, MHC C-terminal, and MHC class II β domains were used to search the six-frame translation of each genome. The domain matches in the 6-frame translation were transformed back to genomic coordinates and the α_1 , α_2 , Ig and C-terminal domains within the model matches were identified. In each species, we found thousands of matches to these domains (summarized in Supplementary Table 1). For example in the opossum genome, we found 2127 matches to the α_1 domain, 3571 matches to the α_2 domain, 5028 matches to the Ig domain and 5546 matches to the MHC C-terminal domain. The majority of these matches had low scores. However, both isolated and clustered high scoring matches were also apparent (Figure 1B). Genomic features matching the expected structure of an MHC class I gene, that is a chain of α_1 , α_2

and Ig domains and optionally a C-terminal domain on the same strand and at intron-like distances (for example Figure 1C) were identified by aligning a canonical model of an MHC class I gene, taking match score and the gaps between domains into account (Figure 1D and methods for details). Once again, a variety of HMMs were tested (e.g. PFAM, SUPFAM, and custom models based on the protein search results; see Methods for details). The custom models were adopted as the most sensitive.

From the 388,409 domain matches across all species, the genome search identified 361 genomic features possessing the MHC class I gene structure (summarized in Table 1; Supplementary Table 2 for details). These included 26 putative MHC class I genes in the human genome, 49 in mouse, and 40 in the opossum. Again, searches of the negative controls identified no MHC class I genes, as one would expect. These genomic features included annotated genes and both annotated and unannotated pseudogenes. Merging the protein and genome searches produced a total of 449 MHC class I genes and proteins across the species searched (Supplementary Table 3), including a total of 33 in human, 55 in mouse and 47 in the opossum.

The most dramatic differences between the results of searching annotated class I proteins and an unbiased search of the whole genome arose in the marsupials and monotremes. The annotation of the opossum genome (Ensembl Release 75) contains 28 MHC class I genes, but 40 putative MHC class I genes (genomic features with structural similarity to MHC class I genes) were identified in the sensitive genome search results. Seven of the annotated proteins were missed in the genome search, as the corresponding loci lack Ig

domains. Fifteen of the loci identified by the genome search were unannotated. Five of these contained in-frame stops, including opossum *CDI*, *UH*, and a *MIC*-like gene. These in-frame stops may be due to sequencing errors in the draft opossum genome, polymorphisms in the individual sequenced or the fact that our model does not take splice sites into account and may erroneously include short segments of intronic sequence in the domain matches, resulting in the genomic feature going out of frame. Consequently, we retained these genes in our analyses. Thus, a total of 47 putative MHC class I genes were identified. A similar pattern emerged in other marsupial and monotreme genomes.

Phylogenetic analysis

To annotate these genes and understand the evolutionary relationship between them, we inferred the phylogenetic relationships between all MHC class I genes identified in the selected vertebrates using a Markov Chain Monte Carlo (MCMC) method on the JTT+IGF model. Four MCMCs were run (see Supplementary Figure 1 for traces of posterior probability) and the consensus tree from the last 500 steps of each run was taken to represent the evolutionary history of the genes (Figure 2A). Additionally, a smaller phylogeny consisting of just human and opossum class I genes and the mouse *Mill* genes was also inferred by maximum likelihood (Supplementary Figure 2).

While support in parts of the trees is low, the phylogenies provide a number of insights into the evolution of MHC class I genes in vertebrates. The large tree provides additional evidence for the previous observation that the non-classical MHC class I gene family *MRI* is found only eutherians and marsupials¹⁶. Similarly, it suggests that the *FCGRT*,

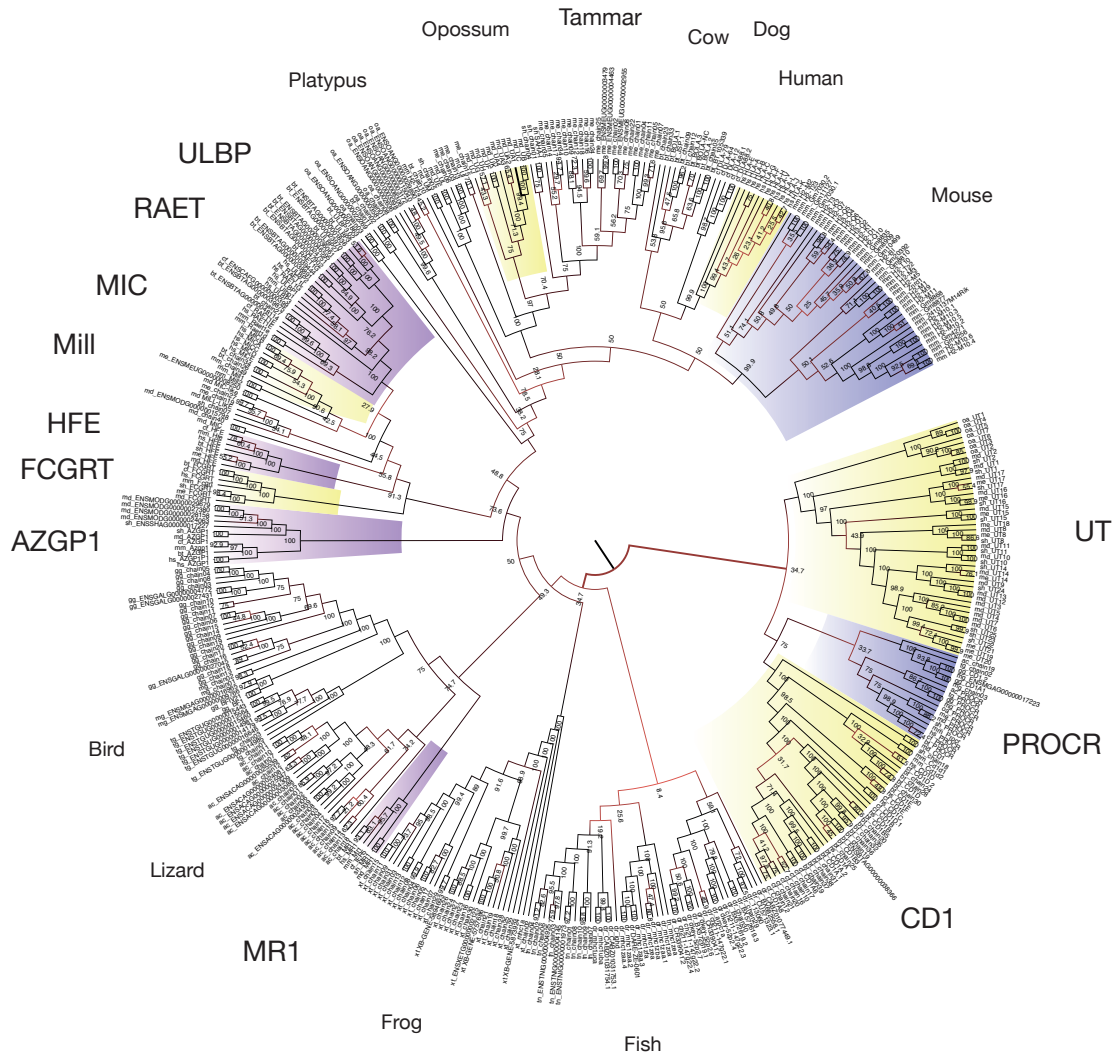


Figure 2. Phylogeny of class I_s predicted in representative species spanning the jawed vertebrates estimated by MCMC on the JTT+IG model. Numbers at nodes represent the frequency with which that split is observed. Gene families are labelled around the outside. The species label shows the location of classical MHC class I for each species or group of species. Key gene families or species' classical class I genes are highlighted in colour.

HFE and *AZGPI* gene families are specific to eutherians and marsupials. It demonstrates that the *PROCR* gene family is found across the amniotes. It suggests that MIC is duplicated in opossum (md_chain40), though this contains in-frame stops. The small tree supports the previous observation that marsupials may have a member of the *ULBP* gene family (ENSMODG00000015798)¹⁷. It identifies a possible expansion of *AZGPI* in opossum (ENSMODG00000024063, ENSMODG00000027380, ENSMODG00000028158, and ENSMODG00000029679). The phylogenies also reveals two new opossum MHC class I genes that are located in the MHC, but have not previously been identified, which we have denoted *UA3* and *UA4*. These appear to be closely related to *UA1* and *UA2*.

Strikingly, the phylogenetic tree identifies an extensive and entirely novel clade of MHC class I genes in marsupials and monotremes, which we have named *UT*. There are 17 *UT* family genes identified in the opossum genome, 9 in tammar wallaby, 13 in the Tasmanian devil and 7 in the platypus. The numbering of *UTs* is based on location in the gene cluster in the opossum and clear orthology, or lack of it in other marsupials. Platypus *UTs* are numbered independently as these appear to form a distinct clade. This is highlighted by the *UT* gene tree (Supplementary Figure 3), which was estimated using maximum likelihood with the JTT+IGF model and reconciled with the species tree using NOTUNG. No *UTs* were identified outside of the marsupials and monotremes in our searches.

Chromosomal location

The *UT* family of MHC class I genes is encoded in a gene cluster on chromosome 1 in the opossum genome (Figure 3). This region is approximately 460 kilobases in size. Interestingly, the cluster is located at an evolutionary breakpoint and is flanked by genomic regions that share synteny with different chromosomes in human (chr2 and chr20) and mouse (chr6 and chr2). The tammar wallaby genome assembly (Meug1.0) is highly fragmented and scaffolds are not mapped to chromosomes. Fluorescence In-Situ Hybridization (FISH) shows that the *UT* gene cluster is also located on chromosome 1 in the tammar wallaby genome (Figure 4), as predicted by conserved synteny between the tammar and opossum¹⁸. Interestingly, the FISH also shows a signal on the tammar Y chromosome. As all marsupial genomes sequenced were female, this locus was not detected in genome-wide searches and the significance of this signal is not yet understood. Based on the digital karyotype of the Tasmanian devil¹⁹, the *UT* gene family is also located on chromosome 1. In platypus, the *UT* genes are located on unmapped scaffolds and we were unable to map them using FISH.

Sequencing and gene expression

Of the 17 putative opossum *UT* genes, the expression of 8 genes, consisting of *UT4*, *UT5*, *UT6*, *UT8*, *UT9*, *UT10*, *UT15*, and *UT17*, was confirmed in opossum thymus using RT-PCR (Supplementary Figure 4). Predicted sequences obtained from our sensitive search method were confirmed using RT-PCR to obtain amplicon sequences from within exons 2 and 3 (Supplementary Table 4). A further 4 *UT* loci, *UT2*, *UT3*, *UT7*, and *UT16*, were

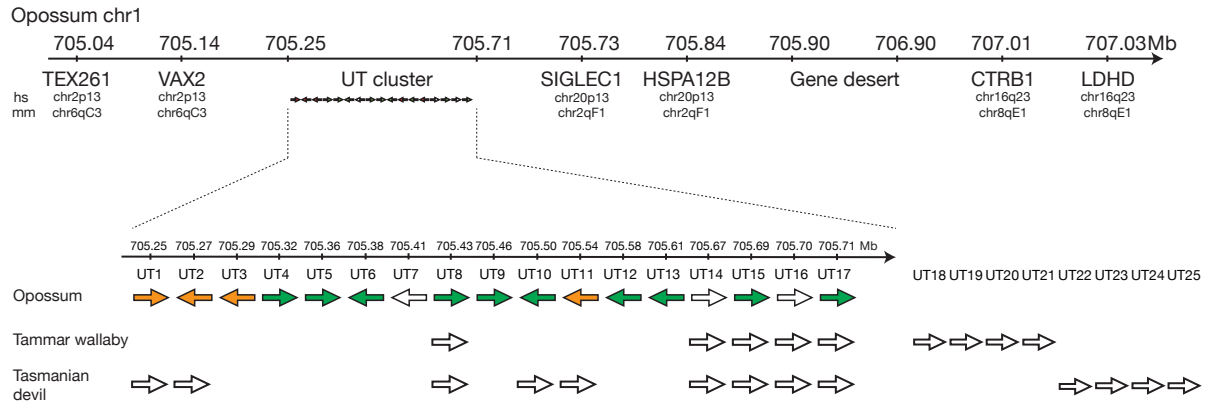


Figure 3. Comparative map of the *UT* cluster. (A) Genomic region containing the *UT* cluster in opossum showing the non-synteny of flanking genes between the opossum and human/mouse genomes. (B) Comparative map of *UT* cluster in opossum, tammar wallaby and Tasmanian devil. Green arrows represent *UT* genes found transcribed, orange for genes not transcribed, and white for unknown transcription

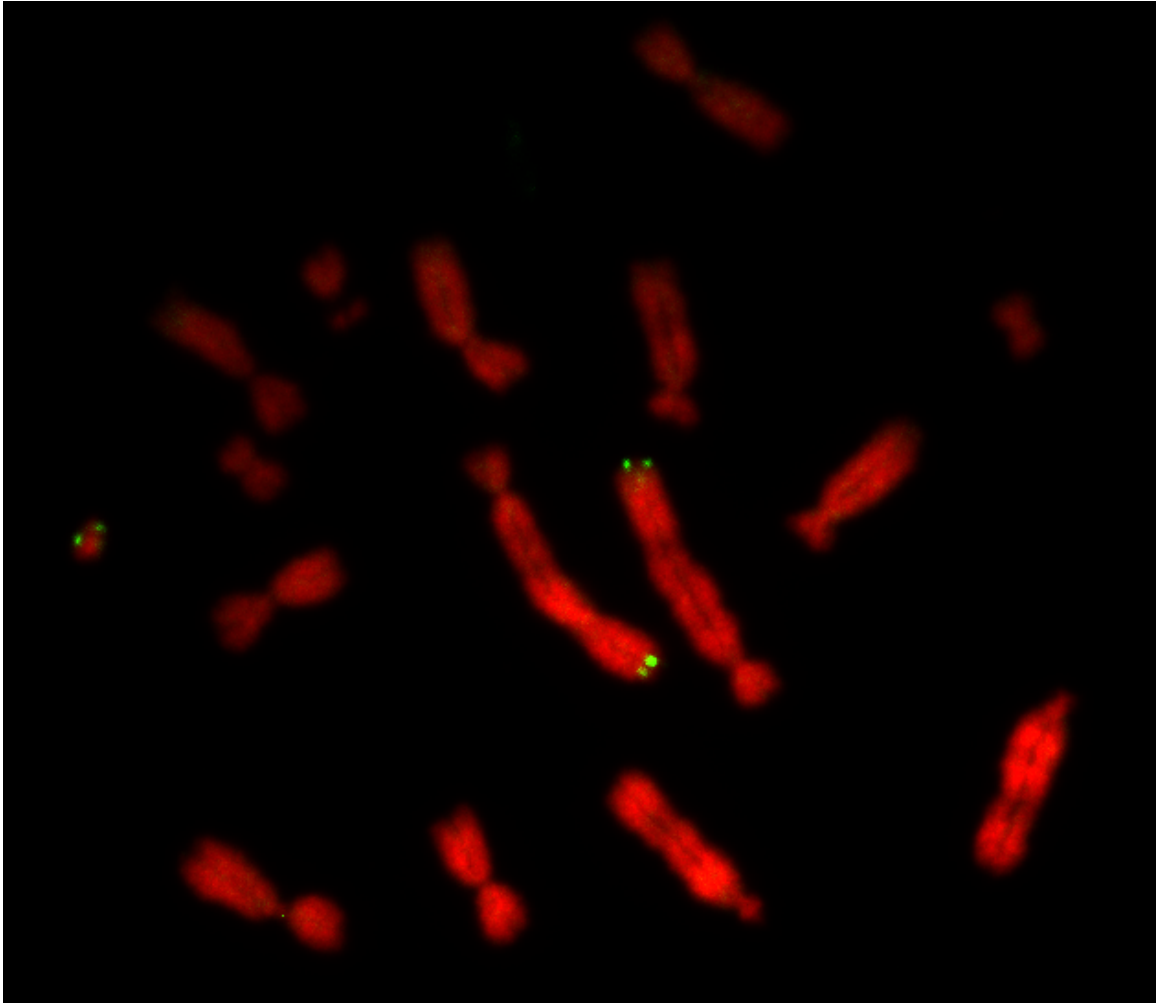


Figure 4. Fluorescence In-Situ Hybridisation showing location of *UT* cluster on chr1 in the tammar wallaby. A signal is also observed on the Y chromosome.

confirmed as expressed in Roche 454 sequencing data from an opossum thymus cDNA library (data not shown).

Transcription of tammar wallaby *UT1*, *UT2*, *UT3*, *UT4*, *UT5*, and *UT8* was confirmed in 454 data from thoracic and cervical tammar thymus cDNA libraries. There was support for Tasmanian devil *UT2*, *UT8*, and *UT11* in 454 cDNA data from devil spleen, but no UTs were detected in a lymph node library.

Limited transcriptome sequence data is available from immune tissues of other species of marsupial. An EST (id: 161106CS44009845FFFFFF) from brushtail possum immune tissues with homology to *meUT2* was also identified, providing support for the existence of functional *UTs* in the possum. No *UT* transcripts were detected in BLASTN searches of a small bandicoot thymus EST library. No platypus immune tissues transcriptome data was available.

Homology mapping

To investigate the function of *UT* family members we predicted the protein structure of selected *UTs* (opossum *UT4*, *UT5* and *UT8*) using homology modeling with the I-TASSER method²⁰. Protein structures from the Protein Data Bank (PDB) that were closest to the predicted models comprised both classical and non-classical MHC class I genes from chicken, cow, mouse and human (Supplementary Table 5; Figure 5A for an annotated sequence alignment of 7 of the top matches). The structure of the classical chicken MHC class I protein B21 (3BEV²¹) was the best match for *UT8* and appeared in

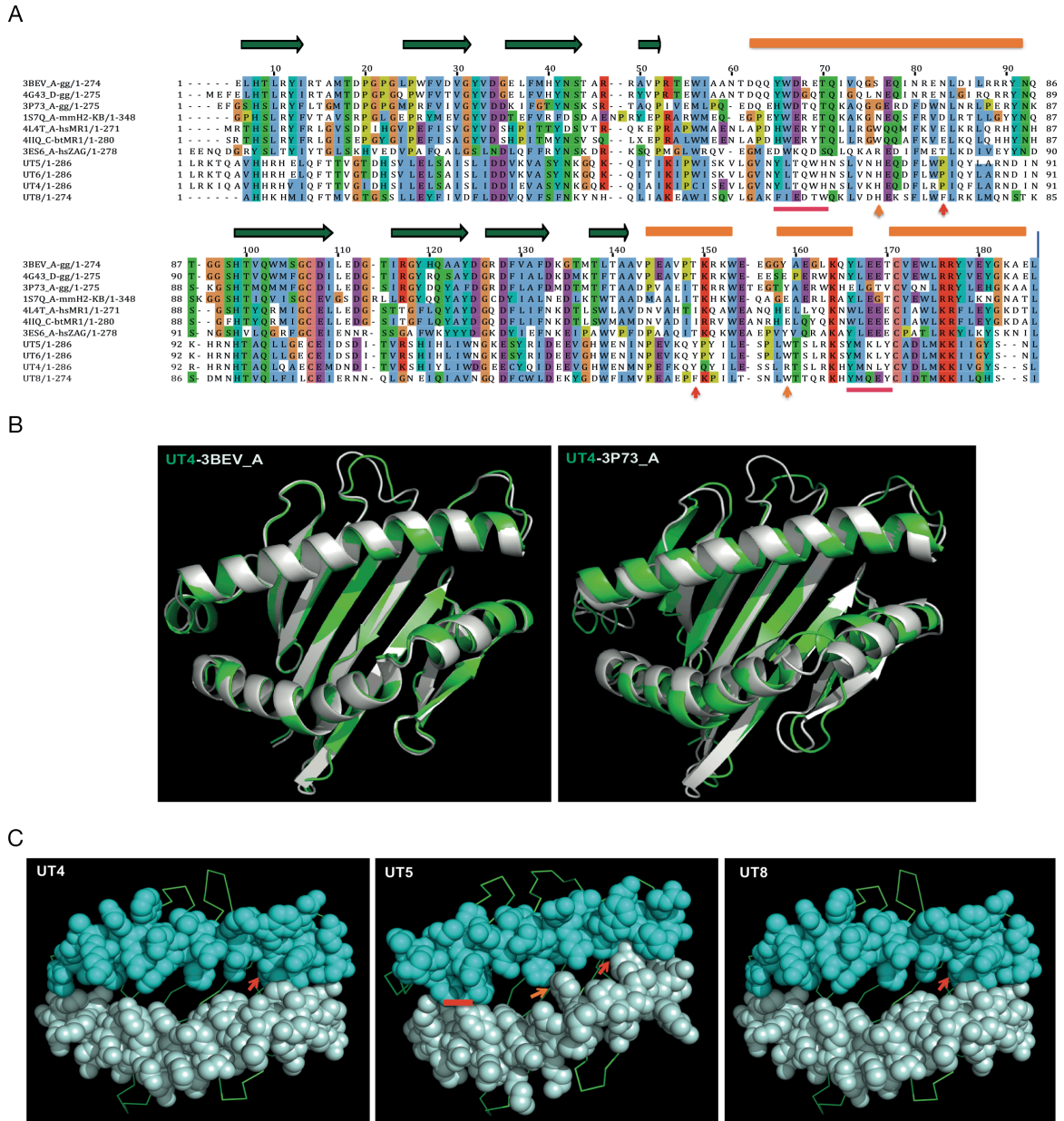


Figure 5. Predicted structure of UT proteins. (A) Sequence alignment of the α_1 and α_2 domains of *UT4*, *5*, *6*, *8* with *7* of the top 10 structural analogs from Protein Data Bank (PDB) identified by I-TASSER. Orange bars show α -helices. Green arrows highlight β -strands. Major differences between the *UTs* and templates are indicated by arrows, or red lines, and the consequence of these on the protein structures are shown in Figure 5C. (B) Overlay of the backbone of the peptide binding groove of *mdUT4* with its top 2 structural analogs, 3BEV_A (*ggB21*) and 3P73_A (*ggYF1*). (C) Superposed model structures of *mdUT4*, *5* and *8* with the *B21* template shows the antigen-binding groove is open, but possibly short. Filled sphere view shows the α -helices and ribbons show the β -sheets of the peptide-binding platform on the modeled protein structures. The residues indicated with arrows in Figure 5A cause the binding grooves to be short or narrow (e.g. *UT4*: the distance between Pro81 C γ to Tyr144 OH is 3.0Å; *UT5*: the distance between Pro81 C β to Tyr144 OH is 2.9Å, and a close hydrophobic contact between $_{64}$ LTQW $_{67}$ and $_{161}$ MNLY $_{154}$; *UT8*: the distance between Phe75 C ϵ to Phe138 C ξ is 2.9Å).

the top 5 templates for all *UTs* examined. The backbone structural alignment of *UT4* with 3BEV and 3P73²², the top 2 structural analogs for *UT4* and *UT8*, are shown in Figure 5B.

The peptide-binding grooves of *UT4*, 5 and 8 are shown in Figure 5C.

Discussion

MHC class I molecules have been historically defined by their function. The classical MHC class I typically presents peptide fragments derived from antigens to CD8⁺ cytotoxic T lymphocytes. This particular function is ubiquitous across the jawed vertebrates and is likely the primordial function of the class I protein. However it is clear that members of the MHC class I family have evolved to perform other functions, often in lineage specific ways. Therefore, a more appropriate definition of an MHC class I molecule is its unique structure which is a heterodimer of an α -chain paired with β 2-microglobulin. The MHC class I α -chain is composed of three extracellular domains. The α_3 domain is an immunoglobulin domain, a protein fold that predates the origin of jawed vertebrates in evolution. The origins of the α_1 and α_2 domains that make up the antigen-binding groove are more enigmatic and appear unique to the MHC molecules. Searching the genomes of jawless vertebrates and invertebrates failed to uncover genes encoding α_1 - and α_2 -like domains, shedding no light on their evolutionary origin.

The diversity of functions that MHC class I molecules have evolved to perform demonstrate the plasticity of this protein structure. For example, FcRN, which functions as an IgG receptor in mammals, does not bind the Fc region using the antigen-binding groove. Rather that groove is fairly closed and the IgG binds to a combination of the outer face of the α_2 and β 2-microglobulin domains²³. Such functional plasticity of a protein structure leads to the question of what other roles these molecules may have evolved to perform and how divergent they may have become. The results presented here

demonstrate that highly divergent genes, based on nucleotide sequence can encode proteins that fold to produce the MHC class I α -chain structure. This is not to say that the *UT* genes are a product of convergent evolution. Indeed, the phylogenetic analyses place them squarely within the extended MHC class I family. The phylogenetic position of these genes, close to *CDI* and *PROCR*, suggests they are non-classical MHC class I genes, but diversity and expression also need to be assessed.

The presence of the *UT* family of MHC class I genes in both marsupials and monotremes is consistent with their being ancient and present in the last common ancestor of all living mammals. The marsupials and eutherians (placental mammals) most recently shared a common ancestor approximately 165 million years ago²⁴. It appears that after the divergence of these two lineages the *UT* family was lost in the eutherians, likely prior to the radiation of the modern eutherians. Without knowing the function of *UT* genes it is difficult to speculate on why they were dispensable in the eutherians. However this is not the first case where mammal specific immune system genes have been lost in the eutherians. Both marsupials and monotremes have orthologues of a uniquely mammalian T cell receptor called TCR μ , however this locus has been lost in the eutherian lineage^{25,26}. Given the classical role of MHC class I molecules interacting with TCR it is an intriguing possibility that there is a functional connection between the *UT* molecules and TCR μ , whereby *UT* present antigen to TCR μ^+ T cells. Hence the loss of one may have resulted in the loss of the other in eutherians. While possible this would not be consistent with current models of how TCR μ chains interact with antigen in an MHC independent manner²⁷. Similar to *UT* loci, the *TCR μ* cluster is located in a region of the mammalian

genomes that have a break in synteny between marsupials and monotremes²⁸. It may be these gene families were independently lost due to being in regions of the genome subject to instability or rearrangement.

Other working hypotheses on the function of UT molecules are based on models predicting the structure of the region corresponding to an antigen-binding groove. Structurally, the UT proteins are most similar to the chicken B21 MHC class I molecule (3BEV), which binds peptide promiscuously²¹. The presence of some hydrophobic residues in the α -helices may make the effective binding size short or narrow. This may suggest the UTs present small peptide fragments, but is also consistent with a structure where the space between the α_1 and α_2 domains is relatively closed or alternatively, UT molecules may be involved in presenting hydrophobic antigens such as lipids. Marsupials have a homologue of the CD1 molecule that is normally involved in the presentation of glycolipids and lipoprotein antigens²⁹. However, the marsupial gene is single copy and not orthologous to any of the known *CD1a*, *b*, *c*, *d*, or *e* genes found in eutherians³⁰. Furthermore, in the opossum *M. domestica* *CD1* is a pseudogene³⁰. It is possible that there has been less pressure to retain or diversify the *CD1* family in marsupials due to some functional overlap with the *UT* genes.

The region of chromosome 1 containing the *UT* cluster does not correspond to one of the four MHC paralogous regions. These paralogous regions are the remnant of the two rounds of whole genome duplication that occurred during vertebrate evolution³¹. In humans these regions are located on chromosomes 1, 6, 9, and 19, with chromosome 6

containing the *bona fide* MHC region. In the opossum they are located on chromosomes 1, 2, and 3, with two paralogous regions on chromosome 2, one being the MHC proper³². The paralogous region on opossum chromosome 1 corresponds to human chromosome 9 and is syntenic to, but not identical to, the opossum *UT* region. Therefore, the extant *UTs* are the product of the novel expansion of diverging MHC class I genes in the marsupials and monotremes and likely originate from the duplication of an MHC class I gene in the ancestral mammal.

Conclusion

Using a novel, boutique method to search the annotated proteins and genomes of a selection of species spanning vertebrate life for MHC class I genes with high sensitivity, we identified a new class I gene family, the *UTs*. *UT* family members are encoded in gene clusters on chromosome 1 of the opossum, tammar wallaby and Tasmanian devil genomes, and are present but have not been mapped in platypus. The region is located in a synteny break between marsupial and the human genome. Homology modeling suggests *UT* genes have an open but short antigen-presenting groove, raising the possibility that they may present peptide epitopes or non-peptide fragments.

References

- 1 MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401**, 921-923 (1999).
- 2 Gorer, P. A. The genetic and antigenic basis of tumour transplantation. *Journal of Pathology* **44**, 691-697 (1937).
- 3 Trowsdale, J. The MHC, disease and selection. *Immunol Lett* **137**, 1-8 (2011).
- 4 Lessard, C. J. *et al.* The genomics of autoimmune disease in the era of genome-wide association studies and beyond. *Autoimmun Rev* **11**, 267-275 (2012).
- 5 Pietrangelo, A. Hereditary hemochromatosis. *Biochim Biophys Acta* **1763**, 700-710 (2006).
- 6 Baker, K. *et al.* Immune and non-immune functions of the (not so) neonatal Fc receptor, FcRn. *Semin Immunopathol* **31**, 223-236 (2009).
- 7 Kelley, J., Walter, L. & Trowsdale, J. Comparative genomics of major histocompatibility complexes. *Immunogenetics* **56**, 683-695 (2005).
- 8 Ohno, S. *Evolution by gene duplication*. (Springer-Verlag, 1970).
- 9 Miska, K. B. *et al.* Analysis of a marsupial MHC region containing two recently duplicated class I loci. *Mammalian genome : official journal of the International Mammalian Genome Society* **15**, 851-864 (2004).
- 10 Belov, K. *et al.* Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. *PLoS Biol* **4**, e46 (2006).
- 11 Deakin, J. E., Siddle, H. V., Cross, J. G., Belov, K. & Graves, J. A. Class I genes have split from the MHC in the tammar wallaby. *Cytogenet Genome Res* **116**, 205-211 (2007).
- 12 Zou, Z., Nomura, M., Takihara, Y., Yasunaga, T. & Shimada, K. Isolation and characterization of retinoic acid-inducible cDNA clones in F9 cells: a novel cDNA family encodes cell surface proteins sharing partial homology with MHC class I molecules. *Journal of biochemistry* **119**, 319-328 (1996).
- 13 Nomura, M. *et al.* Genomic structures and characterization of Rael family members encoding GPI-anchored cell surface proteins and expressed predominantly in embryonic mouse brain. *Journal of biochemistry* **120**, 987-995 (1996).

- 14 Cosman, D. *et al.* ULBPs, novel MHC class I-related molecules, bind to CMV glycoprotein UL16 and stimulate NK cytotoxicity through the NKG2D receptor. *Immunity* **14**, 123-133 (2001).
- 15 Radosavljevic, M. *et al.* A cluster of ten novel MHC class I related genes on human chromosome 6q24.2-q25.3. *Genomics* **79**, 114-123 (2002).
- 16 Tsukamoto, K., Deakin, J. E., Graves, J. A. & Hashimoto, K. Exceptionally high conservation of the MHC class I-related gene, MR1, among mammals. *Immunogenetics* **65**, 115-124 (2013).
- 17 Kondo, M. *et al.* Comparative genomic analysis of mammalian NKG2D ligand family genes provides insights into their origin and evolution. *Immunogenetics* **62**, 441-450 (2010).
- 18 Deakin, J. E. *et al.* Genomic restructuring in the Tasmanian devil facial tumour: chromosome painting and gene mapping provide clues to evolution of a transmissible tumour. *PLoS Genet* **8**, e1002483 (2012).
- 19 Murchison, E. P. *et al.* Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell* **148**, 780-791 (2012).
- 20 Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* **9**, 40 (2008).
- 21 Koch, M. *et al.* Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. *Immunity* **27**, 885-899 (2007).
- 22 Hee, C. S. *et al.* Structure of a classical MHC class I molecule that binds "non-classical" ligands. *PLoS Biol* **8**, e1000557 (2010).
- 23 Martin, W. L., West, A. P., Jr., Gan, L. & Bjorkman, P. J. Crystal structure at 2.8 Å of an FcRn/heterodimeric Fc complex: mechanism of pH-dependent binding. *Molecular cell* **7**, 867-877 (2001).
- 24 O'Leary, M. A. *et al.* The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* **339**, 662-667 (2013).
- 25 Parra, Z. E. *et al.* A unique T cell receptor discovered in marsupials. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9776-9781 (2007).
- 26 Wang, X., Parra, Z. E. & Miller, R. D. Platypus TCRmu provides insight into the origins and evolution of a uniquely mammalian TCR locus. *J Immunol* **187**, 5246-5254 (2011).

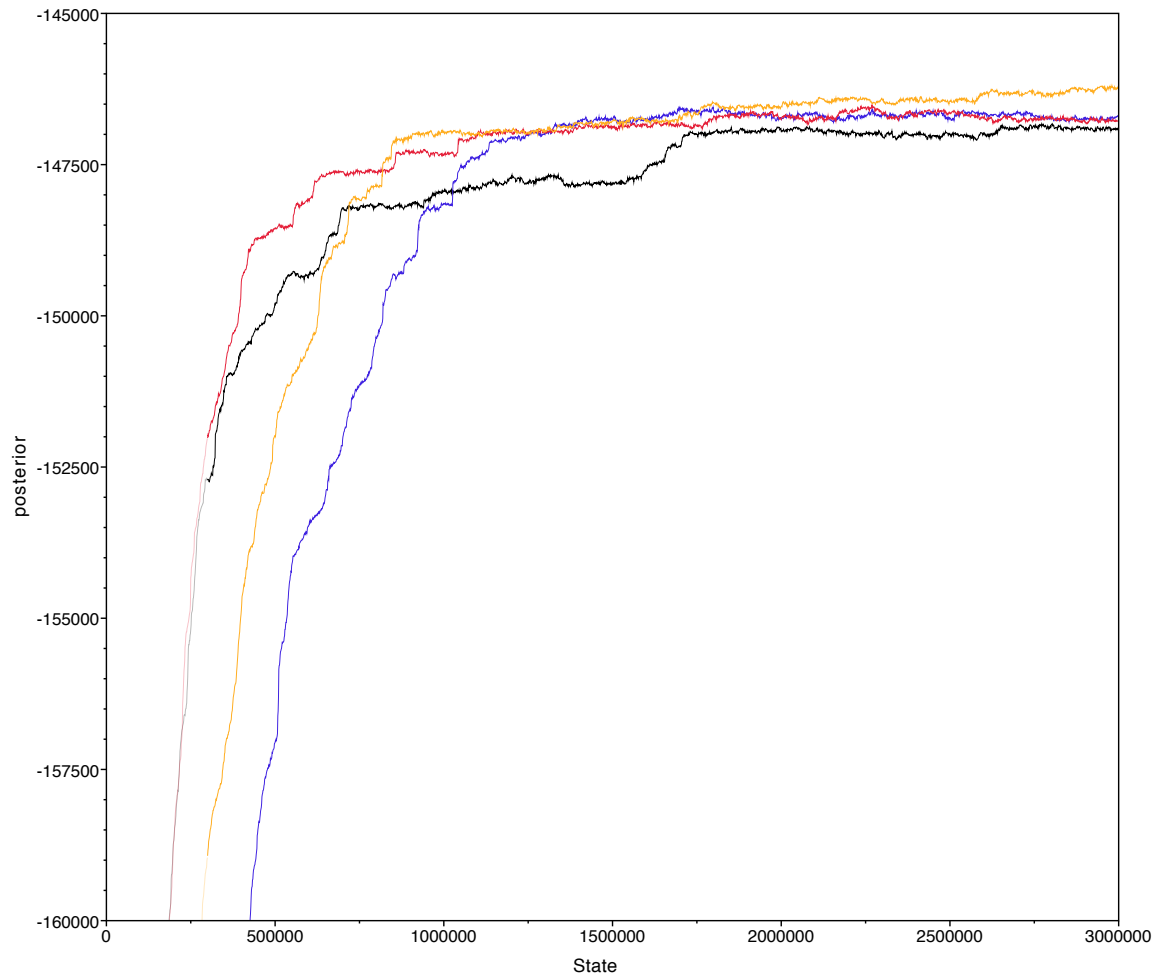
- 27 Flajnik, M. F., Deschacht, N. & Muyldermans, S. A case of convergence: why did a simple alternative to canonical antibodies arise in sharks and camels? *PLoS Biol* **9**, e1001120 (2011).
- 28 Parra, Z. E. *et al.* Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica*. *BMC genomics* **9**, 111 (2008).
- 29 Brigl, M. & Brenner, M. B. CD1: antigen presentation and T cell function. *Annual review of immunology* **22**, 817-890 (2004).
- 30 Baker, M. L. & Miller, R. D. Evolution of mammalian CD1: marsupial CD1 is not orthologous to the eutherian isoforms and is a pseudogene in the opossum *Monodelphis domestica*. *Immunology* **121**, 113-121 (2007).
- 31 Kasahara, M. The 2R hypothesis: an update. *Curr Opin Immunol* **19**, 547-552 (2007).
- 32 Belov, K. *et al.* Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome Res* **17**, 982-991 (2007).
- 33 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
- 34 Mouse Genome Sequencing Consortium *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562 (2002).
- 35 Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803-819 (2005).
- 36 Bovine Genome Sequencing Analysis Consortium *et al.* The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522-528 (2009).
- 37 Mikkelsen, T. S. *et al.* Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167-177 (2007).
- 38 Renfree, M. B. *et al.* Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol* **12**, R81 (2011).
- 39 Warren, W. C. *et al.* Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175-183 (2008).

- 40 International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716 (2004).
- 41 Warren, W. C. *et al.* The genome of a songbird. *Nature* **464**, 757-762 (2010).
- 42 Alfoldi, J. *et al.* The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587-591 (2011).
- 43 Hellsten, U. *et al.* The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**, 633-636 (2010).
- 44 Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498-503 (2013).
- 45 Jaillon, O. *et al.* Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**, 946-957 (2004).
- 46 Smith, J. J. *et al.* Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nature genetics* **45**, 415-421, 421e411-412 (2013).
- 47 Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-2195 (2000).
- 48 Mewes, H. W. *et al.* Overview of the yeast genome. *Nature* **387**, 7-65 (1997).
- 49 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**, 539 (2011).
- 50 Sievers, F. & Higgins, D. G. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods in molecular biology* **1079**, 105-116 (2014).
- 51 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).
- 52 Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences : CABIOS* **8**, 275-282 (1992).
- 53 Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology* **10**, e1003537 (2014).

- 54 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164-1165 (2011).
- 55 Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *Journal of computational biology : a journal of computational molecular cell biology* **7**, 429-447 (2000).
- 56 Deakin, J. E. *et al.* Physical map of two tammar wallaby chromosomes: a strategy for mapping in non-model mammals. *Chromosome Res* **16**, 1159-1175 (2008).
- 57 Wong, E. S. *et al.* Transcriptomic analysis supports similar functional roles for the two thymuses of the tammar wallaby. *BMC Genomics* **12**, 420 (2011).
- 58 Baker, M. L. *et al.* Analysis of a set of Australian northern brown bandicoot expressed sequence tags with comparison to the genome sequence of the South American grey short tailed opossum. *BMC Genomics* **8**, 50 (2007).

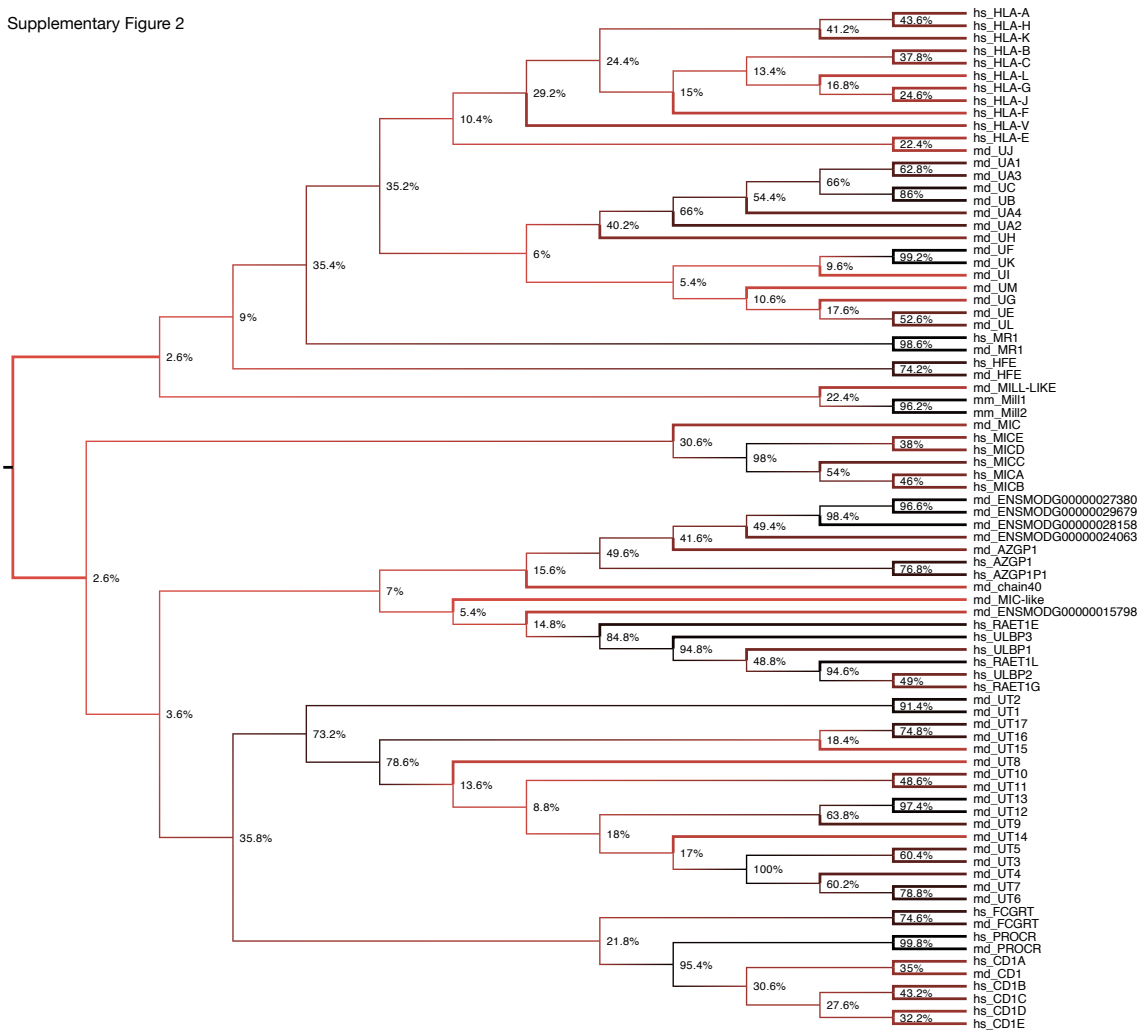
Supplementary Figures & Tables

Supplementary Figure 1



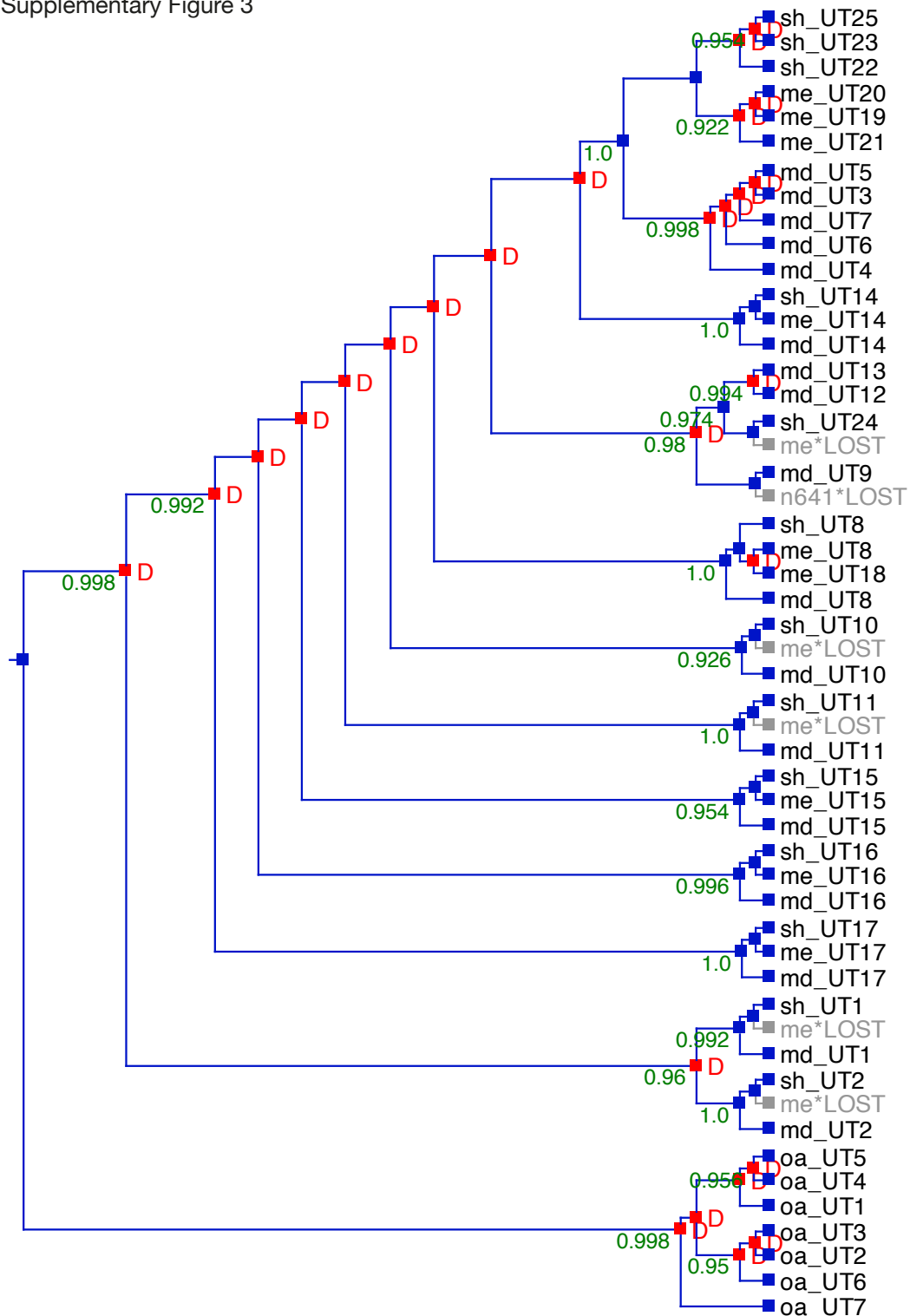
Supplementary Figure 1. The posterior probability of the phylogeny of 449 predicted MHC class I peptides generated by 4 MCMCs using BEAST2, started from random trees and sampled every 1000 steps.

Supplementary Figure 2



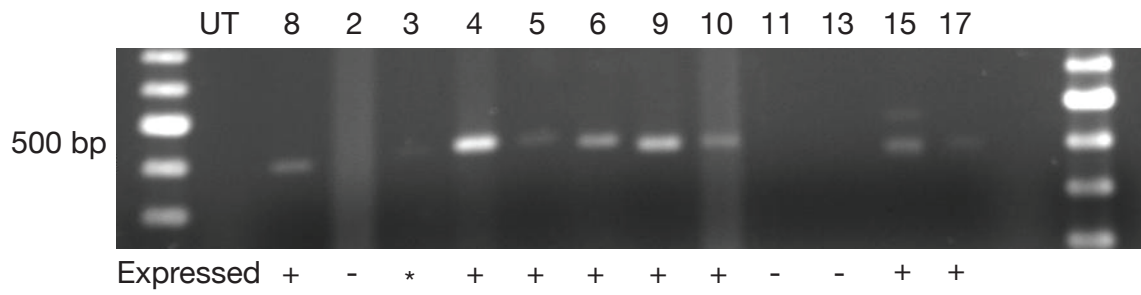
Supplementary Figure 2. Maximum likelihood tree (JTT+IGF model) of selected proteins including *UT* family members. Numbers at nodes indicate bootstrap support.

Supplementary Figure 3



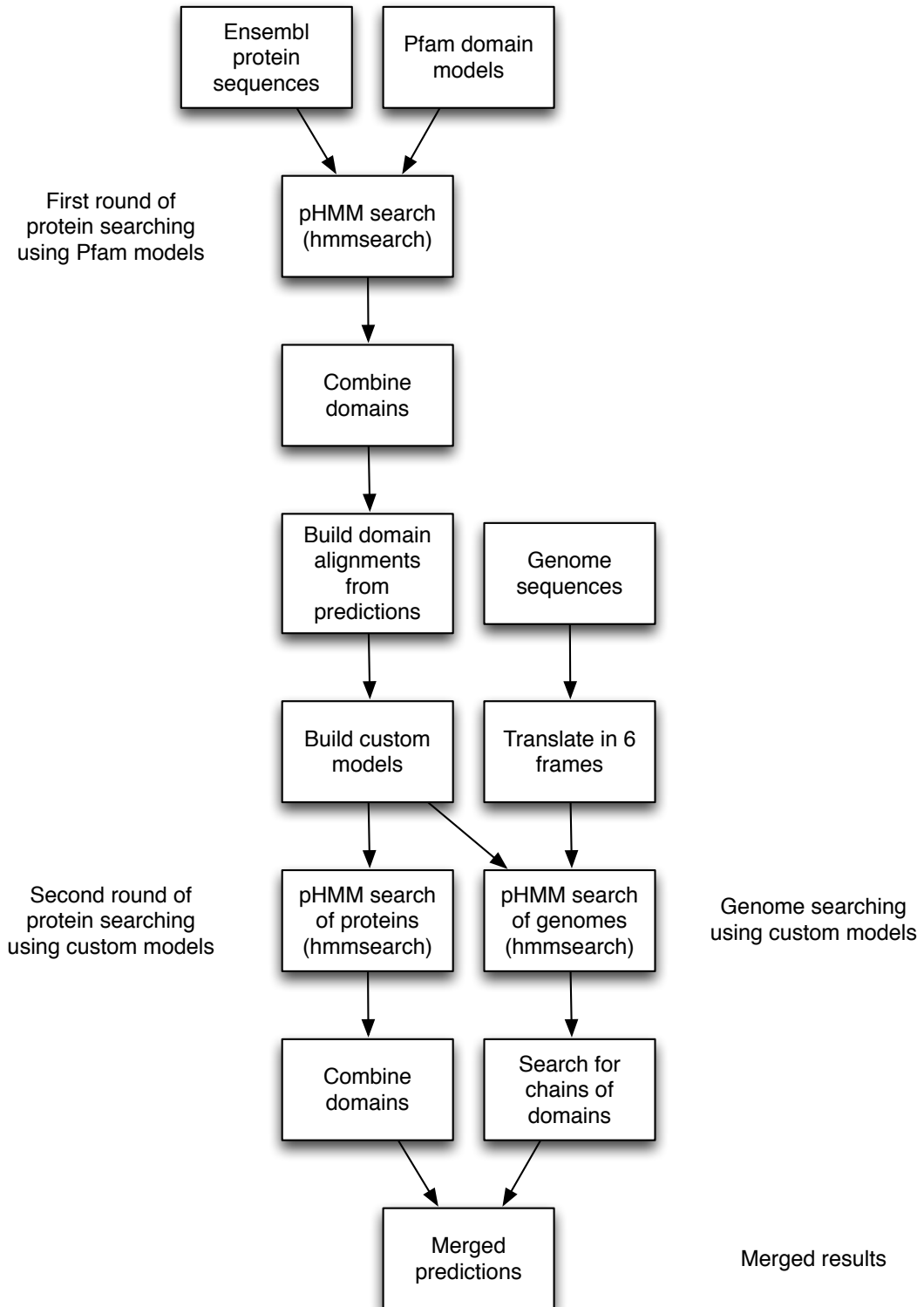
Supplementary Figure 3. *UT* gene family tree was estimated by maximum likelihood using the JTT+IGF model and reconciled with the species tree using NOTUNG. Predicted gene losses are shown in grey. Predicted duplications are indicated by a “D” at internal nodes. Bootstrap support is shown in red.

Supplementary Figure 4



Supplementary Figure 4. Confirmation of expression by RT-PCR in the opossum thymus.

Supplementary Figure 5



Supplementary Figure 5. Sensitive protein and genome search method workflows.

Supplementary Table 1. Number of domain matches found in the genomes of each species using custom profile hidden Markov models.

Species	Number of domain matches			
	Class I APD	Ig	C-terminal	Class II beta
Human	8468	17738	25836	24437
Mouse	7718	19763	25160	24902
Cow	9752	17066	33669	22055
Dog	6663	18066	22088	21248
Opossum	2127	3571	5028	5546
Tammar wallaby	212	295	53	38
Tasmanian devil	3728	4147	4093	6992
Platypus	795	4651	5480	5469
Chicken	3075	8752	10813	11979
Zebrafinch	4069	10490	13807	14411
Turkey	2833	7236	9760	10959
Green anole	4528	8813	9732	14282
Frog	1541	4996	2673	8766
Zebrafish	4674	9489	10040	16523
Tetraodon	930	2887	4448	3706
Lamprey	263	927	658	328
Sea squirt	606	588	403	1537
Fruitfly	644	779	1455	1762
Yeast	108	107	118	178
	62734	140361	185314	195118

Total class I domains	388409
-----------------------	--------

Supplementary Table 2. Genomic features of predicted MHC class I genes found in the sensitive genome search.

Please see pages 145-157 and publication “Papenfuss, A.T., Feng, Z.P., Krasnec, K.V., Deakin, J.E., Davoren, C.J., Baker, M.L., and Miller, R.D. Marsupials and monotremes possess a novel family of MHC class I genes that is lost from the eutherian lineage” for Supplementary Table 2 information

Supplementary Table 3. Predicted MHC class I sequences.

Please see pages 158-210 and publication “Papenfuss, A.T., Feng, Z.P., Krasnec, K.V., Deakin, J.E., Davoren, C.J., Baker, M.L., and Miller, R.D. Marsupials and monotremes possess a novel family of MHC class I genes that is lost from the eutherian lineage” for Supplementary Table 3 information

Supplementary Table 4. Opossum *UT* exon 2 and 3 amplicon sequences.

Please see pages 211-212 and publication “Papenfuss, A.T., Feng, Z.P., Krasnec, K.V., Deakin, J.E., Davoren, C.J., Baker, M.L., and Miller, R.D. Marsupials and monotremes possess a novel family of MHC class I genes that is lost from the eutherian lineage” for Supplementary Table 4 information

Supplementary Table 5. The pairwise backbone Root Mean Square Deviation (Å) between the α_1 and α_2 domains of the opossum *UT4*, *UT5*, and *UT8* modeling structures, several of the top 10 closest structural analogs identified using I-TASSER, and selected classical and non-classical MHC class I proteins from human and mouse.

Gene symbol	PDB code	Opossum UTs			Comment
		<i>UT4</i>	<i>UT5</i>	<i>UT8</i>	
mdUT4	-	0	1.716	0.936	
mdUT5	-	1.716	0	1.449	
mdUT8	-	0.936	1.449	0	
ggYF1	3P73*	2	1.79	1.953	
ggB4	4G43*	1.02	1.368	1.22	
ggB21	3BEV*	1.041	1.969	0.501	Binds 11mer peptide
ggB21	2YF6*	3.356	1.884	0.871	Binds 10mer peptide
btMR1	4IIQ*	2.387	2.308	2.308	
mmH-2Kb1	1S7Q*	2.492	2.979	2.959	
hsMR1	4L4T*	3.399	3.402	3.296	
hsAZGP1	3ES6*	3.121	3.941	3.462	
hsFcRn	1EXU	1.941	2.881	4.288	
hsHFE	1A6Z	2.657	2.977	2.919	
hsMICA	1HYR	4.519	6.291	5.534	
hsMICB	1JE6	6.989	7.94	8.283	
hsULBP	1KCG	22.309	7.634	4.986	
hsCD1a	1ONQ	12.601	13.225	4.042	
hsCD1b	1GZQ	7.31	5.466	3.49	
hsPROCR	1L8J	6.17	6.669	4.545	

*From top 10 templates for all the UTs

Supplementary Table 6. Overgo sequences used to isolate tammar wallaby and platypus BACs containing *UT* loci.

Gene symbol	Overgo A	Overgo B	BACs
me_UT21	CATGTGGGTCAGTTCACTGCAGTG	GAGAATGGATTGTGCCCACTGCAG	289L16, 370J4
me_UT20	CACCACAAGCACATAGCTCAGTTC	TTGTGCCCACTGCAGTGAAGTGAAG	289L16, 370J4
me_UT18	GTCTGAGGATCAGGATCAAAGAAC	GTCCTCTCCCTAGAATGTTCTTTG	No BACs
me_UT16	CACCCTTGAGATGTGTTTCACAGC	CTTTGTTGTACCTACTGCTGTGAA	334C11, 352G17, 432E9
me_UT19	GTTTGTGGGTCAGTTCACTGCAGT	TTTTGTGCTTGTGCCCACTGCAGT	289L16, 370J4
me_UT14	ATTTCACTGCAGTGGGCACAGCCA	AAGTTCAGCAGGGAGCTGGCTGTG	334C11, 352G17, 432E9
me_UT15	CAAGGTCTCTCAAACCTCATAGTC	CATCCACAGAAGTATGACTATGA	334C11, 352G17, 432E9
me_UT21	TAGGCCCATGAATGATGGATTTTC	CACTAGACAGCAGATAGAAAATCC	289L16, 370J4

Supplementary Table 7. RT-PCR primers for opossum *UT* exons 2 and 3.

Gene	Exon 2/3 primers
mdUT2	5' ACAGCCTGGAGATCCAGTTCA 3' 5' GTTCCAGGACAATCTTCAGAAA 3'
mdUT3	5' ATTCAGTCCACCACAGCCATA 3' 5' TTTTCCTCATCAAGTCAACACAA 3'
mdUT4	5' ACCACAGGCATGTGATTCAAGT 3' 5' GCTTGAATATCCAACAATTTTCTTC 3'
mdUT5	5' ATTCAGTCCACCACAGGCATA 3' 5' TTATTTTCTGATCCCGTCAAC 3'
mdUT6	5' ATTCAGTCCACCACAGACATGA 3' 5' TGATTTTCTTCATCAAGTCAGCA 3'
mdUT8	5' GTTCTGCAGCTCACCACAAAC 3' 5' CATGGTGTCAATACAATATTCTTGC 3'
mdUT9	5' CATTTAGAATACCACTGGAATGAAG 3' 5' ACATCGACACAATATTGCTTAACG 3'
mdUT10	5' AGGCTTGAGGGTGAATTCATTG 3' 5' TGCAGGATTTTCTTATCATTTC 3'
mdUT12	5' ACAGATTACCACAGGCATGAAG 3' 5' TACTGCAGAATTTTCTTCATCA 3'
mdUT13	5' CAATCACAAGCATGAATTGC 3' 5' GAAGGATTTTCTGCATGACTCC 3'
mdUT15	5' CCACAGGCATGACTTCTATTTTC 3' 5' CATGCCATCAATACAATCTTCC 3'
mdUT17	5' CCATCACAGCCATGAGATGTTT 3' 5' TCATTGGCTGGACACAATAGTC 3'

CHAPTER 3

THE *UT* FAMILY OF MHC CLASS I LOCI UNIQUE TO NON-EUTHERIANS MAMMALS HAVE LIMITED POLYMORPHISM AND TISSUE SPECIFIC PATTERNS OF EXPRESSION IN THE OPOSSUM

By Katina V. Krasnec¹, Anthony T. Papenfuss^{2,3,4,5}, and Robert D. Miller¹

¹Center for Evolutionary and Theoretical Immunology, Department of Biology,
University of New Mexico, Albuquerque, New Mexico, United States of America

²Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, 1G
Royal Pde, Parkville 3052, Australia

³Department of Medical Biology, University of Melbourne, Australia

⁴Peter MacCallum Cancer Centre, East Melbourne 3002, Australia

⁵Sir Peter MacCallum Department of Oncology, University of Melbourne, Australia

Conceived and designed experiments: KK AP RM. Performed the experiments: KK.

Analyzed the data: KK RM. Wrote the paper: KK RM

Manuscript formatted for a planned submission to PLoS One

Abstract

Recently a new family of Major Histocompatibility Class I genes was discovered in marsupial and monotreme mammals. These loci were designated the *UT* genes. Seventeen *UT* loci were identified as a tandem array on chromosome 1 of a model marsupial species, the gray short-tailed opossum *Monodelphis domestica*. Of these 17 genes, 12 are transcribed, with most having restricted tissue distribution. Tissues where *ModoUT* genes were most frequently found are the thymus and skin. The *ModoUT* genes have limited polymorphism and little evidence of positive selection. One of the loci, *ModoUT8*, was chosen for further analysis due to its conservation amongst marsupials and generic characteristics. *ModoUT8* transcription is limited to developing $\alpha\beta$ thymocytes, while being absent from mature peripheral $\alpha\beta$ T cells in lymphoid tissues. The overall characteristics and features of *ModoUT* genes including low polymorphism and limited distribution make it likely that the molecules encoded by *ModoUT* genes perform a function other than antigen presentation.

Introduction

The Major Histocompatibility Complex (MHC) class I gene family encodes cell surface proteins that share a unique, well-conserved structure. The typical MHC class I molecule is a heterodimer of an α chain paired with β 2-microglobulin [1]. The α chain contains three extracellular domains (α 1, α 2, and α 3), along with a transmembrane domain and a cytoplasmic tail [2]. The α 1 and α 2 domains form a structure unique to MHC class I proteins by combining to form a groove comprised of β -sheets with two α helices [3]. The α 3 domain is an immunoglobulin superfamily type domain. So far, genes encoding MHC class I α chains have only been found in the gnathostomes [4].

In spite of the conserved domains, molecules sharing the MHC class I structure perform a diverse set of functions. The initial role uncovered for MHC class I proteins was in the presentation of antigenic peptides to cytotoxic, CD8⁺ T cells [5]. Antigenic peptides derived from proteolysis of self or pathogen derived proteins can be bound in the groove created by the α 1 and α 2 domains and "presented" to CD8⁺ T cells by forming a ligand for the T cell receptor (TCR) [6]. Because this was the first role discovered for MHC class I molecules it is usually termed their "classical" role. Peptide presentation is the most ubiquitous role for MHC class I molecules in gnathostomes and is likely their ancestral function [7]. MHC Class I molecules involved in peptide presentation, such as human HLA-A and -B, are usually highly polymorphic and ubiquitously expressed [8].

Some MHC Class I molecules have evolved to present non-peptide antigens to T cells. One example is the CD1 molecules that present glycolipids to T cells [9]. Even more

wide ranging roles for MHC class I molecules have evolved such as the neonatal Fc receptor (FcRn) that binds and transports IgG. FcRn transports IgG across the placenta in some mammals, and is expressed in the mammary tissue in others for depositing IgG in milk [10]. Although maintaining the traditional MHC class I domain structure, FcRn is unable to bind peptides and the groove formed by the $\alpha 1$ and $\alpha 2$ domains is occluded and does play a role in binding IgG [11, 12]. Other examples of divergent MHC class I roles include the HFE molecule that associates with the transferrin receptor, reducing the receptor's affinity to load transferrin molecules bound with iron, and the murine M10 molecule that binds and serves as a chaperone for vomeronasal organ olfactory receptors [13, 14]. In contrast to MHC class I genes encoding molecules that present peptides, those encoding "non-classical" molecules such as M10, FcRn, and CD1 are typically less polymorphic and often have tissue specific patterns of expression.

The gray short-tailed opossum, *Monodelphis domestica*, is among the more established model marsupial species, and the first to have a sequenced genome [15, 16]. Many of the MHC class I genes in the genome of *M. domestica* have been annotated [17-22]. Using traditional homology based *in silico* search methods, eleven genes encoding the class I α chain have been uncovered in the *M. domestica* MHC region, along with two clear MHC class I pseudogenes [17, 19, 20, 23]. There are three presumed peptide-presenting MHC class I genes in *M. domestica*; *ModoUA1*, *UA3*, and *UA4*. *ModoUA1* and *UA3* display many traits common to peptide presenting MHC genes, such as ubiquitous expression throughout tissues and high levels of polymorphism, *ModoUA4* has lower levels of polymorphism but remains ubiquitously expressed [19, 23]. Many of the other known

opossum MHC class I genes encode molecules likely to have non-classical functions. *ModoUG*, for example, can be expressed in three different alternatively spliced mRNA forms [20]. In these forms, a short cytoplasmic tail has been found that does not have traditional phosphorylation sites. *ModoUJ* and *UM* also have several alternative mRNA isoforms, including one possible *ModoUJ* soluble isoform [17]. A majority of the opossum class I genes had limited polymorphism consistent with possible non-classical function.

The functional plasticity of the MHC class I structure makes these molecules good models to study the evolutionary relationship between structure and function. The availability of whole genome sequences from a variety of gnathostome lineages provides the information needed to discover new MHC class I genes and potentially new roles for these molecules. Recently, a new family of MHC class I loci called *UT* was found, and appears to be restricted to marsupials and monotremes [24]. The phylogenetic distribution is consistent with having emerged early in mammalian evolution and subsequently lost in the eutherian lineage. A more in-depth analysis of these genes revealed that some were transcribed in immune tissues of the opossum, tammar wallaby, brushtail possum and Tasmanian devil. The goal of this study was to characterize the *UT* genes in opossum, *M. domestica*, further.

Materials and Methods

Tissue Transcription of *ModoUT*

ModoUT gene transcription was examined using the Illumina generated transcriptome data from 19 adult *M. dometica* tissues, publically available at OpossumBase (<http://opossumbase.org>). Tissues included kidney, liver, lung, spleen, colon, skin (ear pinna), intestine, stomach, diaphragm, heart, skeletal muscle, tongue, pancreas, thyroid, brain, eye, adipose, testes, and ovary. In addition, a thymus transcriptome was generated using the Roche 454 sequencing platform. BLAST searches of the databases were performed using the predicted nucleotide sequence of exons 2, 3 and 4 that encode the regions corresponding to the $\alpha 1$, $\alpha 2$, and $\alpha 3$ domains from all *ModoUT* genes identified previously [24]. Sequences in the assembled RNA databases that were over 180 bp and 98% or greater nucleotide sequence identity to a known *UT* gene were scored as transcripts of that gene. In order to verify the accuracy of the transcription sites, any transcripts of *ModoUT* genes uncovered using BLAST search were aligned to all the previously identified predicted sequences. Alignments were then assembled into a phylogeny using MEGA 5.0 and a maximum-likelihood tree was generated for basic identification of transcripts to a specific *ModoUT* in order to avoid incorrect sites of transcription due to sequence identity between the *ModoUT* genes [25].

Isolation of full-length coding sequences

Two methods were used to isolate full-length coding sequences of transcribed *ModoUT* loci: Rapid Amplification of cDNA Ends (RACE) polymerase chain reactions (PCRs)

and assembly using available transcriptome sequence databases. The latter included OpossumBase and the opossum thymus 454 transcriptome. This approach allowed for assembly of a *ModoUT* gene in the absence of obtaining the full 3' end via RACE PCR, or to assemble without PCR amplification and sequencing. Full-length coding sequences for *ModoUT4*, *5*, *6*, and *8* were isolated entirely using the RACE method, while *ModoUT9* and *10* were obtained using a combination of sequences generated by RACE and from the transcriptome databases. RACE PCR were performed using the SMARTer RACE cDNA Amplification Kit (Clontech, Mountain View, CA) with Advantage HF-2 high-fidelity Taq polymerase (Clontech, Mountain View, CA) following the manufacturer's recommended protocol. PCR parameters for all primers were 94°C for 30 seconds, followed by 72° C for 3 minutes, repeated 5 times, then 94°C for 30 seconds, 70° C for 30 seconds, and 72° C for 3 minutes repeated 5 times, followed by 94°C for 30 seconds, 68° C for 30 seconds, and 72° C for 3 minutes, repeated 25 times. If no PCR products were visible on an agarose gel, a secondary PCR amplification was done by an additional 5 cycles at 94°C for 30 seconds, 68° C for 30 seconds, and 72° C for 3 minutes. The primers were designed within either the predicted exon 2 to amplify the 5' end or exon 3 to amplify the 3' ends [24]. If it was not possible to design both 5' and 3' primers according to guidelines of the amplification kit, the same sequence was used for the forward and reverse primers. The PCRs were performed on cDNA made from total thymus RNA extracted from a 6-week-old male *M. domestica*. The 5' and 3' RACE PCR was performed in a single step for *ModoUT4*, *5*, *9*, and *10* with specific primers (Table 1), or as a nested PCR for *ModoUT6* and *ModoUT8* using additional nested specific primers. PCR products were cloned, and sequenced using BigDye Terminator v3.1 Cycle

Sequencing Kit (Invitrogen, Grand Island, NY). Any sequences obtained that did not span the full length of the 3' end were supplemented using 454 transcripts. The nucleotide sequences from the 5' and 3' ends were then aligned against the MonDom5 opossum genome assembly (GenBank accession number AAFR03000000) to identify the exons present in each sequence [15].

Full-length coding sequences for *ModoUT2*, *3*, *15*, *16*, and *17* were determined using transcripts from tissues transcribed from OpossumBase or from thymus 454. Transcribed *ModoUT* loci were assembled and aligned against the opossum genome in Sequencher 5.0 (Gene Codes, Ann Arbor, MI). In some instances, full-length genes could be found from a single transcript from the transcriptome, while others required multiple transcripts to generate a contig that covered all of the exons. Accession numbers for full-length RNAs encoding the class I α chains of *ModoUT2*, *3*, *4*, *5*, *6*, *8*, *9*, *10*, *15*, *16*, and *17* including alternative mRNA splice variants are KP125495-KP125507.

Polymorphism Analysis

Polymorphism was determined only for the *ModoUT* genes that are transcribed in the thymus. Genomic DNA was isolated from *M. domestica* liver or spleen tissue that was provided by the Southwest Foundation for Biomedical Research (SFBR, San Antonio, TX) and the Museum of Southwestern Biology (MSB), University of New Mexico [17, 20]. DNA was extracted using a standard phenol/chloroform extraction protocol. The SFBR tissues were from captive-bred animals of Brazilian populations 1 and 2 [26]. Tissues from the MSB were from wild-caught *M. domestica* collected from five different

sites in Bolivia. These samples and their origins have been published previously [17, 20]. PCR products using primers listed in Table 1 were amplified using the Advantage HF 2 PCR kit (Clontech, Mountain View, CA), with the following parameters for all primers: 94°C for 1 minute, 33 cycles of 94°C 30 seconds and 62°C for 4 minutes, followed by a 7 minute extension at 68°C.

For loci where intron 2 was less than 400 bp (*ModoUT2*, *UT3*, and *UT5*), exon 2, intron 2, and exon 3 were amplified as a single product with primers placed at the start of exon 2 and end of exon 3. For loci where intron 2 was greater between 800 and 1400 bp (*ModoUT4*, *UT6*, *UT7*, *UT8*, and *UT15*), exon 2, intron 2, and exon 3 were amplified as two overlapping fragments with primers nested within intron 2. Three loci (*ModoUT9*, *UT10*, and *UT17*) had intron 2 greater than 3Kb and exons 2 and 3 were amplified separately.

PCR products were cloned for sequencing using the TOPO TA cloning kit (Life Technologies, Grand Island, NY). A minimum of five clones per individual per locus were isolated and sequenced using the illustra TempliPhi kit (GE Life Sciences, Pittsburgh, PA) and BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies, Grand Island, NY), respectively. Sequences were analyzed and edited using Sequencher 5.0 (Gene Codes, Ann Arbor, MI) and BioEdit [27]. Accession numbers for novel allele sequences reported are: *ModoUT2*, KP174147-KP174156; *ModoUT3*, KP221763-KP221785; *ModoUT4*, KP174157-KP174166; *ModoUT5*, KP221786-KP221799; *ModoUT6*, KP174167-KP174180; *ModoUT7*, KP221744-KP221762; *ModoUT8*,

Table 1: List of gene specific primers sets used for amplification of genomic DNA for PCR and cDNA for RACE for *M. domestica* UT genes

Gene	Exon 2 F	Exon 2 R	Intron 2 F	Intron 2 R	Exon 3 F	Exon 3 R	Length	Overlap
Modb1T2	TGCTCTCAGCTCTCCACA					ATCTTCAGAAAATGTTGGGCACT	689bp	
Modb1T3	ACCACAGCCUATATAGTGCAGTTC					TCAMGTCCACACAAATAGAGCTTCA	829bp	
Modb1T4	CCCTCTATAAATCTAGTCCACCACA		CAATCATGGGGGTTAGAAACAT	TCGCCAAAATCTTAGTGCAGTTC		TCACACAAATTTCTTCAATCAAGTTC	812bp	
Modb1T5	CTATAATTCAGTCCACACAGAG					CTGACCATATATCCCTTATGCTTC	742bp	207bp
Modb1T6	CCCTCTATAAATTCAGTCCACCACA		GTCCCATGCTCTTTTTCAGTTC	TCTCAGTGTATGGCCATGTTGCC		TTCCCTTAAGTTTGGAAATACCCAAATG	859bp	
Modb1T7	CACACTATAACTTCATGTAAGTGTTC		GCTGCTTGGAGGAGTAAAGGAT	GACTTTCAGAGTTTGGCCAAAAGG		CCTATAAATTCAGTCCACCACAGAGG	794bp	248bp
Modb1T8	GTCTGGAGCTCAGCCACAAC		CTCAATCCACCAGTGTGTGTTTC	GGGTAGGAGGCTTCATGTCAGTTC		CTTGGCATGTAGTGGCTTCTCTTG	753bp	154bp
Modb1T9	CCCATATATGACCTGCTTCTCT	GTGGAGGTGGAAATAGGTGAGG			TTAGCTGGGCCATAATACTTCC	GGGMAAATATGACTGTGTACTCTCTG	788bp	98bp
Modb1T10	CCCTACCTTTCGTGCAATCAG	GGACAAAGGCTTTTTCAGGATATG			TTCAATTCGAAATCTGATTCGA	GATTTGGTAGAGGCTTGGTCTTT	604bp	
Modb1T15	GETTTCACCTCTGCTCTCAGC		ACCTCGAAGGTTCATCTAGTCCA	TGTCCTTTTGAAACTTAGGGCAAT		TCATGTTGAAATGCTGTGAGAAC	853bp	138bp
Modb1T17	CTTTCGTGCTTTAGGCCATC	CCCTCTCTATCATTTTATACC			CCAGAGAACCAATACAGTCCAGAT	TGTTTGGACTGTAGGGACTTCTCTC	870bp	

Gene	5' RACE R	5' RACE R nested	3' RACE F	3' RACE F nested	overlap
Modb1T4	TTCAATGTTCTCCAGTGGCCCAACC		GAAACCACACAGCCCAAGCTCCAG		89bp
Modb1T5	TGTGCCAACTCCAGTCAACTGTCCC		GGGACAGTTGACTGTGAGTTGGACACA		0bp
Modb1T6	AGACTGTGTCCAGAGGGGAGCTCTC		CTGGAGAGTCCCTCTGTGGACCACTCT		0bp
Modb1T8	GTGTGCCACAGGTTACTTGTCA	TGGATTCATTTCTAGCTGTT	GGAAAGTCCCTCTTGGAAATACT	TCATAGAGGACACTTGGCCAGAA	303 bp 119 bp

Modo, *Monodelphis domestica*, RACE, rapid amplification of cDNA end.

KP174181-KP174197; *ModoUT9*, KP245840, KP245841, KP245858-KP245860; *ModoUT10*, KP245842-KP245847; *ModoUT15*, KP174198-KP174203; and *ModoUT17*, KP245848-KP245857.

Evidence of selection on the *ModoUT* loci was determined using a maximum-likelihood estimation of the dN/dS ratio using the CODEML module of the Phylogenetic Analysis by Maximum Likelihood (PAML) package [28]. The regions encoding the $\alpha 1$ and $\alpha 2$ domains of all *ModoUT* alleles obtained from the polymorphism study were analyzed for evidence of positive selection by looking at non-synonymous/synonymous substitution ratios. This was done using an initial Maximum Likelihood phylogenetic tree of sequences which was then analyzed with four site-specific models, M1a – nearly neutral, M2a – positive selection, M7 - β , and M8 - β , and ω , in CODEML. A Bayes Empirical Bayes analysis was performed on each model to infer the location of the codons under selection and the ω values for each of these sites [29].

Fluorescent In-situ Hybridization (FISH)

Thymus and spleen tissue from six-week old opossums were collected for cell suspensions. The tissues were manually separated into single cell suspensions that were washed and resuspended in RPMI 1640 media with HEPES (Sigma-Aldrich, St. Louis, MO). Erythrocytes were removed by Ficoll gradient centrifugation. Lymphocytes were resuspended in 1X PBS (Sigma-Aldrich, St. Louis, MO) and counted on a hemocytometer. Cells at a concentration of 2×10^6 /ml were deposited on non-coated Shandon Cytoslides using a Cytospin (Thermo-Fisher Scientific, Waltham, MA) and used

immediately or stored frozen at -80°C . Custom Stellaris FISH probes (Biosearch Technologies, Petaluma, CA) were designed using the constant regions of T cell receptors (TCRs) and untranslated region (UTRs) for T cell identification. The slides were probed and hybridized with a set of 48 singly labeled oligonucleotide probes, each 20 nucleotides in length that were designed from the full-length sequences of *ModoUT8*. A Quasar® 570 fluorophore was used for *ModoUT8* and a Quasar® 670 fluorophore for TCR α or TCR β , following the Stellaris FISH protocol for frozen tissue (Stellaris, Biosearch Technologies, Petaluma, CA)[30]. Slides were then stained using wash buffer consisting of 5 ng/mL of DAPI dilactate (Sigma-Aldrich, St. Louis, MO) and mounted using Vectrashield® Mounting Medium (Vector Labs, Burlingame, CA). Slides were imaged on a Nikon Ti Eclipse inverted fluorescent microscope using the 100x oil immersion lens. Z-stacks images of each fluorophores were merged together and edited using NIS-Elements Imaging Software (Nikon Instruments, Melville, NY).

Results

Analysis of ModoUT Transcription

To investigate transcription of *ModoUT* genes in a variety of tissues, public transcriptome data from the OpossumBase public database was downloaded and searched. This database contains transcriptome profiles of 19 different tissue types. Since OpossumBase lacked information on the thymus, a good candidate tissue for immune gene expression, a separate thymus transcriptome database was generated using the Roche 454 platform.

The initial analyses focused on the diversity of *ModoUT* loci transcribed in each tissue type. Of the 17 *ModoUT* genes, the most common sites for transcription are the thymus and the skin (Figure 1). Twelve out of 17 *UT* loci are transcribed in the opossum thymus, and eleven out of 17 are transcribed in the skin. For all other adult tissues, the colon and thyroid contained the greatest variety of transcribed *ModoUT* genes with four each. In contrast, the testes, eye, lung, diaphragm, skeletal muscle, pancreas, stomach, and heart each only transcribe a single *ModoUT* locus. Brain and kidney were the only adult tissues investigated that lacked any *ModoUT* transcripts.

Next the breadth of transcription of each *ModoUT* locus among the various tissues was examined. Most of the *ModoUT* genes demonstrated restricted tissue transcription patterns. *ModoUT1*, *11*, and *14* were each found transcribed in only a single tissue, the testes, liver, and skin, respectively. Similarly, *ModoUT2*, *6*, and *16* were only transcribed

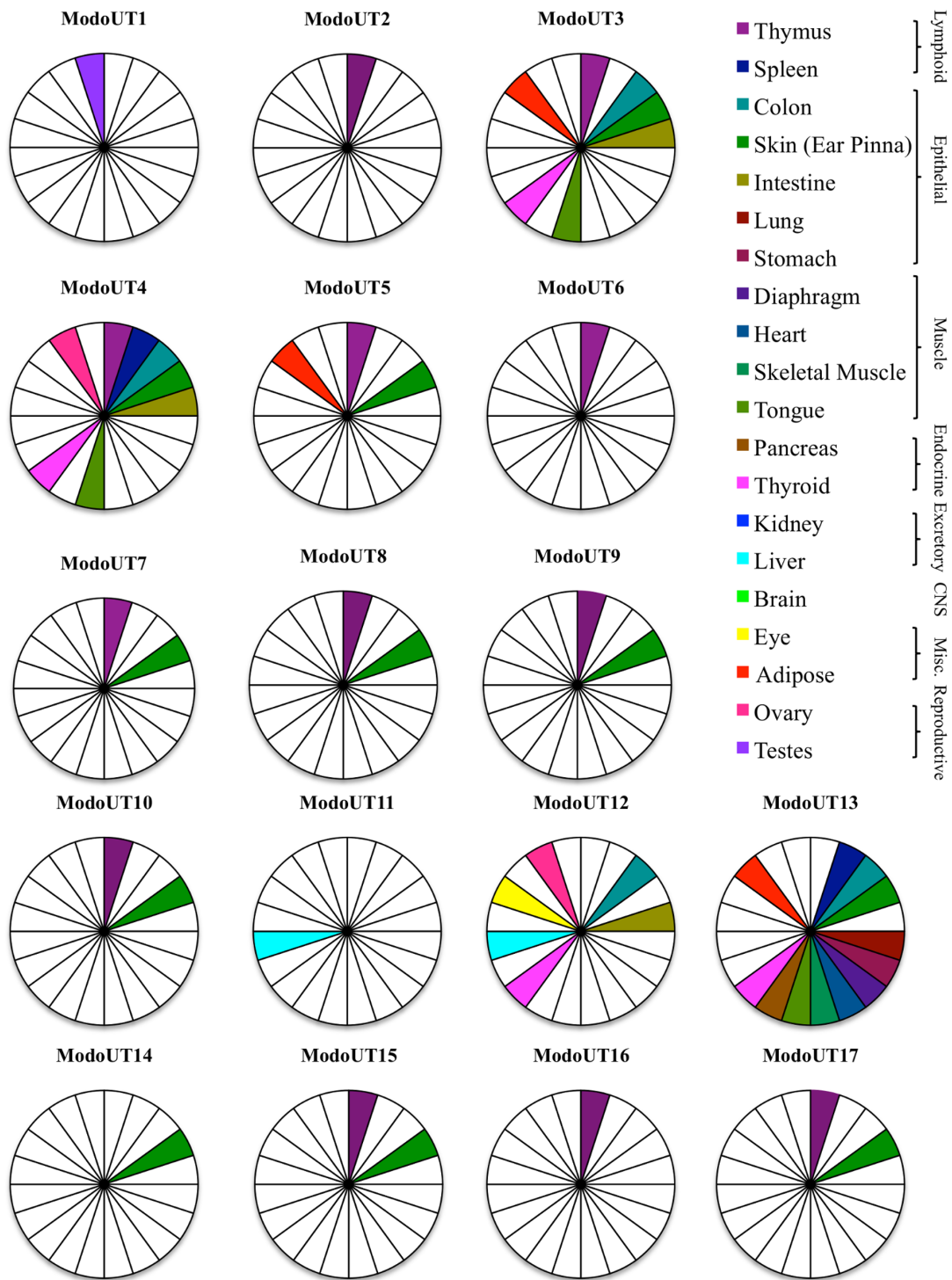


Figure 1: Differential expression of *ModoUT* genes across 20 different tissue types. Filled pie wedges represent the detection of a *ModoUT* transcript in either a Roche 454 dataset or from Illumina sequences found at www.OpossumBase.org with a 99% sequence identity at 180 or more base pairs. Tissue types examined included lymphoid, epithelial, muscle, endocrine, excretory, reproductive, CNS, and miscellaneous that included the eye and adipose.

in the thymus. In contrast, *ModoUT13* was found most broadly transcribed in 16 different tissues (Figure 1). Of the 17 known *UT* loci in the opossum genome, only *ModoUT7* is thought to be a pseudogene due to an in-frame stop codon near the beginning of exon 3. In spite of being a putative pseudogene, it is transcribed in both thymus and skin.

ModoUT Gene Organization

To determine the complete structure of *ModoUT* genes, a combination of RACE PCR and the sequence from the Illumina and Roche 454 databases was utilized. Sequences from all sources were aligned to the opossum whole genome and contigs representing full-length mRNA sequences for the twelve *ModoUT* genes transcribed in the thymus, allowing for the determination of gene boundaries and alternative splice variants (Figure 2). The size range of the full-length *ModoUT* genes ranged from 5,453bp (*ModoUT6*) to 16,053bp (*ModoUT9*). *ModoUT2*, 4, 6, and 16 have similar lengths (5,453-7,484bp) while *ModoUT3*, 5, 8, 9, 10, 15, and 17 are larger in size from 9,672bp to 16,053. As is typical of MHC class I loci, the number of exons per locus varied between five and seven (Figure 2). The variation in gene size does not correlate with phylogenetic relationship based on coding sequence, as highly related groups like *UT3*, *UT4*, *UT5*, *UT6*, and *UT7* range from 5,453bp to 12,219bp [24].

Previous work in *M. domestica* has revealed that a majority of the MHC class I loci within the MHC region were transcribed to generate alternative splice variants with open reading frames (ORFs) [17, 20]. Only *ModoUT2* and *UT16* generated variants with

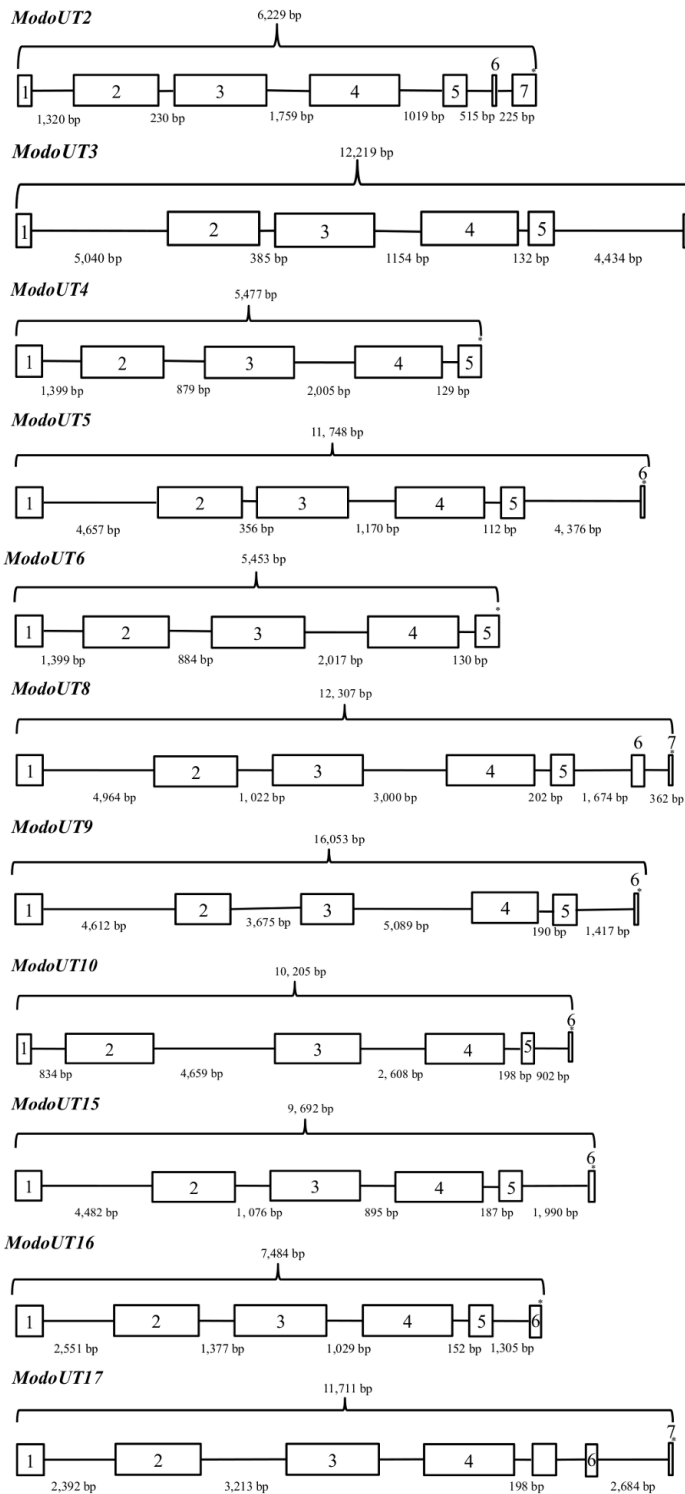


Figure 2: Gene organization and exon composition of *ModouT2*, *UT3*, *UT4*, *UT5*, *UT6*, *UT8*, *UT9*, *UT10*, *UT15*, *UT16*, and *UT17* mRNAs. Asterisks indicate relative location of stop codons.

ORFs other than the full length (Figure 3). The *ModoUT2* splice variant would encode a truncated $\alpha 3$ domain due to using an alternative splice site within exon 4, whereas the *ModoUT16* variant would be translated without an $\alpha 3$ domain due to splicing out exon 4 altogether.

Polymorphism and Evidence of Positive Selection

Levels of polymorphism in the *ModoUT* genes were investigated using both captive bred animals from Brazilian stock as well as wild caught Bolivian animals. MHC class I genes typically have the highest level of polymorphism in exons 2 and 3, which encode the $\alpha 1$ and $\alpha 2$ domains, respectively. Genomic DNA was amplified for exon 2, intron 2, and exon 3 from eight captive bred Brazilian *M. domestica* and 19 wild caught individuals from Bolivia. Alleles were counted for all nucleotide variants, whether in coding or non-coding regions of the gene. The number of nucleotide alleles for each of the *UT*'s varied substantially in comparison with previous polymorphism studies on MHC Class I genes in *M. domestica*, as they had consistently greater numbers of genomic alleles found [17, 20] The lowest number of alleles were found in *ModoUT9* and *ModoUT10*, with three and four alleles respectively (Supplemental Table 1). *ModoUT3* and *ModoUT7* had the greatest number of alleles, with 23 and 19, but much of the variation was found in the intron. No individual was found with more than two alleles at any locus lending validity to the genotyping. The majority of the alleles for each locus were unique to either the Brazilian or Bolivian populations, with the exception of one

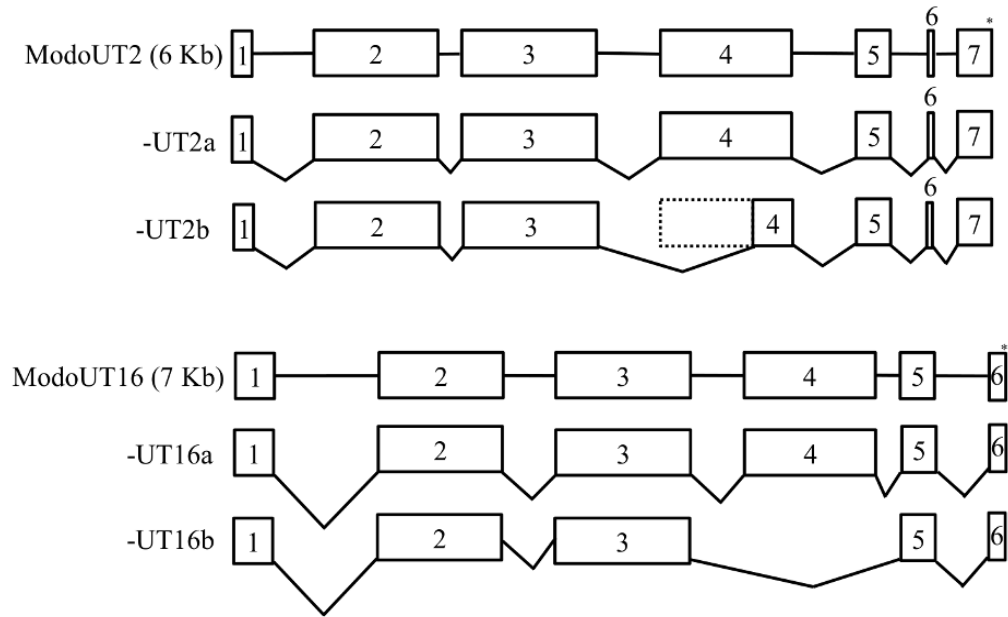


Figure 3: *ModoUT2* and *ModoUT16* are transcribed into alternatively spliced isoforms that differ with the loss or truncation of exon 4. Asterisks indicate relative location of stop codons.

allele in both *ModoUT7* and *ModoUT10* that were found in both populations. When examining alleles at the protein level, the numbers dropped 3 to 4-fold, indicating that a majority of the nucleotide diversity at each of the loci was found in intron 2 or were synonymous mutations in exons. The number of protein alleles ranged between one and six for exons 2 and 3, with an average of 3.4 alleles for exon 2 and three for exon 3. Most of the polymorphism found in *ModoUT* genes was limited to one to five amino acid changes between alleles. The exception to this is *ModoUT5*. While some of the alleles found in *ModoUT5* do not differ substantially to the sequence found in the reference genome, ten of the alleles contain an additional insertion of three nucleotides in the intron, and another allele, *ModoUT5.08*, has a 26 base pair long insertion in the intron (not shown). No insertions were seen in the exons of the *ModoUT5* alleles, but substantial polymorphism between the alleles was seen. A large 23 base pair insertion and two base pair insertion were also seen in intron 2 of two *ModoUT7* alleles. Premature stop codons were found in one allele of *ModoUT5* in exon 2 and two alleles in exon 3, as well as all alleles of exon 3 in *ModoUT7* and one allele of *ModoUT17*.

Heterozygosity was common across the UT genes and most frequently seen in the wild-caught Bolivian individuals. Of the captive Brazilian animals, those from Population 2 were homozygous in two-thirds of the *ModoUT* genes analyzed for polymorphism. The high percentage of homozygosity in captive-bred is likely the result of some partial inbreeding [26]. Only four and six animals respectively were found to be homozygous at *ModoUT3* and *ModoUT5*, while *ModoUT9*, *ModoUT10*, and *ModoUT15* had twenty or more animals that were homozygous. For most of the *ModoUT* genes, approximately half

of the animals were heterozygous. Animals collected from the Porvenir region of Bolivia displayed the greatest amount of heterozygosity, with the highest numbers in *ModoUT3*, *ModoUT5*, *ModoUT6*, *ModoUT7*, and *ModoUT8*.

Overall, there is amino acid conservation across the different *ModoUT* genes. The conservation can be seen across the $\alpha 1$ domain, at a leucine in position 54 and glutamic acid in position 71 are shared across all alleles (Supplemental Figure 1). Similar levels of conservation are seen across all alleles in the $\alpha 2$ domain. A shared threonine and glutamine are seen at positions 7 and 9, a cysteine at position 14, a glycine at position 32, and a proline at position 50. (Supplemental Figure 2) A comparison of the number of hydrophobic amino acids using all the *ModoUT* alleles to CD1 and other MHC class I genes was made. Analysis reveals nearly double the number of hydrophobic amino acids across the *ModoUT* $\alpha 1$ and $\alpha 2$ domains with 41 sites, when compared to classical MHC class I from both eutherians and marsupials, which had approximately 21. The number of hydrophobic sites in *ModoUT* genes was comparable to those of CD1 from both eutherians and marsupials.

Positive selection was inferred for the $\alpha 1$ and $\alpha 2$ domains of all *ModoUT* alleles individually with the Pairwise Analysis of Maximum Likelihood method to determine which sites were under selection [28]. Only *ModoUT5* had more than a single site with evidence of being under positive selection, and these were all in the $\alpha 1$ domain (Figure 4). *ModoUT6* had a single site in the $\alpha 2$ domain that appeared to be under selection (not

shown). All other *ModoUT* loci revealed no evidence of positive selection acting on their evolution.

Expression of ModoUT8 in thymic but not peripheral $\alpha\beta$ T cells

ModoUT8 was chosen as a model *UT* for further expression analysis as it was the first *ModoUT* uncovered, it demonstrates a limited transcription pattern like most opossum *UT* genes, it has clear orthologues in other marsupial species, and has the greatest structural similarity of all *ModoUT* molecules to the chicken MHC class I protein [24]. To detect mRNA transcription of *ModoUT8* in specific cell lineages, fluorescent in-situ hybridization (FISH) was performed utilizing opossum thymus and spleen. *ModoUT8* and T cell receptor β chain (TCR β) mRNA were imaged simultaneously, allowing the visualization of the location of each of the *ModoUT8* or TCR mRNA molecules within a cell. In the thymus, *ModoUT8* co-localized with TCR β (Figure 5) and TCR α (not shown), demonstrating that it is transcribed in $\alpha\beta$ thymocytes. Consistent with the transcriptome databases (Figure 1) and RT-PCR experiments (not shown), *ModoUT8* transcripts were not detected in mature, splenic $\alpha\beta$ T cells (Figure 5).

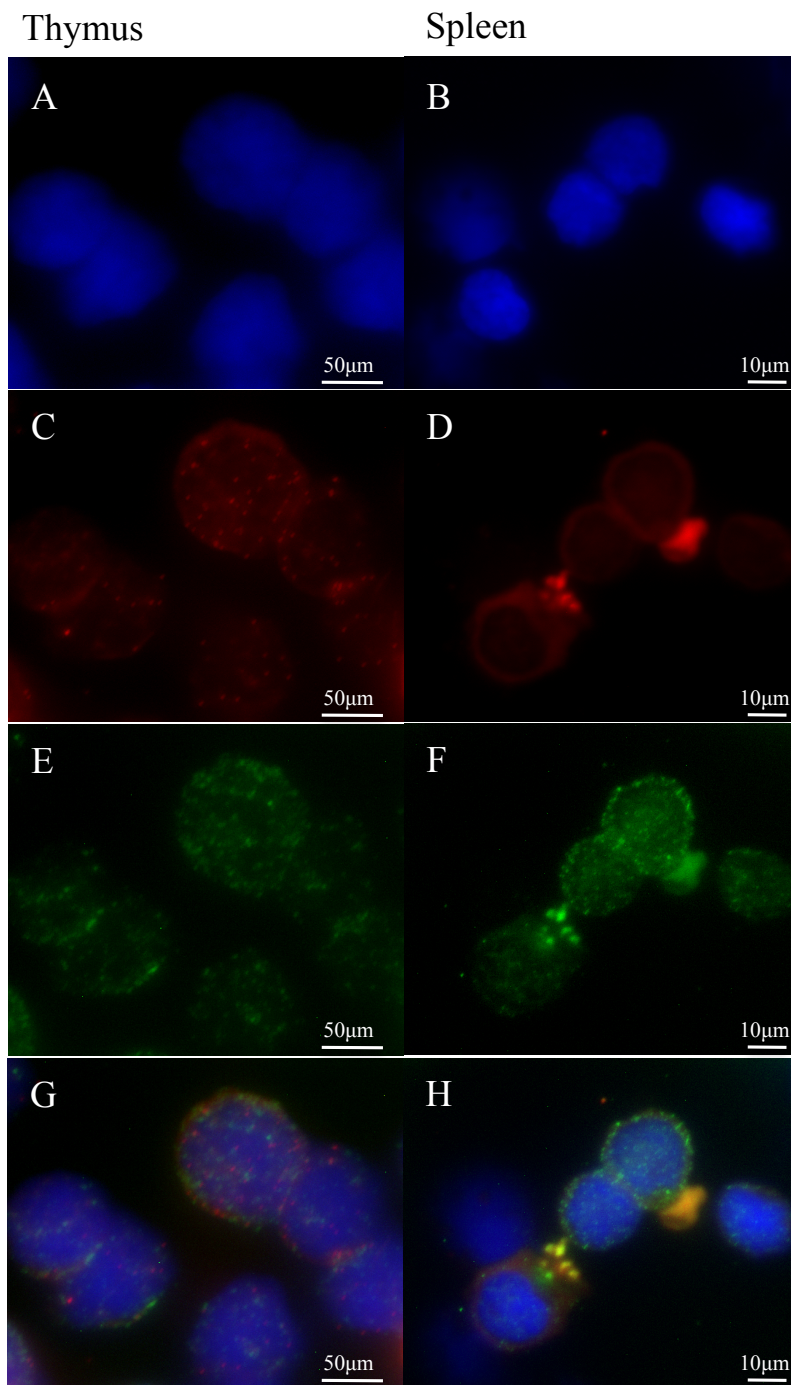


Figure 5: Fluorescent *in-situ* hybridizations of *ModoUT8* with TCR α or TCR β in the spleen and thymus. **A&B:** DAPI nuclear staining in the spleen and thymus allows visualization of lymphocytes. **C&D:** *ModoUT8* probes hybridizes to the thymus but not the spleen. Red points indicate evidence of transcription in the cell. Large areas of fluorescence are background and associated with binding of probes to dead cells. **E&F:** TCR β probes bind to cells in both the spleen and thymus. **G&H:** Merged images of DAPI, *ModoUT8*, and TCR β give no evidence of *ModoUT8* transcription in the spleen, while *ModoUT8* co-localized with TCR β in the thymus.

Discussion

An unexpected result to emerge from the analyses of marsupial and monotreme genomes was the presence of immune related genes that are absent from eutherian genomes. One example is the *UT* family of MHC class I genes [24]. Another was a non-conventional TCR locus, *TCR μ* [31]. Both *UT* and *TCR μ* genes are present in marsupial and monotreme genomes, consistent with their presence in the last common ancestor of all living mammals [24, 32-34]. Neither of these gene families has been found in any eutherian mammal or non-mammalian genome to date, consistent with their being unique to mammals but lost in the eutherians [24, 32]. Why genes such as the *UT* family or *TCR μ* have been lost in the eutherian lineage is a matter of speculation and can only be understood when their function in the species where they are found is determined. Towards this end, this study was convened to further characterize the *UT* family in a model marsupial.

The opossum *UT* family of genes can be generally described as having limited tissue expression, low levels of polymorphism, and little evidence for positive selection. *ModoUT5* is an exception in that it had several sites under selection in the $\alpha 1$ domain. These characteristics are usually indicative of MHC Class I molecules that have non-classical roles. The genes that encode the *ModoUT* molecules do not have the ubiquitous pattern of expression like the polymorphic *ModoUAI* locus in *M. domestica*. *ModoUAI*, along with its nearest related genes *ModoUA3* and *UA4*, is likely involved in peptide presenting functions based on its similarity to classical MHC class I loci. Thirteen of the

ModoUT genes are transcribed in three or less tissue types, indicative of a restricted transcription pattern. It is noteworthy that two of the *ModoUT* genes, *ModoUT12* and *ModoUT13*, have six or more sites of transcription, but do not have transcription in the most common tissue, the thymus. This may indicate that the *ModoUT12* and *ModoUT13* loci may be playing a different role than the *UT* loci that are found transcribed in the thymus and skin. The majority of *ModoUT* genes are found in the thymus and skin, and thus may have other roles than those found across more numerous tissues. All *ModoUT* loci do have numerous hydrophobic residues in the binding groove as well as structurally resembling the chicken B21 MHC class I molecules, which binds peptides promiscuously [24]. The hydrophobic residues as well as the possibility of non-specific peptide binding due to structure suggest they may be able to present small peptides within the relatively closed binding groove, or they bind hydrophobic antigens such as lipids. A more specific role for *ModoUT* in the thymus may be for T cell selection and maturation. Alternatively, *ModoUT* genes may also just be ectopically expressed in the thymus for negative selection against self-molecules and not serve in T cell selection.

The role of *ModoUT* in skin is less clear. The high levels of *ModoUT* expression in the skin may be the result of the *UT* molecules playing a role in innate immunity. The skin serves as not only a barrier against many pathogens, but both cells within the skin and secreted molecules can protect against pathogens. Potentially, *ModoUT* molecules may also play a protective role in development. The altricial birth and development of marsupials and monotremes can leave the developing young immunologically immature and susceptible to pathogens. The presence of the *UT* molecules on the surface of the skin

of undeveloped young may help to provide protective benefits in a pathogen-rich environment. Close examination of the other epithelial tissues beyond skin did not find a correlation between epithelial tissues and transcription of *ModoUT*'s. Therefore, the abundance of expression of *ModoUT* genes in the skin is likely not the result of high levels of *ModoUT* transcription of the genes in epithelial cells, but rather, *ModoUT* transcription in skin cells only for potentially serving a protective role after altricial birth.

Except for *ModoUT2* and *UT16*, there was no evidence for alternative splice variants generated by the *ModoUT* loci. This is in contrast to the non-classical MHC class I loci encoded within the opossum MHC where alternative splicing was a common characteristic [17, 20]. Unfortunately, this result provides little insight into the function of the *ModoUT* gene products other than they are likely to be expressed in a form with all three extracellular α -domains and there does not appear to be secreted or soluble forms of the molecules.

Minimal polymorphism and little evidence for positive selection on the *ModoUT* genes follows a similar pattern as other *M. domestica* non-classical MHC class I genes. The lack of polymorphism is consistent with *ModoUT* molecules likely having a function other than peptide presentation. In examination of the *in-situ* hybridizations, *ModoUT8* cannot be found in the spleen, but does co-localize with TCR α and TCR β in the thymus. This suggests that perhaps the mature α/β T cells found in lymphatic tissues like the spleen do not express *ModoUT8*, but *ModoUT8* is transcribed in immature α/β T cells in the thymus, perhaps indicating that it may play a role in T cell maturation.

The sequencing of the $\alpha 1$ and $\alpha 2$ domains of the *UT* genes revealed a high degree of hydrophobic residues, nearly double that of other marsupial and eutherian MHC class I, but comparable to CD1 [24]. This characteristic remained intact with the analysis of additional *UT* alleles. Some instances of polymorphism did occur at hydrophobic sites, but the majority of the hydrophobic sites were highly conserved across all of the *UT* loci. This likely indicates a need for conservation of the hydrophobic sites, potentially serving as invariant sites for binding lipids.

The characterization of the *UT* genes in *M. domestica* contributes new insights into genes that are being uncovered using novel methods, but also provides greater knowledge into the evolution and development of marsupials and their immune system and how they differ from eutherian mammals. Future work done on *ModoUT* genes will possibly allow for the identification of the function of this large family of genes and the immune or non-immune role they may play in marsupials.

References

1. Yeager, M. and Hughes, A.L. (1999) Evolution of mammalian MHC: natural selection, recombination, and convergent evolution. *Immunol. Rev.* 167:45-58
2. Bjorkman, P. and Parham, P. (1990) Structure, Function, and Diversity of Class I Major Histocompatibility Complex Molecules. *Annu. Rev. Biochem.* 59:253-288.
3. Bjorkman, P.J, Saper, M.A., Samraoui, B., Bennett, W.S., Stromingers, J., and Wiley, D.C. (1987) Structure of the human class I histocompatibility antigen, HLA-A 2. *Nature.* 329(8):506-512.
4. Kelley, J., Walter, L., and Trowsdale, J. (2005) Comparative genomics of the major Histocompatibility complexes. *Immunogenetics.* 56:683-695.
5. Schwartz, R.H. (1985) T-Lymphocyte Recognition of Antigen in Associate with Gene Products of the Major Histocompatibility Complex. *Annu Rev Immunol.* 3:237-261.
6. Bjorkman, P.J, Saper, M.A., Samraoui, B., Bennett, W., Stromingers, J., and Wiley, D.C. (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature.* 329(8):512-518.
7. Kasahara M. (1997) New insights into the genomic organization and origin of the major histocompatibility complex: Role of chromosomal (genome) duplication in the emergence of the adaptive immune system. *Hereditas.* 127:59-65.
8. Klein, J., Sato, A., and Nikolaidis, N. (2007) MHC, TSP, and the Origin of Species: From Immunogenetics to Evolutionary Genetics. *Annu. Rev. Genet.* 41:281-304.
9. Beckman, E.M. Porcelli, S.A., Morita, C.T., Behar, S.M., Furlong, S.T. and Brenner, M.B. (1994) Recognition of a lipid antigen by CD1-restricted alpha beta+ T cells. *Nature.* 372(6507):691-4694.
10. Story, C.M., Mikulska, J.E., and Simister, N.E. (1994) A Major Histocompatibility Complex Class I-like Fc Receptor Cloned from Human Placenta: Possible Role in Transfer of Immunoglobulin G from Mother to Fetus. *J. Exp. Med.* 180: 2377-2381.
11. Ghetie, V., and Ward, E.S. (1997) FcRn: the MHC class I-related receptor that is more than an IgG transporter. *Immunology Today.* 18(12):592-598.

12. Ghetie, V., and Ward, E.S. (2000) Multiple roles for the major histocompatibility complex class I-related receptor FcRn. *Annual Review of Immunology*. **18**(1):739-766.
13. Dulac, C. and Torello, A. (2003) Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nat. Rev. Neurosci.* **4**(7):551-562.
14. Feder J.N. et al. (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature Genetics*. **13**(4):399-408.
15. Mikkelsen et al. (2007) Genome of the marsupial *Monodelphis domestica* reveals innovations in non-coding sequences. *Nature*. **447**(7141):167-177.
16. Samollow, P. (2006) Status and application of genomic resources for the gray, short-tailed opossum, *Monodelphis domestica*, an American marsupial model for comparative biology. *Australian Journal of Zoology*. **53**(3):173-196.
17. Baker, M.L., Melman, S.D., Huntley, J. and Miller, R.D. (2009) Evolution of the opossum MHC: Evidence for diverse alternative splice patterns and low polymorphism among class I genes. *Immunology*. **128**:e418-e431.
18. Baker, M.L. and Miller, R.D. (2007) Evolution of mammalian CD1: marsupial CD1 is not orthologous to eutherian isoforms and is a pseudogene in the opossum *Monodelphis domestica*. *Immunology*. **121**(1):113-121.
19. Belov, K., Deakin, J.E., Papenfuss, A.T., Baker, M.L., Melman, S.D., Siddle, H.V., Gouin, N., Goode, D.L., Sargeant, T.J., Robinson, M.D., Wakefield, M.J., Mahony, S., Cross, J.G.R., Benos, P.V., Samollow, P.B., Speed, T.P., Graves, J.A.M. and Miller, R.D. (2006) Reconstructing an ancestral mammalian immune supercomplex from a marsupial MHC. *PLoS Biology*. **4**(3):e46.
20. Gouin, N., Wright, A.M., Miska, K.B., Parra, Z.E., Samollow, P.B., Baker, M.L. and Miller, R.D. (2006) Modo-UG, a marsupial nonclassical MHC class I locus. *Immunogenetics*. **58**:396-406.
21. Miska, K.B., Wright, A.M., Lundgren, R., Sasaka-McClees, R., Osterman, A., Gale, J.M. and Miller R.D. (2004). Analysis of a marsupial MHC region containing two recently duplicated class I loci. *Mamm. Genome*. **15**:851-864.
22. Miska, K.B. and Miller, R.D. (1999). Marsupial MHC class I: classical sequences from the opossum, *Monodelphis domestica*. *Immunogenetics*. **50**:89-93.
23. Krasnec, K.V., Sharp, A.R., Williams, T.L., and Miller, R.D. (2014) The MHC class I region of the opossum *Monodelphis domestica* revisited. Manuscript in preparation

24. Papenfuss, A.T., Feng, Z.P., Krasnec, K.V., Deakin, J.E., Davoren, C.J., Baker, M.L., and Miller, R.D. (2014) Marsupials and monotremes possess a novel family of MHC class I genes that is lost from the eutherian lineage. Manuscript in preparation.
25. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetic Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* **28**(10):2731-2739.
26. VandeBerg, J.L. and Robinson, E.S. (1997) The Laboratory Opossum (*Monodelphis domestica*) in Biomedical Research In Saunders, N.R. and Hinds, L.A. (Eds.) *Marsupial Biology: Recent Research, New Perspectives* (Ch 14, pp 238-253) University of New South Wales Press.
27. Hall, TA. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* **41**:95–98.
28. Yang, Z. (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**(8):1586-1591.
29. Yang, Z., Wong, W.S.W., and Nielsen, R. (2005) Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Mol. Biol. Evol.* **22**(4):1107-1118.
30. Raj, A., Van Dan Bogaard, P., Rifkin, S.A., Van Oudenaarden, A., and Tyagi, S. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Meth.* **5**(10):877-879.
31. Parra, Z.E., Baker, M.L., Schwartz, R.S., Deakin, J.E., Lindblad-Toh, K., and Miller, R.D. (2007) A unique T cell receptor discovered in marsupials. *Proc. Natl. Acad. Sci. USA.* **104**:9776-9781.
32. Parra, Z.E., Baker, M.L., Hathaway, J., Lopez, A.M., Trujillo, J., Sharp, A., and Miller, R.D. (2008) Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica*. *BMC Genomics.* **9**:111.
33. Parra, Z.E., Mitchell, K., Dalloul, R.A. and Miller, R.D. (2012). A Second TCR δ Locus in Galliformes Uses Antibody-like V Domains: Insight in the evolution of TCR δ and TCR μ Genes in Tetrapods. *Journal of Immunology.* **188**:3912-3919.
34. Wang, X., Parra, Z.E., and Miller, R.D. (2011) Platypus TCR μ Provides Insight into the Origins and Evolution of a Uniquely Mammalian TCR Locus. *J. Immunol.* **187**:5246-5254.

Supplementary Table 1: Number of genomic alleles found across *ModoUT* loci in different *M. domestica* populations. The numbers of alleles varies between 2 and 23, but overall low levels of polymorphism were found.

Animal ¹	Source ²	UT2	UT3	UT4	UT5	UT6	UT7	UT8	UT9	UT10	UT15	UT17
C7585	1	01/01	01/02	01/01	01/01	01/01	01/02	01/02	01/01	01/01	01/01	01/01
C6076	1	01/02	03/04	01/02	01/02	01/01	03/03	03/04	02/02	01/02	01/02	01/05
D0214	1	01/01	03/05	01/01	01/01	01/01	01/01	03/03	01/01	01/01	01/01	01/01
D3888	1	03/04	03/06	03/04	03/04	02/03	n.d.	05/06	02/02	03/05	02/02	02/06
D4368	1	02/02	03/07	05/05	01/02	01/02	04/05	05/06	02/02	02/03	03/04	01/11*
D2718	1	01/01	03/08	01/01	01/05	01/01	01/01	03/03	01/01	01/01	01/01	02/02
D2920	2	01/02	03/03	02/02	01/04	04/04	06/06	04/04	02/02	03/03	02/02	02/06
D5468	2	02/02	03/09	05/05	01/04	05/05	07/07	06/06	02/02	03/03	05/05	03/03
23169	Brecha Tres	01/05	10/11	06/07	06/06	06/06	08/08	07/08	03/03	02/02	06/06	04/07
21696	Rio Limon	06/06	n.d.	07/07	06/07	07/08	09/10	04/09	03/03	04/04	06/06	04/04
12653	Porvenir	06/06	10/10	07/07	06/07	06/07	09/11	08/08	03/03	02/02	06/06	04/04
12664	Porvenir	06/06	10/12	07/08	06/06	08/09	08/09	10/11	03/03	02/02	06/06	04/04
23170	Brecha Tres	06/06	12/13	07/07	08/09	08/08	12/12	08/08	03/03	02/02	06/06	04/04
12663	Porvenir	06/07	10/12	08/09	06/09	07/08	10/13	08/09	03/03	02/02	06/06	04/08
12668	Porvenir	06/08	12/14	07/09	08/10	06/09	14/15	08/12	03/03	02/02	06/06	04/04
12669	Porvenir	06/09	n.d.	n.d.	06/07	06/06	15/16	10/13	03/03	01/02	06/06	08/09
221692	S.R. de A.	08/10	15/15	06/07	06/08	10/11	17/17	08/08	03/03	02/02	06/06	04/08
12670	Porvenir	06/08	11/16	07/10	06/06	10/10	11/13	08/09	03/03	01/02	06/06	04/08
12538	Tita	06/08	17/17	07/07	06/06	07/07	09/09	08/08	03/03	01/02	06/06	04/08
12350	Santiago de Chiquitos	n.d.	15/18	07/07	08/11	08/12	11/18	14/15	03/03	02/02	06/06	04/10
12571	Porvenir	n.d.	10/19	07/07	06/12	06/06	15/19	07/16	03/03	02/02	06/06	04/04
12555	Porvenir	06/06	10/12	07/07	06/12	09/13	12/17	08/16	03/03	02/02	06/06	04/04
12622	Porvenir	07/07	10/20	07/09	06/09	06/08	15/20	08/09	03/03	02/02	06/06	04/04
12580	Porvenir	06/06	12/21	07/07	06/13	06/08	11/15	16/16	03/03	02/02	06/06	04/04
12578	Porvenir	n.d.	18/22	07/09	06/13	06/09	14/21	08/17	03/03	02/02	06/06	04/04
12579	Porvenir	06/08	15/21	07/07	06/09	08/09	09/15	10/16	03/03	02/02	06/06	04/04
12623	Porvenir	n.d.	15/23	07/09	06/14	09/14	17/22	08/08	03/03	01/02	06/06	04/04

¹ Identifier numbers for individual *M. domestica* of Brazilian origin from the Southwest Foundation for Biomedical Research (SFBR) breeding colony or for wild-caught Bolivian animals in the frozen tissue collections at the Division of Genomic Resources, Museum of Southwestern Biology (MSB), University of New Mexico, respectively.

²Source refers to the Brazilian SFBR breeding population (Population 1 or 2) from which animal is descended from or the site of local collection of wild Bolivian *M. domestica* from the MSB collection. n.d. indicates animals whose genotype was not determined at the indicated loci, * an unconfirmed but likely allele.

α1 domain

```

10      20      30      40      50      60      70      80      90
UT2.01  -ALHSLEIQFTGMAITDDLPDIIFVWLDSSQLLFSYDTPQSKEVLLKLGCGAYPPLKRLMEKLKEQLRRGGEDDLRRFCAYMVSYNETWG
UT2.03  -ALHSLEIQFTGMAITDDLPDIIFVWLDSSQLLFSYDTPQSKEVLLKLGCGAYPPLRRLMEKLKEQLRRGGEDDLRRFCAYMVSYNETWG
UT2.05  -ALHSLEIQFTGMAITDDLPDIIFVWLDSSQLLFSYDTPQSKEVLLKLGCGAYPPLKRLMEKLKEQLRRGGEDDLRRFCAYMVSYNETWG
UT2.09  -ALHSLEIQFTGMAITDDLPNIIFVWLDSSQLLFSYDTPQSKEVLLKLGCGAYPPLKRLMEKLKEQLRRGGEDDLRRFCAYMVSYNETWG
UT3.01  --HSHIGQFTAVGTAHSLLELSAINFLNDIEVGSYNKAHQIIVKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT3.04  --HSHIGQFTAVGTAHSLLELSAINFLNAIEVGSYNKAHQIIVKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT3.15  --HSHIGQFTAVGTAHSLLELSAINFLNDIEVGSYNKAHQIIVKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT3.18  --HSHIGQFTAVGTAHSLLELSAINFLNDIEVGSYNKAHQIIVKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT4.01  ----RHVIOFTTVGIDHSLLELSAISLIDDEIVASYNKGQKQIAIKIPCISEVLGVNYLTQWHSNLVKH-EQDFLRPIQFLARNDINRH-
UT4.02  ----RHVIOFTTVGIDHSLLELSAISLIDDEIVASYNKGQKQIAIKIPCISEVLGVNYLTQWHSNLVKH-EQDFLRPIQFLARNDINRH-
UT4.03  ----RHVIOFTTVGIDHSLLELSAISLIDDEIVASYNKGQKQIAIKIPCISEVLGVNYLTQWHSNLVKH-EQDFLRPIQFLARNDINRH-
UT4.04  ----RHVIOFTTVGIDHSLLELSAISLIDDEIVASYNKGQKQIAIKIPCISEVLGVNYLTQWHSNLVKH-EQDFLRPIQFLARNDINRH-
UT5.01  SVHHRHMGQLTGVTGTAHSLLELSAISLIDDEIVVSYNKAHQIDLKIPWISKALGVKYLTELHDLLVKQ-EQDFLRWTIEIVERNDSHNS
UT5.02  SVHHRHMGQLTGVTGTAHSLLELSAISLIDDEIVVSYNKAHQIDLKIPWISKALGVKYLTELHDLLVKQ-EQDFLRWTIEIVERNDSHNS
UT5.03  SVHHRHMGQLTGVTGTAHSLLELSAISLIDDEIVVSYNKAHQIDLKIPWISKALGVKYLTELHDLLVKQ-EQDFLRWTIEIVERNDSHNS
UT5.04  SVHHRHMGQLTGVTGTAHSLLELSAISLIDDEIVVSYNKAHQIDLKIPWISKALGVKYLTELHDLLVKQ-EQDFLRWTIEIVERNDSHNS
UT5.08  SVHSHIGQFTAVGTAHSLLELSAINFLNDIEVGSYNKAHQIIVKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT5.10  SVHHRHMGQLTGVTGTAHSLLELSAISLIDDEIVVSYNKAHQIDLKIP*ISKALGVKYLTELHDLLVKQ-EQDFLRWTIEIVERNDSHNS
UT6.01  SVHHRHELOFTTVGTHHSLLELSAISLIDDEIVASYNKGQKQITIKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT6.02  SVHHRHELOFTTVGTDHSLLELSAISLIDDEIVASYNKGQKQITIKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT6.03  SVHHRHVIQFTTVGTDHSLLELSAISLIDDEIVASYNKGQKQITIKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT6.06  SVHHRHELOFTTVGTDHSLLELSAISLIDDEIVASYNKGQKQITIKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT6.12  SVHHRHELOFTTVGTDHSLLELSAISLIDDEIVASYNKGQKQITIKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT7.01  SVHHRHELOFTTVGTDHSLLELSAISLIDDEIVASYNKGQKQITIKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT7.05  SVHHRHELOFTTVGTDHSLLELSAISLIDDEIVASYNKGQKQITIKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT7.06  SVHHRHELOFTTVGTDHSLLELSAISLIDDEIVASYNKGQKQITIKIPWISKALGVNIIKKINLLVEH-EQHFRWTIQFLSRNDTNHNS
UT8.01  --HHKRMIOFTTVGTGSSLLEFIVDFLDDVQVFSFNKYNHQLIAKEAWISQVLGAKFIEETWQKLVVDH-EKSFLEWLRKLMQNSTKSD-
UT8.02  --HHKRMIOFTTVGTGSSLLEFIVDFLDDVQVFSFNKYNHQLIAKEAWISQVLGAKFIEETWQKLVVDH-EKSFLEWLRKLMQNSTKSD-
UT8.03  --HHKRMIOFTTVGTGSSLLEFIVDFLDDVQVFSFNKYNHQLIAKEAWISQVLGAKFIEETWQKLVVDH-EKSFLEWLRKLMQNSTKSD-
UT8.11  --HHKRMIOFTTVGTGSSLLEFIVDFLDDVQVFSFNKYNHQLIAKEAWISQVLGAKFIEETWQKLVVDH-EKSFLEWLRKLMQNSTKSD-
UT8.14  --HHKRMIOFTTVGTGSSLLEFIVDFLDDVQVFSFNKYNHQLIAKEAWISQVLGAKFIEETWQKLVVDH-EKSFLEWLRKLMQNSTKSD-
UT9.01  ----KFTAVSTDRSLDDFTGTAVIDDQWISYKLNQIAVAVDWDSDVLDGDFTEEMCHLTKSH-ERDAYSILHFTLGNDDTKTE-
UT10.01 -GHRHLEGEFTIAGTRHSLFNFTIIVHVDVLLVCTYDKGNQTVFNCAWITRALGKDFENKKEFNLSA-EMTFRWLSFFIKKDTKSN-
UT10.02 -GHRHLEGEFTIAGTRHSLFNFTIIVHVDVLLVCTYDKGNQTVFNCAWITRALGKDFENKKEFNLSA-EMTFRWLSFFIKKDTKSN-
UT15.01 ---HRHDFYFTATGSSGSLDFTIVNSVVDVQVCSYHKEKHQLVMKEDDMYQALGASEIAQKQKLLDL-EKNFQWAKNLIKNEAKSG-
UT15.06 ---HRHDFYFTATGSSGSLDFTIVNSVVDVQVCSYHKEKHQLVMKEDDMYQALGASEIAQKQKLLDL-EKNFQWAKNLIKNEAKSG-
UT17.01 ---HSHHEMFFETAVGTTKTLDDFTLVSEFLDDVEIVFYNKKQQQFMKEAWVSHALGADFDIEDIQOKLKYN-EIDYLWALQNVKNDKRGE-
UT17.02 ---HSHHEMFFETAVGTTKTLDDFTLVSEFLDDVEIVFYNKKQQQFMKEAWVSHALGADFDIEDIQOKLKYN-EIDYLWALQNVKNDKRGE-

```

α2 domain

```

10      20      30      40      50      60      70      80      90
UT2.01  ---ELQTTQVTVFCELDGDIQVDSRMLVAFDGETVQCLNEQARWVIKKPEVTRFCNYLDDSVQWHR---ILAEECPFLKIRANS---
UT2.04  ---ELQTTQVTVFCELDGDIQVDSRMLVAFDGETVQCLNEQAGWVIKKPEVTRFCNYLDDSVQWHR---ILAEECPFLKIRANS---
UT2.08  ---ELQTTQVTVFCELDGDIQVDSRMLVAFDGETVQCLNEQAGWVIKKPEVTRFCNYLDDSVQWHR---ILAEECPFLKIRANS---
UT2.10  ---ELQTTQVTVFCELDGDIQVDSRMLVAFDGETVQCLNEQAGWVIKKPEVTRFCNYLDDSVQWHR---ILAEECPFLKIRANS---
UT3.01  ---GNHTAQLLAECEIDNNITVKSHIYLIWDGVEYFRIDEEVGHWEDINPEFKEQY-HILESFFWTTLRKRKYMKLYCVDLKGRTI---
UT3.03  ---GNHTAQLLAECEIDNNITVKSHIYLIWDGVEYFRIDEEVGHWENINPEFKEQY-HILESFFWTTLRKRKYMKLYCVDLKGRTI---
UT4.01  ---NHTAQLQAECEMDNDITVKSHIYLIWDGEECYFRIDEEVGHWENINPEFKEQY-YILESSLRTSLRKHYNMNYCVDLMKKIVGYSSKG
UT4.06  ---NHTAQLQAECEMDNDITVKSHIYLIWDGEECYFRIDEEVGHWENINPEFKEQY-YILESSLRTSLRKHYNMNYCVDLMKKIVGYSSKG
UT5.01  SPTGNHTAQLLAECEIDNDITVKSHIYLIWDGEECFRIDEEDGHWENINPEFKEQY-HILGSPFWTTLRKRKYMKLYCVDGIRKI---
UT5.02  SPTGNHTAQLLAECEIDNNITVKSHMHLIWDGEECFRIDEESGHWENINPEFKEQY-HTLERPFWTSLRKYMKLYCVDGIRKI---
UT5.04  SPTGNHTAQLLAECEIDNNITVKSHMHLIWDGEECFRIDEESGHWENIKPEFKEQY-HTLERPFWTSLRKYMKLYCVDGIRKI---
UT5.08  SPTGNHTAQLLAECEIDNNITVKSHIYLIWDGVEYFRIDEEVGHWENINPEFKEQY-HILESFFWTTLRKRKYMKLYCVDLMRKI---
UT5.09  SPTGNHTAQLLAECEIDNNITVKSHMHLI*DGEECFRIDEESGHWENIKPEFKEQY-HTLERPFWTSLRKYMKLYCVDGIRKI---
UT5.13  SPTGNHTAQLLAECEIDNNITVKSHMHLIWDGEECFRIDEESGHWENIKPEFKEQY-HTLERPFWTSLRKYMKLYCVDGIRKI---
UT6.01  ---NHTAQLLGECEIDNDITVRSIHILWNGKESYRIDEEVGHWENINPEVKQYP-YILESPLWTSLRKYMKLYCADLMKKI---
UT6.06  ---NHTAQLLGECEIDNDITVRSIHILWNGKESYRIDEEVGHWENIKPETKQYP-YILESPLWTSLRKYMKLYCADLMKKI---
UT6.07  ---NHTAQLLGECEIDNDITVRSIHILWNGKESYRIDEEVGHWENINPEVKQYP-YILESPLWTSLRKYMKLYCADLMKKI---
UT8.01  ---NHTVQLFILCEIERNQQLGNETQIAVNGQDFCWLDEKYGDWFMVPEAEFFK-PILTSNLQWTQRKH---
UT9.01  ---NNTMQYILGCELDLDDIKLNSKIRIGLDGETMLEDEHQRHWVQSPRAENLK-PIVESLFWNLRRK---
UT9.03  ---NNTMQYILGCELDLDDIKLNSKIRIGLDGETMLEDEHQRHWVQSPRAENLK-PIVESLFWNLRRK---
UT10.01 ---NHTIQLYGCELDNDIYIGSQIQFAVDGEDFIRNDKMEHWIVTKPEAEHFR-TIAESTLVNLRKVMKHYC---
UT15.01 ----HTLQVFNCEVDKDIIVSSHQFALDGEDFCQMDQEPHHRMAMKPEAQHFK-FLWDSAFKRRVRHYIQEDC---IDG---
UT15.06 ----HTLQVFNCEVDKDIIVSSHQFALDGEDFCQMDQEPHHRMAMKPEAQHFK-FLWDSAFKRRVRHYIQEDC---IDG---
UT17.01 ---ENHTVQI-HNCHLDQDIHVSSRVCAVDGEDFCQMDQEQISHWVAKKPEAEKIV-LLESEKIYCSKKVKRYLEDYC---VQPM---
UT17.02 ---ENHTVQI-HNCHLDQDIHVSSRVCAVDGEDFCQMDQEQISHWVAKKPEAEKIV-LLESEKIYCSKKVKRYLEDYC---VQPM---
UT17.03 ---ENHTVQI-HNCHLDQDIHVSSRVCAVDGEDFCQMDQEQISHWVAKKPEAEKIV-LLESEKIYCSKKVKRYLEDYC---VQPM---
UT17.04 ---ENHTVQI-HNCHLDQDIHVSSRVCAVDGEDFCQMDQEQISHWVAKKPEAEKIV-LLESEKIYCSKKVKRYLEDYC---VQPM---
UT17.05 ---ENHTVQI-HNCHLDQDIHVSSRVCAVDGEDFCQMDQEQISH*VAKKPEAEKIV-LLESEKIYCSKKVKRYLEDYC---VQPM---

```

Supplemental Figure 1 and 2: Comparison of identity across *ModoUT* alleles in the α1 and α2 domains. Yellow shading represents conserved hydrophobic regions, and red circles show three sites of positive selection in *ModoUT5* alleles that are found in the hydrophobic regions.

CHAPTER 4

THE MHC GENOMIC REGION OF THE OPOSSUM *MONODELPHIS* *DOMESTICA* REEXAMINED

By Katina V. Krasnec^{1*}, Alana R. Sharp^{1*}, Tracey L. Williams, and Robert D. Miller¹

¹Center for Evolutionary & Theoretical Immunology, Department of Biology, University of New Mexico, Albuquerque, NM, 87131, USA

*These authors contributed equally

ARS, KVK, and RDM drafted manuscript. ARS and KK generated and figures. TLW performed sequencing reactions.

Manuscript to be submitted as Brief Communication to Immunogenetics

Abstract

The gray short-tailed opossum *Monodelphis domestica* is one of the few marsupial species for which a high quality whole genome sequence is available and the Major Histocompatibility Complex (MHC) region has been annotated. Previous analyses revealed only a single locus within the opossum MHC region, designated *Modo-UA1*, with the features expected for encoding a functionally classical class I α -chain. Nine other class I genes found within the MHC are highly divergent and have features usually associated with non-classical roles. The original annotation however was based on an early version of the opossum genome assembly. More recent analyses of allelic variation in individual opossums revealed too many *Modo-UA1* sequences per individual to be accounted for by a single locus found in the genome assembly. Analyses of a later generation assembly, MonDom5, revealed the presence of two additional loci, now designated *Modo-UA3* and *UA4*, in a region that was expanded and more complete than in the earlier assembly. *Modo-UA1*, *UA3* and *UA4* are all transcribed, although *Modo-UA4* transcripts are rarer. *Modo-UA4* is also relatively non-polymorphic. Evidence presented support the accuracy of the later assembly and the existence of three related class I genes in the opossum, making opossums more typical of mammals by having multiple apparent classical MHC class I loci.

Introduction

The Major Histocompatibility Complex (MHC) is among the most gene dense and polymorphic regions of the vertebrate genome (Beck et al. 1999; Gaudieri et al. 2000). The MHC encodes molecules involved in the processing and presentation of antigens to T cells and has been associated with disease resistance, transplant rejection, and reproductive success (Kumánovics et al. 2003). Classical MHC class I molecules, such as human HLA-A and B, are typically responsible for peptide presentation to cytotoxic CD8⁺ T cells. They are generally polymorphic and ubiquitously expressed on most nucleated cells. Non-classical class I genes, on the other hand, are often less polymorphic, restricted in their tissue expression, and have evolved to perform a variety of roles (Hofstetter et al. 2011).

The study of distantly related species has provided important insight into the evolution of the MHC. Marsupials represent a lineage of mammals that diverged from the eutherians (placental mammals) over 150 million years ago and have provided insights into the evolution of mammalian MHC (Baker et al. 2009; Belov et al 2006, Gouin et al. 2006). The gray short-tailed opossum (*Monodelphis domestica*) is arguably one of the better-developed model marsupial species and has been useful for comparative immunology and biomedical research (Samollow 2008). *M. domestica* is used as a model for melanoma, developmental biology, spinal cord injury, and genetics of cholesterol regulation (Samollow 2008). Significantly it is one of the few non-eutherian mammals for which a well-annotated whole genome sequence is available (Mikkelsen et al. 2007).

The opossum MHC, a 3.95 Mb region on chromosome 2, is as gene rich and complex as that of mouse and human (Belov et al. 2006). However its organization differs from that of eutherians in that the class I and II genes are interspersed at one end of the MHC, rather than being separated by the class III region (Belov et al. 2006). Eleven class I loci were identified within the opossum MHC region (Belov et al. 2006). One, *Modo-UAI*, appeared to be highly polymorphic and ubiquitously transcribed and likely involved in peptide presentation (Miska and Miller 1998; Gouin et al 2006; Belov et al. 2006; Baker et al. 2009). *Modo-UAI* also appeared to account for all the known MHC class I transcripts known at the time (Miska and Miller 1998). Six of the class I loci (*Modo-UE*, *UI*, *UG*, *UJ*, *UK*, and *UM*) are transcribed but have features of genes encoding molecules with non-classical functions (Baker et al., 2009). *Modo-UF* and *UL* have complete ORF but transcripts have yet to be identified in any tissue tested. A gene designated *Modo-UA2* due to its similarity to *UAI*, along with *Modo-UH*, are pseudogenes due to incomplete open reading frames (Belov et al., 2006). Two other loci, *Modo-UB* and *UC*, appear related to *Modo-UAI* but are located outside MHC (Miska et al. 2004, Belov et al 2006).

The number of classical MHC class I loci can vary between species and vertebrate lineages. Chickens and *Xenopus*, for example, only have a single classical class I locus each (Flajnik et al. 1999; Gunther and Walter 2001; Kaufman 1999; Renard et al. 2001; Shiina et al. 1999). Mammals, however, tend to have multiple classical MHC class I loci. Therefore the presence of only a single classical class I chain locus in the opossum seemed unusual. Unfortunately, the previous analysis was based on an early version of

the opossum genome assembly, MonDom2, and gaps in the area containing corresponding to the MHC were present (Belov et al 2006).

Recent genotyping studies of both captive bred and wild caught *M. domestica* uncovered too many *Modo-UAI* "like" sequences per individual to be accounted for by a single locus (Table 1). Here we reexamined the opossum MHC genomic region using a later assembly than that used for the earlier annotation (Belov et al 2006; Mikkelsen et al. 2007).

Table 1: Alleles from both captive and wild-caught *M. domestica*

Animal	Source	Alleles
46	Brazil	01/02/11
47	Brazil	03 ¹ /04/05/27/28
50	Brazil	03 ¹ /05 ¹ /09 ¹ /12 ¹ /23
96	Brazil	03 ² /05 ² /21 ² /24/25/26
97	Brazil	03 ² /05 ² /12
97.6	Brazil	03
97.7	Brazil	03/05
97.9	Brazil	03/05
110	Brazil	03/05
110.3	Brazil	03
136	Brazil	03/12
139	Brazil	03/12
D4368 ³	Brazil	03/05/09/13/14
21696 ³	Bolivia	04/08/10/15/16
12653 ³	Bolivia	06/07/17
12538 ³	Bolivia	07/18/19/20/21/22

¹Genomic DNA sequences confirmed with targeted PCR on splenic cDNA from same animals

²Genomic DNA sequences confirmed with 454 sequencing on RNA from same animals

Materials and Methods

Whole genome sequence analysis

Sequence analyses were performed using the MonDom5 *M. domestica* genome assembly available (GenBank accession number AAFR03000000) (Mikkelsen et al. 2007). Homology searches using the BLAST algorithm and *Modo-UAI* cDNA sequences (accession number AF125540.1) were performed. Intron-exon boundaries were identified by determining splice sites using cDNA sequences and with the GENSCAN 1.0 algorithm (Altschul et al. 1997; Burge and Karlin 1997; Miska and Miller 1999).

RNA and DNA extraction and cDNA synthesis

Opossum adult spleen tissue or whole embryonic tissues were collected from captive-bred *M. domestica* and stored overnight at 4°C in RNALater (Invitrogen, Carlsbad, CA). RNA was extracted using the Trizol protocol (Invitrogen, Carlsbad, CA) and stored at -80°. Complementary DNA was synthesized from RNA using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA). Genomic DNA was extracted from spleen or liver tissue from six captive-bred adult animals with the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA). Genomic DNA was also isolated from liver and spleen tissues from one Brazilian and three wild-caught Bolivian animals provided by the Division of Genomic Resources, Museum of Southwestern Biology (MSB) at the University of New Mexico. All live animal protocols were approved under the University of New Mexico Institutional Animal Care and Use Protocol No. 10-100413-MCC.

PCR and sequencing

The region of *Modo-UA1*, *UA3*, and *UA4* loci corresponding to exon 2, intron 2, and exon 3 were amplified from genomic DNA by PCR using the primers 5'-AGGGAGAGCTGCCCTGAGCC-3' and 5'-AGTCTTTCCTTCTCTCTGCGATGCT-3' which flank the exons and the Advantage HF 2 PCR kit (Clontech, Mountain View, CA). PCR conditions were 94°C for 1 minute, 33 cycles of 94°C for 30 seconds and 62°C for 4 minutes, followed by a 7 minute extension at 68°C. The sequences of full-length transcripts were determined from overlapping clones amplified from a *M. domestica* thymus library using primers (5'-ATTAACCCTCACTAAAGGGA-3' and 5'-TAATACGACTCACTATAGGG-3') flanking the cloning site in the Uni-ZAP XR vector paired with primers in exons 2 and 3 (*Modo-UA1-F*, 5'-CGACCAGCAGTTCGTGCGCT-3'; *Modo-UA1-R*, 5'-CACTCTCTCTGCGAAGCTCCTCTC-3'; *Modo-UA3-F*, 5'-GCGGCCTGGATGGACAAGATGG-3'; *Modo-UA3-R*, 5'-AGTCTTTCCTTCTCTCTGCGATGCT-3'; *Modo-UA4-F*, 5'-GGGCGGCCTGGATGGACAAG-3'; *Modo-UA4-R*, 5'-ATCTCTCTGCGATGCTCCTCTCTG-3'). Targeted PCR were performed on splenic cDNA from one genotyped Brazilian animal with two primer pairs in exons 2 and 3 (5'-GATATTGCCACACTACCATGTCCCG-3' and 5'-TGCAGGTA CT TCTTCAGCCACTGC-3'; 5'-TTTCCACACTGCCATGTCCCG-3' and 5'-GCCACTGCAGGCAGGTCTCC-3').

All PCR products were cloned using the pCR4-TOPO TA kit following manufacturers recommended protocol (Invitrogen, Carlsbad, CA). Clones were sequenced with Big Dye Terminator v3.1 Cycle Sequencing Kit (Invitrogen, Carlsbad, CA) (Holmes and Quigley 1981). Sequences were analyzed and edited using Sequencher 5.0 (Gene Codes, Ann Arbor, MI). Sequence identity was determined by BLAST comparison and nucleotide alignment with the MonDom5 assembly. Exon and intron structures were determined via the GENSCAN algorithm (<http://genes.mit.edu/GENSCAN.html>).

Sequence analyses

Phylogenetic analyses were performed based on nucleotide alignments corresponding to exons 2 and 3 of MHC class I genes from *M. domestica* and other representative vertebrate species. Nucleotide sequences were first translated into protein and aligned using ClustalW in BioEdit (Hall, 1999) to preserve codon positions. Aligned sequences were then converted back to nucleotides for tree construction using MEGA 5.05 (Tamura et al. 2011). The tree was constructed using neighbor joining statistical method using the Tajima-Nei Model with 1000 bootstrap replicates. Accession numbers for sequences used in the phylogenetic analyses were: Human *HLA-A* (U03862.1), *HLA-B* (X91749.1), *HLA-Cw* (U06487.1), *HLA-F* (BC009260.2), *HLA-G* (M32800.1), FcRn (NM_004107.4), HFE (NM_000410.3), MICA (NM_000247.1), and MICB (AK314228.1); mouse *H-2K(b)* (U47328.1), and FcRn (D37874.1); Grey short-tailed opossum *Modo-UB* (NM_001079820.1), *Modo-UC* (NM_001079819.1), *Modo-UE* (NM_001171835.1), *Modo-UG* (DQ138606.1), *Modo-UI* (NM_001171837.1) *Modo-UJ* (NM_001171836.1), *Modo-UM* (NM_001171834.1), *Modo-I* (AF135040.1), *Modo-3* (AF125540), *Modo-10*

(AF125542), and *Modo-16* (AF125543); Northern brown bandicoot *Mhc-1* (DQ927302.1); brushtail possum *Trvu-UB* (AF359509.1), and FcRn (AF191647.1); Red-necked wallaby, *Maru-UA* (L04950.1), and *Maru-UB* (L04952.1); Rhesus macaque *Mamu-A* (AF157401.1), and *Mamu-B* (NM_001048245.1).

Divergence time between *Modo-UA1*, *Modo-UA3*, and *Modo-UA4* was estimated using the Jukes-Cantor model in MEGA5 (Jukes and Cantor 1969; Tamura et al. 2011). The number of substitutions between *Modo-UA1* and *Modo-UA4* across introns 1-5 was estimated at 0.0390 substitutions/site (over 2642 sites) divided by the marsupial-specific rate of neutral nucleotide substitution of $6.8-8.9 \times 10^{-9}$ substitutions/site/year based on previous analysis of opossum class I loci (Miska et al. 2004)

Evidence of selection was performed using exon 2 and 3 sequence of the alleles from *Modo-UA1*, *UA3*, and *UA4* (GenBank Accession No. JX661062-JX661088). Sequences were aligned using Clustal X (Larkin et al. 2007) and analyzed using MEGA (Tamura et al. 2011) where a maximum likelihood tree was generated. The alignment and tree were input into the Phylogenetic Analysis through Maximum Likelihood (PAML) program (Yang, 2007) for maximum likelihood analysis of to determine sites of positive selection. The analysis used both Naïve Empirical Bayes and Bayes Empirical Bayes to determine sites of positive selection, and any site determined to be under selection had p values of 0.05 or lower.

Sequence deposition

Sequence data for *Modo-UA1*, *Modo-UA3*, and *Modo-UA4* described in this paper have been deposited in GenBank (<http://www.ncbi.nlm.nih.gov/Genbank>) under accession numbers JX661062 through JX661088.

Results & Discussion

This reanalysis revealed two additional loci encoding MHC class Ia chains along with a previously unrecognized *TAP2A* gene. These were located in a 390 Kb region flanked by the non-classical class I *Modo-UJ* and class II *DXA* genes (Fig. 1). The two additional class I genes are located in gaps in the earlier MonDom2 assembly and were designated *Modo-UA3* and *UA4*. The *Modo-UA1* locus retained its original designation. In comparing the full-length sequences of each of the genes with each other, *Modo-UA1* and *Modo-UA3* have the greatest levels of sequence identity at 93.09%, while comparing *Modo-UA1* to *Modo-UA4* is 88.39%. A comparison of *Modo-UA3* to *Modo-UA4*, there is a sequence identity of 89.77%, indicating that *Modo-UA4* is the most different of the three.

Confirmation of *Modo-UA1*, *UA3*, and *UA4* loci presence in the genome of *M. domestica* was made using genomic DNA to ensure discovery was not another assembly artifact. Genomic DNA from eight opossums, five of which were from a captive colony with Brazilian founders and the remaining three were wild-caught Bolivian animals (Baker et al. 2009). Primers located in intron 1 and exon 3 were used to amplify a 734 bp product containing exon 2, intron 2, and most of exon 3 by PCR. PCR products were then cloned and sequenced from each individual. Five or six unique alleles were amplified from six of the eight individuals, consistent with there being at least three *UA* loci, supporting the later genomic assembly (not shown).

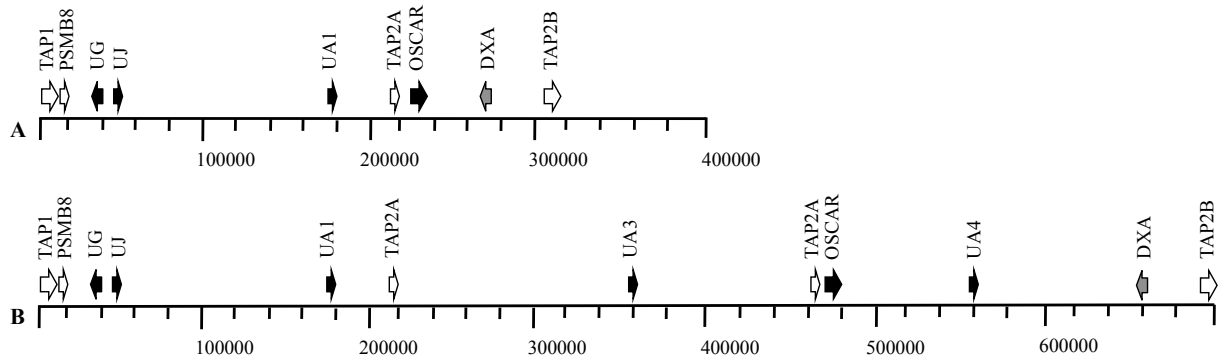


Figure 1. A: *Monodelphis domestica* MHC map derived from Belov et al. 2006. B: *M. domestica* updated MHC map derived from coordinates relative to opossum build 2.2 (MonDom5) base pair 270,288,868 on chromosome 2. Black arrows represent Class I genes, grey arrow are Class II genes, and white arrow are antigen presenting genes.

RACE-PCR was performed on the thymic cDNA library to investigate the complete gene structure for *Modo-UA1*, *Modo-UA3*, and *Modo-UA4*. Transcripts that correspond to all eight exons for each locus were amplified. Overall, the *Modo-UA1*, *UA3*, and *UA4* loci have similar intron/exon structure that is typical of MHC class I genes (Figure 2). While the exon length was constant between loci, the intron lengths varied considerably, notable was the variation in intron 1 length. Alignment of full length *Modo-UA1*, *UA3*, and *UA4* cDNA sequences reveal substantial similarity between the *Modo-UA* loci and other MHC class I genes (Figure 3).

Phylogenetic analyses using marsupial and eutherian class I sequences revealed *Modo-UA1*, *UA2*, and *UA3* form a single clade sister to that containing *Modo-UB* and *UC* (Figure 4). Three of MHC class I cDNA sequences, *Modo-3*, 10 and 16, reported previously and thought to be from *Modo-UA1*, are actually more related to *Modo-UA3* (Miska and Miller, 1999). Another cDNA, *Modo-1*, appears to be transcribed from *Modo-UA1*. The topology of the *Modo-UA* clade is consistent with *Modo-UA1* and *UA3* being a more recent gene duplication and the duplications that gave rise to the three *Modo-UA* genes having occurred after the divergence from the locus that duplicated to form *Modo-UB* and *UC*. These interpretations are also supported by the divergence times estimated for the three *UA* loci. Using estimates of divergence rates for opossum MHC class I genes established previously (Miska et al. 2004), the *Modo-UA1* and *Modo-UA3* duplications estimated as 6.35-8.31 million years ago (MYA) (0.0564 substitutions/site), whilst that for *Modo-UA3* and *Modo-UA4* is 7.56-9.89 MYA (0.0672 substitutions/site).

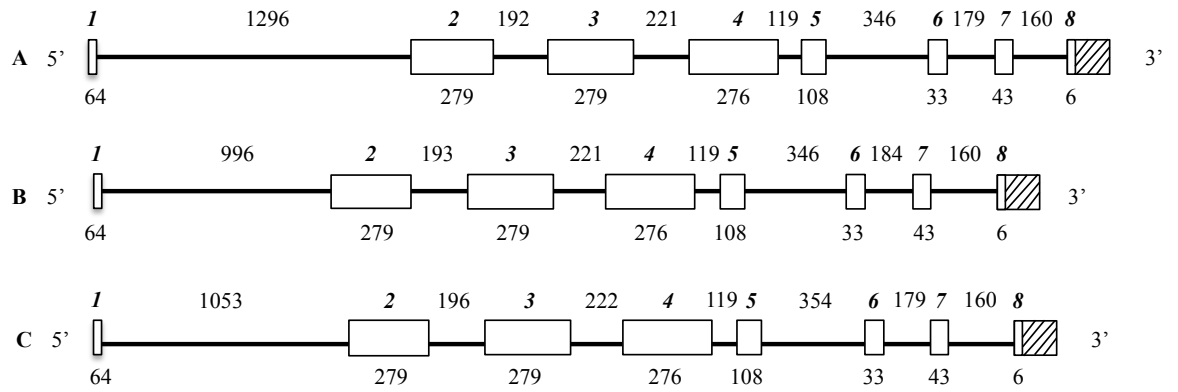


Figure 2. Intron and exon structure of A: *Modo-UAI*, B: *ModoUA3*, C: *ModoUA4*.

```

          Leader peptide
          -20          -10
          .....|.....|.....|.....|.....
ModoUA1      ...MKPSLLSLFVLGVVALTETRA
ModoUA3      .....
ModoUA4      ...-V-----K---
ModoUB       ...-EL-----M-----Q---
ModoUC       ...-EL-----
ModoUA2      .....-F-----
ModoUE*01    ...-GFLV---LL--ILV---I--
ModoUI*01    .....MFF--LF--TLV--K--W-
ModoUK*12    ...-ES--VVLSSL--EA--M--P--IK-
ModoUMf1*01  .....ML--F--VSL-----W-
HLA-A        MAV-A-RT-V-LLS-AL---Q-W-

α1 domain
          10          20          30          40          50          60          70          80          90
ModoUA1      GSHSMRYFYTSMRSP...ELGFSQFISVGYVDDQQFVRFDSSESQRAEPRAAWMDQDEEDRNYWEGQTQINRRTAQITRVLDLETALGYNOSRG
ModoUA3      .....H-A-----T-----M-----K-----EP-----E-----Y-KN---Y-RN---LR-----Q-
ModoUA4      ---L--CH-T-----I-A-----M-----K-E---PD---RN-H-C-SN---AQ---K--IL---Q-
ModoUB       .....H-A-----T-----M-----K-----EP-----E-----Y-KN---Y-RN---LR-----Q-
ModoUC       .....H-A-----T-----M-----K-----EP-----E-----Y-KN---Y-RN---LR-----Q-
ModoUA2      .....F-----L-----L-----KV-D...R---S-----YG-N---L-----
ModoUE*01    ---LK---AA-Q...G-AEPR-TA-V-----VLS---D-P-SI-TKTP-IE-EM...S---LE-R...EATLRY-MC-QKVPV---H-E-
ModoUI*01    ---KS--V-I--S...Q-EPRKVV---A-IEY--H-AN--V---Q-HK-G...QD-Q---R-A-AET-NY--G---LRN---K-
ModoUK*12    SF--L-Q---Q---L...GRKP-----L--ML--GN-PN--E---P-L-----QD---KNSR-S-E---TFELG-QNLV---E-
ModoUMf1*01  ---ID--VVTA-RLG-EERW-TL-----H-NNA--ST-IG-P-IELET...PD---REKRHL-KDS-NCPMS-QNLHFN---HD
HLA-A        -----V---GR-EPR--A-----T-----DAA-R-M---P-IE-EG...PE--D-E-RKVKHS-TH---G-LR-----EA

α2 domain
          100          110          120          130          140          150          160          170          180
ModoUA1      G.LHTIQRYGCEVHPDGSFRKGFYQLAYDGRDYIALHRETLTWTAAADPGAENTKRKWEAERSFAERVKAYLEETCVQWVKKYLQMGEDVLLRT
ModoUA3      .....M-----SR-----H-----D-----T-----I-----R-D-V-VE---L---K-----
ModoUA4      ---I---S-AE-L-----F---H-----T---S-----P-K-I--SY---K--TH-L-V---K-----
ModoUB       ---I---S-E-----F---H-----S---S-----T-----P-K-I--SY---KE-TH-L---E-K-----
ModoUC       ---L---L-----D-H---H-----D-----T-----GQ-----L---K-----
ModoUA2      ---V--Y--QA---FSG---SR--A-YGF---LT-DTG--R-V--A--L-N---DORI--NW-I---E---HSLQR--EN-KER---A
ModoUE*01    ---V-VF-TLI---ISS---KR-IL-H---L---S-DM--Y---V-V-LI---SDGK-LK-G---E---LH-H-ED-K-S---
ModoUI*01    ---Y--LV---SHNRM--R--E--Y--Q---S-DF--S---IP-L-S-H---M-PNIAQ-Q-T---KN---LY---ET-NET---A
ModoUK*12    -GV--L---FGN---STF-LLYG---Q-KLS-DP---N-I-S-SV-LSI-L-LD-L-P---W-VH-TV---LLRH-EK-KET---A
ModoUMf1*01  ---S--L---CD-GS-WR-LR-YH-Y---K---KEDLRS---MA-QT--H---AH-V---QWR---G---E-LRR--EN-KET-Q---
HLA-A        ---S--L---CD-GS-WR-LR-YH-Y---K---KEDLRS---MA-QT--H---AH-V---QWR---G---E-LRR--EN-KET-Q---

α3 domain
          190          200          210          220          230          240          250          260          270          280
ModoUA1      DPPSARVSRHSGPDGEVSLRCRAQGFYPAEISLTLWRLDGEEQLQDTEFIETRPGGDTPQKWAAVAMAPGQEDRYSCRVOHEA.LAQPLSLRWE
ModoUA3      .....
ModoUA4      .....M-----
ModoUB       ---T--R-----H-----M-----
ModoUC       ---V--T-Y-----H-----
ModoUA2      .....
ModoUE*01    ---T--TSS---T-K---D---V---E-----A-----E-LS-S-HK-T---G-PE--F-K---
ModoUI*01    ---A-Q-TH-A-SK---I-Q---S---A-----A-----SLTS---EE-I-QI---E--PE--T-K---
ModoUK*12    G---TH--ISHE-K-M---W-RD---D---T---HN-----T-----G---PE--T-K---
ModoUMf1*01  ---VQ-T--ITS-S--I-K--W---I---N---I---AY-D-----E-PF-N-EK-T---G-PE--F-K---
HLA-A        -A-KTHMTH-AVS-H-AT--W-LS---T--Q---D-T---LV---A-----VVPFG--Q--T-H---G--PK--T---

Transmembrane domain
          290          310
ModoUA1      PEAPSVVWVIGVTAGVLVLTAA...VVAGAVIL
ModoUA3      ---S-L-----A-----
ModoUA4      ---S-L-----
ModoUB       ---S-L-----
ModoUC       ---S-----
ModoUA2      ---S-L--A-----M-----
ModoUE*01    SQFS--IRLSA--ITIL-L-AAV...-II-V-VW
ModoUI*01    -QTS-I-ICG-II-.IA-LI...-V-W
ModoUK*12    -Q-S-IFIF-V-IIAGFLFALII...-T-W
ModoUMf1*01  LQS-FTGYMI-IM-F--L-LTS...-I--I--W
HLA-A        -SSQPTIP---II-GLVLFGAVITGA-VA--MW

Cytoplasmic tail
          320          330          340          350          360          370          380          390          400          410          420          430
ModoUA1      RR...RNSGGKGGAYVPAADKDSAQG.SDVSLTVTA*.....
ModoUA3      .....*.....
ModoUA4      .....*.....
ModoUB       .....Q-----L-V-A*.....
ModoUC       .....A*.....
ModoUA2      .....*.....
ModoUE*01    -K...NT-DS-R-S-TAT-SN-----F-ERT*.....
ModoUI*01    -K...Q--D--E-N-IQT--N--T---FI*.....
ModoUK*12    --LIGVED--ITFRHHAVIVHRKLSIVEDSGITFRHHAVIVHRSMQCTPKAQ-IILCGIVQRTQGYSSLRFLSPVDPDLWFLSVALIERSFEVQGVKVEHPLPFLSSFSIKHS*
ModoUMf1*01  KK...CI-----N-ALTSGS-----I-ER*.....
HLA-A        ---KS-DR---S-TQG-SS-----ACKV*.....

```

Figure 3. Alignment of protein translation of *Modo-UA1*, *Modo-UA3*, *Modo-UA4*, and select opossum and human MHC genes. Dashes indicate sequence identity, dots indicate gaps, and asterisks represent stop codons. Accession numbers for sequences used are Modo-UB, Modo-UC, and Modo-UA2, AAFR03000000 ; Modo-UE, NM_001171835.1; Modo-UI, NM_001171837.1; Modo-UK, EU886706.1; Modo-UM, EU886712.1; HLA-A, AAA03603.

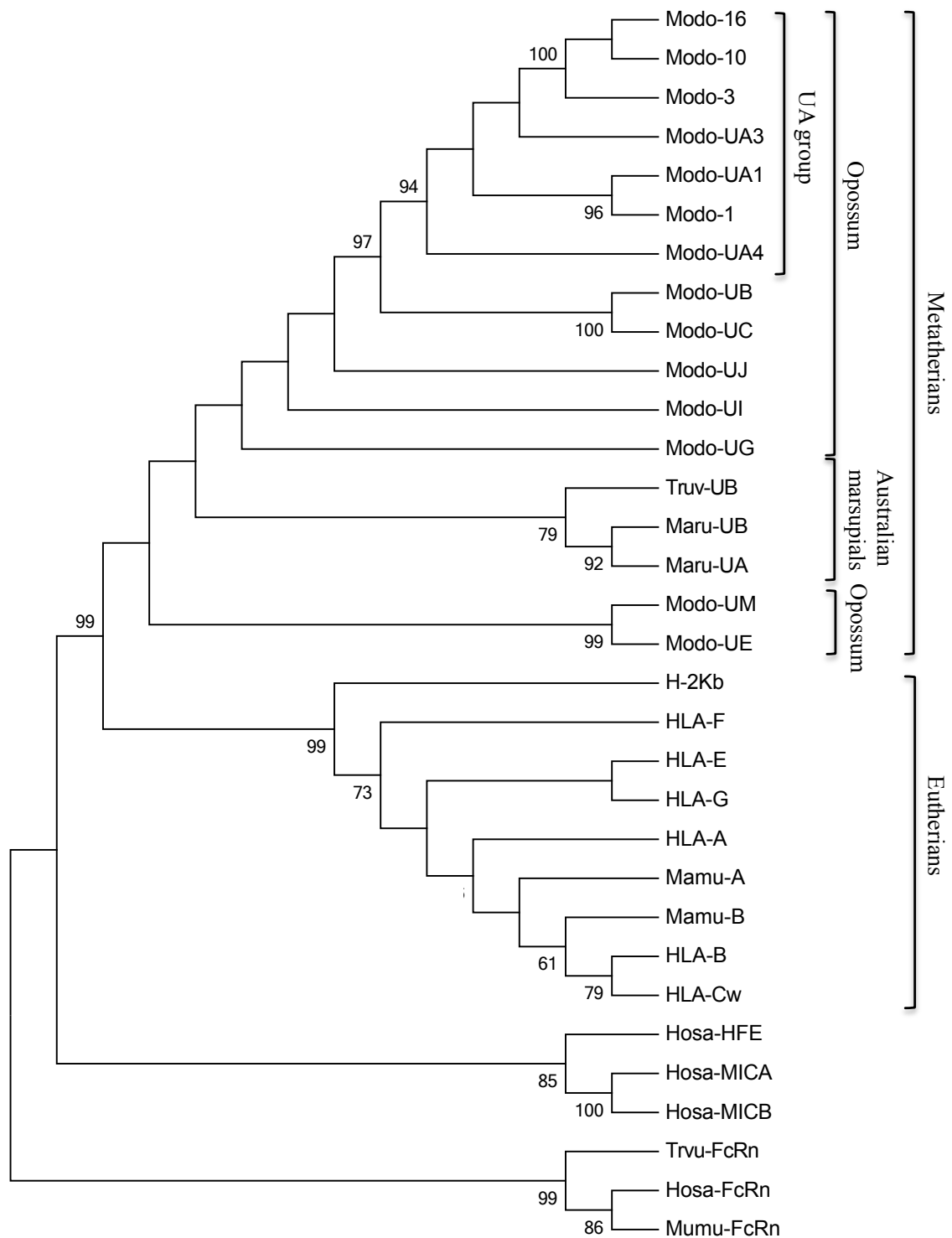


Figure 4. Phylogenetic relationship among MHC loci. Phylogenetic tree based on nucleotide alignments corresponding to exons 2 and 3 of MHC class I from *M. domestica* (Modo) and representative vertebrate species. The tree was constructed using neighbor joining statistical method using the Tajima-Nei Model with 1000 bootstrap replicates using MEGA 5.05 (Tamura et al. 2011).

By comparison, the duplication that gave rise to *Modo-UB* and *UC* occurred less than 6 MYA.

Previously, it was thought that the *Modo-UA1* locus was highly polymorphic (Gouin et al 2006). It is likely that the polymorphism identified was due by the presence of multiple related loci. Sequencing of the *Modo-UA* loci from eight individual *M. domestica* from both captive-bred Brazilian animals and wild-caught animals from Bolivia revealed that *Modo-UA1* and *UA3* alleles are too similar to be distinguished from one another. *Modo-UA4* is sufficiently divergent from both *Modo-UA1* and *UA3* that alleles for this locus could be distinguished. A total of six different *Modo-UA4* alleles were uncovered from both the captive and wild-caught, while a total of 22 *Modo-UA1/UA3* alleles were identified. An analysis of the *Modo-UA4* $\alpha 1$ and $\alpha 2$ domain sequence, which could be analyzed independently, revealed no evidence of sites under positive selection (data not shown).

In summary, a reanalysis of the *M. domestica* MHC genomic region using a later whole genome assembly revealed that a majority of the gene content and gene order remained unchanged from what was published previously (Belov et al 2006). The region differed substantially due to an additional 390 kb of sequence not present in the original assembly, located in a region between the TAP1 and TAP2B genes. Present were three additional genes, two encoding class I a chains, designated *Modo-UA3* and *Modo-UA4*, and one encoding a second TAP2A. *Modo-UA3* and *Modo-UA4* were so named in recognition of their sequence homology to *Modo-UA1* and the pseudogene *Modo-UA2*. Phylogenetic

analyses place the three loci in a clade with *Modo-UA2* and the MHC class I-like genes *Modo-UB* and *UC*, located outside of the MHC (Miska et al. 2004). As with other rapidly-evolving class I genes, these genes are nested within a larger *M. domestica* clade and are not orthologous to other species' class I genes (Belov et al. 2007; Miska and Miller 1999). Previous annotations and analyses of the opossum genome depicted the MHC as more akin to nonmammalian vertebrates' genomes, with only a single classical class I gene (Flajnik and Kasahara 2001; Kulski et al. 2002). This revisiting of a marsupial MHC reveals a class I gene organization more akin to eutherian mammals' multilocus class I system.

References

- Altschul SF, Madden TL, Schäffer AA, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:3389-402.
- Baker ML, Melman S, Huntley J, Miller R (2009) Evolution of the opossum major histocompatibility complex: evidence for diverse alternative splice patterns and low polymorphism among class I genes. *Immunology* 128:418-431.
- Beck S, Geraghty D, Inoko H, et al. (1999) Complete sequence and gene map of a human major histocompatibility complex. *Nature* 478:476-481.
- Belov K, Deakin JE, Papenfuss AT, et al. (2006) Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. *PLoS Biology* 4:e46. doi: 10.1371/journal.pbio.0040046
- Belov K, Sanderson CE, Deakin JE, et al. (2007) Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome Research* 17:982-991.
- Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* 268:78-94.
- Flajnik MF, Kasahara M (2001) Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* 15:351-62.
- Flajnik MF, Ohta Y, Greenberg AS, et al. (1999) Two Ancient Allelic Lineages at the Single Classical Class I Locus in the *Xenopus* MHC. *Journal of Immunology* 163:3826-3833.
- Gaudieri S, Dawkins RL, Habara K, et al. (2000) SNP Profile within the Human Major Histocompatibility Complex Reveals an Extreme and Interrupted Level of Nucleotide Diversity. *Genome Research* 10:1579-1586.
- Gouin N, Wright AM, Miska KB, et al. (2006) Modo-UG, a marsupial nonclassical MHC class I locus. *Immunogenetics* 58:396-406. doi: 10.1007/s00251-006-0115-4
- Gunther E, Walter L (2001) The major histocompatibility complex of the rat (*Rattus norvegicus*). *Immunogenetics* 53:540-542.
- Hall, TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium Series* 41:95-98.

- Hofstetter AR, Sullivan LC, Lukacher AE, Brooks AG (2011) Diverse roles of non-diverse molecules: MHC class Ib molecules in host defense and control of autoimmunity. *Current Opinion in Immunology* 23:104-110.
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: HN M (ed) Mammalian Protein Metabolism. Academic Press, New York, pp 21-132
- Kaufman J (1999) Co-evolving genes in MHC haplotypes: the “rule” for nonmammalian vertebrates? *Immunogenetics* 50:228-36.
- Kulski JK, Shiina T, Anzai T, et al. (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunological Reviews* 190:95-122.
- Kumánovics A, Takada T, Lindahl K (2003) Genomic organization of the mammalian MHC. *Annual Review of Immunology* 21:629-657.
- Larkin MA, Blackshields G, Brown NP, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.
- Mikkelsen TS, Wakefield MJ, Aken B, et al. (2007) Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447:167-77. doi: 10.1038/nature05805
- Miska KB, Miller R (1999) Marsupial Mhc class I: classical sequences from the opossum, *Monodelphis domestica*. *Immunogenetics* 50:89-93.
- Miska KB, Wright AM, Lundgren R, et al. (2004) Analysis of a marsupial MHC region containing two recently duplicated class I loci. *Mammalian Genome* 15:851-64. doi: 10.1007/s00335-004-2224-4
- Renard C, Vaiman M, Chiannikulchai N, et al. (2001) Sequence of the pig major histocompatibility region containing the classical class I genes. *Immunogenetics* 53:490-500.
- Samollow PB (2008) The opossum genome: insights and opportunities from an alternative mammal. *Genome Research* 18:1199-1215.
- Shiina T, Tamiya G, Oka A, et al. (1999) Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region. *Proceedings of the National Academy of Sciences* 96:13282-13287.
- Tamura K, Peterson D, Peterson N, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology* 113:1530-4. doi: 10.1093/molbev/msr121

Yang, Z. (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24(8):1586-1591.



The pathology and pathogenicity of a novel *Haemoproteus* spp. infection in wild Little Penguins (*Eudyptula minor*)



B.L. Cannell^{a,*}, K.V. Krasnec^b, K. Campbell^c, H.I. Jones^d, R.D. Miller^b, N. Stephens^c

^a School of Veterinary and Life Sciences, Murdoch University, South Street, Murdoch, WA 6150, Australia

^b Center for Evolutionary & Theoretical Immunology, Department of Biology, University of New Mexico, Albuquerque, NM 87131, United States

^c Department of Veterinary Pathology, School of Veterinary and Life Sciences, Murdoch University, South Street, Murdoch, WA 6150, Australia

^d Microbiology and Immunology, School of Pathology and Laboratory Medicine, M502, University of Western Australia, Nedlands, WA 6009, Australia

ARTICLE INFO

Article history:

Received 7 February 2013

Received in revised form 5 April 2013

Accepted 13 April 2013

Keywords:

Little Penguins

Haemoproteus

Wildlife

Haemosporidian parasites

ABSTRACT

One hundred and thirty four Little Penguin (*Eudyptula minor*) carcasses found since 2004 in south west Australia were necropsied. The livers and spleens from ten of the penguins exhibited varying degrees of multifocal, randomly scattered areas of necrosis and varying numbers of parasites were associated with these areas. Hepatomegaly and splenomegaly were noted in many of these ten cases. Necrosis and parasites were also observed in the cardiac muscle of four of the cases and in the lung tissue in one of the penguins. Using PCR, the parasites were positively identified in four of the cases as *Haemoproteus* spp. and morphologically identical tissue stage parasites associated with histopathological changes were observed in all ten dead penguins. This is the first study to demonstrate both the *in situ* presence of the *Haemoproteus* parasite in any member of the Sphensicidae family and mortality due to its presence. We postulate the involvement of anomalous environmental conditions in a potential increase in local vectors.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The largest colony of Little Penguins (*Eudyptula minor*) in Western Australia, found on Penguin Island, has the highest conservation status of all the major Little Penguin colonies in Australia and New Zealand (Dann et al., 1996). However, they face an increasing range of natural and anthropogenic threats, such as reduction in food resources, collisions with watercraft, hyperthermia, pollution, and introduced predators (Cannell, 2004). These threats, along with parasites and infectious diseases, are

known causes of mortality for Little Penguins throughout Australia and New Zealand (Obendorf and McColl, 1980; Harrigan, 1992; Norman et al., 1992; Clarke and Kerry, 1993; Hocken, 2000a; Cannell unpubl. data).

The prevalence of the haemosporidian parasites of the genera *Plasmodium*, *Haemoproteus* and *Leucocytozoon* is generally low in seabirds (Peirce, 2005; Quillfeldt et al., 2011), and worldwide, penguins appear to be more likely to be infected by *Plasmodium* than by species of *Haemoproteus* or *Leucocytozoon* (Levin et al., 2009; Quillfeldt et al., 2011). Infections of avian malaria, caused by *Plasmodium relictum*, *P. elongatum* and *P. cathemerium*, have been observed in many captive populations of penguins, often resulting in high mortality (Fleischman et al., 1968; Bennett et al., 1993a; Clarke and Kerry, 1993; Graczyk et al., 1994; Jones and Shellam, 1999a; Valkiunas, 2005;

* Corresponding author. Tel.: +61 421 641 050.

E-mail addresses: B.Cannell@murdoch.edu.au, belindacannell@bigpond.com (B.L. Cannell).

Bueno et al., 2010). *Haemoproteus* spp. in penguins have only been reported so far in a study of haemoparasites of Galapagos penguins (*Spheniscus mendiculus*), where a single individual, out of 362 tested, was positive (Levin et al., 2009). *Leucocytozoon* spp. have been observed in Fiordland crested penguins (*Eudyptes pachrhynchus*), African penguins (*Spheniscus demersus*) yellow-eyed penguins (*Megadyptes antipodes*), and Macaroni penguins (*Eudyptes chrysolophus*) (Valkiunas, 2005; Peirce et al., 2005) and were associated with increased regional chick mortality with evidence of severe, disseminated megaloschizont formation in multiple tissues in the latter species (Hill et al., 2010). Haemoparasites have rarely been observed in Little Penguins (Jones and Shellam, 1999a,b), though a “malaria-like parasite” was noted to have caused the mortality of two wild Little Penguins from New South Wales in 2000 (Rose, 2001) and *Leucocytozoon takawi* was transferred from a Fiordland crested penguin into a juvenile Little Penguin (Allison et al., 1978).

Avian haemosporidians have a sexual phase that develops in a vector, a blood-sucking dipteran, and an asexual phase in birds (Valkiunas, 2005). The abundance and geographic range of vectors and hence vector-borne disease may be impacted by climate change (Jones and Shellam, 1999b; Harvell et al., 2002), while migratory birds are an important factor in the distribution of parasites and diseases (Jourdain et al., 2007) due to their movement over large areas. In addition to this normal large scale movement, the geographic range of many bird species is also affected by changes in climate. In Western Australia, tropical birds have moved into areas that once were the domain of temperate birds (Dunlop and Wooller, 1986; Wooller et al., 1991), and haemosporidian parasites have been reported in a greater number of species of tropical compared to temperate seabirds (Quillfeldt et al., 2011). The expansion of tropical seabirds onto temperate islands such as Penguin Island could therefore result in the introduction or rise in prevalence of such parasites in endemic species.

Since 2004, 134 necropsies have been performed on many of the Little Penguins found dead on the South-western Australian coastal foreshores as well as Penguin Island itself. It is extremely rare to find wild birds that have died of haemoparasitic infections (Valkiunas, 2005), as most dying birds are predated upon (Bennett et al., 1993b; Ladds, 2009). In this paper, we report on the first known cases of haemoproteosis caused by *Haemoproteus* spp. in wild Little Penguins in Western Australia, and describe the pathology and pathogenicity.

2. Materials and methods

2.1. Specimen collection and gross examination

Necropsies were performed on 10 dead Little Penguins found on Penguin Island, Western Australia (32°18' S, 115°41' E) or at various locations in the southwest of Western Australia (Fig. 1) between November 2006 and May 2012. Eight of the dead penguins were frozen for 10–90 days prior to necropsy (following an unspecified interval between death and the discovery of the carcass), and two were necropsied within 24 h of discovery and

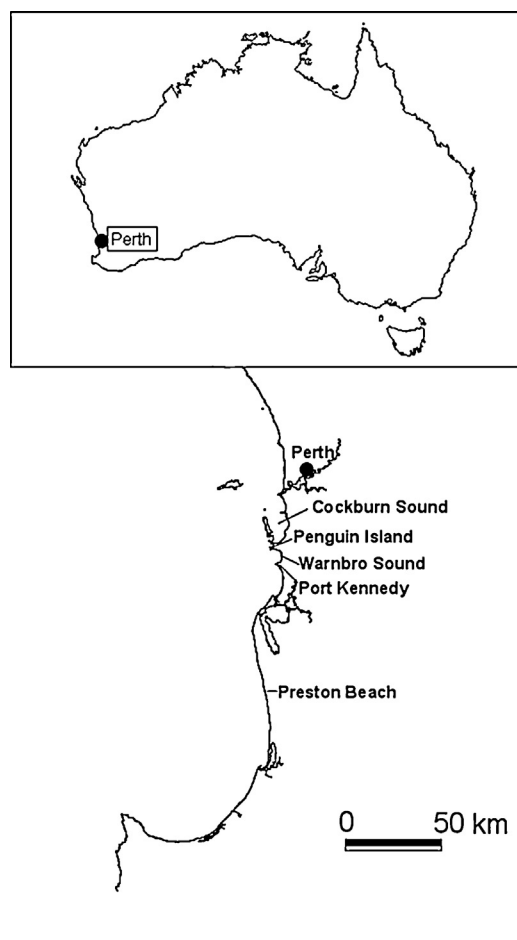


Fig. 1. The location of dead penguins in Western Australia: on Penguin Island and the foreshores of Cockburn Sound, Warnbro Sound, Port Kennedy and Preston Beach. Penguin Island is approximately 50 km south of Perth, the capital city of Western Australia.

subsequent refrigeration, though the post-mortem interval was unknown. Gross external and internal examinations were performed and the body was weighed in all but one case. The liver and spleen were weighed in eight of the cases and abdominal fat pad in seven cases. The bodies were weighed using an Avery platform scale (resolution 10 g) and internal organs were weighed using an A&D EK-410i scale (resolution 0.01 g). The mass of each of the organs was compared with that reported for Little Penguins from New Zealand (Hocken, 2000b). However, penguins from the Penguin Island colony are known to be larger than those from New Zealand (Wienecke, 1993; Dann et al., 1996), and as Hocken (2000b) found a consistent organ-to-body mass ratio in penguins, the liver to body mass ratio was used to demonstrate hepatomegaly. The ratio of abdominal fat to body mass was also used to demonstrate the body condition of the penguins. Histopathology was performed on various organs, providing autolysis did not preclude such examination; all carcasses were scored according to their degree of autolysis.

2.2. Screening of organs for tissue stages of haemosporidians

Tissue samples from the lung, spleen, liver and heart were fixed in formalin, cut into 4 μm thick sections and stained with haematoxylin and eosin. 2 μm thick sections stained with Martius Scarlet Blue were used to obtain photomicrographs of the organisms, which appeared more prominent using this technique. The samples were examined using light microscopy, including 1000 \times magnification with oil immersion. Impression smears were made for cytological examination of the liver and spleen in the final case, by rolling the cut surface of the fresh tissues onto glass slides, air drying, then staining with Wright–Giemsa.

2.3. Molecular screening for haemosporidians and sequence analysis

DNA was extracted and purified from the liver of each bird using a Qiagen DNeasy Blood and Tissue Kit. Using polymerase chain reaction (PCR) a fragment of approximately 480 bp of the mitochondrial cytochrome-*b* (cyt-*b*) gene was targeted for *Plasmodium* and *Haemoproteus* using a nested PCR strategy. For each bird, at least three independent PCR amplifications were performed before being declared negative. Avian blood samples known to be positive for either *Haemoproteus* or *Plasmodium* were used as positive controls for the PCR. Samples were first amplified using the primers HaemNF (5'-CATATATTAAGAGAATTATGGAG-3') and HaemNR2 (5'-AGAGGTGTAGCATATCTATCTAC-3') described in Waldenström et al. (2004). A second nested PCR was performed using HaemNF1 (5'-CATATATTAAGAGAAITATGGAG-3') [I = a universal base, inosine] and HaemNR3 (5'-ATAGAAAGATAAGAAATACCATTCC-3') (Hellgren et al., 2004). PCR were set up as 15 μl reactions containing 50 ng template DNA, 0.75 units of AmpliTaq Gold (Applied Biosystems, Carlsbad, California), 1.5 μl of MgCl₂, 0.2 mM of each dNTP, and 0.5 mM of each primer. The HaemNF and HaemNR2 PCR were run using a programme of 94 °C for 8 min followed by 35 cycles of 94 °C for 30 s, 50 °C for 30 s, and 72 °C for 45 s, and a final extension at 72 °C for 10 min. The conditions for HaemNF1 and HaemNR3 are identical except for an annealing temperature at 52 °C. An additional positive control to assess DNA quality was performed using avian-specific cyt-*b* primers as described in Cicero and Johnson (2001).

PCR products were cloned using the pCR4-TOPO TA kit following the manufacturers recommended protocol (Invitrogen, Carlsbad, CA). Cloning rather than direct sequencing of PCR products was performed initially to avoid problems that may arise due to presence of mixed infection. Multiple clones from each positive PCR amplification were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Invitrogen, Carlsbad, CA). Sequences from independent clones obtained from the same bird were always identical indicating absence of variation due to PCR and/or cloning artefacts. Sequences obtained were viewed and manually edited using Sequencher ver 5.0 (GeneCodes, Ann Arbor, MI) and aligned using ClustalX (Larkin et al., 2007). Sequences were compared to

those in the GenBank database using the BLAST algorithm (<http://www.ncbi.nlm.nih.gov>). Novel sequences generated in this study have been deposited in GenBank under accession numbers KC121053–KC121056.

Parameters for phylogenetic analysis and rooting were based on that of Outlaw and Ricklefs (2011). A maximum likelihood phylogeny was generated using MEGA 5.0 (Tamura et al., 2011) and a GTR+ gamma model of nucleotide substitution with 1000 bootstrap iterations, and was rooted between mammalian *Plasmodium* and avian *Plasmodium*. Accession numbers for additional sequences used in the phylogenetic analyses are: Avian *Haemoproteus* spp.: *Spheniscus mediculus*, JX679087 and KC121057; *Heteromyias albuspectularis*, AY714147; *Turdus merulus*, DQ630013; *Sylvia atricapilla*, GU784854; *Malimbus rubricollis*, HQ386243; *Quelea quelea*, EF117230; *Fringilla coelebs*, DQ368340; *Phylloscopus trochilus*, AF254972; unknown avian host, EF032812. Avian *Parahaemoproteus* spp.: *Dumetella carolinensis*, GU252002; *Vireo olivaceus*, GU252005; *Falco sarverius*, GQ141558. Avian *Plasmodium* spp.: *Seiurus aurocapillus*, HM222481; *Parus major*, DQ658590 and JQ778277; *Passer domesticus*, AF069611. Mammalian *Plasmodium* sp.: *Gorilla gorilla*, GU045322; *Grammomys surdaster*, DQ414646; *Mandrillus sphinx*, JF923750.

3. Results

3.1. History and body condition of penguins cases

Three of the ten cases were adults that had previously been banded or microchipped. These marked penguins had been observed 11–45 days prior to being found dead (Table 1). Of the other seven cases, one was an unmarked fledgling and six were unmarked adults, none of which had previously been observed (Table 1). Penguin 11/624 had just completed rearing two chicks, which were last observed in the nest on 14th September 2011, i.e. approximately 2 months prior to 11/624 being found dead.

The body condition of the penguins was variable; however none were in poor condition, all having an appreciable layer of subcutaneous adipose tissue and a reasonable to adequate intra-abdominal fat pad. The penguins of known ID were all lighter than their average body mass, however the body mass of 11/624 had actually increased from the time it was last observed alive (Table 1).

3.2. Gross pathological findings

3.2.1. Liver

Hepatomegaly with rounded lobar edges was evident in several of the penguins. The livers were weighed in eight penguins and ranged from 58.6 to 79.7 g (Table 2), with an average of 65.8 \pm 4 g for males and 65.7 \pm 3.3 g in females. This is notably greater than that of healthy Little Penguins in New Zealand (males: 42.4 \pm 14.6 g, females: 34.2 \pm 10.9 g; Hocken, 2000b). The liver mass, as a percentage of body mass, averaged 5.3 \pm 0.5 for males and 5.5 \pm 0.3 in females (Table 3). In this study, the ratio of liver to body mass for all penguins (5.4 \pm 0.3) was substantially greater than that obtained by Hocken (2000b) for healthy Little Penguins (3.67 \pm 0.8). Thus, all eight

Table 1
Details for each of the penguins.

Case no.	Sex	Area found	Date found dead	Date last seen alive	Body mass – dead (g)	Body mass – last seen alive	Av. body mass (g)
06/1172 ^a	M	PI	8/11/06	n/a	Unknown	n/a	n/a
08/1075 ^a	M	PI	28/7/08	n/a	1300	n/a	n/a
12/021 ^a	M	CS	6/10/11	n/a	1300	n/a	n/a
11/627 ^a	M	PI	3/11/11	23/9/11	1160	1220	1323 (3)
11/624 ^a	M	PI	6/11/11	25/10/11	1370	1210	1410 (24)
11/628 ^a	F	PB	8/11/11	23/9/11	1225	1450	1450 (1)
12/020 ^a	F	PK	8/11/11	n/a	1230	n/a	n/a
11/626 ^b	F	PI	9/11/11	n/a	1105	n/a	n/a
12/338 ^a	M	WS	30/04/12	n/a	1190	n/a	n/a
12/362 ^a	M	PI	30/05/12	n/a	1240	n/a	n/a

Area found – PI = Penguin Island, PB = Preston Beach, PK = Port Kennedy, CS = Cockburn Sound, WS = Warnbro Sound. The number of times each penguin was weighed to determine the average body mass is noted in parentheses (note an average was not possible for 11/628).

^a Adult.

^b Fledgling.

Table 2
Mass of the liver, spleen and abdominal fat pad in each case. Note that the organs were not weighed in all cases.

Case no.	Liver mass (g)	Spleen mass (g)	Abdominal fat pad
06/1172 ^a	Not weighed	Not weighed but splenomegaly noted	Not weighed
08/1075 ^a	69.4	Not weighed	Not weighed
11/627 ^a	79.7	5.0	8.3
11/624 ^a	Not weighed	9.6	Not weighed
11/628 ^a	63.7	4.2	16.1
11/626 ^b	61.2	7.5	5.4
12/020 ^a	72.1	3.8	8.3
12/021 ^a	58.6	3.3	4.1
12/338	62.4	6.9	6.7
12/362	59.0	5.9	15.0

^a Adult.

^b Fledgling.

individuals in which the liver was weighed had demonstrable hepatomegaly. Pinpoint flat white to off-white spots were disseminated over the hepatic capsular surfaces in 11/626, 11/627, 11/628, and 12/362 (Fig. 2). A focal area of metallic yellow speckling was noted on the capsular surface of the right liver lobe of 08/1075 (interpreted as urate precipitate).

3.2.2. Spleen

The spleen was enlarged in all cases (8/10) in which it was measured (Table 2). The average for both males (6.1 ± 1.0 g) and females (5.2 ± 1.2 g) was notably greater

than that found by Hocken (2000b) for healthy males (1.5 ± 1.0 g) and females (1.2 ± 0.7 g). Pinpoint white to off-white spots were scattered throughout the splenic parenchyma and over its capsular surface in the case of 11/624.

3.2.3. Abdominal fat pad

The abdominal fat pad was weighed in six of the penguins and ranged from 4.1 to 16.1 g (Table 2), with an

Table 3
Liver mass as a percentage of body mass.

Case no.	Liver	Gender
08/1075 ^a	5.4	Male
11/627 ^a	6.9	Male
12/021 ^a	4.5	Male
12/338	5.2	Male
12/362	4.8	Male
Mean \pm SD	5.3 ± 0.5	
11/626 ^b	5.5	Female
11/628 ^a	5.2	Female
12/020 ^a	5.9	Female
Mean \pm SD	5.5 ± 0.3	
All penguins	5.4 ± 0.3	

^a Adult.

^b Fledgling.

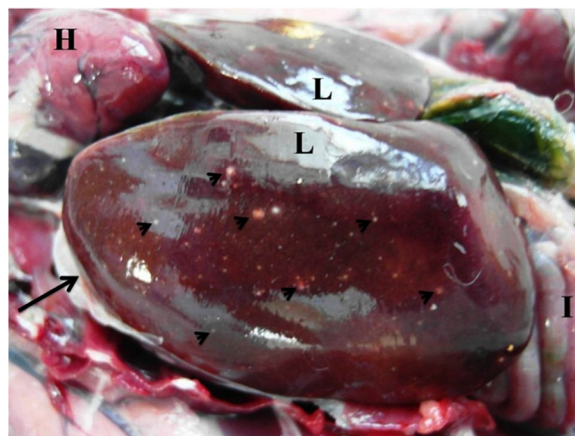


Fig. 2. Liver, penguin 12/362. H = heart, L = liver, I = intestines. Note the rounded hepatic lobar edges (arrow) and the multitudinous disseminated white to off-white flat pinpoint capsular discolourations (arrowheads).

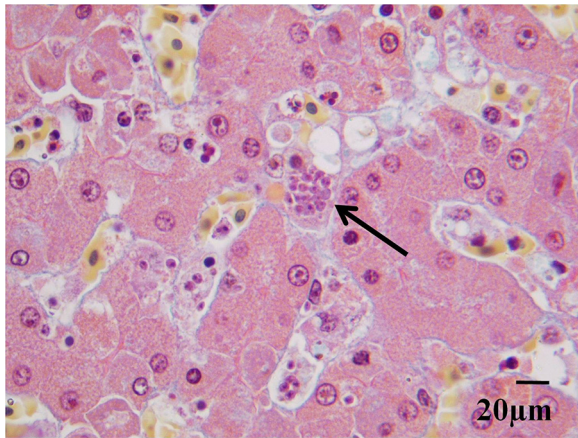


Fig. 3. Liver, penguin 12/362. 400× magnification, 2 μm thick, Martius Scarlet Blue stain. The arrow indicates a meront containing numerous cytomeres at the periphery of a necrotic focus.

average of 8.6 ± 2.3 g for the males and 9.9 ± 3.2 g for the females. The mass of the abdominal fat pad as a percentage of body mass varied from 0.3 to 1.3%. The average for the males was $0.7 \pm 0.2\%$ and $0.8 \pm 0.2\%$ for the females. This was lower than that for healthy males ($1.4 \pm 0.9\%$) and slightly lower than that for females ($0.9 \pm 0.8\%$) (Hocken, 2000b).

3.3. Histopathological findings

The degree of autolysis ranged from minimal to moderate-marked and rendered histopathological examination of some tissues difficult in some cases (however parasites were obvious in all ten cases).

The livers from all penguins exhibited varying degrees of multifocal, randomly scattered areas of hepatocellular necrosis (Table 4), which in the worst affected cases was multifocal to coalescing. Associated with these areas were varying numbers of parasites, present in all ten cases. The protozoa were present both extracellularly (scattered amongst necrotic tissue) and intracellularly within macrophages/Kupffer cells, hepatocytes and endothelial cells (the latter was difficult to confirm, given the host cell was significantly enlarged/distorted; however they were often in an intracellular location immediately adjacent sinusoids) at the periphery of the necrotic areas. Each parasite varied from 2 to 4 μm diameter and was round to ovoid with a dark basophilic (in haematoxylin and eosin stained sections) paracentral to eccentric nucleus and pale amphophilic cytoplasm (cytomere). When present intracellularly they were often clustered in small groups (approximately 20 μm diameter) of up to 18 cytomeres contained within a thin walled vacuole (meronts containing cytomeres) (Fig. 3). Increased numbers of macrophages were scattered throughout the liver (particularly associated with areas of necrosis) in five cases, with two of these also exhibiting heterophilic infiltrates (Table 4). Scattered hepatocytes and Kupffer cells (particularly at the periphery of necrotic areas) contained small amounts of intracytoplasmic pigment in four cases (Table 4). This material appeared to be comprised of two components

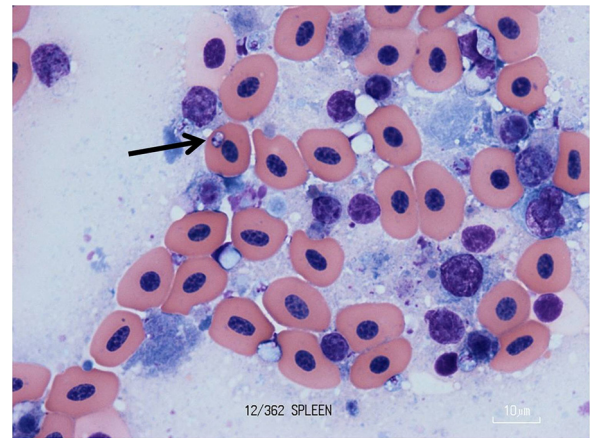


Fig. 4. Splenic impression smear, penguin 12/362. 1000× (oil immersion) magnification, Wright–Giemsa stain. The arrow indicates an intraerythrocytic merozoite prior to its development into a young gametocyte.

mixed together in varying amounts, depending on the area within the section. The first was opaque, golden-brown and granular (in haematoxylin and eosin stained sections), staining bright blue with Perl's Prussian Blue histochemistry, confirming its identity as haeme pigment (from erythrocyte degradation, given the extensive areas of necrosis with accompanying haemorrhage seen); rather than volutin pigment (which can be present in gametocytes of *Haemoproteus* spp.). This is because the latter is intraerythrocytic and also (being protein bound) does not react positively with Perl's Prussian Blue histochemistry. The second was refractile, appearing pale golden to tan (in haematoxylin and eosin stained sections; it was negative for Perl's Prussian Blue); it occasionally (but not reliably) polarised with the use of a polarising lens. Its morphology was interesting in that it was not granular, but rather linear and occasionally formed a lining around apparently clear vacuoles within the affected hepatocytes and Kupffer cells, often in close proximity to individual extracellular merozoites/cytomeres. Its exact identity is unknown; however it was felt it may represent the residual hyaline walls of ruptured meronts having released merozoites/cytomeres.

Multifocal areas of splenic parenchymal necrosis of varying size were observed in all penguins, the worst affected cases exhibited multifocal to coalescing necrosis. Cytomeres identical to those described in the liver were observed scattered extracellularly throughout the necrotic debris. Single to small groups (up to 8) of identical organisms, possibly meronts, were observed intracellularly within both macrophages as well as what appeared to be reticuloendothelial cells on the periphery of the necrotic areas. Rare intraerythrocytic organisms (consistent with intraerythrocytic merozoites prior to their development into young gametocytes) were observed in an impression smear of the spleen of 12/362 (Fig. 4). Accompanying the necrosis and intrasplenic parasites was a moderate to marked, multifocal to coalescing histiocytic and heterophilic (the latter often appearing degenerate) inflammatory infiltrate in five cases, with another two cases exhibiting solely a histiocytic infiltrate (Table 4).

Table 4
Summary of histopathological findings.

Case no.	Degree of autolysis	Hepatocellular necrosis	Pigment in Kupffer cells/hepatocytes	Splenic necrosis	Cardiac necrosis	Histiocytic infiltrate		Heterophilic infiltrate	
						Spleen	Liver	Spleen	Liver
06/1172	Minimal	✓		✓	✓				✓
08/1075	Moderate	✓		✓	✓	✓			
11/627	Moderate-marked	✓		✓		✓	✓	✓	
11/624	Mild-moderate	✓	✓	✓		✓	✓	✓	✓
11/628	Moderate-marked	✓		✓	✓	✓	✓	✓	
12/020	Mild	✓	✓	✓		✓	✓	✓	
12/021	Moderate	✓	✓	✓					
11/626	Moderate	✓		✓		✓		✓	
12/338	Moderate	✓	✓	✓					
12/362	Minimal	✓		✓	✓	✓	✓		✓

Pigment of mixed morphology as previously described in the liver was observed within the cytoplasm of scattered macrophages in the spleen of one case (12/338).

The lungs of 08/1075 and 11/624 exhibited increased numbers of inflammatory cells, predominantly macrophages; however they appeared to be confined within circulation, rather than being representative of a true inflammatory infiltrate. It is unknown whether this was related to a leukocytosis, as pre-mortem haematology was not performed; alternatively, it could have occurred secondary to stress-related leukocyte sequestration within the pulmonary vasculature. Rare parasites resembling merozoites ($\leq 1 \mu\text{m}$) were observed within alveolar macrophages in the case of 06/1172. Rare merozoite-like organisms were also found within the pulmonary interstitium in three cases (06/1172, 08/1075 and 11/624), but were not observed in the other cases.

Cardiac muscle from four cases (06/1172, 08/1075, 11/628 and 12/362) exhibited rare small focal to multifocal areas of cardiac myocyte degeneration and necrosis. 06/1172 additionally exhibited a focal histiocytic and lymphoplasmacytic periarteritis. Associated with these foci of necrosis and inflammation were sparse individual parasites resembling merozoites (each 1–2 μm), which appeared largely extracellular and scattered throughout the affected tissue.

3.4. PCR results and sequence analysis

Of the ten Little Penguin samples obtained from 2006 to 2012, only four (11/626, 12/020, 12/021 and 12/338) were positive for haemoparasite *cyt-b* (Fig. 5). When multiple PCR products from each positive bird were cloned and sequenced, each bird was found to harbour a single haemoparasite *cyt-b* sequence (not shown). When compared between birds, the sequences from 11/626, 12/021 and 12/338 were identical. The sequence from 12/020 differed by only a single nucleotide (not shown). When compared to public databases of *cyt-b* genes, the sequences were found to have 86–97% nucleotide identity to sequences from parasites previously identified as *Haemoproteus* subgenus *Parahaemoproteus*, and 80–84% identity to *Haemoproteus* subgenus *Haemoproteus*. The isolated sequences only shared 79–83% nucleotide identity to avian *Plasmodium* isolates. Further confirmation of the

Haemoproteus spp. nature of the sequences was established using phylogenetic analysis (Fig. 6). The sequences obtained from the Little Penguin samples were aligned to representative haemosporidian parasite *cyt-b* sequences obtained from GenBank. A phylogeny based on these alignments reveals that the Little Penguin sequences are clearly within a clade containing avian *Haemoproteus* spp., distinct from *Plasmodium* spp. Within the *Haemoproteus* clade they are most related to the *Haemoproteus* of the subgenus *Parahaemoproteus* as would be expected based on the nucleotide sequence identity. Previously isolated *Haemoproteus* sequences from Galapagos penguins were also in the *Parahaemoproteus* subgenus, however these are distinct from those detected in the Little Penguins.

4. Discussion

Haemoproteus spp. were identified by PCR in four dead penguins in 2011 and 2012, and morphologically identical tissue stage parasites were observed in all ten dead penguins associated with significant pathological changes sufficient to result in acute death. It is worth noting that parasites were not seen in histologically normal tissue, only in areas affected by significant pathological changes. Given that the birds were still in adequate to good body condition (as evidenced by adequate subcutaneous and intra-abdominal adipose); it is unlikely that the penguins died from starvation. This, combined with the fact that significant hepatic and splenic necrosis associated with *Haemoproteus* spp. was seen indicates that disseminated haemoproteosis was responsible for their deaths. These constitute only the second recorded cases of *Haemoproteus* spp. identified in any member of the Spheniscidae family (Levin et al., 2009), however this report is the first to demonstrate the presence of the parasite *in situ* associated with overt pathological changes. Moreover, it is the first time that *Haemoproteus* spp. infection has been implicated in the mortality of Little Penguins.

Although all ten birds tested here for haemosporidian infection using DNA analysis exhibited similar histopathological findings, six of the birds were negative by PCR. Previous investigators have reported similar false negatives using PCR based methodology (Richard et al., 2002; Beadell et al., 2004; Valkiunas et al., 2006). Inability to amplify haemosporidians from DNA extracted from

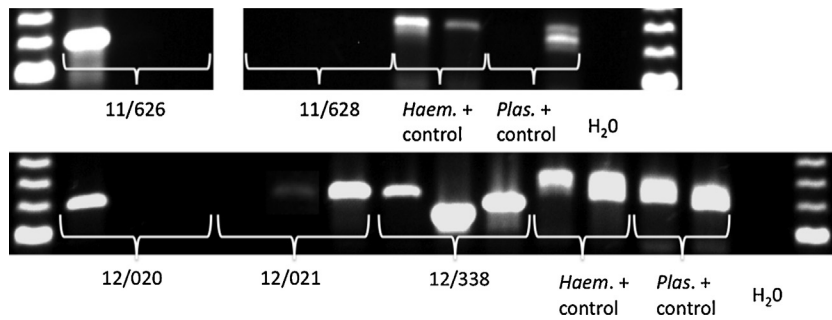


Fig. 5. Visualisation of PCR results of the four positive (11/626, 12/020, 12/021 and 12/338) and one representative negative (11/628) birds. Each sample was run in triplicate except for the positive controls, which were run in duplicate. H₂O was included as a negative control.

microscopically positive samples may be the result of degradation of the DNA sample over time; however this does not seem to be the case here since positive controls performed support the quality of the DNA. Alternatively, due to variation in tissue sampling, parasite DNA may be underrepresented in some samples. Microscopy alone can also result in false negatives due to low infection levels and often does not provide species level information on the parasite. For example, it is difficult or impossible to

use the morphology of the preerythrocytic tissue stages alone for accurate diagnosis of *Haemoproteus* spp. infection (Atkinson, 2008). Hence the value in the use of both molecular as well as microscopic methods in identifying infection (Valkiunas et al., 2008a) is supported by the results presented here.

The molecular methodology used in this study has been used previously to identify both *Haemoproteus* and *Plasmodium* spp. (Hellgren et al., 2004; Waldenström et al.,

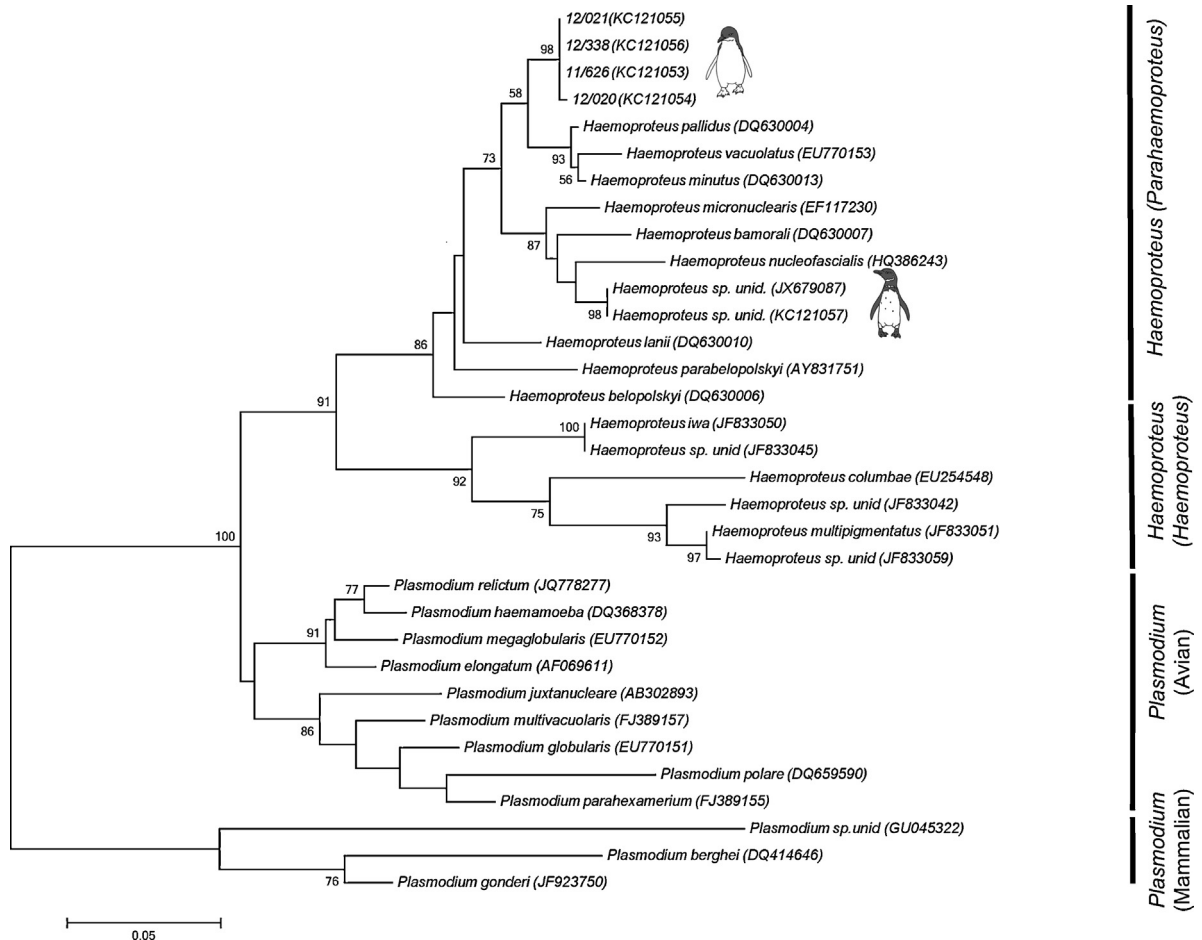


Fig. 6. Phylogenetic tree based on mitochondrial cytochrome *b* sequences of haemosporidians obtained from the liver of Little Penguins along with representatives obtained from GenBank.

2004). Indeed, the close phylogenetic relationship between *Plasmodium* and *Haemoproteus* spp. makes it difficult to design primers specific for one that would not amplify the other (Beadell and Fleischer, 2005). Only *Haemoproteus* was identified in the Little Penguins tested here, however it is impossible to rule out the presence of other infections based on negative results. Mixed infections can result in preferential amplification of DNA from the parasite with the greatest level of parasitemia and/or tissue burden (Valkiunas et al., 2006). Therefore PCR based methods can underestimate the occurrence of mixed *Haemoproteus* and *Plasmodium* spp. infections. One method to reduce such underestimation is by sequencing multiple independent clones from more than one independent PCR reaction, increasing the chances of identifying dual infections, as was performed here.

As the primers used only amplify the *Plasmodium* or *Haemoproteus* spp. *cyt-b* gene, we cannot rule out the possibility that the penguins could have also harboured *Leucocytozoon* spp. infections. In addition, the clinical observations of splenomegaly, hepatomegaly and histiocytic infiltrates in the current study not only correlate with previous reports of *Haemoproteus* spp. infections (Atkinson and Van Riper III, 1991; Ladds, 2009) but also those of *Plasmodium* and *Leucocytozoon* spp. infections (Fix et al., 1988; Graczyk et al., 1994; Valkiunas, 2005; Ko et al., 2008). However, the intraerythrocytic stage observed in 12/362 (noting that this case was PCR negative but histologically positive), did not deform the host cell nucleus, as would be expected of even young *Leucocytozoon* gametocytes (Valkiunas, 2005). Furthermore, the intralesional haemosporidians observed in the six “PCR-negative” cases were morphologically identical by light microscopy to that of the four PCR positive cases. However, many of the necropsy findings could also apply to infection with *Babesia* spp., and concurrent infection with *Babesia peircei* cannot be excluded. This piroplasm has been identified from Little Penguins in Australia (Peirce, 2000), and from Jackass Penguins (*Spheniscus demersus*) in South Africa (Erlé et al., 1993). Schizonts and trophozoites of *B. peircei* are difficult to differentiate from early intraerythrocytic stages of haemoproteids, the main difference being the absence of pigment in species of *Babesia* (Homer et al., 2000). However, refractile hyaline pigment of unknown significance was present in the livers of four out of the ten birds in this study, including one which was negative for *Haemoproteus* spp. by PCR (11/624). In addition, no organisms consistent with *Babesia* were noted in the splenic impression from 12/362 (P. Irwin, pers comm.), even though intraerythrocytic merozoites were present in this case. It should be noted that the molecular techniques that we used for haemosporidia do not detect *Babesia* spp.; therefore, the potential remains for mixed infections. However, ticks are vectors of *Babesia* spp., and although they are usually associated with nesting bird colonies, none were found on Little Penguins or in their nests in this study. Therefore we conclude that a diagnosis of *Haemoproteus* spp. is likely the case for all 10 dead penguins.

Haemoproteus spp. are normally considered benign in birds (Bennett et al., 1993b; Quillfeldt et al., 2010), although it has also been postulated that heavy infections of

Haemoproteus spp. may delay bird migration (Valkiunas and Iezhova, 2001). In addition, tissue stages of the parasite have been reported to cause necrosis of pectoral muscle fibres surrounding megaloschizonts in several avian species (Atkinson and Forrester, 1987; Atkinson, 2008; Olias et al., 2011); pneumonia-like symptoms (Valkiunas, 2005) and haemorrhage and necrosis associated with cyst rupture (Ferrell et al., 2007). However, all but one of these cases were from captive or laboratory animals, and the apparent lack of pathogenicity in the wild could be artefactual, arising from sampling birds that are in apparent good condition (*i.e.* aclinical/asymptomatic), and not those found ailing or dead (Valkiunas, 1998). Indeed, Ladds (2009) notes that the severity of the disease may be underestimated in wild birds.

The exact life cycle and pathogenesis of the *Haemoproteus* spp. identified is unknown; however it is assumed that they follow the general scheme of the haemoproteid life cycle as elaborated by Valkiunas (2005). Unlike the members of the genus *Plasmodium*, to which they are closely related, haemoproteids undergo merogony (asexual reproduction) in tissues, rather than circulating erythrocytes. It is these developing preerythrocytic tissue stages that have, albeit in a limited number of cases, been reported to be pathogenic, causing significant myonecrosis (Ferrell et al., 2007; Atkinson, 2008; Olias et al., 2011) as they mature and rupture to release merozoites capable of invading erythrocytes. No gross signs of skeletal myonecrosis were noted in our cases, and unfortunately skeletal muscle was not examined histologically. Four of the ten cases did exhibit small foci of cardiac myonecrosis associated with sparse intralesional merozoites.

No haemoparasites were previously found in blood smears taken from penguins on Penguin Island in 1992 (Jones and Shellam, 1999a), nor from any of the other 124 penguins from the Perth region that have been necropsied since 2004. This, and the fact that only a limited number of preerythrocytic tissue stages (and rare intraerythrocytic merozoites in the case of the spleen of 12/362) were seen in all individuals, suggests that these cases represent the outcome of lethal abortive development of *Haemoproteus* spp. in a non-adapted penguin population (Valkiunas, 2011). This is similar to the cases reported by Olias et al. (2011) in which non-adapted exotic parrots of various species in Europe died acutely; with PCR identifying parasites with 99–100% homology to *Haemoproteus* spp. known to be highly prevalent in wild European songbirds (without causing overt disease or mortality).

It is assumed, similar to the situation in human malaria (particularly that associated with *Plasmodium falciparum*) (Clark et al., 2006); that the birds, having had heavy parasite burdens (*i.e.* disseminated infection), died acutely as a result of the excessive release of pro-inflammatory cytokines (in particular TNF and IL-1; “cytokine storm”) secondary to the damage caused by the tissue stages (and therefore before gametocytes were able to develop), leading to systemic inflammatory response syndrome (SIRS). It is unclear whether these birds had a significant anaemia which might also have contributed to disease and death by way of exacerbating tissue ischaemia. It is also unclear how much of a role the pathogen-associated molecular pattern

(PAMP) glycosylphosphatidylinositol anchor (GPI) plays in the pathogenesis of *Haemoproteus*-associated disease; this highly conserved molecule is present in the membranes of many parasitic protozoa and elicits the synthesis of pro-inflammatory cytokines as well as nitric oxide (Ropert and Gazzinelli, 2000); thus having been implicated in human malaria via its recognition by host cell toll-like receptor 2 (TLR2) (Oakley et al., 2011).

It is probable that the dead penguins found elsewhere originated from Penguin Island, as shown by previous radio and satellite tracking data (Cannell unpubl. data). The penguin located near Garden Island (12/021) could have been either from the colony at Garden Island or from Penguin Island, as penguins from both colonies forage in Cockburn Sound (Cannell unpubl. data). As penguins are unlikely to be bitten by competent vectors of *Haemoproteus* spp. while at sea, and the penguins do not land on the mainland, they were probably infected whilst at their colony. The competent vectors of *Haemoproteus* spp. include biting midges (Ceratopogonidae) and louse flies (Hippoboscidae) (Valkiunas, 2005). The presence of louse flies on Little Penguins has rarely, if ever, been observed (Cannell pers. obs.), with fleas, lice and ticks being the listed ectoparasites on these birds (Obendorf and McColl, 1980). In contrast, biting midges are likely to be on Penguin Island and are therefore the most likely competent vector for *Haemoproteus* spp.

Apart from the 10 cases in this study, no haemosporidians have been previously found in Little Penguins in this region (Cannell unpubl. data, Jones and Shellam, 1999a), nor from colonies elsewhere in Australia and New Zealand (Quillfeldt et al., 2011). However, the distribution, density, biology, physiology and transmission rates of vectors are influenced by changes in climate and habitat (Harvell et al., 2002; Rogers and Randolph, 2006). In addition, the growth rate of the haemoparasite can also change with altered climatic conditions (Harvell et al., 2002). An anomalously strong La Niña from September 2010 and throughout 2011 was responsible for above average sea surface temperatures along the West Australian coast (Pearce et al., 2011; Pearce and Feng, 2013) and high tides causing inundation of coastal wetlands (City of Mandurah, 2011, 2012). It also affected other climatic conditions, with the mean minimum and maximum daily temperature well above average throughout the year resulting in the hottest year on record, with a record number of days $\geq 32^\circ\text{C}$. This temperature increase was combined with above average rainfall during the austral spring as well as heavy rainfall in the first two weeks of December (beginning of the austral summer) (Australian Bureau of Meteorology). It is therefore possible that these climatic changes contributed to a sudden increase of vectors present on the island and hence in the cases of avian haemosporidiosis documented.

Even if prevailing climatic conditions are suitable for an increased presentation of the haemoparasite, it must first be available to be transmitted and for it to shift hosts. There is evidence to suggest that some *Haemoproteus* spp. have a low host specificity and thus can be transmitted between birds belonging to different families (Fallon et al., 2005; Križanauskienė et al., 2006). However, Beadell et al. (2004) found that the stability of jumps between different host families in tropical songbirds is low, though this

could be due to the type of birds they sampled as well as their isolation. Fallon et al. (2005) found that one lineage of *Haemoproteus* was found in different bird families, dependent on the geographic region. Furthermore, the families of birds in one region were migrants that overwintered in another region. There are several species of migratory birds on Penguin Island. For example, bridled terns (*Sterna anaethetus*), caspian terns (*Hydroprogne caspia*) and crested terns (*S. bergii*) breed on Penguin Island, but migrate from the island after breeding. The caspian and crested terns are likely to move to wetland areas within the localised region (Cannell, 2004), whereas bridled terns migrate to the Sulawesi Sea for the austral winter (Dunlop and Johnstone, 1994). Further research needs to be undertaken to determine if any of these species are reservoir hosts for the *Haemoproteus* spp. found in the Little Penguins.

The dramatic increase in penguins presenting with disseminated, fatal haemosporidiosis is of concern. It is known that island hosts tend to be more susceptible to pathogens compared to mainland hosts due to lower exposure to pathogens (Levin et al., 2011 and refs within Valkiunas, 2005). In addition, the prevalence of *Haemoproteus* parasites has been shown to increase in other bird species when stressed (Quillfeldt et al., 2010). While the penguins that died from *Haemoproteus* spp. infections were not in poor body condition, the average weight of the fat pad as a percentage of body weight was lower than that identified by Hocken (2000b), particularly for the males. However, it is unknown if the Hocken (2000b) study included penguins in a premoult or moult stage, which would carry a large reserve of fat. Nevertheless, in this study, the infected penguins for which we have body-mass data were lighter than their mean body mass, and a greater than average number of penguins were found to have died as a result of starvation in 2011 (Cannell unpubl. data). It is therefore likely that the infected penguins in the present study were in a state of stress and perhaps thus more vulnerable to succumbing to the infection.

Inclusion of the *Haemoproteus* sequences from Little Penguins continues to support two main haemoparasite clades, one containing mammalian *Plasmodium* and the other avian parasites. The avian clade is further subdivided into *Plasmodium* and *Haemoproteus* clades. Previous classification divided the genus *Haemoproteus* into two subgenera, *Haemoproteus* and *Parahaemoproteus*, that form sister clades in phylogenetic analyses (Valkiunas et al., 2008b; Levin et al., 2011). Martinsen et al. (2008) found *Haemoproteus* and *Parahaemoproteus* each form distinct monophyletic clades based on limited gene sequence analysis and suggested raising *Parahaemoproteus* to be a distinct genus. The addition of the sequences isolated from the Little Penguins continue to support the conclusion that *Parahaemoproteus* and *Haemoproteus* are distinct sister clades with the penguin parasites in the *Parahaemoproteus* subgenus.

5. Conclusion

Fatal infection by haemosporidian parasites, identified by molecular means in conjunction with histopathological findings as *Haemoproteus* spp., is reported in four

of ten Little Penguins in Western Australia. Additionally, the other six cases exhibited identical histopathological findings with the presence of morphologically identical intralesional parasites. One of these also had rare intraerythrocytic merozoites. This population has previously been free of infection. Recent changes in climate, ocean currents and the presence of other species of birds which could introduce *Haemoproteus* spp. to the area could all be contributing factors ultimately leading to these infections. Further studies are needed to ascertain the epidemiology and pathology of this parasite; and indeed, to investigate the possibility of mixed apicomplexan infections in this population.

Acknowledgements

We would like to thank various departments within the Department of Environment and Conservation (DEC), who have funded the necropsies (WA Marine Monitoring Program, Species and Communities Branch, Marine Science Program) and the retrieval of many of the dead penguins (WA Marine Monitoring Program). DEC had no role in the study design, analysis and interpretation of data, writing the report or decision to submit the article. We would also like to thank all those people who have found, and informed the corresponding author about the dead penguins. Additionally, we would like to thank Daithi Murray (School of Veterinary and Life Sciences, Murdoch University) for performing the DNA extractions, Gerard Spoelstra and Michael Slaven, technologists in histology (School of Veterinary and Life Sciences, Murdoch University), for their invaluable assistance in preparing the histological sections, Victoria Hansen (University of New Mexico) for the penguin drawings used in Fig. 6, Dr Cheryl Moller (Resident in Clinical Pathology, School of Veterinary and Life Sciences, Murdoch University) for Fig. 4, and Dr Peter Irwin (Associate Professor, School of Veterinary and Life Sciences, Murdoch University) for his expertise and advice on *Babesia*. We also thank Dr Gediminas Valkiunas (Nature Research Centre, Lithuania) for his very constructive comments on the manuscript. DNA samples were transferred from Western Australia to the Museum of Southwestern Biology at the University of New Mexico under U.S. Department of Agriculture permit no. 102547. The Museum of Southwestern Biology at the University of New Mexico provided assistance in importing the samples to the US. All DNA based analysis utilised the Molecular Biology Facility at the University of New Mexico which is supported by grant no. P20 GM103452-09 from the National Institutes of Health Institutional Development Award program.

References

- Allison, F.R., Desser, S.S., Whitten, L.K., 1978. Further observations on the life cycle and vectors of the haemosporidian *Leucocytozoon tawaki* and its transmission to the Fiersland crested penguin. *N. Z. J. Zool.* 5, 371–374.
- Atkinson, C.T., Forrester, D.J., 1987. Myopathy associated with megaloschizonts of *Haemoproteus meleagridis* in a wild turkey from Florida. *J. Wildl. Dis.* 23, 495–498.
- Atkinson, C.T., Van Riper III, C., 1991. Pathogenicity and epizootology of avian haematzoa: Plasmodium, Leucocytozoon, and Haemoproteus. In: Loye, J.E., Zuk, M. (Eds.), *Bird-Parasite Interactions: Ecology, Evolution and Behaviour*. Oxford University Press, New York, NY, pp. 19–48.
- Atkinson, C.T., 2008. *Haemoproteus*. In: *Parasitic Diseases of Wild Birds*. Wiley-Blackwell, Ames, IA.
- Australian Bureau of Meteorology. <http://www.bom.gov.au/climate/current/annual/wa/archive/2011.perth.shtml>
- Beadell, J.S., Gering, E., Austin, J., Dumbacher, J.P., Peirce, M.A., Pratt, T.K., Atkinson, C.T., Fleischer, R.C., 2004. Prevalence and different host-specificity of two avian blood parasite genera in the Australo-Papuan region. *Mol. Ecol.* 13, 3829–3844.
- Beadell, J.S., Fleischer, R.C., 2005. A restriction enzyme-based assay to distinguish between avian hemospordians. *J. Parasitol.* 91, 683–685.
- Bennett, G.F., Bishop, M.A., Pierce, M.A., 1993a. Checklist of the avian species of *Plasmodium* Marchiafava & Celli, 1885 (Apicomplexa) and their distribution by avian family and Wallacean life zones. *Syst. Parasitol.* 26, 171–179.
- Bennett, G.F., Peirce, M.A., Ashford, R.W., 1993b. Avian Haematzoa: mortality and pathogenicity. *J. Nat. Hist.* 27, 993–1001.
- Bueno, M.G., Lopez, R.P.G., de Menezes, R.M.T., Costa-Nascimento, M. de J., de Castro Lima, G.F.M., de Sousa Araujo, R.A., Guida, F.J.V., Kirchgatter F K., 2010. Identification of *Plasmodium relictum* causing mortality in penguins (*Spheniscus magellanicus*) from Sao Paulo Zoo, Brazil. *Vet. Parasitol.* 173, 123–127.
- Cannell, B.L., 2004. Distribution of the major marine fauna in the Perth metropolitan area (Yanchep to Mandurah). Report No. MMS/CWC/LNE/MMP/SEMP. SIMP-79. Department of Conservation and Land Management, Perth, Australia.
- Cicero, C., Johnson, N.K., 2001. Higher-level phylogeny of new world vireos (Aves: Vireonidae) based on sequences of multiple mitochondrial DNA genes. *Mol. Phylogenet. Evol.* 20 (1), 27–40.
- City of Mandurah, 2011. Environmental Health Services Mosquito Management Program Annual Report 2010/11. 32 pp.
- City of Mandurah, 2012. Environmental Health Services Mosquito Management Program Annual Report 2011/12. 40 pp.
- Clarke, J.T., Kerry, K.R., 1993. Diseases and parasites of penguins. *Korean J. Pol. Res.* 4, 79–96.
- Clark, I.A., Budd, A.C., Alleva, L.M., Cowden, W.B., 2006. Human malarial disease: a consequence of inflammatory cytokine release. *Malar. J.* 5, 85 (Review).
- Dann, P., Cullen, M., Weir, I., 1996. National Review of the Conservation Status and Management of Australian Little Penguin Colonies: Final Report. The Australian Nature Conservation Agency, Melbourne, Australia.
- Dunlop, J.N., Johnstone, R.E., 1994. The migration of bridled terns *Sterna anaethetus* breeding in Western Australia. *Corella* 18, 125–129.
- Dunlop, J.N., Wooller, R.D., 1986. Range extensions and the breeding seasons of seabirds in south-western Australia. *Rec. West. Aust. Mus.* 12, 389–394.
- Erlé, R., Huchzermeyer, F.W., Bennett, G.F., Brossy, J.J., 1993. *Babesia peircei* sp. nov. from the Jackass penguin. *S. Afr. J. Zool.* 28, 88–90.
- Fallon, S.M., Bermingham, E., Ricklefs, R.E., 2005. Host specialization and geographic localization of avian malaria parasites: a regional analysis in the Lesser Antilles. *Am. Nat.* 165, 466–480.
- Ferrell, S.T., Snowden, K., Garner, M., Marlar, A.B., Lung, N.P., 2007. Fatal hemoprotozoal infections in multiple avian species in a zoological park. *J. Zool. Wildl. Med.* 38, 309–316.
- Fix, A.S., Waterhouse, C., Greiner, E.C., Stoskopf, M.K., 1988. *Plasmodium relictum* as a cause of avian malaria in wild-caught Magellanic penguins (*Spheniscus magellanicus*). *J. Wildl. Dis.* 24, 610–619.
- Fleischman, R.W., Sladen, W.J.L., Melby, E.C., 1968. Malaria (*Plasmodium elongatum*) in captive African penguins (*Spheniscus demersus*). *J. Am. Vet. Med. Assoc.* 153, 928–935.
- Graczyk, T.K., Cranfield, M.R., McCutchan, T.F., Bicknese, E.J., 1994. Characteristics of naturally acquired malaria infections in naive juvenile African black-footed penguins (*Spheniscus demersus*). *Parasitol. Res.* 80, 634–637.
- Harrigan, K.E., 1992. Causes of mortality of Little Penguins *Eudyptula minor* in Victoria. *Emu* 91, 273–277.
- Harvell, C.D., Mitchell, C.E., Ward, J.R., Altizer, S., Dobson, A.P., Ostfeld, R.S., Samuel, M.D., 2002. Climate warming and disease risks for terrestrial and marine biota. *Science*, 296.
- Hill, A.G., Howe, L., Gartrell, B.D., Alley, M.R., 2010. Prevalence of *Leucocytozoon* spp., in the endangered yellow-eyed penguin *Megadyptes antipodes*. *Parasitology* 137, 1477–1485.
- Hellgren, O., Waldenström, J., Bensch, S., 2004. A new PCR assay for simultaneous studies of *Leucocytozoon*, *Plasmodium*, and *Haemoproteus* from avian blood. *J. Parasitol.* 90, 797–802.
- Hocken, A.G., 2000a. Cause of Death in Blue Penguins (*Eudyptula m. minor*) in North Otago, New Zealand. *N. Z. J. Zool.* 27, 305–309.

- Hocken, A.G., 2000b. Internal organ weights of the blue penguin *Eudyptula minor*. N. Z. J. Zool. 27, 299–304.
- Homer, M.J., Aguilar-Delfin, I., Telford III, S.R., Krause, P.J., Persing, D.H., 2000. Babesiosis. Clin. Microbiol. Rev. 13, 451–469.
- Jones, H.I., Shellam, G.R., 1999a. The occurrence of blood-inhabiting protozoa in captive and free-living penguins. Polar Biol. 21, 5–10.
- Jones, H.I., Shellam, G.R., 1999b. Blood parasites in penguins, and their potential impact on conservation. Mar. Ornithol. 27, 181–184.
- Jourdain, E., Gauthier-Clerc, M., Bicolout, D.J., Sabatier, P., 2007. Bird migration routes and risk for pathogen dispersion into western Mediterranean wetlands. Emerg. Infect. Dis. 13, 365–372.
- Ko, K.-N., Kang, S.-C., Jung, J.-Y., Bae, J.-H., Kim, J.-H., 2008. Avian malaria associated with *Plasmodium* spp. infection in a penguin in Jeju Island. Korean J. Vet. Res. 48, 197–201.
- Križanauskienė, A., Helligren, O., Kosarev, V., Sokolov, L., Bensch, S., Valiunas, G., 2006. Variation in host specificity between species of avian hemosporidian parasites: evidence from parasite morphology and Cytochrome B Gene sequences. J. Parasitol. 92, 1319–1324.
- Ladds, P., 2009. Pathology of Native Australian Wildlife. CSIRO Publishing, Victoria, Australia.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23 (21), 2947–2948.
- Levin, I.L., Outlaw, D.C., Vargas, F.H., Parker, P.G., 2009. Plasmodium blood parasite found in endangered Galapagos penguins (*Spheniscus mediculus*). Biol. Conserv. 142, 3191–3195.
- Levin, I.L., Valkiūnas, G., Santiago-Alarcon, D., Cruz, L.L., Iezhova, T.A., O'Brien, S.L., Hailer, F., Dearborn, D., Sreiber, E.A., Fleischer, R.C., Ricklefs, R.E., Parker, P.G., 2011. Hippoboscids-transmitted *Haemoproteus* parasites (Haemosporida) infect Galapagos Pelecaniform birds: evidence from molecular and morphological studies, with a description of *Haemoproteus iwa*. Int. J. Parasitol. 41, 1019–1027.
- Martinsen, E.S., Perkins, S.L., Schall, J.J., 2008. A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera); evolution of life-history traits and host switches. Mol. Phylogenet. Evol. 47, 261–275.
- Norman, F.I., Du Guesclin, P.B., Dann, P., 1992. The 1986 “wreck” of Little Penguins *Eudyptula minor* in Western Victoria. Emu 91, 369–376.
- Oakley, M.S., Gerald, N., McCutcheon, T.F., Aravind, L., Kumar, S., 2011. Clinical and molecular aspects of malaria fever. Trends Parasitol. 27 (10), 442–449 (Review).
- Obendorf, D.L., McColl, K., 1980. Mortality in Little Penguins (*Eudyptula minor*) along the coast of Victoria. J. Wildl. Dis. 16, 251–260.
- Olias, P., Wegelin, M., Freter, S., Gruber, A.D., Klopffleisch, R., 2011. Avian malaria deaths in parrots, Europe. Emerg. Infect. Dis. 17, 950–952.
- Outlaw, D.C., Ricklefs, R.E., 2011. Rerooting the evolutionary tree of malaria parasites. Proc. Natl. Acad. Sci. 108 (32), 13183–13187.
- Pearce, A., Lenanton, R., Jackson, G., Moore, J., Feng, M., Gaughan, D., 2011. The “Marine Heat Wave” off Western Australia during the summer of 2010/11. Fisheries Research Report No. 222. Fisheries Research Division, Perth, Western Australia.
- Pearce, A., Feng, M., 2013. The rise and fall of the “marine heat wave” off Western Australia during the summer of 2010/2011. J. Mar. Syst. 111–112, 139–156.
- Peirce, M.A., 2000. A taxonomic review of avian piroplasms of the genus *Babesia* Starcovici, 1893 (Apicomplexa: Piroplasmorida: Babesiidae). J. Nat. Hist. 34, 317–332.
- Peirce, M.A., 2005. A checklist of the valid avian species of *Babesia* (Apicomplexa: Piroplasmorida), *Haemoproteus*, *Leucocytozoon* (Apicomplexa: Haemosporidia), and *Hepatozoon* (Apicomplexa: Haemogregarinidae). J. Nat. Hist. 39, 3621–3632.
- Peirce, M.A., Greenwood, A.G., Stidworthy, M.F., 2005. *Leucocytozoon* in captive penguins. Vet. Rec. 157, 819–820.
- Quillfeldt, P., Martinez, J., Henniscke, J., Ludynia, K., Gladbach, A., Masello, J.F., Riou, S., Merino, S., 2010. Haemosporidian blood parasites in seabirds – a comparative genetic studies of species from Antarctic to tropical habitats. Naturwissenschaften 97, 809–817.
- Quillfeldt, P., Arriero, E., Martinez, J., Masello, J.F., Merino, S., 2011. Prevalence of blood parasites in seabirds – a review. Front. Zool. 8, 1–10.
- Richard, A., Sehgal, R.N.M., Jones, H.I., Smith, T.B., 2002. A comparative analysis of PCR-based detection methods of avian malaria. J. Parasitol. 88, 819–822.
- Rogers, D.J., Randolph, S.E., 2006. Climate change and vector-borne diseases. Adv. Parasitol. 62, 346–381.
- Roport, C., Gazzinelli, R.T., 2000. Signaling of immune system cells by glycosylphosphatidylinositol (GPI) anchor and related structures derived from parasitic protozoa. Curr. Opin. Microbiol. 3, 395–403.
- Rose, K., 2001. May. Australian Registry of Wildlife Pathology Summary of Cases. Wildlife Diseases Association, Australasian Section Newsletter.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
- Valkiūnas, 1998. Haematzoa of wild birds: peculiarities in their distribution and pathogenicity. Bull. Scand. Soc. Parasitol. 8, 39–46.
- Valkiūnas, G., Iezhova, T.A., 2001. A comparison of the blood parasites in three subspecies of the Yellowtail Wagtail *Motacilla flava*. J. Parasitol. 87, 930–934.
- Valkiūnas, G., 2005. Avian Malaria Parasites and other Haemosporidia. CRC Press, Florida.
- Valkiūnas, G., Bensch, S., Iezhova, T.J., Križanauskienė, A., Helligren, O., Bolshakov, C.V., 2006. Nested cytochrome b polymerase chain reaction diagnostics underestimate mixed infections of avian haemosporidian parasites: microscopy is still essential. J. Parasitol. 92, 418–422.
- Valkiūnas, G., Iezhova, T.A., Križanauskienė, A., Palinauskas, V., Sehgal, R.N.M., Bensch, S., 2008a. A comparative analysis of microscopy and PCR-based detection methods for blood parasites. J. Parasitol. 94, 1395–1401.
- Valkiūnas, G., Iezhova, T.A., Loiseau, C., Chasar, A., Smith, T.B., Sehgal, R.N.M., 2008b. New species of haemosporidian parasites (Haemosporida) from African rainforest birds, with remarks on their classification. Parasitol. Res. 103 (5), 1213–1228.
- Valkiūnas, G., 2011. Haemosporidian vector research: marriage of molecular and microscopical approaches is essential. Mol. Ecol. 20, 3084–3086.
- Waldenström, J., Bensch, S., Hasselwuijst, D., Östman, Ö., 2004. A new nested polymerase chain reaction method very efficient in detecting plasmodium and haemoproteus infections from avian blood. J. Parasitol. 90, 191–194.
- Wienecke, B.C., 1993. The size and breeding patterns of Little Penguins *Eudyptula minor* in Australia: a comparative study. PhD Thesis Murdoch University, Western Australia.
- Wooller, R.D., Dunlop, J.N., Klomp, N.I., Meathrel, C.E., Wienecke, B.C., 1991. Seabird abundance, distribution and breeding patterns in relation to the Leeuwin Current. J. R. Soc. West. Aust. 74, 129–132.

1219	lizard	>chain16 Position:GL344250.1:24152-28586(+) Name=ac_chain16 GeneID=ENSACAG00000006471 ProteinID=ENSACAP00000006333 Score=239.0 E-value=9.9e-69 Length=4435 Comment=Overlaps single annotation & in proteins
1220	lizard	alpha1 24152 24403 ... 1 81 62.6 1.30E-17 +
1221	lizard	alpha2 27182 27460 ... 89 180 121.8 3.80E-35 +
1222	lizard	alpha2 28330 28586 ... 1 88 54.2 2.00E-17 +
1223	lizard	>chain17 Position:GL343777.1:59265-67362(-) Name=ac_chain17 GeneID=ENSACAG00000004068 ProteinID=ENSACAP00000004029 Score=194.3 E-value=2.4e-55 Length=8098 Comment=Overlaps single annotation & in proteins
1224	lizard	alpha1 67120 67362 ... 1 80 32.2 1.30E-08 -
1225	lizard	alpha2 61314 61589 ... 89 180 113.9 7.90E-33 -
1226	lizard	lg 59265 59522 ... 1 88 48.2 2.30E-15 -
1227	lizard	>chain18 Position:GL343619.1:393831-405607(+) Name=ac_chain18 GeneID=ENSACAG00000003922 ProteinID=ENSACAP00000003807 Score=186.2 E-value=4.6e-53 Length=11777 Comment=Overlaps single annotation & in proteins
1228	lizard	alpha1 393831 394073 ... 1 79 79.9 9.30E-23 +
1229	lizard	alpha2 399902 399123 ... 1 87 180 54.2 3.80E-15 +
1230	lizard	lg 405350 405607 ... 1 88 52.1 1.30E-16 +
1231	lizard	>chain19 Position:2:189407192-189410738(-) Name=ac_chain19 GeneID=ENSACAG00000011395 ProteinID=ENSACAP00000011331 Score=157.3 E-value=3.2e-44 Length=3547 Comment=Overlaps multiple annotations; one in proteins
1232	lizard	alpha1 189407192 189407431 ... 1 80 43.1 7.40E-12 +
1233	lizard	alpha2 189409106 189409387 ... 89 180 74 5.20E-21 +
1234	lizard	lg 189410466 189410738 ... 1 88 40.2 8.20E-13 +
1235	x_tropicalis	>chain01 Position:GL173531.1:22383-54316(-) Name=xt_mhc1a GeneID=ENSGT000000017065 ProteinID=ENSGTPO00000037199 Score=318.5 E-value=5.6e-91 Length=31394 Comment=Overlaps single annotation & in proteins
1236	x_tropicalis	alpha1 54029 54316 ... 1 92 124.3 2.00E-95 -
1237	x_tropicalis	alpha2 24637 24927 ... 86 180 134.5 2.00E-38 -
1238	x_tropicalis	lg 22383 22640 ... 1 88 59.7 1.40E-18 -
1239	x_tropicalis	>chain02 Position:GL172781.1:692190-695424(-) Name=xt_chain02 GeneID=ENSGT000000017938 ProteinID=ENSGTPO00000038903 Score=276.9 E-value=1.3e-78 Length=3235 Comment=Overlaps single annotation & in proteins
1240	x_tropicalis	alpha1 695140 695424 ... 1 91 93.7 2.30E-26 -
1241	x_tropicalis	alpha2 693372 693647 ... 89 180 126 6.40E-36 -
1242	x_tropicalis	lg 691290 692447 ... 1 88 57.2 8.90E-18 -
1243	x_tropicalis	>chain03 Position:GL173193.1:381302-386177(-) Name=xt_chain03 GeneID=ENSGT000000009073 ProteinID=ENSGTPO00000019930 Score=264.1 E-value=6.6e-75 Length=4876 Comment=Overlaps single annotation & in proteins
1244	x_tropicalis	alpha1 385944 386177 ... 1 80 84.6 1.10E-23 -
1245	x_tropicalis	alpha2 383798 384070 ... 89 180 119.7 4.60E-34 -
1246	x_tropicalis	lg 381302 381559 ... 1 88 59.8 1.30E-18 -
1247	x_tropicalis	>chain04 Position:GL173193.1:354444-357204(-) Name=xt_chain04 GeneID=ENSGT000000009076 ProteinID=ENSGTPO00000007870 Score=257.4 E-value=7.3e-73 Length=2761 Comment=Overlaps single annotation & in proteins
1248	x_tropicalis	alpha1 356971 357204 ... 1 80 83.5 2.50E-23 -
1249	x_tropicalis	alpha2 355712 355981 ... 90 180 115.2 9.70E-33 -
1250	x_tropicalis	lg 354444 354701 ... 1 88 58 7.00E-18 -
1251	x_tropicalis	>chain05 Position:GL173063.1:196677-971978(-) Name=xt_chain05 GeneID=ENSGT0000000031111 ProteinID=ENSGTPO00000035926 Score=257.1 E-value=7.6e-73 Length=5202 Comment=Overlaps single annotation & in proteins
1252	x_tropicalis	alpha1 97145 971978 ... 1 80 71.7 7.40E-20 -
1253	x_tropicalis	alpha2 970069 970350 ... 88 180 125.4 9.30E-36 -
1254	x_tropicalis	lg 96677 967034 ... 1 88 60 1.10E-18 -
1255	x_tropicalis	>chain06 Position:GL173063.1:979645-982170(-) Name=xt_chain06 GeneID=ENSGT0000000031629 ProteinID=ENSGTPO00000035924 Score=256.9 E-value=7.9e-73 Length=2526 Comment=Overlaps single annotation & in proteins
1256	x_tropicalis	alpha1 981937 982170 ... 1 80 72.7 3.90E-20 -
1257	x_tropicalis	alpha2 980268 980549 ... 88 180 120.1 6.50E-34 -
1258	x_tropicalis	lg 979645 979902 ... 1 88 64.1 5.60E-20 -
1259	x_tropicalis	>chain07 Position:GL173193.1:1938-6154(-) Name=xt_chain07 GeneID=ENSGT000000009064 ProteinID=ENSGTPO00000019874 Score=252.8 E-value=1.6e-71 Length=4217 Comment=Overlaps single annotation & in proteins
1260	x_tropicalis	alpha1 5915 6154 ... 1 80 67.5 1.30E-18 -
1261	x_tropicalis	alpha2 4307 4585 ... 88 180 125.9 6.70E-36 -
1262	x_tropicalis	lg 1938 2195 ... 1 88 59.4 1.80E-18 -
1263	x_tropicalis	>chain08 Position:GL173193.1:87977-91470(-) Name=xt_chain08 GeneID=ENSGT000000009068 ProteinID=ENSGTPO00000019887 Score=248.3 E-value=4e-70 Length=3494 Comment=Overlaps single annotation & in proteins
1264	x_tropicalis	alpha1 91237 91470 ... 1 80 85.7 5.50E-20 -
1265	x_tropicalis	alpha2 90479 90751 ... 89 180 111.6 1.20E-31 -
1266	x_tropicalis	lg 87977 88234 ... 1 88 58.2 4.40E-18 -
1267	x_tropicalis	>chain09 Position:GL173193.1:259437-268566(-) Name=xt_chain09 GeneID=ENSGT0000000033941 ProteinID=ENSGTPO00000038336 Score=244.7 E-value=4e-69 Length=9130 Comment=Overlaps single annotation & in proteins
1268	x_tropicalis	alpha1 268333 268566 ... 1 80 83.2 3.00E-23 -
1269	x_tropicalis	alpha2 264625 264888 ... 92 180 102.3 6.40E-29 -
1270	x_tropicalis	lg 259437 259694 ... 1 88 59.2 2.10E-18 -
1271	x_tropicalis	>chain10 Position:GL173193.1:425428-427883(-) Name=xt_XB-Gene-5894693 GeneID=ENSGT000000016466 ProteinID=ENSGTPO00000019932 Score=243.3 E-value=1.1e-68 Length=2456 Comment=Overlaps single annotation & in proteins
1272	x_tropicalis	alpha1 427650 427883 ... 1 80 67.5 1.30E-18 -
1273	x_tropicalis	alpha2 426799 427071 ... 90 180 117.1 2.70E-33 -
1274	x_tropicalis	lg 425428 425685 ... 1 88 58.7 3.00E-18 -
1275	x_tropicalis	>chain11 Position:GL173193.1:298049-301990(-) Name=xt_chain11 GeneID=ENSGT000000009077 ProteinID=ENSGTPO00000038337 Score=241.7 E-value=3.3e-68 Length=3942 Comment=Overlaps single annotation & in proteins
1276	x_tropicalis	alpha1 301757 301990 ... 1 80 85.1 8.00E-24 -
1277	x_tropicalis	alpha2 300142 300405 ... 92 180 98 1.30E-27 -
1278	x_tropicalis	lg 298049 298306 ... 1 88 58.6 2.10E-18 -
1279	x_tropicalis	>chain12 Position:GL173193.1:58340-69297(-) Name=xt_chain12 GeneID=None ProteinID=None Score=237.9 E-value=3.7e-67 Length=10958 Comment=No overlapping annotations
1280	x_tropicalis	alpha1 69028 69297 ... 1 89 75 8.10E-21 -
1281	x_tropicalis	alpha2 67653 67925 ... 89 180 101.4 1.20E-28 -
1282	x_tropicalis	lg 58340 58597 ... 1 88 61.5 3.80E-19 -
1283	x_tropicalis	>chain13 Position:GL173193.1:16498-36539(-) Name=xt_chain13 GeneID=None ProteinID=None Score=235.8 E-value=1.8e-66 Length=20042 Comment=No overlapping annotations
1284	x_tropicalis	alpha1 36270 36539 ... 1 89 79.7 3.20E-22 -
1285	x_tropicalis	alpha2 19082 19360 ... 87 180 95.7 5.50E-27 -
1286	x_tropicalis	lg 16498 16755 ... 1 88 59.5 1.70E-18 -
1287	x_tropicalis	>chain14 Position:GL176203.1:8516-10992(-) Name=xt_XB-Gene-5879324 GeneID=ENSGT000000015263 ProteinID=ENSGTPO00000033387 Score=235.6 E-value=3e-66 Length=2477 Comment=Overlaps single annotation & in proteins
1288	x_tropicalis	alpha1 10759 10992 ... 1 80 72.2 5.30E-20 -
1289	x_tropicalis	alpha2 9585 9860 ... 89 180 110.5 2.60E-31 -
1290	x_tropicalis	lg 8516 8773 ... 1 88 52.9 2.20E-16 -
1291	x_tropicalis	>chain15 Position:GL173193.1:280451-286573(-) Name=xt_chain15 GeneID=ENSGT0000000032875 ProteinID=ENSGTPO00000019918 Score=234.4 E-value=5.4e-66 Length=6123 Comment=Overlaps single annotation & in proteins
1292	x_tropicalis	alpha1 286340 286573 ... 1 80 85.7 5.50E-20 -
1293	x_tropicalis	alpha2 284289 284552 ... 92 180 90.9 1.60E-25 -
1294	x_tropicalis	lg 280451 280708 ... 1 88 57.8 5.90E-18 -
1295	x_tropicalis	>chain16 Position:GL173193.1:549443-557399(-) Name=xt_chain16 GeneID=ENSGT0000000033834 ProteinID=ENSGTPO00000019884 Score=228.4 E-value=2.6e-64 Length=7957 Comment=Overlaps single annotation & in proteins
1296	x_tropicalis	alpha1 557160 557399 ... 1 81 78.3 8.60E-22 -
1297	x_tropicalis	alpha2 550621 550890 ... 90 180 89.2 5.20E-25 -
1298	x_tropicalis	lg 549443 549700 ... 1 88 60.9 5.90E-19 -
1299	x_tropicalis	>chain17 Position:GL174797.1:5412-7994(-) Name=xt_chain17 GeneID=ENSGT000000005015 ProteinID=ENSGTPO0000003705 Score=228.1 E-value=3.6e-64 Length=2583 Comment=Overlaps single annotation & in proteins
1300	x_tropicalis	alpha1 7836 7994 ... 1 53 52.2 4.70E-14 -
1301	x_tropicalis	alpha2 6873 7154 ... 88 180 117.4 2.20E-33 -
1302	x_tropicalis	lg 5412 5669 ... 1 88 58.5 3.50E-18 -
1303	x_tropicalis	>chain18 Position:GL173193.1:198714-203900(-) Name=xt_chain18 GeneID=ENSGT0000000033980 ProteinID=ENSGTPO00000019901 Score=224.4 E-value=4.2e-63 Length=5187 Comment=Overlaps single annotation & in proteins
1304	x_tropicalis	alpha1 203610 203900 ... 1 96 64.5 1.00E-17 -
1305	x_tropicalis	alpha2 200770 201036 ... 90 180 100.3 2.60E-28 -
1306	x_tropicalis	lg 198714 198971 ... 1 88 59.6 1.60E-18 -
1307	x_tropicalis	>chain19 Position:GL173193.1:181934-186043(-) Name=xt_chain19 GeneID=ENSGT0000000030073 ProteinID=ENSGTPO00000063559 Score=219.2 E-value=1.4e-61 Length=4110 Comment=Overlaps single annotation & in proteins
1308	x_tropicalis	alpha1 185774 186043 ... 1 89 55.5 4.70E-15 -
1309	x_tropicalis	alpha2 184523 184789 ... 90 180 104.1 1.90E-29 -
1310	x_tropicalis	lg 181934 182191 ... 1 88 59.6 1.60E-18 -
1311	x_tropicalis	>chain20 Position:GL173193.1:168702-175190(-) Name=xt_XB-Gene-5757686 GeneID=ENSGT000000014445 ProteinID=ENSGTPO00000060129 Score=219.0 E-value=1.7e-61 Length=6489 Comment=Overlaps multiple annotations; one in proteins
1312	x_tropicalis	alpha1 171960 175190 ... 1 96 66.4 2.80E-18 -
1313	x_tropicalis	alpha2 169694 169960 ... 90 180 92.6 5.00E-26 -
1314	x_tropicalis	lg 168702 168959 ... 1 88 60 1.20E-18 -
1315	x_tropicalis	>chain21 Position:GL173193.1:130442-134120(-) Name=xt_chain21 GeneID=ENSGT000000009089 ProteinID=ENSGTPO00000019892 Score=216.8 E-value=7.2e-61 Length=3679 Comment=Overlaps single annotation & in proteins
1316	x_tropicalis	alpha1 133830 134120 ... 1 96 59.8 2.50E-16 -
1317	x_tropicalis	alpha2 131363 131629 ... 90 180 98.1 1.10E-27 -
1318	x_tropicalis	lg 130442 130699 ... 1 88 58.9 2.60E-18 -
1319	x_tropicalis	>chain22 Position:GL173063.1:992422-996364(-) Name=xt_chain22 GeneID=ENSGT0000000030689 ProteinID=ENSGTPO00000059118 Score=215.5 E-value=2e-60 Length=3943 Comment=Overlaps single annotation & in proteins
1320	x_tropicalis	alpha1 996128 996364 ... 1 80 59.5 3.10E-16 -
1321	x_tropicalis	alpha2 994765 995034 ... 89 180 98.4 9.70E-28 -
1322	x_tropicalis	lg 992422 992679 ... 1 88 57.6 6.60E-18 -
1323	x_tropicalis	>chain23 Position:GL173193.1:444157-447998(-) Name=xt_chain23 GeneID=ENSGT0000000034352 ProteinID=ENSGTPO00000058545 Score=207.6 E-value=3.8e-58 Length=3842 Comment=Overlaps single annotation & in proteins
1324	x_tropicalis	alpha1 447759 447998 ... 1 81 60.4 1.70E-16 -
1325	x_tropicalis	alpha2 445747 446016 ... 90 180 86.5 3.20E-24 -
1326	x_tropicalis	lg 444157 444414 ... 1 88 60.7 7.00E-19 -
1327	x_tropicalis	>chain24 Position:GL173193.1:512159-517175(-) Name=xt_chain24 GeneID=ENSGT0000000033504 ProteinID=ENSGTPO00000058304 Score=206.1 E-value=1e-57 Length=5017 Comment=Overlaps single annotation & in proteins
1328	x_tropicalis	alpha1 516906 517175 ... 1 89 68.2 8.10E-19 -
1329	x_tropicalis	alpha2 514907 515176 ... 90 180 79.1 4.80E-22 -
1330	x_tropicalis	lg 512159 512416 ... 1 88 58.8 2.70E-18 -
1331	x_tropicalis	>chain25 Position:GL173193.1:585862-595649(-) Name=xt_chain25 GeneID=ENSGT0000000030785 ProteinID=ENSGTPO00000053844 Score=198.9 E-value=1.9e-55 Length=9788 Comment=Overlaps single annotation & in proteins
1332	x_tropicalis	alpha1 595509 595649 ... 1 47 51.7 6.30E-14 -
1333	x_tropicalis	alpha2 589070 589045 ... 89 180 93 3.70E-26 -
1334	x_tropicalis	lg 585862 586119 ... 1 88 54.2 8.10E-17 -
1335	x_tropicalis	>chain26 Position:GL173193.1:530734-535327(-) Name=xt_chain26 GeneID=ENSGT0000000032895 ProteinID=ENSGTPO00000061810 Score=195.7 E-value=2.4e-54 Length=4594 Comment=Overlaps single annotation & in proteins
1336	x_tropicalis	alpha1 535058 535327 ... 1 89 67.2 1.70E-18 -
1337	x_tropicalis	alpha2 532891 533160 ... 90 180 79.8 3.10E-22 -
1338	x_tropicalis	lg 530734 530985 ... 1 88 48.7 4.60E-15 -
1339	x_tropicalis	>chain27 Position:GL173193.1:490297-495395(-) Name=xt_chain27 GeneID=None ProteinID=None Score=191.1 E-value=4.1e-53 Length=5099 Comment=No overlapping annotations
1340	x_tropicalis	alpha1 495126 495395 ... 1 89 67.4 1.50E-18 -

1463	zebrafish	alpha1	11300253	11300414	..	1	55	[49.2	2.30E-14	+
1464	zebrafish	alpha2	11300605	11300886	..	88	180]	38.4	3.50E-11	+
1465	zebrafish	lg	11300985	11301251	..	1	88	[40.5	1.30E-13	+
1466	zebrafish	>chain17 Position=25:11216200-11217255(+)	Name=dr_si:ch211-147g22.4		GeneID=ENSDARG00000097275	ProteinID=ENSDARPO0000129196	Score=127.2	E-value=2.3e-37	Length=1056	Comment=Overlaps single annotation & in proteins	
1467	zebrafish	alpha1	11216200	11216343	..	1	49	[40.5	8.40E-12	+
1468	zebrafish	alpha2	11216610	11216888	..	89	180]	48.9	2.80E-14	+
1469	zebrafish	lg	11216989	11217255	..	1	88	[37.8	9.60E-13	+
1470	zebrafish	>chain18 Position=25:11082577-11085035(+)	Name=dr_BX572619.1		GeneID=ENSDARG00000046057	ProteinID=ENSDARPO0000067714	Score=123.0	E-value=3.6e-36	Length=2459	Comment=Overlaps multiple annotations; one in proteins	
1471	zebrafish	alpha1	11082577	11082726	..	1	51	[44	7.80E-13	+
1472	zebrafish	alpha2	11084375	11084653	..	89	180]	39.6	1.60E-11	+
1473	zebrafish	lg	11084769	11085035	..	1	88	[39.4	2.90E-13	+
1474	zebrafish	>chain19 Position=25:11281133-11284387(+)	Name=dr_si:ch211-147g22.3		GeneID=ENSDARG00000096830	ProteinID=ENSDARPO0000128887	Score=120.9	E-value=1.3e-35	Length=3255	Comment=Overlaps single annotation & in proteins	
1475	zebrafish	alpha1	11281133	11281408	..	1	95	[43.9	8.60E-13	+
1476	zebrafish	alpha2	11283740	11284006	..	89	180]	35.4	2.70E-10	+
1477	zebrafish	lg	11284121	11284387	..	1	88	[41.6	5.60E-14	+
1478	zebrafish	>chain20 Position=25:11378000-11380505(+)	Name=dr_si:key-52p2.7		GeneID=ENSDARG00000097766	ProteinID=ENSDARPO0000128245	Score=119.3	E-value=5e-35	Length=2506	Comment=Overlaps single annotation & in proteins	
1479	zebrafish	alpha1	11378000	11378149	..	1	51	[44.5	5.40E-13	+
1480	zebrafish	alpha2	11379840	11380118	..	89	180]	37.8	5.50E-11	+
1481	zebrafish	lg	11380299	11380505	..	1	88	[37	1.70E-12	+
1482	zebrafish	>chain21 Position=8:47095167-47096314(+)	Name=dr_CU571323.1		GeneID=ENSDARG00000016227	ProteinID=ENSDARPO000002240	Score=115.1	E-value=1.2e-33	Length=1148	Comment=Overlaps single annotation & in proteins	
1483	zebrafish	alpha1	47095167	47095424	..	1	88	[53.9	9.10E-16	+
1484	zebrafish	alpha2	47095620	47095895	..	89	180]	28.6	2.90E-08	+
1485	zebrafish	lg	47096045	47096314	..	1	88	[32.6	4.40E-11	+
1486	zebrafish	>chain22 Position=25:11342613-11350193(+)	Name=dr_si:key-52p2.4		GeneID=ENSDARG00000096977	ProteinID=ENSDARPO0000128999	Score=109.3	E-value=3.8e-32	Length=7581	Comment=Overlaps single annotation & in proteins	
1487	zebrafish	alpha1	11342613	11342780	..	1	57	[43.1	1.40E-12	+
1488	zebrafish	alpha2	11349554	11349820	..	89	180]	25.9	1.80E-07	+
1489	zebrafish	lg	11349927	11350193	..	1	88	[40.3	1.50E-13	+
1490	zebrafish	>chain23 Position=25:11315123-11316131(+)	Name=dr_si:ch211-147g22.2		GeneID=ENSDARG00000051712	ProteinID=ENSDARPO0000067872	Score=107.1	E-value=1.5e-31	Length=1009	Comment=Overlaps single annotation & in proteins	
1491	zebrafish	alpha1	11315123	11315284	..	1	55	[45.2	3.50E-13	+
1492	zebrafish	alpha2	11315492	11315770	..	89	180]	19.2	1.70E-05	+
1493	zebrafish	lg	11315865	11316131	..	1	88	[42.7	2.60E-14	+
1494	zebrafish	>chain24 Position=25:11257981-11263809(-)	Name=dr_CR339041.2		GeneID=ENSDARG00000051710	ProteinID=ENSDARPO00000084819	Score=105.0	E-value=1e-30	Length=5829	Comment=Overlaps single annotation & in proteins	
1495	zebrafish	alpha1	11263579	11263809	..	1	76	[42.3	2.50E-12	-
1496	zebrafish	alpha2	11259351	11258629	..	89	180]	29.2	1.90E-08	-
1497	zebrafish	lg	11257981	11258253	..	1	88	[33.5	2.20E-11	-
1498	zebrafish	>chain25 Position=25:11063812-11066177(-)	Name=dr_BX572619.2		GeneID=ENSDARG00000086127	ProteinID=ENSDARPO0000067869	Score=101.0	E-value=1.1e-29	Length=2366	Comment=Overlaps single annotation & in proteins	
1499	zebrafish	alpha1	11066001	11066177	..	1	60	[56.9	1.20E-16	-
1500	zebrafish	alpha2	11064293	11064427	..	89	130]	3.2	0.94	-
1501	zebrafish	lg	11063812	11064078	..	1	88	[40.9	9.80E-14	-
1502	zebrafish	>chain26 Position=25:11363689-11372081(+)	Name=dr_si:key-52p2.6		GeneID=ENSDARG00000051713	ProteinID=ENSDARPO0000067874	Score=98.1	E-value=1e-28	Length=8393	Comment=Overlaps multiple annotations; one in proteins	
1503	zebrafish	alpha1	11363689	11363888	..	1	70	[27.9	4.50E-08	+
1504	zebrafish	alpha2	11365741	11366007	..	89	180]	33.8	8.20E-10	+
1505	zebrafish	lg	11371815	11372081	..	1	88	[36.4	2.70E-12	-
1506	zebrafish	>chain27 Position=3:23763318-23765350(-)	Name=dr_BX682558.1		GeneID=ENSDARG00000055813	ProteinID=ENSDARPO0000072675	Score=95.1	E-value=5.1e-28	Length=2033	Comment=Overlaps single annotation & in proteins	
1507	zebrafish	alpha1	23765204	23765350	..	1	48	[39.6	1.60E-11	-
1508	zebrafish	alpha2	23763867	23763962	..	87	118]	12.5	0.0016	-
1509	zebrafish	lg	23763318	23763587	..	1	88	[43	2.00E-14	-
1510	zebrafish	>chain28 Position=25:11071627-11074602(+)	Name=dr_BX572619.3		GeneID=ENSDARG00000087161	ProteinID=ENSDARPO000010996	Score=84.6	E-value=7.5e-25	Length=2976	Comment=Overlaps single annotation & in proteins	
1511	zebrafish	alpha1	11074339	11074602	..	1	89	[13.3	0.00097	-
1512	zebrafish	alpha2	11072049	11072318	..	92	180]	30.7	7.00E-09	-
1513	zebrafish	lg	11071627	11071902	..	1	88	[40.6	1.10E-13	-
1514	zebrafish	>chain29 Position=8:47125558-47128530(-)	Name=dr_chain29		GeneID=None	ProteinID=None	Score=83.9	E-value=1.7e-24	Length=2973	Comment=No overlapping annotations	
1515	zebrafish	alpha1	47128387	47128530	..	1	49	[37.7	5.90E-11	-
1516	zebrafish	alpha2	47125931	47126005	..	156	180]	11.2	0.0039	-
1517	zebrafish	lg	47125558	47125827	..	1	88	[35	7.50E-12	-
1518	zebrafish	>chain30 Position=25:11228370-11229374(+)	Name=dr_chain30		GeneID=None	ProteinID=None	Score=70.7	E-value=1.1e-20	Length=1005	Comment=No overlapping annotations	
1519	zebrafish	alpha1	11228370	11228633	..	1	88	[31.8	3.10E-09	+
1520	zebrafish	alpha2	11228895	11228993	..	138	180]	1.2	3.7	+
1521	zebrafish	lg	11229108	11229374	..	1	88	[37.7	9.80E-13	+
1522	zebrafish	>chain31 Position=3:34038903-34043015(-)	Name=dr_lghd		GeneID=ENSDARG00000096248	ProteinID=None	Score=21.7	E-value=1.1e-05	Length=4113	Comment=Overlaps single annotation; not in proteins	
1523	zebrafish	alpha1	34042932	34043015	..	1	28	[2.4	1.7	-
1524	zebrafish	alpha2	34039305	34039361	..	157	180]	0.1	7.7	-
1525	zebrafish	lg	34038903	34039169	..	1	88	[19.2	8.10E-07	-

Supplementary Table 3. Predicted MHC class I sequences.

>md_UA1 Chain=chain01 Position=2:269726572-269727822(-) GeneID=ENSMODG00000010596
ProteinID=ENSMODP00000013275 Score=358.1 E-value=6.7e-108 Length=1251 Comment=Overlaps single
annotation & in proteins
PGTMKPSLLSLFVLGVVLTETRAGSHSMRYFHTAMSRPELGDSQFITVGYVDDQQFVRF
DSSSESQRMPEPRAAWMDKMEDEEPPNYWEEQTQIYRKNAQIYRRNLETLRGYYNQSQGGLH
TIQRMYGCEVHPDGSFRKGFYQLAYDGRDYIALHRETLTWTAAADPGAENTKRKWEAERSI
AERYKAYLEETCLOWLKYLQMGKDVLLRTPPSARVSRHSGPDGEVSLRCRAQGFYPAE
ISLTWLRDGEELQDTEF IETRPGGDGTFOKWAAVAMAPGQEDRYSCRVOHEALAQPLSL
RWGTPSSGLSLFPSSPEASSLWVIVGVTAGVLALVTAVVAGAVILRRRNSGRE

>md_UA3 Chain=chain02 Position=2:270107254-270108503(-) GeneID=ENSMODG00000014450
ProteinID=ENSMODP00000017908 Score=352.0 E-value=4.5e-106 Length=1250 Comment=Overlaps single
annotation & in proteins
MKPSLLSLFVLGVVLTETRAGSHSMRYFYTSMSRPELGDSQFISVGYVDDQQFVRFDSS
SESQRAEPRAAWMDQMEEDRNYWEGQTQINRRTAQITRVLDLETALGYYNQSRGGLHTIQ
RMYGCEVHPDGSFRKGFYQLAYDGRDYIALHRETLTWTAAADPGAENTKRKWEAERSFAER
VKAYLEETCQVQVKYLQMGEDVLLRTPPSARVSLRCRAQGFYPAEISLTWLRDGEELQ
QDTEF IETRPGGDGTFOKWAAVAMAPGQEDRYSCRVOHEALAQPLSLRWEPEAPSVVWVIV
GVTAGVLVLTAVVAGAVILRRRNSGGKGGAYVPAADKDSAQGSVSLTVA

>md_UC Chain=chain03 Position=2:38025176-38026415(+) GeneID=ENSMODG00000003927
ProteinID=ENSMODP00000004814 Score=339.8 E-value=1.8e-102 Length=1240 Comment=Overlaps single
annotation & in proteins
MELSLLSLFLVGVVLTETRAGSHSMRYFHTAMSRPELGDSQFITVGYVDDQQFVRFDSS
SESQRMPEPRAAWMDKMEDEEPPNYWEEQTQIYRKNAQIYRGNLETLRGYYNQSHGGLHTIQ
RMIGCEVSPPEGSFRKGFYQFAYDGHDIYIALHSETLSWTAADTGAENTKRKWEPEKSIAS
YKAYLEKECTHWLKKYLEMGKDVLLRTPPSVTRVTRYSGPDGEVSLRCRAHGFYPAEISL
TWLRDGEELQDTEF IETRPGGDGTFOKWAAVAMAPGQEDRYSCRVOHEALAQPLSLRWE
PEASSVWVIVGVTAGVLVLTAVVAGAVILRRRNSGGKGGAYVPAADKDSAQGSVSLTA
TGEA

>md_UB Chain=chain04 Position=2:38000208-38001442(+) GeneID=ENSMODG00000003914
ProteinID=ENSMODP00000004796 Score=335.9 E-value=2.5e-101 Length=1235 Comment=Overlaps single
annotation & in proteins
MELSLLSLFLMGLVVALTQTRAGSHSMRYFHTAMSRPELGDSQFITVGYVDDQQFVRFDSS
SESQRMPEPRAAWMDKMEDEEPPNYWEEQTQIYRKNAQIYRRNLETLRGYYNQSQGGLHTIQ
RMIGCEVSAEGLRKGFFYQFAYDGHDIYIALHTETLSWTAADTGAENTKRKWEPEKSIAS
YKAYLEKECTHWLKKYLMGKDVLLRTPPSARVTRHRGPDGEVSLRCRAHGFYPAEISL
MWLRDGEELQDTEF IETRPGGDGTFOKWAAVAMAPGQEDRYSCRVOHEALAQPLSLRWE
PEASSLWVIVGVTAGVLVLTAVVAGAVILRRRNSGGKGGAYVPAADKDSAQGSVSLTA
TGEA

>md_UA4 Chain=chain05 Position=2:269928699-269929953(-) GeneID=None ProteinID=None Score=335.4
E-value=3.7e-101 Length=1255 Comment=No overlapping annotations
SISGSHSLRYCHTTMSRPELGDSQIIAVGYVDDQQFVRFDSSSESQRMPEPRAAWMDKME
EDPDYWERNTHICRSNAQIAQVDLKTALILYNYQSQGGLHTIQSMMGCEVSRDGSFRKGF
YQLAYDGHDIYIALDRETLTWTAAADTGAENTKRKWEAERSIAERRKDYVEVECVQWLKYL
QMGKDVLLRTRVSRHSGPDGEVSLRCRAQGFYPAEISLMWLRDGEELQDTEF IETRP
GDGTFOKWAAVAMAPGQEDRYSCRVOHEALAQPLSL

>md_UG Chain=chain06 Position=2:270248895-270250095(+) GeneID=ENSMODG000000023743
ProteinID=ENSMODP000000031501 Score=326.5 E-value=2.3e-98 Length=1201 Comment=Overlaps multiple
annotations; one in proteins
AQFYRVLDLETLRGYFNQSEEGVHTLQRLFGCEVSPDGSFKRFSFYQYGDGHDIY

>md_MR1 Chain=chain07 Position=2:75227311-75239120(+) GeneID=ENSMODG00000006527
ProteinID=ENSMODP00000008100 Score=319.6 E-value=2.1e-96 Length=11810 Comment=Overlaps single
annotation & in proteins
MIFLPHLLMTSATQGGDARTHSLRYFRLGLSDSNQGMPEFISVGYVDSHPITSYDSNGRQ
KMPQASWMEENLGDHWEKYTQLLRGWQOTFKIELRALQNHYNHTGGFHLYQRMIGCELL
EDGSTTGFLQYAYDGKDFIVFNKESLSWIAMDNVARLTKQAWEANRRELRYQKNWLETEC
IAWLKFLDFGKDLQRTETPLLSGCKKSSTGITTLICKAYGFPEITMTWIKNGELI
TQIEHGDILPSGDGTQYTWVS IDIDPQSKDHYSQVEHNNFLKVLHVPVELKTIISLSPY
VEIVSGVIIIASFLIGLGFIVYKRKQSELKEVGYVPTPVK

>md_UF Chain=chain08 Position=2:270337954-270339246(-) GeneID=ENSMODG00000014336
ProteinID=ENSMODP00000017918 Score=317.1 E-value=9.5e-96 Length=1293 Comment=Overlaps multiple
annotations; one in proteins
MESVVLSSLLLEAVMLPEIKASFHSLRYFQTSMSLPGRKPQFISVGYVDDLQFMLFDGNSP
NQREEPRAVLDQMDQDQYWEKNSRISRETAQTFFEVGLQNLVYNYQSEGLHTYQRLVGC
EVSHNRMFRRGFEQYAYDQDYISLDFETLSWTAADIPSLNSKHKWEMPNIAQRQKTYL
EKNCVQWLYKYLEIGNETLLRADPPSVRVSRRHSGPDGEVSLRCRAQGFYPAEISLTWLRD
GEEQLQDTEF IETRPGGDGTFOKWAAVAMAPGQEDRYSCRVOHEALAQPLSLRWEPEAPS
VWVIVGVTAGVLVLTAVVAGAVILRRRNSGGKGGAYVPAADKSSQGSVSLIATG

>md_UE Chain=chain09 Position=2:269513803-269515199(-) GeneID=ENSMODG00000014624
ProteinID=ENSMODP00000018269 Score=316.4 E-value=1.5e-95 Length=1397 Comment=Overlaps single
annotation & in proteins
FLTMEFFVLSLLLLGLTTLTETWAGKCFHSLKYFYTMSLPELIEPRFTAVVYVDDQQIL
SFDSASQSMERPRTQWIKVEPPDYWERVTRISREDSQCNRMLQKVSTNYNHSEGGVHT
YQRQAGCEVFSNGSFSRGFAQYGFDRDYLTLDGTGLRWVAADAGALNNKRWKWEADQRIA
ENWKIYLEGECVHSLQRYLENGKERLLRADPPSARVTRHTSSDGEVTLKCRADFPYPAEI
SLVWLREGEEQLQDTEFIETRPAGDGTQKWAAVEMLSGSEHKYTCRVQHEGLPEPLFLK
WGKDGESQFSSIGLSAGVITILLLLAAVIIGVVVWRKNTSGREGSYPATASNDSAQGS
VSFTARSKI

>md_UL Chain=chain14 Position=2:269554750-269556130(-) GeneID=ENSMODG00000014624
ProteinID=ENSMODP00000018269 Score=287.7 E-value=7.3e-87 Length=1381 Comment=Overlaps single
annotation & in proteins
SDLGPHSLKYFYTTMSLPELIEPRFTAVVYVDDQQILSFDSASQSMERPRTQWIKVEPP
DYWERVTRISREDSQCNRMLQKVSTNHCCVEVFNKSGVSHGFVQFAYDGYDIKLDTEFL
RWTALDVRALNIKRWKADPNAAFVKIYLEEECVRWLQRHLENGKETLLRTARVTRHTR
SDEVTLRCAQDFYPAEISLTLWRDGEQEQLQDTEFIETRPAGDGTQKWAAVEMLSGS
EHRYTCVQVQHEGLPEPLFL

>md_UJ Chain=chain10 Position=2:270235977-270237289(-) GeneID=ENSMODG00000023744
ProteinID=ENSMODP00000031502 Score=314.4 E-value=5.1e-95 Length=1313 Comment=Overlaps single
annotation & in proteins
LETLQGLYNQSEGGIHLQKMGFCEVSANGNFRRGFKRFAYDGH DYL

>md_UA2 Chain=chain11 Position=2:268510859-268512046(-) GeneID=ENSMODG00000014839
ProteinID=ENSMODP00000018565 Score=302.9 E-value=1.5e-91 Length=1188 Comment=Overlaps single
annotation & in proteins
PGAMKPSLPSLFVLGVVALTETRAGEEGSVFHHCHGPGPEVRDSQFISVGYVDDQQVFRF
DSSSESQRLELRAAWMDKVDDEYWERQTQISRRTAQIYGVNLETLLGGYVNSRGGHLTLO
RMLGCEVHPDGSFRKGFQDQAYDGHDIYIALDRETLTWTAAADTGAENTKRWKAERSFAEG
QKAYLEETCVQWLKYLQMGKDVLLRTDPPSARVSRHSGPDGEVSLRCRAQGFYPAEISL
TLWRDGEQEQLQDTEFIETRPAGDGTQKWAAVAMAPGQEDRYSCRVQHEALAQPLSLRWE
PEASSLWVIAGVTAGVLMVMTAVVAGAVILRRRNSSGGKGGAYVPAADKDSAQGSVSLTV
TGEA

>md_UK Chain=chain12 Position=2:270355967-270357857(-) GeneID=ENSMODG00000014336
ProteinID=ENSMODP00000017918 Score=298.0 E-value=4.7e-90 Length=1891 Comment=Overlaps multiple
annotations; one in proteins
LSSAFHSLRYFQTSMSLGRKPKQFISVGYVDDLQFMLFDGNSPNQREEPAPWLDQMDQD
YWEKNSRISRETAQTFEVLQNLLVYVYVNSGEGGLHTYQRVLGCEVSHNRMFRRGFEQYA
YDGDQYISLDFETLSWTAADIPSLNSKHKWEEMPNIAQRQKTYLEKNCVWLYKYLEIGN
ETLLRAARVTHHISHEGKVMLRCWARDYPADISLIWLRDGEQEQLHNTEFIETRPAGDGT
FQKWATVAMAPGQEDRYSCRVQHEGLPEPLTL

>md_UI Chain=chain13 Position=2:270297448-270298687(-) GeneID=ENSMODG00000014352 ProteinID=None
Score=293.5 E-value=8.6e-89 Length=1240 Comment=Overlaps single annotation; not in proteins
YISGSHSKSYFVTIMSSPELQEPKRVVVGVDAAQIEYFDSDSHANQRVEPRAQWMMHKMGQ
DYWQQGTRIARAETQNYRVGLETLRNYYNQSKGGVHVFTLIGCEISSDGSFKRGIHQ
AYDGLDIYISLDFEYTYTWAAVVALITKRWKESDGKFLKRGKAYLEEECVQWLKHLELDG
KDSLRLTAQVTHHAGSKGEVILQCRAQGFYPSIISLAWLRDGEQEQLQDTEFIETRPAGDGT
TFQKWAASVLTSGQEYIICIQHEGLPEPLTL

>md_UH Chain=chain16 Position=2:270376723-270378000(-) GeneID=ENSMODG00000014352 ProteinID=None
Score=268.8 E-value=4.1e-81 Length=1278 Comment=Overlaps single annotation; not in proteins
SVSASHSLRYFSTAVSRPELGDTRYISVGYVDDQQVFRFDSDEGQREQESRAPWMDNMDQ
EDPDYWEKSTRISRNTQIYGVNLETLRGYYNESQGGGAGVHTVQRMMSGCEVSTDCRFMR
GFYQYG*DGRDHRLDSGTLTWTMDTGAENTKRWKWEADKFLERYKAYLEDYVAYLHT
YLEMGKDILLRTHLLSE*PSIGGEVTLRCGAQGFYPADISLIWLRDGEQEQLHNTEFIET
PGDGTQKWAAVAMALGQEDRYSCPQREGLPEPLTL

>md_UM Chain=chain15 Position=2:269490678-269492082(-) GeneID=None ProteinID=None Score=270.2
E-value=8.2e-82 Length=1405 Comment=No overlapping annotations
ILSGSHSMRYIDTVVTAAPRLGLERWFTLVGYVDDQQVFRFHNSNNSQSTIEGAPWIELE
TPDYWEREKRLKDSQHWSGVHTLQRMYGCEVFNKSGVSTFFLLYDGDQKLSLDPETL
NWLASDSVALSIKLLDADRS PAERWVHLTVTCVQWLRLHLEKKEKTTLLRTVQVTRHT
SDSEVILKRAWGFYPAEISLIWLRNDEEQLQDTEYIDTRPGDGTQKWAAVEMPFNGE
EKYTCRVQHEGLPEPLFL

>md_HFE Chain=chain17 Position=2:251413457-251416783(+) GeneID=ENSMODG00000008651
ProteinID=ENSMODP00000010764 Score=260.4 E-value=8.4e-79 Length=3327 Comment=Overlaps multiple
annotations; one in proteins
GASHPDLGLPSFMFLGVYDEHLFVCDHESKAKSWGPWLTKEEDEFWVRLTQSCLKGWDHM
FIIDLWTIMDNHSGQGSQSHILQVILGCEFLVEDNRTRGFYKGYDGDYTLFFHPTMNT
ALQPEAQATKQEWEMNKIWKQHAYLERDCKPQLQRYLKIIESEILNKKCPRLRVVTHQV
TDEGVTLRCAQNFSPVNIITLSWLWDGKQVVTQGTGLGDIHPSGDTGQMTVAITPGE
WRYACQVHPLDQPLIMTWEPSPSWT

>md_AZGP1 Chain=chain18 Position=1:225700462-225706840(+) GeneID=ENSMODG00000002928
ProteinID=ENSMODP00000003567 Score=250.1 E-value=1.3e-75 Length=6379 Comment=Overlaps single
annotation & in proteins
MGLLTTFVLLLMFTGTTMLKESGIRCYSLIYEDVARSHPRPGQSNFTNTGYLNGQVIFYQY
DSKNQRAVPSPPWPKVKGMHDWEKESQFQKRREDFVLDNMQNILDYNDGNGSHIFQGRF
GCKLCGDNFIKGFWKQYDGRDFIEFNTDEKPSWIPLDPAAHMIKQRWEASPGAVYRAKA
YLEEEICIGTLRRYIEYGTHTLRQLIPPTVKLSHNSLLEGNLTMKCLAHFYPKEIELNWL
WNGKVQEKKVGEIQPIGDGTYQTWITAEIPISSKKNYSCHVVHESSFESLVVNWAKE

>md_MILL-LIKE Chain=chain19 Position=4:121301409-121302976(+) GeneID=ENSMODG00000000345
ProteinID=ENSMODP00000000409 Score=209.8 E-value=1.2e-63 Length=1568 Comment=Overlaps single
annotation & in proteins
MEGDWRDWRDLCLFFWSKFLRVQECQNSLHRPASCKVQLSQRFLGIQRALDSGLHSLQY
NFIAISNSDQESFSYALVYLLDDQLFLHYGSKSQLELRSTSLLEVIPPDIWKKESENKLG
MRWKLRLAKITQQSQEKDGSHTLQVILGCDLLKNGKTRGCLEYGDGENFLTFQPETL
TWEDYHYPAAANSIKNDFATDLIDTKLNKAKVEQECPPQLQGYLAVLKEKKASPSLHVTSK
LNNLWVCWAYNIFPGDVKITWLKDGQALNKEQESGILRPSGDGTYQTSVSIYANPKELN
YKCHVEHQGNQTKAIPSGSVRKEFQIIFVNPVIVIIIFNLRE

>md_UT15 Chain=chain20 Position=1:705690258-705693040(+) GeneID=ENSMODG000000011424
ProteinID=ENSMODP000000014281 Score=205.2 E-value=3e-62 Length=2783 Comment=Overlaps single
annotation & in proteins
AHRHRDFYFTATGSSGSLDFIVNSVVDVQLCSYHKEKHQLVMKEDWMMYQALGASEIAQ
KQKLLDLEKNFQWAFKNIKNEAKSGKNHTLQVFNCEVDKIDILVSSHQFALDGEDFC
QMDEQPEHWIAMKPEAQHFKFLWDSAFWKRVRHYIQEDCIDGMNKVLQHSMMRENVVPE
VTVSRVDAPGRVTFSCRATGFYPPSSILLHWEKDGALGVWQESSSGTLPNADATFYLQV
TLELPPGDPGTGYICVVEHSELETPAMFPVPGKPHMERPSVIAMSIILAVIIVVLSGAG

>md_UT8 Chain=chain21 Position=1:705429441-705434274(+) GeneID=ENSMODG000000023068
ProteinID=ENSMODP000000028153 Score=194.6 E-value=4e-59 Length=4834 Comment=Overlaps single
annotation & in proteins
ILDPLFSFLPIHSCSAHHKHMIIQFTMVGTSLSLEYFIVDFLDDVQVFSFNKYNHQLIA
KEAWISQVLGAKFIEDTWQKLVDEKSFVFLRKLKMNSTKSDMNHVQLFILCEIERN
QLGNEIQIAVNGQDFCWLDEKYGDFWIMVPEAEFPKILTSNLWTTQRKHVMQEYCIDTM
KKILQHSSIKKNVPEVTVSYHEAMDGITTLSCSATGFYPPSSILLHWQKKEKKIVVGKES
SSGLLPNADSTFYQRVFTLELPEETGTNYDCVVEHIELGTPKVYPAHVKSSKKRSRTTLT
AILGVVILVLSIIASFIMWHKIKT

>md_FCGRT Chain=chain22 Position=4:412954575-412955578(-) GeneID=ENSMODG000000013649
ProteinID=ENSMODP000000017073 Score=192.7 E-value=2e-58 Length=1004 Comment=Overlaps single
annotation & in proteins
MGSASPSQQLPLRLALFLPLLRVLSAGTSPSLFYQLTAVAAAPPGTSPSFWASGWLGPQLF
LTYSSGSAEPWGAWRWEHQEAWFWEKETWYLKQERLLQEAALRLSKGAQTFQGLVGCQL
NPNNSAYPTSRFALDGMDFLTFDPVARDWLGNSEGAQAVRRHWANETHRADREAQFLLTT
CEPKLRSHLQNGKGNFQWKGMPRCELEATPGQGQSTLTCQAFSFFPPEVKLTFLEGKP
VPEPEKGAEPWPNRDGAFHSQATLQVQRGDEARYSCEVRHPALATPLTVSFEASGLSLAV
IGGLVAACLLILLPIAAVVALVMWKKRGRPAWIFRGRAADDVGLLSAPGP

>md_MIC Chain=chain23 Position=2:267433622-267437852(-) GeneID=ENSMODG000000015810
ProteinID=ENSMODP000000019728 Score=186.5 E-value=7.2e-57 Length=4231 Comment=Overlaps single
annotation & in proteins
MGVIRVPRPLAAAQLGLLLLLQLLQPAQTAAGTSSHYLYRDFAIVSQPSQGSSTYVLG
YLDGQPFVWCRCRQGRAPRAPSQVQGRSWEERLIRSLKTVARLVIDLKSMMGTKETE
KVNNGEAPGSHTLQALAMCEFGQGNQSNPGFWRYGYDGGQELILDLEMLNLTETESKRQEEI
RKYQRANLEQDFLVRTCLHPLRKYLEAGVLTLOKTVPPSVRVTHHGTFSKSTLKCQAF
SFYPPDISLTLWLDGKPTIKGTQESQSVRPGDGTYSYVNVVEVPEEESRYACLVHEQA
LNGTFTGIWEPFPPKLPPIWILSGCLASALVVFVGTIIVFIWRRWRKRRTGFRTKENV

>md_UT17 Chain=chain24 Position=1:705712936-705718320(+) GeneID=None ProteinID=None Score=173.1
E-value=8.8e-53 Length=5385 Comment=No overlapping annotations
CPLAHHSEMFFTAAGTTLDFTLVSLDDVEIVFYNNKQQQVFIKEAWVSHALGADF
IEDIQKLYNEIDYLWALQNWVKNDRGNHTVQIWHNCHLDQDIHVSSRVWCAVDGEDF
CQMDEQISHWVAKKPEAEKIVLLSEKIYCSKKVQRYLEDYCVPMRKVLQSNVTVSRYDD
PDGHITLSTARGFYPRISILLHWEKDEQLDVGQESSGILPNTDATFHLQITLKLQSSD
SGTGYTCVVEHNELETPAVY

>md_UT5 Chain=chain25 Position=1:705355518-705357856(+) GeneID=ENSMODG000000011387
ProteinID=ENSMODP000000014240 Score=169.9 E-value=7.8e-52 Length=2339 Comment=Overlaps single
annotation & in proteins
LLELSGISFLDDVEVLSYNKAHKQIVFKIPWISKALGVKYLTELHDLVQEQDLRWITIE
IVARNDTNHNHNTAQLLAECEIDNDILVKSIIHLIWDGEECFRIDEEDGHWENVNPEFK
QYQHILGSLPWTTLRKYQMKIYCVDGIRKIIGHSSIRDNVVPEAVSQHVSPEGSIIILSC
IASGFYPRISILMHWERNKGLGIWGNSSSGLTPNMDSTFYLVKVTLELPPEDSGAGYTCVV
EHSELKSPAIYSVPRKPTMNKSWILALGITLAVILLSCAGAFIWKKKTG

>md_UT3 Chain=chain26 Position=1:705293594-705295946(-) GeneID=ENSMODG00000011316 ProteinID=ENSMODP00000014153 Score=169.4 E-value=1.2e-51 Length=2353 Comment=Overlaps single annotation & in proteins
NSVHSHHIGQFTAVGTAHSLLELSAINFLNDIEVGSYNKAHKQIIVKIPWISKALGVNNI
IKKINLVEHEQHFRWTIQLFRNDTNHNRNHTAQLLAECEIDNNTIVKSHIYLIWDGVE
YFRIDEEVGHWENINPEFKEYQHILESPFWTTLRKRKRYMKLYCVDLMRKIVGYSSIRDNVV
PEVAVSQHVSPEGSIIILSCIASGFYPRSIILMSWEKYKGLGIWGNESSTGLPNMDSFYLL
KVTLLELPPEDSGAGYTCVVEHSELKSPAIYSVPGKSTVEKSWVLELGLVLAIIILLSCAG
AIIWKKR

>md_UT10 Chain=chain27 Position=1:705495099-705503179(-) GeneID=ENSMODG00000011403 ProteinID=ENSMODP00000014257 Score=169.1 E-value=1.4e-51 Length=8081 Comment=Overlaps single annotation & in proteins
GHRLEGEFIAVGRHSLFNFTIIHVIDDVLVCTYDKGNKQTVFNAGAWITRALGKDFTE
KKEFNLSAETYFRWLLSFFIKKDKTSNKNHTIQLYGECELDNDIYIGSQIQFAVDGEDFI
RTNDKMEHWIVTKPEAEHFRTIAESTLGVNLRKKVMKHYCAEMIRKILHYSNLKENVAPE
VTVSRHDVSDGRVAFNCATGFYPRSIIMMRWEKGGQLGVWQESSSTGLPNADSTFYLLQI
TLELPPGDSGAGYACVVEHNKLQAPAVHPGESWVVLVGLLTAIILVLSCSGAFILWKKTK
GMVYRRQGEKTNCPKRS

>md_UT16 Chain=chain28 Position=1:705702206-705705425(+) GeneID=None ProteinID=None Score=153.2 E-value=No Length=3220 Comment=No overlapping annotations
PYLAYHSLMCFATVGTTKSLLDFTMVGSMNGVQGSFYGKKNQQLVIKESWVSQALGAQY
IEEKROKLVNIEIDFLWALQNWIQNDTKGHSNHTVQFWDHCHLDGDIHMSHFYAVDVG
ETFCGINELKHVWAMKPEAECEFRPFWEVIFPKYKIKRMEEDCIEPLRKVLKYSSSSGT
LPNADATFYLLQVTLLELPPGDPGTGYTCVVEHSELETPAMF

>md_UT9 Chain=chain29 Position=1:705459481-705469054(+) GeneID=None ProteinID=None Score=143.2 E-value=7.4e-44 Length=9574 Comment=No overlapping annotations
CHLEYHWNVEKFTAVSTDRSLLDFTGLIAVIDDIQWISYKLNHQAIVKVDWISDLVGTDF
IEEMCHLIKSHERDGNNTMQYILGCELDKLNKIRIGLDGETMIEFDEHQHVVQVS
PRAENLKPIVESLFWTNLRKRYVKQYCVVQKVIQNSVVSRRHDSFNGTITFSCLVTGF
YPKFIMLIWKKANEQAIWKGESSSTGLPNADGTFYLRITLELPGDVGAGYTCVVEHREL
KKTIIY

>md_UT7 Chain=chain30 Position=1:705406068-705408767(-) GeneID=None ProteinID=None Score=141.0 E-value=2.9e-43 Length=2700 Comment=No overlapping annotations
YNSVHHRHELQFTTVGTTHSVLELSVLSLIDDVVKVASYKKGQKQITIKIPWISEVLGVNY
VTQLHDLNHNVEQDFLWSSTGNCTVQLLAECEIDNDITVKSHIYLIWNGKESYRIDEEVG
HWENIKPETKQYQYILESPWTSLRKHYMKL*CADLMRKIIIEYSMAVVSQHVSPEDSIIILS
CIATGFYPHSILMPWEKNGDLGIWGNESSTGLPNMDSFYLLKVTLLELPPEDTGAGYSCV
VEHRELKSPAMYSK

>md_UT6 Chain=chain31 Position=1:705380659-705384355(-) GeneID=None ProteinID=None Score=132.2 E-value=1.3e-46 Length=3697 Comment=No overlapping annotations
YNSVHHRHELQFTTVGTTHSVLELSAISLIDDVEVASYKKGQKQITIKIPWISEVLGVNY
LTQWHNSLVNHEQDFLWSSTGNHTAQLLGECEIDNDITVRSIHILWNGKESYRIDEEVG
HWENINPEVKQYPILESPLWTSLRKSYMPLYCADLMKKIIGYSVAVSQHVSPEDSIIILS
CIASGFYPRSIILMHWERNKGLGIWGNESSTGLPNMDSFYLLKVTLLELPPEDSGAGYTCV
VEHSELISPAMY

>md_UT11 Chain=chain32 Position=1:705540409-705547747(-) GeneID=None ProteinID=None Score=131.1 E-value=2e-40 Length=7339 Comment=No overlapping annotations
CFTDYHRHEVQFIAVGKTSLLDYKILHVIDGVNVCIDYKKNKQLRVKEAWISLALGEEFI
KEKKKVIWESEVYFHLALKFLLQNDTNHTIQILAVCELDGDIETVTSQVRIAIDGEAFFQV
DDQADQWDFLKLPAKQFKSLLKSHFWTDLRQTMKQYCVNMMRKILQYSVTVSHHDDPDG
NITLSCLATGFYPCSIQLHWEKNEQLGVWQERSSTGLPNADDTFYLLQVTLLEVPDSTVT
GYTCVVQHSALEMPVTY

>md_UT14 Chain=chain33 Position=1:705669598-705673727(+) GeneID=None ProteinID=None Score=124.1 E-value=3.9e-38 Length=4130 Comment=No overlapping annotations
HYSQGHRYAFYFTAVGTSNSLNFMTSFDIDDIQLCSYHKLNTGTWSSVPCANFSEKMQ
SVLLNQEENFRLIHYLEQNGTKSESSHTLQLFADCELEDDIQVDSHVHLAWDGEDLFWR
DEQQGYWIFLKPMGYLLKPAVESPFWESVRRNMQKCYCFDIMRKILKSSVTVSRHDALGG
IITLFCATGFYPHSILLGWKKGYEHAIWKGESSGILPNADGTFYLLQVTLLELPPGDPGI
GYTCVVEHTELKTPAVY

>md_UT4 Chain=chain34 Position=1:705317276-705319984(+) GeneID=None ProteinID=None Score=123.8 E-value=3.5e-38 Length=2709 Comment=No overlapping annotations
YNLHVHRRHVIQFTTVGIDHSILELSAISLIDDIEVASYNKGQKQIAIKIPICISEVLGVNY
LTQWHNSLVKHEQDFLRSSTGNHTAQLQAECEMDNDITVKSHIYLIWDGEECYQIDEEVG
HWENMNPEFKQYQYILESSLRTSLRKHYMNLVYCVLDMKKIVGYSMAVVSQHVSPEDSIIILS
CIATGFYPRSIILMRWEKNGKGLGIWGNESSTGLPNMDSFYLLKVTLLELPLEDSGAGYTCV
VEHSELTPAMY

>md_UT13 Chain=chain35 Position=1:705610150-705616325(-) GeneID=ENSMODG00000011414 ProteinID=ENSMODP00000014272 Score=121.6 E-value=1.8e-37 Length=6176 Comment=Overlaps single annotation & in proteins

HELQFIAVGTSGQLELSAIAFIDDIQWASYGKPLQQIKVKHAWISEALGRQFLKEMQYQ
MTDQEEGYHRFIQHLTRNDRSKNHRTLQFYLNCELNEDIPLKYHVKYGFDGEDFIETNEK
GKWKVLHHWAIGSEIFFESPAVFEQFNKIYICIGVMQKILQKSSLKENRPPDMYVSHQEFF
NGTITFSCTATGFYQAIQMSWKKADNGTVVGKEDSSGILPNSDDTFYLQISLEVQPGEP
RTGYACVVDHSQLLEAPAVL

>md_UT2 Chain=chain36 Position=1:705270450-705273246(-) GeneID=None ProteinID=None Score=118.0
E-value=2.1e-36 Length=2797 Comment=No overlapping annotations
CSPALHSLLEIQFTGMAITDDLPDIIFNVWLDSQLLFSYDTQSKEVLLKLGAYPPLKNRL
MEKLKEQLRRGGEDSPTLEQTQVTVFCELDGDIVDSRMLVAFDGETVCQLNEQNARWV
IKKPEVTRFCNYLDDSVQWHRILAEECTFLKIVLELFTVTVSHDDAPDGRFILFCTARGF
YPRPIRLHWEKDGQMGVWRKERSSTLGNADATFYLVVLELHSDRDPGDGYACVVEHYEL
RMPAIY

>md_UT12 Chain=chain37 Position=1:705577166-705583437(-) GeneID=None ProteinID=None Score=109.2
E-value=1.2e-33 Length=6272 Comment=No overlapping annotations
CYLAYHKHEMFKTAVGTSKGLLDLIGVAFIDGIQWASYKSSQKIVVVKPTHIYEALGENF
LKDMDNEMKQDERGYHHFIGNHTMQLYLDCELEDETHLSSLVKYGFDGEDFIEIDEHGKW
KVLHPWAHSFEDFYEGPDVAKLLNKRYCIGAMLKILQKSVYVSRQEFNPTISFSCATG
FYSQSIILMHKKGVDAILGKESSSNILPNYDDTFYCRISLEIQPGDSGTGYACVVNHS
LETPTVY

>md_UT1 Chain=chain38 Position=1:705246778-705258091(+) GeneID=None ProteinID=None Score=76.9
E-value=2.6e-23 Length=11314 Comment=No overlapping annotations
SLSDLHSLQLQTTMAAYGDLDPDIFFNGCLDDHLILSYDSHSKSPGTGIHTAQALLRCE
KGILVNSLVVAFDGEDMCQLEDEEKQWTVRKPEAEDFCPFWKDPFWIKAAQVDCPFPLS
LLLQIVTPEVTLQHVGRAVLSCPAAGF*FHSVFMHWKKGDEQLIQGMEKSS

>md_CD1 Chain=chain39 Position=2:168357284-168363550(-) GeneID=None ProteinID=None Score=59.9
E-value=7.5e-19 Length=6267 Comment=No overlapping annotations
QSMVSCWLEDELETHKCNQRTDTIDFLWPWAQTGFSAEDREELQTNLQVYVVGFTGEVSPR
DFQVAFQGGKGFWSFGQDSWKPLPGAESIAQHSSILNPDQSVQIIVQRHLLYLPQLR
GLLDKQWKTLELQIVK*TYFCTRVADIALSVSGFYPKSIQVTKWIKNGEQSGTQTNLVL
NSHSTWWLNVSLIMEAEKAGGLACRVKHSLSGSKDLIQ

>md_chain40 Chain=chain40 Position=2:267606293-267607759(+) GeneID=None ProteinID=None
Score=53.8 E-value=1.5e-16 Length=1467 Comment=No overlapping annotations
ESSGTSILQYDFTSVSGLPGQPAYMALGYINGCISINYDSEVKRAELRAEIESTVLLLE
TKEHLFLKROEQDIRWLSILGYNKLSGHH*LGHGVQGMTPDSHTIQLSHGCELGIQPGG
DGIYQS*ENVEVPSRDEL*YICLVEHQRLNQLTLM

>md_ENSMODG00000027380 Chain=None Position=1:225336071-225344637(+) GeneID=ENSMODG00000027380
ProteinID=ENSMODP00000038711 Score=93.0 E-value=7.3e-27 Length=0 Comment=Not detected in genome
search
MLTYITFRIGMMRPLFTANFLFLLSGTMSQRPRIRCDVLHLKDFGFSNPQDGHYSFKNEV
YLNQVYTYDYSKSKRAVPQPEWQNVENWNEVSQKEREELIMKDLLEEIRNLNTDNTDI
YTTSONVACQLCSNDAFRMSGYSLVEGKEFSKSKRVFTLATQDPAEAMMKHQSKQDPMF
HGYQEKKCIEILQYRNYRDAH

>md_PROCR Chain=None Position=1:388590506-388590571(+) GeneID=ENSMODG00000001848
ProteinID=ENSMODP00000002258 Score=93.8 E-value=4.1e-27 Length=0 Comment=Not detected in genome
search
MLPLLSLFSWVIYQCGAMKGSQSFLMLQISHFRDPSSVQFVGNASLGLGATHTELEGGGHN
ITIQLKPLESPEHWQTKKQLLNYPFEFQVLVQLVNKERGVAFPLLRCSLGLCELPDQ
QEAHVFFFEVALNGSSVFSFQPEKGPDPHPIELYDFTLKQLNTYNRTRFELLEFLQETCVGF
LKQHGNNRVRGVPVQVRSYTMILGIMGVFAISSVAVGIFLCTGGKWS

>md_ENSMODG00000024063 Chain=None Position=1:224303763-224308941(+) GeneID=ENSMODG00000024063
ProteinID=ENSMODP00000032790 Score=104.8 E-value=2.3e-30 Length=0 Comment=Not detected in
genome search
MMEPLVTAIFLLLSGTTMSKEPNLGHSCQDSLAYRDMALIYSGAEHDLFTNKAFNNNE
LVYKYDSESQTAKPQPGWENVENWGEISKLQKERGDFALENVELIKKTFGYRDDYNTGI
YGDVCKNNVIKVRWYTYDSDKPLITFDNTNGPAWKALSREGKILKKKWEEDPSAPNRAKS
YLEEDCPKKLQYKKNYKNTN

>md_ENSMODG00000028158 Chain=None Position=1:225082268-225092663(+) GeneID=ENSMODG00000028158
ProteinID=ENSMODP00000040872 Score=129.4 E-value=1.2e-37 Length=0 Comment=Not detected in
genome search
MMRLLFTAIFLILISGTLMFQHTAFNCYFLTYKDTIFSKPQTGNYSFKNAAYINGHTIYI
YDNSHKRAVPQPGWYLENWDEVSQKQKREDLIMKNLQEIIMNSNEGTFTYTYEYITCCIC
PYKYLKVIIRIFYINGEKFKITNKNMLAQASHNPAEEMIKQKLDKHALNQLNAYLEKQC
PAKLQRYKTSYYARLPRVRGN

>md_ENSMODG00000029679 Chain=None Position=1:224712002-224719873(+) GeneID=ENSMODG00000029679
ProteinID=ENSMODP00000038748 Score=90.6 E-value=3.7e-26 Length=0 Comment=Not detected in genome
search
MLTYNSRFRGMMKPLFTAIFLFLLSGTTMSQHPRTCDVLQYKNMGYSNPQTRGRYSFKNEV
NLNERTVYVYDSSSKRAVPQPGWENVENWNVKVSQKEREELIMKDLQEIIRNLNTDTGT

YTVTHTFDCQLCPNDAFAISGYNLFEGNEFSKNFKRIFTLATQDPAEMMKQSQRYPML
HGYQETQCQVETLQKYLNYRDAH

>md_MIC-like Chain=None Position=2:267557192-267558003(-) GeneID=ENSMODG00000015796
ProteinID=ENSMODP00000019705 Score=195.6 E-value=3.1e-57 Length=0 Comment=Not detected in
genome search
PASLQYDLTVLSQAEEPRQPLYVALGYINDQLFLHYDGENENSRAELQKPWNDEEGRKV
WERVTQGLEEKKGKELRMLIQDILGQNNQSSHTLQATLGCCELQGNDSRTRGFWRFRFDGQDF
ITFSPEKLSWISVHPGAQKIKEKWEDWFQMKLQEEYCSIRLQRYLATWKGILERT

>md_ENSMODG00000015798 Chain=None Position=2:267518869-267520166(+) GeneID=ENSMODG00000015798
ProteinID=ENSMODP00000019708 Score=127.2 E-value=5.2e-37 Length=0 Comment=Not detected in
genome search
LLFLSSSDFSTLRLNLTAYFGWNKTSSYIISGFINDNKFLTYDSKSRHVDFKSCCEKADI
QNLKEENKELRDIVSNIALRNGENGSSHTLQVTYGCKRINGDVKGFWHYIYDQKFLSFS
LEDGSWGITGPEAEEAKKILQKSHSQKVKAYILGDPCADLKKCLKYQECVKKTGIEHLE
PGLPTHISQELPSFCLWFLSWEGDTE

>me_chain01 Chain=chain01 Position=Scaffold270567:38-1383(-) GeneID=ENSMEUG00000016222
ProteinID=ENSMEUP00000014833 Score=352.2 E-value=1.5e-96 Length=1346 Comment=Overlaps single
annotation & in proteins
PHSMRYFYTLTRPGLGEPFLSVGYMDDQQFVRFSDSDSPGQVEPRAAWMEGVQEEEPG
YWERETQISRNTQIFRESLENIRGYFNQASAGVHTFORMYGCEVSPDLTFRGFQEHAY
DGTDYIALDPTLWTAAVPPAVSTKRWAEHEGTKAYLEETCVTWVKKYLEMGKETLM
RTHPLPVTHTLTPYGEVTLRCRAQDFYPKEISLWLRDGEHPQDTEFIETRPAGDGT
QKWADVHTSGEERYTCRVQHEGLSEPLTLQW

>me_chain02 Chain=chain02 Position=Scaffold10370:39445-41903(+) GeneID=None ProteinID=None
Score=347.8 E-value=5.9e-91 Length=2459 Comment=No overlapping annotations
HVPGPSLSLRYFNTAMTRPGLGEPFLAVGYVDDQQFMSFSDSDSPSQGLEPRVAMMERVEQ
EEPEFWERRSGICKAETQFFRGTLESLLSNQSSGGGAGVHTIQRNGCEVSPDLTFRGF
FIQYAYDGDYLDLDTTEFYKWTAAVPEAVNIKQWEADGNIVKRWKAYLNEECVRRLLKY
LEAGKETLMRKARVTRHTAPHGEVTLRCRAQDFYPKEISLWLRDGEHPQDTEFIETRP
AGDGTQKWADVMTSGQEDKYTCQVQHEGLSEPLTLGNSDAQGSDISLTVKG

>me_chain03 Chain=chain03 Position=GeneScaffold_9817:27277-29655(-) GeneID=ENSMEUG00000007297
ProteinID=None Score=347.4 E-value=8.7e-91 Length=2379 Comment=Overlaps single annotation; not
in proteins
PDSGSHSLKYFYAAMSQPLAKPRFIAVTVDDQVLSFSDSDRESQSTPEPTPWIEQEIP
DYWERETRISREATQRYRMCLQKVSSYNNHSEGGVHTYQRLSGCEVFSNRSFSGRFVQY
AYDMDYLALDTEFLSWIAGNAGALNNKRWKWEADQRIAKYWKGYMEECVWLRYLENG
KETLLRIVRVTHTSSDGEVTLRCRAQDFYPKAEISLWLRDGEHPQDTEFIETRPAGD
TFQKWAAGVMSLGSSEQKYTCRVQHEGLPEPLFLPGTDSAQGSDVSLTAKG

>me_chain04 Chain=chain04 Position=Scaffold312982:286-2910(-) GeneID=None ProteinID=None
Score=344.2 E-value=6.7e-90 Length=2625 Comment=No overlapping annotations
HVPGPSLSLRYFDTGVTRPGLGEPFLSVGYVDDQFVGFSDSDSPG*REEPRAAWIERVEQ
EPPGYWEQQTTRTSRANMPIAREGLENLRGYFNQSKGGVHTVQAMYGCEVSPDLTFRGF
IQFAYDGDYLDLDRDPTLWTAAVPPAVNTKREWEAERSFAEGWKAYLEEKCVLWVKKYL
EMGKEALMRTARVTHHTAPYGMQLRCRAQDFYPKEISLWLRDGEHPQDTEFIETRP
AGDGTQKWAADVMTSGQEGRYTCRVQHEGLSEPLTLGNSDAQGSDMSLTAAG

>me_chain05 Chain=chain05 Position=Scaffold33820:13576-15767(-) GeneID=None ProteinID=None
Score=329.9 E-value=1.2e-85 Length=2192 Comment=No overlapping annotations
HVLGPHSMRYFAAVMTGPPELWEPYIAVGYVDDQHFLVLFSDSDSPSLRMEPRAVWMEQMDQ
EDPGYWERNTWISRAKTQVYRGNLKTLLSYFNQSSGGGAGVHTFORMSGCEVSPDLTFRK
GFLQHAIDYDQDYLTLDTDTYTWMAAVPPAVNTKRWKWEADRGIVEGYKSYLEETCVLWLLK
HLEMGKELMRTARVTHHTAPSGDVTLRCRAQDFYPKEISLWLRDGEHPQDTEFIETRP
PAGDGTQKWAADVMTSGQEGRYTCRVQHEGLSEPLTLGNSDAQGSDVSLTARG

>me_chain06 Chain=chain06 Position=Scaffold76695:6700-9101(+) GeneID=ENSMEUG00000001234
ProteinID=ENSMEUP00000001136 Score=324.2 E-value=5.9e-84 Length=2402 Comment=Overlaps single
annotation & in proteins
LTLLLGLFLVLRRETWAGSHSLKYFYTTMSRPLAETRFVTVYVDDQVLSFSDSDSDSQS
KEPRMPWIEHDYWERETRISREATQRYRMCLQKVSTYYNHSEGGVHTFQSLSGCEVFSNG
SFRSGFVQYAYDGLDFLALDMEITLWIAAGNAEAVNRKRELEAAKYWKYLEGECVYWLQ
KYLDNGKETLLRTPDPPSVKTRHISSEGEVTLRCRAQDFYPKAEISLWLRDGEHPQDTE
LIETRPAGDGTQKWAAGVMSLGSSEQKYTCRVQHEGLPEPLFLKWEPPQSSITGLSARVIT
ALLLLAAVVIIGVVIRWKNIPGMERRKDGKRESYTTTANNSAQGSDVSLVAKA

>me_chain07 Chain=chain07 Position=Scaffold71957:679-3383(-) GeneID=None ProteinID=None
Score=307.7 E-value=4.2e-78 Length=2705 Comment=No overlapping annotations
HVPGPSLSLRYFHTAVTGPGLGEPFLSVGYVDDQFVRFSDSDSPGQRMPEPRAAWMERVQ
EPPGYWEETRTSRGNAQIYRGLGVHTIQLMSGCEVSPDLRFRQFGLQFAYDGRDYIAL
DTETLWTAAVPPAVSTKRWKWEADWSIAERRKAYVEECVWLRKYLEMGKEVLMRTE*P
ATLTPMGEVTLRCRAQDFYPKEISLWLRDGEHPQDTEPGNSDAQRSDIPLTAKG

>me_chain08 Chain=chain08 Position=Scaffold33251:14389-15623(+) GeneID=None ProteinID=None
Score=307.1 E-value=3.5e-83 Length=1235 Comment=No overlapping annotations
HVSQSHSLKYFHTFVSRPNAEPWFFSVGYVDDQEFVRFDSHSPSQREEPRAAWIEQVEEV
EPGY*EKNTQIHMEANAQSFRENLVHTVQVMHSCLELSELTFRQGRFQDAYDGEDYITLD
TETYTWASMAPAVISKRKWEADRSPAELDKAYIEECCVMWLKYLKMGKEMLETRVTRH
IALDGEQEVILRCWAQDFYFYPKEISLWLRDGEELQNTFEFIEIETRPAGDGTFOKWAAVQMT
SGQEGRYICQVQHEGLSETLTL

>me_chain09 Chain=chain09 Position=Scaffold105202:2204-3461(+) GeneID=None ProteinID=None
Score=295.7 E-value=7.4e-80 Length=1258 Comment=No overlapping annotations
HFPGPMSMRYFLTGVTRPGLGEPYLSVGYVDDQFVSVFSDSDSPGQREEPRAAWMERMEQ
EEPQYWERVQIHRENAQTFRVGLGAGVHTIQSMYGEVSELPETFLRGRQIAYDGRDYI
SLDLETLTWRETPQAVNTKRKWEERSEAEGLKGYLEETXXXXXXXXXXXXXXXXXX
XXXXXXXXPPSVQVTRHTVHEGEVILRCRAQDFYPAEISMIWLRDGEELQDMEFIEIETRPAG
DGTFOKWAAVRITSGQEGKYACRVQHEGLSEPLALKWEPQSKYTWITLGI IAVFILHLVI
LSEPLTL

>me_chain10 Chain=chain10 Position=GeneScaffold_9846:369-1914(-) GeneID=ENSMEUG0000007887
ProteinID=ENSMEUP0000007200 Score=284.1 E-value=2.3e-76 Length=1546 Comment=Overlaps single
annotation & in proteins
VCPHSMRYFDTAMTRPGLGEPYLVGYVDDQLFVSVFSDSDSPGQREEPRAAWMERVEQED
PEYWERTRRARANAQTYRGSLENIRGYFNQAGGVHTLQSMYGEVSELPETFRQGRFYQS
AYDGRDYISLDADLTWRFAVPQAVNTKRKWEERSEAEGLKGYLEETXXXXXXXXXXXXXXXXXX
XXXXXXXXPPSVQVTRHTVHEGEVILRCRAQDFYPAEISMIWLRDGEELQDMEFIEIETRPAG
DGTFOKWAAVRITSGQEGKYACRVQHEGLSEPLALKWEPQSKYTWITLGI IAVFILHLVI
AGVVI

>me_chain11 Chain=chain11 Position=Scaffold49163:19201-21229(-) GeneID=None ProteinID=None
Score=282.8 E-value=1.8e-70 Length=2029 Comment=No overlapping annotations
HVPAPHSRLRYFSAATGPGLEGPRFVSVGYVDDQPFMSFDDTSLSQREEPRAAWMERMKQ
EEPEFWERQTRIHRTAQNRYRVGLGFHTFQRMSCGEVSELPETFRQGRFYQDAYDGDYIAL
DTETLRWTAAPHAVNTKRKWEADRSEAEGWKIYLEECCVQWLKYLKMGKEMLETRVTRH
TCHTTPHE*VTVRCRAQDFYFYPKEISLWLRDGEELQDMEFIEIETRPAGDGTFOKWAAVRITSGQEGKYACRVQHEGLSEPLALKWEPQSKYTWITLGI IAVFILHLVI
AGVVI

>me_chain12 Chain=chain12 Position=Scaffold106256:4826-6083(+) GeneID=None ProteinID=None
Score=276.1 E-value=4.7e-73 Length=1258 Comment=No overlapping annotations
HVPGPMSMRYFNATITRPGFRQPRFLTVGYVDDQLIVSFDTRQCLRSQPAAWMMERMQ
EDPWFWERTEILRGNAQAFRVALQNSGTGVHTLQRMYGEVSELPETFRQGRFYQDAYDGDYIAL
RDYISLDLETLTWRETPQAVNTKRKWEADRSEAEGWKIYLEECCVQWLKYLKMGKEMLETRVTRH
MRTATVTRHTAPSGEVTLRCRAQDFYFYPKEISLWLRDGEELQDMEFIEIETRPAGDGTFOKWAAVRITSGQEGKYACRVQHEGLSEPLALKWEPQSKYTWITLGI IAVFILHLVI
AGVVI

>me_chain13 Chain=chain13 Position=Scaffold40600:14825-17206(-) GeneID=None ProteinID=None
Score=275.6 E-value=1.5e-69 Length=2382 Comment=No overlapping annotations
SHSGSHSLKYFYIIVSRPGLAKPRFIAVTVVDYQVLSFSDSDRESQSTPRTPIQREIA
DYWERETRISREAKQHFPMLQKVSSYNNQSEGGGVHRFQRL*GCEEFNSRFSLSGSMQH
AYDGLDYALDLETPTNWIIVANS GALNDRKREKEDQSVAKYWKHMEECVYWLQRYLENV
ERRVQVTRHTASDWEVTLRC*AQGFYPAEISLWLRDGEELQDMEFIEIETRPAGDGTFOK
WAAVGLMSGREQKYTCRVRHEGLPEPLFLPGNDSAQRSDVSLTARG

>me_chain14 Chain=chain14 Position=Scaffold32785:7456-14156(+) GeneID=None ProteinID=None
Score=256.7 E-value=3.2e-68 Length=6701 Comment=No overlapping annotations
HVPDSHSLRSLSTSVTRPELGEPRFFSVGYVDDQFARFSDVSSAREESRAPWMELVDK
VDPEYWERNTFRHKEAAQKFLARLEILSAYYNDTKEGGAGVHIFQHLGCEVYQNLTFKR
GFFQFAYDGRDFLSLDTETYKWTASVPRALITKLGMEADESPKNDRKAYLEDTCVKARVT
HHTAPNGEVTLKCVTQDFYFYPKEISLWLRDGEELQDMEFIEIETRPAGDGTFOKWAAVQMT
SGQEGKYTCQVQHEGXXXXRTVRR

>me_chain15 Chain=chain15 Position=Scaffold116072:1660-2879(+) GeneID=ENSMEUG00000012186
ProteinID=ENSMEUP00000011105 Score=253.6 E-value=2.4e-67 Length=1220 Comment=Overlaps single
annotation & in proteins
SHSLRYFYTSISRPELGKSRFIAVGYVDDQFVRFYSDSSSQRMPEPRAPWMDQMDLGYWK
TETQHMREHAQDYRVGLENLLRYNNQSEEDGVHTVQFFYGCEVYPNLTFKRGFEQNAIDGQ
DYIALDTETYTWTAAGPEAVTTKHKWEAQSGIMEGKYLEDECEVWLRKYLEIGKEYLN
RTDPPSVQVTRHTAPDEEVTLRCRAQDFYFYPKEISLWLRDGEELQDMEFIEIETRPAGDGT
FOKWAAVGITSGQEGKYACRVQHEGLSEPLTMKWFHFRATVITWITAVVLLLIALVIG
VVIWRNRNSGQDRK

>me_chain16 Chain=chain16 Position=GeneScaffold_10171:5665-9302(+) GeneID=ENSMEUG00000010584
ProteinID=ENSMEUP0000009653 Score=247.5 E-value=5.2e-61 Length=3638 Comment=Overlaps single
annotation & in proteins
LVLFFLVVLLALTLTVTWAGPHSMRYFDTAVTRPGSEPRFLSVGYVDDQFVRFYSDSDSPGQ
REEPRAAWMERMDQEDPEYWERVTRIHRETAQIFRGNLETARGYFNHTAGXXXXXXXXXXXX
XX
ETCVTWLKLKMGKDTLPPSARVTRHTGPFGEVTLRCRAQDFYFYPKDISLWLRDGEELQD
MEFIEIETRPAGDGTFOKWAAVDVTSGQEDKYTCRVQHKGLSEPLTLQWEPQSSPTWLVIG
GIAAALLLPPVVIAGVMMWNRNRSS

>me_chain17 Chain=chain17 Position=Scaffold350269:845-1897(+) GeneID=ENSMEUG00000012694 ProteinID=ENSMEUP00000011572 Score=217.6 E-value=1.1e-56 Length=1053 Comment=Overlaps single annotation & in proteins
SHSLRFFSVSVSGPELGGKPRFFSIDYVDYQQLAGFDSHSASRRVEPLARWLQDGPDPYDWR
NTRHNEVSTHFYTLRQYFNQSESGIHTYQGLSGCEVSPQLTFQRFYQYAYDGLDYISLD
TDTYTTWALVPOAVNTKRVWEAAGVSEERRKAYLEIEICLVWLKYLEMGKETLMRTDPPSA
CVTRHTAPHGEVTLRCQAQDFYFKEISLWLRDGEHPQDTEFIETRPAGDGTQKWA
QMTSGQEDKYTCLVQHQLSEPLNLKWEPPSSPTLVGGIAASLLLLTAVIAGVEIWRKK
NSGG

>me_chain18 Chain=chain18 Position=Scaffold42934:24502-27221(+) GeneID=ENSMEUG00000008531 ProteinID=ENSMEUP00000007792 Score=212.4 E-value=3.6e-55 Length=2720 Comment=Overlaps single annotation & in proteins
QAMIGCEVSPQLTFQRFQMDAYDGDYLVLDLETLTWRAEVPQAVNTKRNWEADEVG
KAYLEETCIVWVKKYLEMGKDTLMRTDPPSARVTRHTGPHGEVTLRCQAQDFYFKEISL
WLRDGEHPQDTEFIETRPAGDGTQKWAADVTSQEGKYSRIRHEGLSEPLTLQWEP
QSSSTWLVVGEIAASLLLLISVIVGVWICIKKYSGRLDWAGSRKGGDYVPSAGNESAVTL
TAKA

>me_chain19 Chain=chain19 Position=GeneScaffold_5406:4803-7043(+) GeneID=ENSMEUG00000005831 ProteinID=ENSMEUP00000005322 Score=209.7 E-value=2.7e-54 Length=2241 Comment=Overlaps single annotation & in proteins
SHSLQYNLLTMTADEESLSYSAQGYLDGELLFHYNSTNQILKQVPSLDGLVETDIWNK
EHDNLKDIDGDFVALSVKQXXX
XXX
PLLNVTSSKKQGEVTLRCWAYNFFPRDIKMTWLRDGRALNQDKQERGLIKPSGDGYTQ
WVSTDVHIEEANYTCHVE

>me_FCGRT Chain=chain20 Position=GeneScaffold_2032:11704-13437(+) GeneID=ENSMEUG00000008659 ProteinID=ENSMEUP00000007917 Score=195.8 E-value=2.2e-49 Length=1734 Comment=Overlaps single annotation & in proteins
HSLFYQLTAVSVAPKGTFRFWASGWLGPQLFLTYSSGGNAEPWGAWRWEPQEPWFWEKET
WYLKTOERLQALSLSKQGAHTFQGLVGCQLNPDNSHHTARYALDGDALLTFDPVSR
VWSGDVVEAWNKESWANESQRAEKDAEFLTTCPQKLSHLKQKGNFHWKEAPEVRLE
ARVSTQSILTCQAFSFFPPELELSFLWKGASGCEGSDHPNGGAFYSRGTLLVFPFGDEAL
YSCVVQHPALQLLTVNF

>me_UT15 me_chain21 Chain=chain21 Position=Scaffold5385:3500-6178(-) GeneID=None ProteinID=None Score=182.6 E-value=7.2e-46 Length=2679 Comment=No overlapping annotations
LAQQGLSNLIVISSVDDVQLCYNQNQLVVMKKAAMWSQALGATAIGQIQKMLINLEKSF
SWAFKNLIKNETNSESTPTENHMLQVFDCELGNDIQVSSHQFALDGEDFYQMDKQLGH
WVAMKPEAQHCKPLWDSALWNLVVKHYIQEDCINGMCKILQYSVTVS*HDAPDGRVTLFC
RATGFCPCSIMLHWERDGVLVGWQESSSGTLLNADATFYLVQVTELESPDPETGYTCVV
EHSELRTPVMF

>me_chain22 Chain=chain22 Position=Scaffold62626:3261-5893(-) GeneID=None ProteinID=None Score=178.7 E-value=8.6e-41 Length=2633 Comment=No overlapping annotations
HVSGSHSLRYFDTAVSRTGLGKPRFIVDDQKFVRFDSDCPGPRQETRAAWMEQMNPDCEW
GQTVWVSMANAQGYFQVTVY*EVSPQLTFKRGFEQKQVYDGREYLTLDTETYSWTASVPOAV
NSKNQKEAEVPTVTRHTDPSGEVILRCRAQDFYFKEIFLWLRDREPPQDTEFIETRPAG
DSTFQKWAADVITSQEGRYTCRVQHEGLSEPLTLPGNDSVQGSVSLTAKG

>me_UT16 me_chain23 Chain=chain23 Position=Scaffold36829:4826-7625(-) GeneID=None ProteinID=None Score=177.7 E-value=1.8e-44 Length=2800 Comment=No overlapping annotations
SQLAYHTLEMCFTAVGTTKSLLDFTMVSSMDGVQGSFYDKKQQLVIKEAWMSQALGAHY
IEKKHQKLVCEINFLWALQNWIQSDTKGDNHTVQFVHDCHELDGDIQVSSHFWYAVDGEA
FCGVDEQLRHVVAMTPEAERFRPFWDVIFPYKMIERYMQEDCVEPLKVKLVQYSVTVS*HD
APDGRVTLFCRATGFYSRSIMLHWEKDGVLGVWQGETSSGTLFNADATFYLVQVTELELPPS
DPGIGYTCVVQHSGLRTPAMF

>me_chain24 Chain=chain24 Position=GeneScaffold_10171:22023-27930(+) GeneID=ENSMEUG00000010584 ProteinID=ENSMEUP00000009653 Score=168.4 E-value=4.3e-42 Length=5908 Comment=Overlaps single annotation & in proteins
PDPGPHSMRYFDTAVTRPGSEPRFLSVGYVDDQQFSRFDTDSFGQREEPRAAWMERMDQE
DPEYWERVTRIHRETAQIFRGNLMKGYLEETCVTWLKKHARVTRHTGPFGEVTLRCRAQD
FYFKEISLWLRDGEHPQDTEFIETRPAGDGSFQKWAADVTSQGEDKYTCRVQHKGLS
EPLTL

>me_chain25 Chain=chain25 Position=GeneScaffold_10234:5091-6704(+) GeneID=ENSMEUG00000003085 ProteinID=ENSMEUP00000002813 Score=160.0 E-value=5.5e-35 Length=1614 Comment=Overlaps single annotation & in proteins
PHSMRYFFTSVTRPRLGEPFRFVSVGYLDXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXX
XXX
XXAPSARVTRHNAPHGEMTLPCQAQDFYFKEISLWLRDGEHPQDTEFIETRPAGDATF
QNWAAVEVTSQEGRYTCRVQHEGLPELLTLQWEPQSSSIWLVGGIVAALLLITAVIAR
VGIWRRRNSGGKGNDSAQGSVSLT

>me_UT17 me_chain26 Chain=chain26 Position=Scaffold136891:6877-11011(+) GeneID=None ProteinID=None Score=155.9 E-value=3.8e-38 Length=4135 Comment=No overlapping annotations
YHLAHSHEFFFTAVGTSKTLFSFTMVSFLDDVEIAFYDKKTQOLEIKEAWVSHALGVDF
INHTVQIWHNCHLDRDIHVSSRIWCAVDGEDFCQMDWEIISLWVAKKPEAEPATLLSEKIY
CSKKVKHYLEDYCVQPMRQVLYSVTVNQHDAPDGNIIILSCTATGFYPRISILLHWEKDAQ
LVVWGQESSGILPNIDATFYLQITLKLQSKDMETGYTCVVEHSELETPAVY

>me_chain27 Chain=chain27 Position=Scaffold28197:17568-19287(-) GeneID=ENSMEUG00000004115 ProteinID=ENSMEUP00000003758 Score=155.7 E-value=5e-38 Length=1720 Comment=Overlaps single annotation & in proteins
SHSLQYNLFAMFSSDGGPPSYSKYLDGELFLHYDSKKHRPEPRSTWLNELVETESWKKED
SNLKEIGQEFRITLTKVMDWKEGSHTFQETLGCCELHDSNCRFWRYGYDGEDLSSNLRFPFL
LKVTQNKQEGEVTLRCWAHNFPGDIKMICLWDGYALIHKDVQPSGDGTYQTWVVSIDVH
IEEANYICHVEHQGRNQITISVPL

>me_UT19 me_chain28 Chain=chain28 Position=Scaffold33001:4408-6729(-) GeneID=None ProteinID=None Score=154.9 E-value=9.9e-38 Length=2322 Comment=No overlapping annotations
YNSAHHKFVQGFTAVGTSTKSVWELNGIIFIDDIELGSYNNHTQQIVVKIPWVSKLMGID
YITQMRNLLVDHEQHLHMMHYLAKNDTNGNHTGQLLADCEIDKDIIVKSHIHLIWDGEE
YYRIDEEVGHWEHLKPEFKRYHHVLDSPFWTDIRKRYMKNYCVLMMKIVGYSVTVSRHV
NPESIILSCTATGFYPSILLRWEKNGKLGWVGKETSTGILPNMDNTFYLQVTLQLPSE
DPGMGYNCVVEHIELKTPVVY

>me_UT8 me_chain29 Chain=chain29 Position=Scaffold57264:2841-7721(+) GeneID=None ProteinID=None Score=147.1 E-value=2.7e-35 Length=4881 Comment=No overlapping annotations
CPAAHKKHVIQFTTAVGTESLLEHVMVDFLDDVQVFSYKNYQKLMMAKEAWISQVLGAKF
IANTQKQLVDHEKSFVFLTMQVSVSCEIEENNQIRKEIQIAVDGGEFCILDKDMGNQII
LMPEALPFKDFLTSILQTTQIEHYMQEYCISSMKNILLHSVTVSHYEAAGMNFSCSAT
GFYPPSILLYWQGEDKILPGKESSGILPNADSTFYQQITIELPPEDMKTIDYDCVVEHI
ELGTPKAY

>me_UT20 me_chain30 Chain=chain30 Position=Scaffold22554:19566-21870(+) GeneID=None ProteinID=None Score=138.6 E-value=6.7e-33 Length=2305 Comment=No overlapping annotations
NNSAHHKHIAQFTAVGTTSHLMELTLISFIDDIELASYNNIQKQIISKIPWATKAVGVDS
ITQIRDLLIYHERHVSMMHYLAKNDTNGNHTGQLLADCEIDKDIIVKSHIHLIWDGEEY
YRIDEEIGQWELLKPEFKKYLHILESPIWTNLRKLYMNQFCVDLMRKIVGYSVTVSRHVS
PEGSIILSCTATGFYPSILLRWEKNGKLGWVGKETSTGILPNMDNTFYLQVTLLELPPED
SGMGYTCVVEHIELKTPAVY

>me_UT18 me_chain31 Chain=chain31 Position=Scaffold3670:19624-23747(-) GeneID=None ProteinID=None Score=135.9 E-value=6.2e-32 Length=4124 Comment=No overlapping annotations
CPVAHQHEVQFTVVGTSNSILDLMVDFVDDIQVFFYDKQNSQLMAKEAWISQVLGAKF
MEKTQKQLVDHEKSFVFLWFLKLRQNDTKSDSTLQVSVIVCELQRDDIDIGNQIKIAVDGEDF
CWLDEIGQWFFIISSETDSFKPTLMSTFWVWQKDYMKYCYGIMKKILQHSVTVLYHETV
NGIILLSCSATGFYPSALLHWQKGGIIVAGKSSSSTLPNADGTFYQRIIEILLPGDT
GTNYDCVVDHIELGEPKVF

>me_UT14 me_chain32 Chain=chain32 Position=Scaffold5385:25353-30701(-) GeneID=None ProteinID=None Score=131.5 E-value=1e-30 Length=5349 Comment=No overlapping annotations
CHSGQHRHFVYFTAVGTASSLLNFTVASFIDDIQLFFYHKLNQFQVKMDWISNVLGANFT
EQMRLLMIHEEHFHLFCYLEHNDTKSESSHILQLFGDCELEDDTEVNNHVLAWDGED
LFQRDEQOQYLIFLKPMPGYILKPAMESTFWGNVRKHYMQKYCFDIRMKILRYSVTVSLHD
ALDGTIILSCTATGFYPRISILLSWEKGYEHAIWGENSSGILPNADGTFYLVTLKLL*G
DPGIAYTCVVEHRELKTPAIY

>me_UT21 me_chain33 Chain=chain33 Position=Scaffold19997:2641-4934(-) GeneID=None ProteinID=None Score=125.5 E-value=5.1e-29 Length=2294 Comment=No overlapping annotations
RNSAHHTHVGQFTAVGTIHSLEINAVSIMDDVELGSYSYAQQKISVKIPWIVELFEVDY
ITQRRHLLVEHEQQARWIMEFLSKNGNHTGQLLADCEIDNGIKINSHIHLIWDGEEYRI
EENGQWEILKPVGKFLHILESFPVTDLRKRYMNEYCVDMVKKITGYKVTVSRHVSPEGN
IILSCIATGFYPHSILLRWEKNGKLGWVWNETSTGILPNVDNTFYLVTLLELPPEDSGMD
YNCVVEHIELKTPAVY

>me_chain34 Chain=chain34 Position=Scaffold34365:7283-8419(-) GeneID=None ProteinID=None Score=107.6 E-value=8e-23 Length=1137 Comment=No overlapping annotations
NARLGEPRFFSVGVYDQLSGCDTDSQGPQEQPQAPWMOGMEKEYFA*DRLAYIALDR
ETYRWTAAVPHAVNTKLRWEADRSAGDRWKAYLEEMCVVWLKK*LEMGKALMRITWLRD
GEEQPQDVELIEMRPAGDGTFOKWAADV

>me_chain35 Chain=chain35 Position=Scaffold1906:88257-88790(+) GeneID=None ProteinID=None Score=49.7 E-value=6e-06 Length=534 Comment=No overlapping annotations
WMEREDPEYWEWNMGNLGRGAGVHTVQCMSCGCELSDLTFQ*GFFQDAYDARVTRHTAPHG
EVTL*CPAQAFYPKEISLTLWLRDGEQVGEKE

>me_ENSMEUG00000002955 Chain=None Position=Scaffold171902:3-1271(+) GeneID=ENSMEUG00000002955 ProteinID=ENSMEUP00000002695 Score=266.7 E-value=2.1e-78 Length=0 Comment=Not detected in genome search
FSTAMTRPGLGEPFLSVGVGDQEFVRFDSDSPQRPEPRAAWMEGLGQEEPGFWERRS

GICKAETQFFRGTLESLLSLNQSSGDVHTIQRHLHGCEVSPDLTFQRGFLLYAYDGEDYLV
LDTETRYRTAAVPEAVHIKQKWEAQEVERWKAYLEEECVLWLKKYLETGKETLMRKDSPA
AQVTRHTAPHGEVTLRCRAQDFYPEISLIWLDRDGEEQPDTEFIETR

>me_ENSMUEG00000003479 Chain=None Position=Scaffold19886:10213-12165(+)
GeneID=ENSMUEG00000003479 ProteinID=ENSMUEP00000003176 Score=203.4 E-value=7.8e-61 Length=0
Comment=Not detected in genome search
VHTFQRMAGCDVSPDLTFQRGFLQYAYYQDYIALDRDLHWTAVVPQAVNTRKRWAEERP
AEKWKGLLEEECVLWVKKYLEMGKEALMRTHPLVTRHTAPHGEVTLRCRAHDFYPKEI
SLTWLDRDGEEQPDTEFIETRPAGDGTFOKWAAVEVTSQEGRYTCRVQHEGLPEPLTLQ
WEPQSSLTWLVGAIAGLLLTAVIVGVGIWKRKHSRKEGDYVPAAGNDSAQGSDISLT
A

>me_HFE Chain=None Position=Scaffold56881:10201-12460(+) GeneID=ENSMUEG00000009762
ProteinID=ENSMUEP00000008889 Score=244.7 E-value=6.1e-72 Length=0 Comment=Not detected in
genome search
SHSLQFLFMGASHPDLGLPSFVALGYVDEHLFVYDHESTRKAKPRGPWRKMMVWTLTQS
LKGWDHMFIIIDLWTIMDNHSSQGGSHILQVLLGCELLVEDNRTRSFWKYGYDGDYLIHF
PETMNWTAQPEAQTTKQEWEMTKIRAKQHRAYLERDCPEKLQSYLETGSEILNKK

>me_PROCR Chain=None Position=Scaffold30751:8059-12290(-) GeneID=ENSMUEG00000012233
ProteinID=ENSMUEP00000011151 Score=101.4 E-value=1.6e-29 Length=0 Comment=Not detected in
genome search
EMLPLLSLFSWVICQGAMEGSQLHILQISHFWDPSSVQFWGNASLGLLTHTTLEGSGHN
ITIQQLKPLESPEHWKQTKGRLLNLYLSQFQLVQVVSERRVDFPLTLRNLGCELLPPDG
KDAHVFVEVALNGSSFSVFOPEKVLWSAAQTTPRAINFTLKQLNNYDQTRFQLLEFLQDT
CVGFLKQHMEVFSKPVSRVSPSYTMLLGIIMGVFTISSVAVGIFLCTGGRWS

>me_ENSMUEG00000004463 Chain=None Position=GeneScaffold_9858:12374-13315(+)
GeneID=ENSMUEG00000004463 ProteinID=ENSMUEP00000004058 Score=217.9 E-value=3.9e-65 Length=0
Comment=Not detected in genome search
VHTLQPMYGEVSPDLRFQRGFRQYAYDGDYIALDRDLHWTAAVPPAVNTRKRWAEERS
FAGKKAYLEEECVLWLRKYLEMGKEALMRTHPLAVTRHTAPHGEVTLRCRAQDFYPKE
ISLTWLRDGEEQPDTEFIETRPAGDGTFOKWAAVEVTSQEGRYTCRVQHEGLSEPLTL
QWEPSSFTWYIVGGIAAA

>me_ENSMUEG00000008250 Chain=None Position=GeneScaffold_9844:13965-14785(-)
GeneID=ENSMUEG00000008250 ProteinID=ENSMUEP00000007529 Score=210.1 E-value=1e-61 Length=0
Comment=Not detected in genome search
PASLQYDLTVVVSQPEPSQPLYMALGYINDQLFLHYDGESESRKVELQKPNYRGRKRVWVER
TQDLEEMGKELRVTLQDILRQNNQSKGSHTLQATLECELQGNDSRGRFRFDGQDFLT
FYREKLSWTAAPSAKKIKEKWRGTGSKRNKVHEEYCSIRLQRYLASWKGILERT

>sh_chain01 Chain=chain01 Position=GL861423.1:192-2390(+) GeneID=ENSSHAG00000000117
ProteinID=ENSSHAP00000000133 Score=380.7 E-value=1.9e-105 Length=2199 Comment=Overlaps single
annotation & in proteins
MGWDLGVRRLRFVHSPGGTRSGGAPRAWPLTPGSRGRPHGVPALSPDRVPPSVSGSHS
MRYFDTAVSRPGLGEPRLAVGYVDDQQVFRFDSASQSEEPRAPWMEKVQDVPDGYWE
RNTQISKENAQISRVDLQTLRGGYNQSEGAGAHTFQRMYGCEVSPELSFRGFLQFAYDG
QDYIALDTEFTLWTAQAQNEAVNTRKRWAEERSYAERDKAYLEETCVLWLRKYLEMGKESL
QRADAPSARVTRHSTPSGEVTLQCRADQDFYFSEISLAWLRDGEEQPDTEFIETRPAGD
TFQKWAAVGVPSQEGRYTCRVQHEGLPEPLTLKWEPESSLPWIVGVLAALLLTAVIA
GAVVWRKKTSGGKGGDYVPAAGNDSAQGSVSLTAKG

>sh_SAHA-I Chain=chain02 Position=GL856884.1:746440-748777(-) GeneID=ENSSHAG00000010776
ProteinID=ENSSHAP00000012589 Score=366.4 E-value=2.2e-101 Length=2338 Comment=Overlaps single
annotation & in proteins
LPTMGSYACSLFLLGALALKETWEGSHSLRYFGTTVSRPGLGEPFFSVGYVDDQQFVGF
DSDSASQRVEPRAPWIEKMENVDRDYWERNTQNSKRNAQISREDLQTLHGYYNQSESGVH
TFQRLVGEIISPDLSFKRGRFDQYAYDGDYIALDTEFTLWTAQAVNEAVNSKHKLEAERSN
AERDKAYLEETCVLWLRKYLEMGKESLQRADPPSVQVTRHSTSNGEVTLQCRADQDFYPAE
ISLTWLRDGEEQPDTEFIETRPAGDGTFOKWAAVGVTSQEGKRYTCRVQHEGLPEPLIL
KWEPESSSPWIVGILLAVAVFLLTVVIVAGAVVWRKKTSGGKGGDYVLAAGSDSAQGSVSL
LSAK

>sh_MR1 Chain=chain03 Position=GL856756.1:514408-537494(+) GeneID=ENSSHAG00000008980
ProteinID=ENSSHAP00000010401 Score=314.5 E-value=7.7e-89 Length=23087 Comment=Overlaps single
annotation & in proteins
MTVAIQEGHARTHSLRYFRLGVSDSTQGIPEFISVGYVDSHPITSYDSIRRQKMPQASWM
EENLGSDDHWEKYTQLLRGWQTFKTELRLVQNHYNHTGGFHTYQRMIGCELLEDGTTGF
LQYAYDGDYIALDTEFTLWTAQAVNEAVNSKHKLEAERSN
DFGKDSFQRTENPLLRGCKKSSLGITTLICRAYGFYFPEITMTWIKNGELIIQIEYGD
ILPSGDGTYTQWISIEIDPQSKDHYFCQVEHNDFLKVLHVPIEPTISPFVEIISGFTVI
VLFILGLGIFVYRRKQSGIKEANYIPTPVKKYRTNPS

>sh_chain04 Chain=chain04 Position=GL857536.1:46569-49389(+) GeneID=ENSSHAG00000002942
ProteinID=ENSSHAP00000003345 Score=309.6 E-value=2e-84 Length=2821 Comment=Overlaps single
annotation & in proteins

MKEKVAVEAARIPRWNEFSAFATEAGPGLGLGGQSWLWLSLGVHSLRYFQTSMSQP
GLTKPRFISVGVVDDQAFVRFDSDSAGLREEMVMWLDKSNQNYWERNRVIWETARTFO
VGLQNLQFYYNQSEGAHYQRLVGCVEVYDWFILRGRFQFAYNGHDYISLDLTLSTWVA
EPTALNSKWKWEAERSIAQRQKEYLEKCVWLKRYLEYGKELLRAPPSVRITHRTHP
DGKVTLLQCRSQDFYPAEISLTLWRDGEQHQDTEFIETRPAGDGTFOKWVAVEVTSQGEQ
NYICRVQHEGLPEPLTLKWEPESSSPWIIIVGVLAAVFLNNAVIAGAVVWRKKTSGGKGGD
YVQAAGNDSAQGSVDVSLTAKA

>sh_chain05 Chain=chain05 Position=GL857227.1:23521-26690(-) GeneID=ENSSHAG0000001982
ProteinID=ENSSHAP0000002229 Score=292.6 E-value=2.5e-79 Length=3170 Comment=Overlaps single
annotation & in proteins
MEHYMLFLFFLAALTQTETWAGCHTMKYFDTIVTGPSSLGLEERWFMTVGVVDDQQFVRFH
NSSGSQITEPRAPWIELETPDYWERETRLKDSQHCPSLQNLHFNYNQSDGGVHTLQR
MYGCEVFPNGSFSSTFFLRYGYDGDQKLSLDPETVIWIASDQVALNLKLETRSLAERW
EVLKTECVQELLRHLEKGGKTHYPLDPPSVRVTRHITSREVFLLKRAWGFYFSEISII
WLKDGEEQIDQTEYIETRPGGDGTFOKWAAVEMPFNEEKYTCLIQHEGLPAPLSLKWEP
QSTSNGLFMGIIAVALLLFTAVVAGVVIWKKSTLGGKGRNYALTSEIDSAQGSVDVSLTAR
G

>sh_AZGP1 Chain=chain06 Position=GL834527.1:2588626-2594335(+) GeneID=ENSSHAG00000017241
ProteinID=ENSSHAP00000020329 Score=250.6 E-value=1.2e-69 Length=5710 Comment=Overlaps single
annotation & in proteins
MMGSLTTFIFILLFMDTTMSKKSGARCYTLIYQDTALSHPGPEEFSFTNTGYLNGQAFYQ
YDKSKQKAIPLSPWDKVEGMNDWEKESQLQKRREDFVLENMQNITDYNDGKGGSHIFQGR
FGCQLCGDNFAKGFWKCQYDGRDFIEFKTTEIPAWIPLDPAADVIKQRWEADPGAVRYR
AYLEEECIGTLRTYLELQKAYLHQLIPPKVNLSHYSAPGGNLVLECLAYDFYPREIELHW
LQDGKIQKKEVGQTIIGNGTQVSVITVNIPLKKKSYSCHVHESLPEAFIVAWDEE

>sh_chain07 Chain=chain07 Position=GL849604.1:1859453-1860996(+) GeneID=ENSSHAG00000015612
ProteinID=ENSSHAP00000018394 Score=228.3 E-value=3.6e-63 Length=1544 Comment=Overlaps single
annotation & in proteins
MAPEKQNKNSDLSLFLFLGSLANAQSHILOYNVFTVIDGELLFVLYLNDLFFYY
DSKSQTEPRPIWLDISIMKTDTWKTENSNKEIIKAFRIIMGKVMQSDQNTDSHTFOEI
LGCCELHKDGSIRGFLRFYDGEDFLIFHSETLTWEAVHPAARSLNFAEAEPEVETKIQA
KLEGGDYCAGLQGYLEEVWKEEKPVFPPLNLTQSKNPKIITLRCRVNFFPGDIKIMWYRN
GDTLSQRDQEGGIIILPSGDGTQYTWVSTDVYSEKSNYCHGEHQGRNQSTSISLGPVSEA
NRKYGIIPALFFGVFAAVHMV

>sh_UT15 sh_chain08 Chain=chain08 Position=GL835378.1:97853-100259(+) GeneID=ENSSHAG00000004293
ProteinID=ENSSHAP00000004897 Score=219.9 E-value=1e-60 Length=2407 Comment=Overlaps single
annotation & in proteins
SAHSHDFHFTAIGRATSLLNFTVISSVDDVQLCYNKQDQVVMKKAWMYEALGAAEIE
RKQKLLISLEKTFQWAFRSLIKNETEKTHMLQVFDCEVDRDQVSSHQFQALDGEDFCQ
MDEQLGHVWAMKPEAQRFKPLWDSALWKRQVKKHFLQEDCVDSMTKVLQYSSMSSLPSVPP
EVTVSRHDAPDDGRVTLFCRATGFYPRISITLHWEREGALGVWQGESSGILPNADATFYL
RVTLELPPSDPGTGYTCVVEHSELETPAVFPVPGKPPMERPWITVLSTAIVILVLSACG
AFIWIWKRKAGVYVWREGKEQEREIMRE

>sh_UT16 sh_chain09 Chain=chain09 Position=GL835378.1:111514-114286(+) GeneID=ENSSHAG00000004579
ProteinID=ENSSHAP00000005233 Score=191.7 E-value=2.6e-52 Length=2773
Comment=Overlaps single annotation & in proteins
HTLEMCFTAVGTTKSLDFTMVSSMDGVQGSFYDKKQQLVIKEAWVSQALGAHYIEKRR
QKLVFSEINFLWALQNWQNDIKGEDNHTVQFVHWDCHLDGDIHVSSHFWYAVDGEAFQCV
DEQLRHVWAMKPEACFKPFWEMIFPYKKIKHYMQEDCPEPLRKLKYSQTSLSLPSVPE
VTVLRHDAPDSDRVTLFCRATGFYPRISITLHWERDQALGVWQGESSGILPNADATFYLQ
VTLELPPSDPGTGYTCVVEHSELETPVIFPEVPEKPSMERRWAMALSILATVILVLSACG
AFNIWREKKAGTYDDLTANGNSNADDGDVIFI

>sh_FCGR1 Chain=chain10 Position=GL849778.1:124944-126410(-) GeneID=ENSSHAG00000004834
ProteinID=ENSSHAP00000005532 Score=184.4 E-value=8.5e-50 Length=1467 Comment=Overlaps single
annotation & in proteins
VSSPCLSPILVGLAFSGCPRGSLGPGGLGAGTTPSLFYQLTAVSSAPRGTPKVVWGSGLGP
QLFLTYSSGGNAEPWGAWRWPEQEAWFWEETWYLKQERLLQALRVSKEGAQIFQGLM
GCQLKPDNSSQPTARFALDGDALLTFDPTRRDWFHGSVEASVNRNWNLSQLAEKVAEF
LLITCPQRLKSHLQRGKHNFWQKEAPEVRAGSHVGPVAVWSTLSCQAFVFPPELELKF
REGKPAELSPGLEPWPNGDGTFFHSRGTQVPSGDEALYSCTVQHPALTESITVNFETPG
RLPLPVRVSLVAGSVLFLACLAAVACMIYRKRGRPALWIFRRRRAGDDVGDVLSAPGS
AQDSSS

>sh_UT17 sh_chain11 Chain=chain11 Position=GL835378.1:126345-131520(+) GeneID=None
ProteinID=None Score=179.6 E-value=8.9e-49 Length=5176 Comment=No overlapping annotations
SHLALHRHEIFFTAIGTTKLLDFILVSYVDDVEIAFYDKKHQQLVIKEAWVSHALGAEF
IEDRQQTILFSEIEYLWALQNWQNDTGNHTVQIWHNCHLDQDIHVSSRIWYAIIDGEDFL
QMQDEQISHWVAKKPEAKPLIPLLEKISCSKMKRYLEDYCIQPMRKILRYSVTVSQHDTF
DGNVLSCTARGFYPRISILLRWEKDEQPVTWGQENSSDILPNIDATFYLQITLKLQSKDT
EPGYTCVMEHSELESAPVY

```

>sh_UT22 sh_chain12 Chain=chain12 Position=GL834773.1:299772-302023(-)
GeneID=ENSSHAG00000007260 ProteinID=ENSSHAP00000008378 Score=174.4 E-value=3.9e-47 Length=2252
Comment=Overlaps single annotation & in proteins
IFSSSLPSHNSVHHRYVGFQFTAVGTYYSLLELSAISFMDEVEVGFYNKENQQLI IKIPWL
TEALGDAYIIQKRNLVYEQHFRWMMHFLAMNDTNHNRNHTVQLLADCEIDNDIKVKS
YIHIWGDGEEYRIDEVEVQWEYLKPEAKQYQHILESPFWDLRKRYMNOFCVDFMCKIMG
YKSIRILIPVPEMVTYSRVNTEGSIILSCTATGFYPRSIQLQWEKNGELGVWGKETSTG
ILPNMDSTFYLRRLTLELPPEDPGMNFTECVVEHVALKTPALYSVPGKLTKEKSFFLALGIV
LAVILLSSCAGAFIVWKKRCT

>sh_UT10 sh_chain13 Chain=chain13 Position=GL835264.1:51864-58802(+) GeneID=ENSSHAG00000003088
ProteinID=ENSSHAP00000003516 Score=168.2 E-value=1.9e-45 Length=6939 Comment=Overlaps single
annotation & in proteins
MNFSWFLKFLIQNDTKKESLSTENHTMQLFIDCELNSGIEVGSRIWFALDGEDFLKTDD
QTNHWTAMKPEAEHIKYMESYFGVKMRDRTIKIYCYDMMRKILEYSSMKENVAPEVTVT
RHDAPDGRVTFNCIATGFYPRSIQLLWKKDEQLGVWGQESSGTLPNADSTFYLRITLLEL
PPGDSGTGYSCVVEHSLKQKPDLVVPEKPAVKRPPWGMALGLLTIIMLLLCAGPFIWKK
KKTG

>sh_UT8 sh_chain14 Chain=chain14 Position=GL835264.1:79331-82841(-) GeneID=ENSSHAG00000003889
ProteinID=ENSSHAP00000004421 Score=166.5 E-value=1.1e-44 Length=3511 Comment=Overlaps single
annotation & in proteins
GVHEMORYGVCSHALVRSGLGKPSCLPMDFLDDVQVLSYDKHTQQLKAKEDWVVQALGAKF
IAKTQQLLDEHEKDFLWFLKMLNETHKRDVNHVQVIFALCEIESDSHIGNEIQIADVGH
EFSVLDDMEGLWIALMSFAQSFLPVLVTLWTTQRELYMQEYCIDTMKKILQHSVINFLP
SVPPEVTVSQYETVDGITLSCSATGFYPRSIILLHWQNGKNIIMAGKQSSSGVLPNADST
FYQRITIELLPEDMGTDYDCVVDHIELGPP

>sh_HFE Chain=chain15 Position=GL856859.1:71173-73362(+) GeneID=ENSSHAG00000003686
ProteinID=ENSSHAP00000004184 Score=163.9 E-value=6.5e-44 Length=2190 Comment=Overlaps single
annotation & in proteins
LFVYYDHESQKAKLQDPWMQKKDKFWMRLTQSLKGDWHIFFVDHNGQSQRILQKVFIFSC
ELLMEDNRTRSFWNXYGDDQDYLLFHQEI MNWTA VQSKSQATKQELMKNIRAKQHRAYL
ESDFPEYLQEIYILSLQIAVLSVTPLVVTRHATHKGTITLRCQALNFSPMTITLSWLWDG
KQLNQGILQDGIYPSSEDTFQVWTAVDITPGEEQRYAYQIEHMGLINPSLCLGHGPLSVT
LATGISTGIIVSIIITTVVIIWKKRNMRSGLIRNYILAEDK

>sh_UT24 sh_chain16 Chain=chain16 Position=GL835264.1:26577-33674(+) GeneID=ENSSHAG00000002090
ProteinID=ENSSHAP00000002357 Score=163.8 E-value=5.2e-44 Length=7098 Comment=Overlaps single
annotation & in proteins
MMNWRRESGFSFSSWLLILVLFALRGIQTGHKNEIKHIAVGTSKTLLDLTVVALIDDIW
WASYFKSNQHMVFKADWIEVLGSLNIEEMHLLINHKEFDQFVHYLTRNDTETEGNHT
LQIQLDCELDGDIQLSSHVKYAFDGEDLIKVDELEGQWVVLNPKARNFKLVNSPFFWTEV
RKRYIKRYCVGAMQKIIGNSNMKKNESPEVYVSQQDYPDGTTLKLSCTATGFYPOPIILLHW
KGTGDAIWGKSSSSTLPNSDDTFYLRISLEIQPGDSVTDYACIVEHSELERPVVYVVP
KPKYKKNFVWVALSILLGVILTVGCLVVFVQWKKRKSVMYC

>sh_UT25 sh_chain17 Chain=chain17 Position=GL835264.1:100680-102932(+)
GeneID=ENSSHAG00000004299 ProteinID=ENSSHAP00000004905 Score=163.4 E-value=6.5e-44 Length=2253
Comment=Overlaps single annotation & in proteins
MDDIEVGSYNSADQITVKIPWISQVIEADYTNQIHHLVLAHEQNTHWLWQFLSNNDTNH
NRNHTAQLLAECEIDNDIKIKSRIHLIWEGEYRIDEVEVQWENLKPEVKKLQYILDSP
FWTNLRKNYMNIEYCVNLLRKRIVGYSSLRDNPVPEVTVSRHVNPEGRIILSCTATGFYPRS
ILLQWKKNKNGELGVWGKETSSGILPNMDSTFYLRRLTLELPPEDSGTGTVCVVEHTELKTPA
VYPVPEKPTLTKPPVLTVSIVLAVILVMSCAGAFIAWKKRKSVMYC

>sh_chain18 Chain=chain18 Position=GL856810.1:337391-341089(-) GeneID=ENSSHAG00000007630
ProteinID=ENSSHAP00000008811 Score=160.7 E-value=4.4e-43 Length=3699 Comment=Overlaps single
annotation & in proteins
MLSCYLPFLLLFSGLSEGWGKTASQEPITFCQLISSFLNDSWVQNLGSGWLGDLETHR
WDLQSTIQFLWPWARGHFSAEQWKKLQSTAVFLISFTRDQDFIKVFRKDYPPVVIQVR
VSYSEGPSVFFQAAFGQTFDMHFQGD SWK PPGAESTSWNISRIINQDQGTDRMLQNL
NHTIPQFVGGLETFQKDIERQVRPEVWLSSTSTTPGQKLLCHVSGFYPKLVRVTWIK
NGQE QPQTSDLLPNSDGTWWIQVILIVEAGNTANLACRVEHSSSLGGQDIIQYWGKTWG
MGIIIAGIVLGLLLIGIYCYLRIVYKRHRRLYEDMM

>sh_UT23 sh_chain19 Chain=chain19 Position=GL835264.1:133457-136209(+) GeneID=None
ProteinID=None Score=152.4 E-value=1.1e-40 Length=2753 Comment=No overlapping annotations
YNSAHRHMGQFTAVGT CNSLLELSVISFMDDVPGVSYNNIYKQINIKIPWIFKALGADY
ITKRDHLLVHEQHFWRWIIPFLSKNDTNGNHTAQLLTDCEIDNDVKKVSHVHLIWDGVEY
YRIDEVQWQNDNLKPEYQYQIILDTFYPWTDLRKRFMSEYCVDLLRKKIIGYSVTVSRHVN
SEVRIILSCTATGFYPRSIILLQWKKNKNGELGVWGKETSSGILPNMDSTFYLRRLTLELPPED
SGTGFTCVVEHTELKTPAVY

>sh_UT14 sh_chain20 Chain=chain20 Position=GL835378.1:71054-77150(+) GeneID=ENSSHAG00000003654
ProteinID=ENSSHAP00000004148 Score=138.6 E-value=2.1e-36 Length=6097 Comment=Overlaps single
annotation & in proteins
MGNWKKRGSLSWKLVLVIFVLKEVQTGQHRHEFYFTAVGTTGSLNFTMVSFIDDIQL

```

CSYHKRNQISGMMECIPSIILSANFSEQMOSALMKHEEDFHRLICYLEQNDTQSEESHILQ
LFVDCHLEDDIQMDSHVQLAWDGEDLSRIDEQQGYWIFLKMGYSLKRGLESAFWETVRK
NYMQEYCYDIMRRILRYSRIKENGESHPSLIFSQEKQISSMLLLWSSS

>sh_UT2 sh_chain21 Chain=chain21 Position=GL834773.1:262941-287202(-) GeneID=ENSSHAG00000007033
ProteinID=ENSSHAP00000008107 Score=138.4 E-value=1.9e-33 Length=24262 Comment=Overlaps single
annotation & in proteins
MGCELLQMRTRRRKGFIPWMPFVLGVFALRETKAALHNLETQFTGLAITDNLDPDIIFNVW
VDGQLLFSYDTQNKELLIKLWAYPPLKNKLMKLEHLQGGEDLRRFCAYWMVSYNE
TWEVQTTQITVFCELDGDIVDSRMRVGFVDFGETVCQLDEQSEGWVTKKSEVTRFCNYLEE
LWQWDRILVDECPFLKIVLELFLKENEPEVTVSRHDGLDGKITLFCCTARGFYPRPIL
LHWEKDGQMGVWVKETSSGTLPNADATFYQLTVELELSDTGDGYTCVVEHCELGVPAVY
PAPGKVTGRKSWELSLISATTILLVSAAFIWRKKDAWGHRSQEHNEEEAV

>sh_UT1 sh_chain22 Chain=chain22 Position=GL834773.1:260795-263307(+) GeneID=ENSSHAG00000006681
ProteinID=ENSSHAP00000007692 Score=117.4 E-value=2.6e-30 Length=2513 Comment=Overlaps single
annotation & in proteins
MEPQWKMQSSSFWMFLFLGIFLLTKTQAEHLSLEFQIIVVLSRIPPDIFFNGLDDHLII
SYNSPSKDLIFKLDNRVNTPMQNFMIQHRYTELLKLETEIDKILQDKTQYYNWTEETHGQ
AFMNCEIDRGIILVNSHMVAFDGEDVCHLDMKEERWIVMKLEDEFFCPYQKDFHWANTIR
GDCTFLLKLLQIVHLKKEKTPPEVTVSRYHSQNGSLIFSLCATGFYPRILLHWEKDGKL
GIWGEESSTGTLPNADSTYYLHVTLKLPPEDPGKGYACLIHSELEKPTVYVVPGETTKR
SSWAVILSSTAAAFIVLSLLAAFIWKKKKTGMTPLSQLECSFPPLFSPAL

>sh_UT11 sh_chain23 Chain=chain23 Position=GL834746.1:1649434-1653202(-) GeneID=None
ProteinID=None Score=92.2 E-value=5.7e-23 Length=3769 Comment=No overlapping annotations
CFPAFHRHEGQFIIVGETSLLDYITLHVIDDVNMCSDYKDRKQLVVKESWMSLALGEEFL
KHKTQSSTENHTIQILAICELDRDIEVKSQICIALDGEDDFQVDNQVHDHVVAVSHRDSL
NGNITLSCLATGFYPRISQLYWEKNGRLGVWQERSSGILPNADATFYQLVTEIPLRDT
YTDYTCVVEHSALEKPVIC

>sh_PROCR Chain=None Position=GL834625.1:3415636-3419909(+) GeneID=ENSSHAG00000018276
ProteinID=ENSSHAP00000021578 Score=84.2 E-value=2.8e-24 Length=0 Comment=Not detected in genome
search
TPREMLLLLFLGWFVCCQGAVEGSGSFLILQISHFLDPFVSQVFWGNASLAGLTHHTLEG
SNHNITIQQLKPLESREDWQQTKEHLLTYLSEFQALVQLVKNKERRVTFPLTLRCSLGCCEL
PPESQDAHVFVEVALNGSSVFSFQPEKALWSACPDQGHPELYNFALKQLNNYNRTRFEL
LEFLQDTCVSFLKQHMDRVSGSVSQVPRSYTMLILGIIMGVFTISSVAVGIFLCTGGRWS

>sh_ENSSHAG00000017227 Chain=None Position=GL834527.1:2571916-2576949(+)
GeneID=ENSSHAG00000017227 ProteinID=ENSSHAP00000020312 Score=164.6 E-value=4.4e-48 Length=0
Comment=Not detected in genome search
MMGALVTVFLLFFSGTVLGDQCDLSMYRDLISSKPPQFSFRNTAYFNGERVIYQYDSNNQK
AVPEPEWKNVQDWDKESQIQKQRGDFAIETLKEIMNSSKKERDHSVNLGRKICELCQNNNA
TSGWSYAYDGNPPIRFDKDKVAWIAEQPAATVIKQKWEENPGAVNRTKEYLEKECIETL
KKFRDYKKAN

>oa_chain01 Chain=chain01 Position=Contig7965:6083-8321(+) GeneID=ENSOANG00000005224
ProteinID=ENSOANP00000008299 Score=327.9 E-value=4e-90 Length=2239 Comment=Overlaps single
annotation & in proteins
METPALPRLLFLFLFLGSGRLPATWARSHFLRYFYTAVSRPGVPAFTAVGVLDLDDQF
VRFDSNRQKVEGLTTWIIQGGQPDYWEQQNQELRGTQQIFLQNLQVALSYNNQSEGGYHS
YQEMYGCELRDGTGKAYDRYGDYDQYITLDDLTLTWTAAANPEAQYTKRKEANKKLL
ELEKAYLQGGQCVYWISEYLKLGDSLNRTEPPSVRVTRHPSQEDNVTLRQALGFYPLA
IRMRWQRDGEDLDRDTEHVEVTRPGDGTQKWTAVVGVSHGQEQRYVVCVVDHDLANPLA
VGWVSDLSPRKAVIMGALATVLIIVTAVLAGVVILRKRPRGEGQSYIIPAASE

>oa_chain02 Chain=chain02 Position=Contig13429:11483-14679(+) GeneID=None ProteinID=None
Score=234.0 E-value=1.9e-62 Length=3197 Comment=No overlapping annotations
TLSGSQSLHYFFLTMSSSGLEVPNFVSVGVVNDKLFVIYDKSESGSHSIQRMFGCEIHED
NSTSAFWQFGYDREDHLLDLDLTLTWISANPLAQDTKKWSVQEVCAQNNKAYLQGLCLT
SLLRYLELQGRDLRDKVRVTRHLAHNGEPMRLRCGAHGFYPRDIRLSWWRDQQLTQETEH
VESRPGDGTYSWGAWEVPSGEEHRYTCRVEHLGLEQLPTV

>oa_chain03 Chain=chain03 Position=Contig11418:5983-9812(-) GeneID=ENSOANG00000001818
ProteinID=ENSOANP00000002884 Score=203.6 E-value=3e-53 Length=3830 Comment=Overlaps single
annotation & in proteins
TAIPFFSLFLGGFVALFPTKRELHSLKFFSFVAVSESPGVPEFGLHVYLDHFLAQYDSE
KQKLFQNKRIENRVGHGYIEKQILQISLWQDFWFSQMQDIMSHYHHSAGSHTFQNVIG
CEIQDNSPKGIFFFYGDYISYDLDKETWTTKDSVAKITEQKWNATEAHRKSELDDK
CVQTFQDLVKYKYLKLDLKERPSGLVTQHTGPNREMILKCVVFSFSSPNIVLSLWLDGEE
MEQDTKHSKARPSGDGTYSQKASVEIIPGKEKEYICRVKHGELSAQFNQAWLEPDSLTI
WVIAFCLVAVL

>oa_UT3 oa_chain04 Chain=chain04 Position=Contig20479:7985-10093(-) GeneID=ENSOANG00000015154
ProteinID=ENSOANP00000023863 Score=188.1 E-value=1.7e-48 Length=2109 Comment=Overlaps single
annotation & in proteins
GVLESLIHLTMHLPRTEHHIGTFFLTVAVSGYDGFELTAIILLDQGMASYNSTVREVV

FSLNWLQVVGQQLVQEKKAELERYEKDFRWGMENWARYLNHSSRTQTQVQLIMGCELDRLG
VQVVSRRFRFAYEGQDVLWLDELKGTWVTAGPAKKQFRHFWEERVVFSRVAEHYVREECPP
LMRMVLHFWYLREHTFSLSPSPVPEVTVTRHDAKDGRIVICLATGFYPPSSILLRWVKDG
EVLGLWGEDSSGTLPNADFSFYVRQTLLEVQGEVDSGYACVVEHSALDGPFFFFGAGRTD
ILPPAAPEKSSWLPMPDQALGVMAAALVLSPAVGI IWKKKKTAGFIPRPLPPTPDGNT
WKGEE

>oa_UT1 oa_chain05 Chain=chain05 Position=Contig21456:3669-10490(-) GeneID=ENSOANG00000012222
ProteinID=ENSOANP00000019341 Score=172.4 E-value=7.2e-44 Length=6822 Comment=Overlaps single
annotation & in proteins
VVRLDGQVLYYDSTRRELVLKLGWLYRALGARRIREKQKQKAYEEDFRWAVENWMSKH
NESGGTHIVQVLVGCMDGDTQVGSFFKFAYNMGDFCWLDELQGGWVAASPAVLRFEFPQ
QESTFWSKGVRSYKKECVTLMRHALRYWQLRELCEWKTPEVTVTRKADPDGSLTSLCL
ATGFYPPSSILLRWVKDGKGLWGEESSSGTLPNADDTFYLRQTLLELRAEVGDSYVCLVE
HSTLGVPTAYPGMSVTPRKPTWWIPWSLALGVAAAAGLVLSPAVGFILWRKKKQG

>oa_UT4 oa_chain06 Chain=chain06 Position=Contig31346:3527-6020(-) GeneID=ENSOANG00000007216
ProteinID=ENSOANP00000011481 Score=164.6 E-value=1.5e-41 Length=2494 Comment=Overlaps single
annotation & in proteins
MRFQNRSGGHTLQLIVGCELDQDIQVGRVFRFAYDGEDFYWLDEQRGVWISSGLAADRY
LPFWEGSPFVSEVRRYKKEECASLMKITLQYWNLIESTPPEVTVTRHDGPDGSLTSLCL
ATGFYPPSSILLRWVKDGKGLWGEESSSGTLPNADDTFYLRQTLLELRAEAGNPSYACLVE
HSILGVPTAFAPLKPWKMPWSLALGMVAAGVLVLSPAVGFILWRKKKTDPERLGHCCS
GGDGGPESR

>oa_UT2 oa_chain07 Chain=chain07 Position=Contig27418:1911-4018(+) GeneID=ENSOANG00000021638
ProteinID=ENSOANP00000027555 Score=162.9 E-value=5.5e-41 Length=2108 Comment=Overlaps single
annotation & in proteins
MEARRRLGFLASFFCLLLGSPSLPQTLEDHHTMYLTAVRGDDGFELTAITMLDGGKI
ASYNSDQIEIVFGQNWLYQVLGRNLIQEMKSDLDISEMDFRWGMENWAQYQNHRSRGSQTV
QVIMGCELDRLGVLVVSRFRFAYEREEMRLRLDELKGAWMTTGOAKKQFGHFWEKRASWFQE
TERYIREECPLMRMILRFWHLRVHTPPEVTVTRRDVKGDRVIFTCLATGFPPSSILLRW
VKDGEVGLWGEESSSGTLPNADSSFYLRQTLLEVREEAGDTGYSCVVEHSTLELPTSNPAP
EKPLWLMWPDQALGVMAASVLVLSPAVGI IWKKKKTDSDGGQIAASEEGMSSTT

>oa_chain08 Chain=chain08 Position=Ultra519:6073147-6073789(-) GeneID=ENSOANG00000004202
ProteinID=ENSOANP000000028190 Score=133.7 E-value=9.1e-32 Length=643 Comment=Overlaps single
annotation & in proteins
MEPLAFPMILLFLPGGCTLPDLSLHFFHYFRVVSQDPPELVPFMAVGVDDQWFMRFGLS
GQSAEAPSAREGRGLGYWQQTQNLQDTVQIFLGNLQISLSYYSQIEGGSQSYQETYG
CQLQANGSPGGLHHPWHAYLSGGLPGAQNTKWKKANRNFKSTWRGSEICIGSRHTCS
TQSLSCQTCGFYKPEIKVKWQWDMAMTQDPEHVETRPSGEG

>oa_UT6 oa_chain09 Chain=chain09 Position=Contig28479:1860-3816(-) GeneID=ENSOANG00000022708
ProteinID=ENSOANP00000029016 Score=80.1 E-value=1e-15 Length=1957 Comment=Overlaps single
annotation & in proteins
PSADHHTGTFFLTTCVCGDDGFELTTIILLDGHMMSYNTNREVVLKVNWLYQVVGKPK
VQEKKAELSYESDLHWMEKLFQYQNHRSRGSQTIQLIKGCELDRLGIQVISRFRLAYEGR
KVQADLPQDCHSGVLSPSHDSGFLKSLRAWELSGVQLASPTSQDLAAEVWRNAFSPSPA
SPEVTVTRHDARDGSIIFSLATGFYPRSIILLHWK

>oa_UT5 oa_chain10 Chain=chain10 Position=Contig34347:2183-4530(+) GeneID=None ProteinID=None
Score=78.2 E-value=3.4e-15 Length=2348 Comment=No overlapping annotations
PQAFLHTRACFTAVGEGIPFGLSMVMLVDRKQVTSYNNTVQELVVCLSWLYQAVGRKL
IQKTEELISYKDFHWGVGHTLQLILGCELN*DIQVSSIFQFAYDGDVTVTRHDGP
DGLTLFCLATKFPSSILLCWVKDGEE

>oa_ENSOANG00000022269 Chain=None Position=Contig46608:551-1719(-) GeneID=ENSOANG00000022269
ProteinID=ENSOANP00000028412 Score=99.4 E-value=9.6e-29 Length=0 Comment=Not detected in genome
search
GSHTFQNVIGEIQKDNSSKGIFFFGYDGEDYISYDLDKETWTTKDSVAKITEQKWNATE
AHRKSELLDKCVQTFQDLVKYKYYILDKGINSCYNYHYVSSTVLSAGFLCTYLIL
LILITV

>oa_UT7 oa_ENSOANG00000022751 Chain=None Position=Contig24929:1584-6219(+)
GeneID=ENSOANG00000022751 ProteinID=ENSOANP00000029074 Score=101.8 E-value=9.9e-30 Length=0
Comment=Not detected in genome search
THTVQLLMGCELDRLGVQVGTFRFRFSCGTGEDFFWLNELQGSWVADEPLGNHFLVWEERDF
LFRVAKHYIQEACAVLLRNILHFWSLTTHEPGPAEVTITLQEPADGYLTLSCATSFYPS
SILLRWVKRGCGRVPSSTLPNANGTYLHQTLLELQAKTGNSDYACLVEHSTLQVPTAY
LEWYLCAYNVPGSVLSAAVGI GLVLSPSVGFILWRKKKTG

>oa_ENSOANG0000001863 Chain=None Position=Contig13429:137-2452(-) GeneID=ENSOANG0000001863
ProteinID=ENSOANP00000002957 Score=253.3 E-value=2.5e-74 Length=0 Comment=Not detected in
genome search
SLIYFNICLFLSIQLPLSPVGSLSLQFFSFGVSEPGPRVPEYIVFGYLDLDDQLIARYDSEK
RRVEARAKWMEKGPWFDSQTLRGLSWQSRVMKYLNTIMGYNHSGGSHTFQHVVGCE
IRGDGSTRGVRYAYDGDQYIFYDTETRTWAANPVAKITKRLWEADTAHEAQRKFDLEH

ECVYFLQKFLKYATKHLGR

>oa_ENSOANG00000008306 Chain=None Position=Contig41088:33-2080(+) GeneID=ENSOANG00000008306
ProteinID=ENSOANP00000013202 Score=185.4 E-value=3.4e-54 Length=0 Comment=Not detected in
genome search
PLFKLGPVSEFTSVGVVDDVDFDINGGQGRAEPRTPWIQDNEGQEYWDQETEILRSQVON
FRVGLQNMYSYNNQSEEGSHTVQGMYGQLEDTTTGEFGMQYGYNGQDYIALDKAMLSW
TAVDARALNTRKRWEDDGTIAKYRKHIRFMLVDFFRRE

>oa_ENSOANG00000002636 Chain=None Position=Contig34233:2-7290(-) GeneID=ENSOANG00000002636
ProteinID=ENSOANP00000004195 Score=150.2 E-value=1.5e-44 Length=0 Comment=Not detected in
genome search
SHWLRYFYTVGSRPGSGVSEFMSLGYVDDVQFVRFEGAPWIQDNEGQEYWDQNTQIFKGW
AQMPFVGLRNLIMGYNNQDKNGERGPAGWSRLPRAPAMPGGVGPTEGNNNDNDNPLDDK
SNSTFVVITLSSSLGESNVTSDELILFPSSKPYLKLTSSESSRSQTGKGGPVSPADPPEVEV
TRHTGPDGEDVSLRCRDLGFYPADIKFKWEREGEDMSQEMEFVGTTRPSGDGNFQKWSVN
VPRGEERKYVCVVEHEGLGQPLAVKWAPVPDASPPIAAIVGAVVVILLV

>oa_ENSOANG00000003581 Chain=None Position=Ultra117:184139-187054(+) GeneID=ENSOANG00000003581
ProteinID=ENSOANP00000005680 Score=306.0 E-value=8.9e-91 Length=0 Comment=Not detected in
genome search
MLRISQLKVGVSRLWGSVQFHLFLFSWKGSQLHYFFLTMSSSSLEVPNFVSVGVVNDLFL
IWDYDRTHLAEARAPWMSAMKAEYWERENQORAWEKVQRAEIQAIMRNLNHSGGSHSIQ
RMFGCEIHDNSTSVFHFHGYDGEDHLLTDLDTLWISANPLAQDTPKKWSEQEVYCYAQN
KAYLQGLCLTSLRLRYLELGRGLNRKVPVVRVTRHLAHNGEPMRLRCWAHGFYPRDIRLS
WWWGQELTQETEHVGSRRPGDGTYSWGAVEVPSGEEHRYTCRVEHLGLERPLTVTWEP
PIDPSPIFTGAVVVLALSIVAAVTWMRRKRKRKRVSDSGSPSHLDCKPHIKLSDLPLIC

>oa_PROCR Chain=None Position=Contig24426:8073-11439(-) GeneID=ENSOANG00000015181
ProteinID=ENSOANP00000023902 Score=46.2 E-value=5.3e-13 Length=0 Comment=Not detected in genome
search
MFHVSHFLNASVVRFWGNATLGDVTHEMEEGGLPPTVROLESPEPPEIWAETLKDLYKAY
LEDFKHLVELLSRERGLSFPLTLTSLGCELPGRGSAHTYEVSLNASASFRFRPTPAL
WDESEWRDLAAFTLRQLNGYNTRYELQHLRVTCVNYLKRKAPHLACEHQTSFPPRSAS
DSTSDRPTLSFHPSEREGARSFTPLVMGIVVGFIFVSGVAVGIFLCTGGGRSS

>oa_ENSOANG00000003172 Chain=None Position=Ultra12:1975819-1976844(-) GeneID=ENSOANG00000003172
ProteinID=ENSOANP00000005033 Score=346.5 E-value=1.6e-102 Length=0 Comment=Not detected in
genome search
STKQARCAPFLPLLCLGAALLPAGRTRSHSLRYFTTGVSRPGVSEFTSVGVVDDVQF
VRFDGDGRRAEPRTPWIRDNEGREYWDQETKIFRSGAQNFWGLQNMIMGYNNQSEERS
HTVQGLYGCQLRENDTPGEGFMQFGYDGRDYIALDKATLSWVADAGALNPKSKWEADGT
IAERDKAYLEDECI EWLRRYMGFGGCLTRADPPKVDVIRHPGPDGEDVSLRCRAFGFYP
ADIKFMWRREGEDVSLQMLVDTRPSGDGNFQKWSVSVPRGEELKYVCIVEHEGLAQPL
AVKWVPDTSLSNKATVGVVIVILLILVAVVGFVIWKRSGPRG

>oa_ENSOANG00000003398 Chain=None Position=Contig21219:1448-6348(+) GeneID=ENSOANG00000003398
ProteinID=ENSOANP00000005392 Score=206.9 E-value=2.2e-61 Length=0 Comment=Not detected in
genome search
GSHTFQWMYGCQLREDDTPGQGFVQLGYDGRDYVALDKATLSWTAADAGALNTRKRWKWEVD
RTEAERWKAYLEGRCIDGLRRLYLGFGGSLTRADPPEVEVTRHTGPDGEDVSLRCRALGF
YPADIKFKWEREGKMSQEMEFVGTTRPSGDGNFQKWSVNVPRGEEQKYVCVVEHEGLAQ
PLAVKWVDPSPPIAAIVGAVVVILLVTAAVVGFVIWKRKKS

>oa_ENSOANG00000029890 Chain=None Position=Contig99779:194-3066(-) GeneID=ENSOANG00000029890
ProteinID=ENSOANP00000029583 Score=106.6 E-value=7.2e-31 Length=0 Comment=Not detected in
genome search
MSIPSAWWILETDVGDHLKDLKRIATGLARDSQDSLRLDLPYYNQTEGGSHSIQTMFGCE
VEDNGTVRNGYQVGYDQDYIALERNHKNMTWAADTVAQITKNMWEANFFFLGCLGFF
MASVKYLRRARHCTKRWGVRLDMGPA

>oa_ENSOANG00000015040 Chain=None Position=Ultra5:6518080-6518628(+) GeneID=ENSOANG00000015040
ProteinID=ENSOANP00000023672 Score=181.5 E-value=4.7e-53 Length=0 Comment=Not detected in
genome search
GLLTLWTEATCSRLLYFTTGVTRPGPGASEFTSEGYVDHVQFVRFAGDGGRRRAEPRTPW
IRDNGGREYWDQETEIFRCGAQNFVGLQNMYSYNNQSKYGYDGRDYIALDKAKLSWTA
DAGALNTRKRWKWEADGTIAERYKGYLEGEICIECLQRYL

>hs_HLA-A Chain=chain01 Position=6:29910524-29913057(+) GeneID=ENSG00000206503
ProteinID=ENSP00000366002 Score=407.3 E-value=1.4e-118 Length=2534 Comment=Overlaps single
annotation & in proteins
MAVMAPTLLLLLSGALALQTWAGSHSMRYFFTSVSRPGRGEPFIAVGVVDDTQFVRF
DSDAASQRMPEPRAPWIEQEGPEYWDQETRNVKAQSQDTRVDLGTLRGYNNQSEAGSHTIQ
IMYCDVGSGRFLRGYRQDAYDGKDYIALNEDLRSWTAADMAAQITKRKWEAAHEAQL
RAYLDGTCVWLRRLYLENGKETLQRTDPPKTHMTHHPI SDHEATLRWCALGFYPAEITLT
WORDGEDQTDTELVEVTRPAGDGTFOKWAAVVVPVSGEEQRYTCHVQHEGLPKPLTLRWEL
SSQPTIPIVGIIAGLVLLGAVITGAVVAAMWRRKSSGEGVKDRKGGSYTQAASSDSAQ
GSDVSLTACKV

>hs_HLA-B Chain=chain02 Position=6:31322257-31324744(-) GeneID=ENSG00000234745
ProteinID=ENSP00000399168 Score=394.1 E-value=9.3e-115 Length=2488 Comment=Overlaps single
annotation & in proteins
MLVMAPRTVLLLLSAALALTEWAGSHSMRYFYTSVSRPGRGEPFISVGYVDDTQFVRF
DSDAASPREEPRAFWIEQEGPEYWDNRNTQYKQAQTDRESLRNLRGYNQSEAGSHTLQ
SMYGCDDVPGDGRLLRGHDQYAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQR
RAYLEGECVWELRRYLENGKDKLERADPPKTHVTHHPISDHEATLRCWALGFYPAEITLT
WQRDGEDQTDTELVEVTRPAGDRTFQKWAAVVVPSEGEQRYTCHVQHEGLPKPLTLRWE
SSQSTVPIVIGIVAGLAVLAVVIGAVVAAMVCRKSSGGKGGYSQAACSDSAQGSVSL
TA

>hs_HLA-H Chain=chain03 Position=6:29855722-29858262(+) GeneID=ENSG00000206341 ProteinID=None
Score=383.2 E-value=1.3e-111 Length=2541 Comment=Overlaps single annotation; not in proteins
SLPGSHSMRYFYTTMSRPRGEPFISVGYVDDTQFVRFSDSDASPREEPRAFWMERE
EYWDRNTQICKAQAQTERENLRALRYNQSEGGGPGSHTMQVMYGCDDVPGDGRFLRGY
QHAYDGKDYIALNEDLRSWTAADMAAQITKRKWEAARRAEQLRAYLEGEFVWELRRYLE
NGKETLQRATHMTHHPISDHEATLRCWALGFYPAEITLTWQRDGEDQTDTELVEVTRPAG
DRTFQKWAAVVVPSEGEQRYTCHVQHEGLPEPLTLGNSAQGSVSLTA*K

>hs_HLA-G Chain=chain04 Position=6:29795814-29798376(+) GeneID=ENSG00000204632
ProteinID=ENSP00000366024 Score=381.0 E-value=8.1e-111 Length=2563 Comment=Overlaps single
annotation & in proteins
MKTTPRMVVMAPRTLFLLLSGALTLTETWAGSHSMRYFSAASRPRGEPFIAMGYVDDT
QFVRFSDSDASCPREPRAPWVEQEGPEYWEETRNTKAHAQTDNRMLQTLRGYNQSEAS
SHTLQWMIQDGLSDGRLLRGYEQYAYDGKDYIALNEDLRSWTAADTAAQISKRKCEAN
VAEQRRAYLEGTCVWELHRYLENGKEMLRADPPKTHVTHHPVDFEATLRCWALGFYPA
EIIITWQRDGEDQTDVELVEVTRPAGDRTFQKWAAVVVPSEGEQRYTCHVQHEGLPEPL
MLRWKQSSLPITIPIMGIVAGLVVLAAVVTGAAVAALWRKSSD

>hs_HLA-C Chain=chain05 Position=6:31237116-31239655(-) GeneID=ENSG00000204525
ProteinID=ENSP00000372819 Score=378.7 E-value=3e-110 Length=2540 Comment=Overlaps single
annotation & in proteins
MRVMAPRALLLLSGGLALTEWACSHSMRYFDTAVSRPGRGEPFISVGYVDDTQFVRF
DSDAASPRGEPRAFWVEQEGPEYWDRETQYKQAQADRVSRLNLRGYNQSEAGSHTLQ
RMSGCDLPGDGRLLRGYDQYAYDGKDYIALNEDLRSWTAADTAAQITQRKLEAARAEQL
RAYLEGTCVWELRRYLENGKETLQRAEPPKTHVTHHPISDHEATLRCWALGFYPAEITLT
WQRDGEDQTDTELVEVTRPAGDRTFQKWAAVVVPSEGEQRYTCHVQHEGLPEPLT
SSQPTIPIMGIVAGLAVLVVLAAGVVTAMCRKSSGGHFLPTGGKGGSCSAACSNSA
QGSDESILITCKA

>hs_HLA-E Chain=chain06 Position=6:30457493-30460218(+) GeneID=ENSG00000204592
ProteinID=ENSP00000365817 Score=368.3 E-value=3.4e-107 Length=2726 Comment=Overlaps single
annotation & in proteins
MVDGTLTLLSSEALALTEWAGSHSLKYFHTSVSRPGRGEPFISVGYVDDTQFVRFDND
AASPRMVPRAFWMEQEGPEYWDRETRSDTAQIFRVNLRRLRGYNQSEAGSHTLQW
GCELGPDGRFLRGYEQFAYDGKDYIALNEDLRSWTAADTAAQISEQKSNDAEAEHQRAY
LEDTCVWELHRYLENGKETLQRAEPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQ
DGEHTQDTELVEVTRPAGDRTFQKWAAVVVPSEGEQRYTCHVQHEGLPEPVTLRWKPASQ
PTIPVIGIAGLVVLSVSGAVVAAVIWRKSSGGKGGYSKAEWSDSAQGSSESHSL

>hs_HLA-F Chain=chain07 Position=6:29691425-29694002(+) GeneID=ENSG00000204642
ProteinID=ENSP00000259951 Score=346.6 E-value=8.7e-101 Length=2578 Comment=Overlaps single
annotation & in proteins
MAPRSLLLLSGALALTDWAGSHSLRYFSTAVSRPGRGEPRIAVEYVDDTQFLRFSD
AAIPRMEPREPFWVEQEGPEYWEWTTGYAKANAQTDRAVLRNLRRLRYNQSEAGSHTLQGMN
GCDMGPDGRLLRGYHQHAYDGKDYIALNEDLRSWTAADTAAQITQRFYEAEEYAEFRY
LEGECLLELRRYLENGKETLQRAEPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQ
DGEHTQDTELVEVTRPAGDRTFQKWAAVVVPSEGEQRYTCHVQHEGLPEPVTLRWKPASQ
PTIPVIGIAGLVVLSVSGAVVAAVIWRKSSGGKGGYSKAEWSDSAQGSSESHSL

>hs_HLA-J Chain=chain08 Position=6:29974567-29977143(+) GeneID=ENSG00000204622 ProteinID=None
Score=340.8 E-value=1.2e-98 Length=2577 Comment=Overlaps single annotation; not in proteins
SPPGSHSMRYFSTAVSWPGRGEPFIAVGYVDDTQFVRFVSDAVSLRMKTRARWVEQEG
EYWDRNTQICKAQAQTDNRNLRALRYNQSEAGTGPYHILQGMFGCDLPGDGRLLRGY
EQYAYDGKDYIALNEDLRSWTAADTAAQITQRKYEANVAEQRRAYLEGTCMEWLRHLE
NGKETLQRAKTHVTHHPPL*T*GITRSWVLFYPAEITLTWQRDGEDQTDQDMEVETRP
TGDTFQKWAAVVVPSEGEQRYTCHVQHEGLPKPLIILPGSQAQGSVSLTACK

>hs_MRI Chain=chain09 Position=1:181018178-181041144(+) GeneID=ENSG00000153029
ProteinID=ENSP00000356552 Score=335.6 E-value=2.4e-97 Length=22967 Comment=Overlaps single
annotation & in proteins
MGELMAFLLPLIIVLMVKHSDSRTHSLRYFRLGVSDFIHGVPEFISVGYVDSHPITTYDS
VTRQKEPRAPWMAENLAPDHWERYTQLLRGWQMFVLEKRLQRHYNHSGSHTYQRMIGC
ELLEDEGSTTGFLQYAYDQDFLIFNKDTLSWLAVDNVAHTIKQAWAENQHLELLYQKNWLE
EECIAWLKRFLEYGKDTLQRTPEPLVRVNRKETFPVGTALFCKAHGFYPPFIYMTWMKNG

EEIVQEIYGDILPSGDGTQAWASIELDPQSSNLYSCHVEHCGVHMVLQVPQSEETIPL
VMKAVSGSIVLVIVLAGVGVLVWRRRPREQNGAIYLPTPDR

>hs_HLA-K Chain=chain10 Position=6:29894427-29897009(+) GeneID=ENSG00000230795 ProteinID=None
Score=291.4 E-value=3.9e-84 Length=2583 Comment=Overlaps single annotation; not in proteins
SPPGTHSIRYFSTAVSRPGRGEPRIYAVGVDDTQFVRFSDAATPRM*PQSPWLEQEGP
EYWRDRSTRNIRPAHRLTRVNLPMGPGSNTLQIMYGCDLGLGRLLRGYEQHANDGKDYIA
RN*DLRSWTAADMAAQITKRKWEAEFEAQIRAYLEGTCTVERLTHMHHSSVSDYKATLRC
WALGFYVPEITLAWQQDGEDQTRDMELLETRPAGDGTQKWAAVVVPVSGEEQRYPCVHQH
EGLPKPLTLPGSDHAQGSVSLTACK

>hs_AZGP1 Chain=chain11 Position=7:99564644-99569639(-) GeneID=ENSG00000160862
ProteinID=ENSP00000292401 Score=276.5 E-value=2.4e-81 Length=4996 Comment=Overlaps single
annotation & in proteins
MVRMVPVLLSLLLLGPAVQENQDGRYSLTYYITGLSKHVEDVPAFQALGSLNDLQFFR
YNSKDRKSQPMGLWRQVEGMEDWKQDSQLQKAREDFMETLKDIVEYYNDSNGSHVQLQR
FGCEIENNRSSGAFWKYYIDGKDYIEFNKEIPAWVPPDPAQITKQKWEAEPVYVQRKA
YLEEECPATLRKYLKYSKNILDRQDPPSVVVTSHQAPGEKKLKLCLAYDFYPGKIDVHWT
RAGEVQEPRLGDLVHNGTYQSWVVAVPPQDTAPYSCHVQHSLSLAQPLVVPWEAS

>hs_HFE Chain=chain12 Position=6:26091059-26093181(+) GeneID=ENSG0000010704
ProteinID=ENSP00000417404 Score=271.1 E-value=7.5e-80 Length=2123 Comment=Overlaps single
annotation & in proteins
MGPRARFALLLLMLLQTAVLQGRLLRSHSLHYLFMGASEQDLGLSLFEALGYVDDQLFV
YDHSRRVEPRTTPWVSSRISSQMMLQLSOLKGDHMTVDVFWTIMENHNHNSKESHVTLQV
ILGCEMQEDNSTEGYWKYGDGQDHLEFCPDTLDWRAAEPRAPWTKLEWERHKIRARQNR
AYLERDCPAQLQQLLELGRVLDQQVPLVVKVTHHVTSSVTTLRCLRALNYQPONITMKWL
KDKQPMDAKEFEPKDVLPNGDGTQGWITLAVPPGEEQRYTCQVEHPGLDQPLIVIEWPS
PSGTLVIGVISGIAVFFVILFIGILFIIILRRKQSGRGMAGHYVLAERE

>hs_AZGP1P1 Chain=chain13 Position=7:99580757-99601225(+) GeneID=ENSG00000214313 ProteinID=None
Score=2.2e-75 Length=20469 Comment=Overlaps multiple annotations; none in
proteins: (AZGP1P1, ENSG00000214313);(RP4-604G5.3, ENSG00000235713)
SFAGHYSLTLYTLGSRSGKGTTHRLQGTVFLNGHAFFHYNSEDRAEPLGPRWHAEGVED
WEKQSQVKAREDFMETLGSALQERFCIEIQNNRSTGAFWKNAVDGKDYIEFNKEIPA
*VPLVPEAQNTKQK*EAEPPVYVQRAKAYLEEECPATLRKYLKYSNILDQVVTSHQAP
GEKKLKLCLAYDFYPGKIDVHWTTRAGEVQEPRLGDLVHGGNGTYLTLWLLVHVPPQDTAP
YSCHVQHSLSLAQPLVVPQSSRLSLTNCW

>hs_HLA-L Chain=chain14 Position=6:30227582-30230118(+) GeneID=ENSG00000243753 ProteinID=None
Score=262.2 E-value=2.7e-74 Length=20469 Comment=Overlaps single annotation; not in proteins
SPPGSHSLRYFSTAVSQPRGEPRIYAVGVDDTEFVRFSDSVPSPRMERRAPWVEQEG
EYWDQETRNAKGAHQIYRVNLRLLRYNQSEAGGPDRLRLRYEQFAYDGDYIALNED
LHSWTAANTAAQISQHKWEADKYSEQVRAVLAHVTOHPISDHEATLRCWALGLYPAEITL
TWQQDGEDQTDTELVEVTRPAGDGTQKWAAVVVPVSGEEQRYMCHVQHEGLPEPLTLPGS
NCAQYSDASHDTCK

>hs_HLA-V Chain=chain15 Position=6:29760204-29770197(+) GeneID=ENSG00000181126 ProteinID=None
Score=241.2 E-value=5.2e-69 Length=9994 Comment=Overlaps multiple annotations; none in
proteins: (HLA-V, ENSG00000181126);(HLA-P, ENSG00000261548)
LLPGFHSRLRYFHTTMSRGRADPRFLSVGDVDDTQCVRLLDSATSPRMERPRAPWMEQEGP
EYWEETGTAKAKAQFYRVNLRLLSGYNNQSERMSVCDVGSGRLLRGLYHQLAYDGDYI
VLNEDLCSLTAADTAAQITQLKWEAARGAEVHAQVTLHPISNYEATLRCWALGFYPLEIT
LTQERDGEDQIQDAEFVETRLAGYRTFQKWAAVVSSGEEKRYTCHVQHEGLPEPLTLPN
RNSTQGSVSLKACK

>hs_FCGR1 Chain=chain16 Position=19:50017132-50041984(+) GeneID=ENSG00000104870
ProteinID=ENSP00000410798 Score=213.1 E-value=1.4e-60 Length=24853 Comment=Overlaps multiple
annotations; one in proteins
MGVPRPQPWALGLLFLPLPGSLGAESHLSLLYHLTAVSSPAPGTPAFVWVSGWLGPPQYLS
YNSLRGEAEPCEAWVWENQVSWYWEKETTDLRIKEKLFLEAFKALGGKGPYTLQGLLGE
LGPDNVTSVPTAKFALNGEEMFNFDLKQGTWGGDWPEALASQRWQQDKAANKELTFLLF
SCPRLRLREHLERGRNLEWKEPPSMRLKARPSSPGFVLTCSAFSFPPELQLRFLRNL
AAGTGGDLFGPNSDGSFHASSSLTVKSGDEHHYCCIVQHAGLAQPLRVELESAPKSSVIV
VGIVIGVLLLTAAAVGALLWRRMRSLPAPWISLRGDDTGVLLPTPGEAQDADLKDENV
IPATA

>hs_CD1A Chain=chain17 Position=1:158224873-158256729(+) GeneID=ENSG00000158477
ProteinID=ENSP00000289429 Score=207.2 E-value=4.3e-59 Length=31857 Comment=Overlaps single
annotation & in proteins
MLFLLPLLAVLPDGNADGLKEPLSFHVWTWIASFYNHSHWQNLVSGWVSDLQTHTWDSN
SSTIVFLCPWSRGNFSNEEWKLETFLFRITIRSEFEGIRRYAHELQFEYFPEIQVTGGCE
LHSGKVSQSFLLQALYQGSDFVFSQNNWLPYPVAGNMAKHFKVLNQNQHENDITHNLLS
DTCPRFLLGLLDAGKAHLQROVKPEAWLSHGSPGPGHQLVCHVSGFYPKPVVMMWRG
EQEQQTQRGDILPSADGTWYLRATLEVAAGEAADLSCRVKHSSLEQDQDIVLYWEHSSV
GFIIAVIVPLLLLLIGLALWFRKRCFC

>hs_CD1B Chain=chain18 Position=1:158299167-158300853(-) GeneID=ENSG00000158485
ProteinID=ENSP00000357150 Score=203.9 E-value=1e-59 Length=1687 Comment=Overlaps single
annotation & in proteins
MLLLPFQLLAVLFPGGNSEHAFQGPSTSFHVIQTSSFTNSTWAQTQSGWLLDLQIHGWDS
DSGTAFILFKPWSKGNFSDKEVAELEEIFRVYIFGFAREVQDFAGDFQMKYPPFEIQGIAGC
ELHSGGAIIVSFLRGAALGGLDFLSVKNASCVPSPGEGSRAQKFCALIIQYQGIMETVRIILL
YETCPRYLLGVLNAGKADLQROVKPEAWLSSGSPGPGRLQLVCHVSGFYKPKPVVWMMWR
GEQEQQGTQLGDILPNANWTWYLRATLDVADGEAAGLSCRVKHSSLEGQDIILYWRNPTS
IGSIVLAIIVPSLLLLLCLALWYMRRRSYQNIIP

>hs_MICA Chain=chain19 Position=6:31378310-31380001(+) GeneID=ENSG00000204520
ProteinID=ENSP00000413079 Score=197.6 E-value=7e-58 Length=1692 Comment=Overlaps multiple
annotations; one in proteins
MGLGPVFLLLAGIFFPAPPAAAAEPHSLRYNLTVLSWDGVSQSGFLAEVHLDDGQPFLLRYD
RQKCRAPQGGWAEDVLGNKTDRETRDLTGNGKDLRMTLAHIKDQKEGLHSLQEIRVCE
IHEDNSTRSSQHFYDGLFSLQNLETEEWTPVQSSRAQTLAMNVRNFKEDAMKTKTHY
HAMHADCLQLRRYLESVVLRRTVPPMVNVTREASEGNIIVTTCRASSFYPRNIILTWR
QDGVSLSHDTQQWGDVLPDNGNTYQTVWATRIRGEEQRFCTCYMEHSGNHSTHPVPSGKV
LVLQSHWQTFHVSVAAGCCYFCYYFLCPLL

>hs_CD1C Chain=chain20 Position=1:158260923-158262657(+) GeneID=ENSG00000158481
ProteinID=ENSP00000357152 Score=195.0 E-value=3.4e-57 Length=1735 Comment=Overlaps single
annotation & in proteins
MLFLQPLLLALLPGGDNADASQEHVSFHVIQIFSVNQSWARQGGSGWLDLQTHGWDS
ESGTIIFLHNWSKGNFSENEELSDLELLFRYFLFGLTREIQDHASQDYSKYFFEVQVKAGC
ELHSGKSPEGFFQVAFNGLDLLSFQNTTWPSPGCGSLAQSVCHLLNHQYEGVTETVYNL
IRSTCPRFLGLLDAGKMYVHRQVRPEAWLSSRPSLGSGLLVCHASGFYKPKPVVWTTM
RNEQEQLGTHGDILPNADGTWYLVQVILEVASEEPAGLSCRVRHSSLGGQDIILYWGHHF
SMNWIALLVIVPLVILVILVILVLFKHCYSQDIL

>hs_MICB Chain=chain21 Position=6:31473384-31475076(+) GeneID=ENSG00000204516
ProteinID=ENSP00000252229 Score=193.5 E-value=1e-56 Length=1693 Comment=Overlaps single
annotation & in proteins
MGLGRVLLFLAVAPFPAAAAEPHSLRYNLMVLSQDGSVQSGFLAEGHLDGQPFLLRYD
RQKRRAPQGGWAENVLGAKTWDTEDELTENGQDLRRTLTHIKDQKGGHLSLQEIIRVCE
IHEDSSTRGRSHFYDGLFSLQNLETOESTVQSSRAQTLAMNVTNFWKEDAMKTKTHY
RAMQADCLQLQRYLKSVAIRRTVPPMVNVTCEVSEGNITVTCRASSFYPRNIILTWR
QDGVSLSHNTQQWGDVLPDNGNTYQTVWATRIRGEEQRFCTCYMEHSGNHSTHPVPSGKA
LVLQSQRTDFPVSAAMPFCVVIILCVPCCKKTSAAEGPELVSLQVLDQHPVGTGDHR
DAAQLGFQPLMSATGSGSTEGT

>hs_CD1D Chain=chain22 Position=1:158151244-158152945(+) GeneID=ENSG00000158473
ProteinID=ENSP00000357153 Score=192.5 E-value=2.2e-56 Length=1702 Comment=Overlaps single
annotation & in proteins
MGCLLFLLLWALLQAWGSAEVPQRLFPLRCLQISSFANSSWTRTDGLAWLGELQTHSWSN
DSDTVRSLKPKWSQGTFSQQWETLQHFIRVYRSSFTRDVKEFAKMLRSLYPLELQVSAGC
EVHPGNASNFFHFAVQGDILSFQGTSWEPTEAPLWVNLAIQVLNQDKWTRVTVQWLL
NGTQPFVSGLLESGKSELKKQVKPAWLSRGPSPGPGRLLLVCHVSGFYKPKPVVWMMWR
GEQEQQGTQPGDILPNADETWYLRATLDVVAGEAAGLSCRVKHSSLEGQDIVLYWGGSYT
SMGLIALAVLACLFLLLIVGFTSRFRKQTSYQGV

>hs_CD1E Chain=chain23 Position=1:158324196-158325894(+) GeneID=ENSG00000158488
ProteinID=ENSP00000357149 Score=165.6 E-value=2.2e-48 Length=1699 Comment=Overlaps single
annotation & in proteins
MLLLFLLFEGLCPCGENTAAPQALQSYHLAAEEQLSFRMLQTSFANHSAHSEGGWLG
DLQTHGWDTVLGTIRFLKPWSHGNSKQELKNLQSLFQLYFHSFIQIVQASAGQFQLEY
FEIQILAGCRMNAPQIFLNMAYQGSDFLSFQGISWEPSPGAGIRAQNICVNLNRYLDIKE
ILQSLLGHTCPRFLAGLMEAGESELKRKVKPEAWLSCGSPGPGRLQLVCHVSGFYKPKPV
WVWMMRGEQEQRGTQRGDVLVNADETWYLRATLDVAAGEAAGLSCRVKHSSLGGHDLIIH
WGGYSIFLILICLTVIVTLVILVVVDSRLKQSSKNILSPHTPSPVFLMGANTQDTKNS
RHQFCLAQVSWIKNRVLLKWKTRLNQLW

>hs_MICE Chain=chain24 Position=6:29712045-29713514(-) GeneID=ENSG00000273340 ProteinID=None
Score=116.5 E-value=1.8e-33 Length=1470 Comment=Overlaps multiple annotations; none in
proteins: (MICE, ENSG00000273340);(HLA-F-AS1, ENSG00000214922)
SSPEPHSLCYNLTVLSQDGSVQSGFLAEGHLDGQFLFLCDRQKGRAGPRGWAEAVLGA
TWDTEDELTENGQELRRTLAHIKQKGGSTRGFRHFYTYGKFLFLSNETQ*WTVPRSS
RTQTLAMNVTNFWEEANAMQTKTHYCPVQADYMQKLW*YKFWRYQKNRYSE*DAQWGLRR
QCHHDVLAISIFYPWNIISLTHWQDEASLSQDAQSRRGVLPNGNGTYQTCMATRIPRGEQR
FTCYMGHRGNHSTHPVPS

>hs_MICC Chain=chain25 Position=6:30382477-30384167(+) GeneID=ENSG00000226577 ProteinID=None
Score=89.2 E-value=5.4e-25 Length=1691 Comment=Overlaps single annotation; not in proteins
FFPRSHSLHYNLTVLSQDGFV*SGFLAEGHLDGQPFLLYDRQKGRAGAGLHSLQEIRVCE
IHEDSSTRGRSHFYDGLFSLQNLETOEWTPVQSSRAQDDAHRVSVGLPNGNGTYQTVW
ATKIPQEEQRATCYVGHSRNHSTYPGLW

>hs_MICD Chain=chain26 Position=6:29938578-29940250(-) GeneID=ENSG00000229390 ProteinID=None
Score=82.3 E-value=1e-22 Length=1673 Comment=Overlaps single annotation; not in proteins
CSPGPHSLHYNLTVLSRDSVQSGFLAEGHLDGQPFLLFNKQKGRAGSRGQWAEAVLGAE
TWDTEFEDLTENGQDLRRALG*LFSLNLETQEWTVSQQSSRTQTFAMNITNF*KEDAMPA
ETHSHPIWVDCRQKLQYLESCDGYHTEVACRIPQGEQRVICSMGYSRNRHSTHPVSS

>hs_RAET1E Chain=None Position=6:150204511-150219238(-) GeneID=ENSG00000164520
ProteinID=ENSP00000349709 Score=170.2 E-value=4.5e-49 Length=0 Comment=Not detected in genome
search
MRRISLTSPPVRLLLFLLLLIALEIMVGGHSLCFNFTIKSLSRPGQPWCEAVQVFLNKNL
FLQYNSDNNMVKPLGLLKKVYATSTWGETQTTLGEVGRDLRMLLCDIKPQIKTSDPSTL
QVEMFCQREAEERCTGASWQFATNGEKSLFDAMNMTWTVINHEASKIKETWKKDRGLEKY
FRKLSKGDCHHLREFLGHWEAMPEPTVSPVNASDIHWSSSLPDRWILLGAFILLVLMG
IVLICVWWQNGEWQAGLWPLRTS

>hs_RAET1L Chain=None Position=6:150339470-150346668(-) GeneID=ENSG00000155918
ProteinID=ENSP00000356310 Score=175.8 E-value=1e-50 Length=0 Comment=Not detected in genome
search
MAAAAIPALLLCLPLLLFLLFGWSRARRDDPHSLCYDITVIPKFRPGRWCAVQGVDEKT
FLHYDCGNKTVTPVSPGLKKNVNTMAWKAQNPVLEVVLDILTEQLLDIQLENYTPKEPLT
LQARMSCEQAEGHSSGSWQFSDGQTFLLFDSEKRMWTTVHPGARKMKEKWENDKDVAM
SFHYISMGDCIGWLEDFLMGMDSTLEPSAGAPLAMSSGTTQLRATATTLILCCLLIILPC
FILPGI

>hs_ULBP1 Chain=None Position=6:150285143-150294846(+) GeneID=ENSG00000111981
ProteinID=ENSP00000229708 Score=177.9 E-value=2.5e-51 Length=0 Comment=Not detected in genome
search
MAAAASPFAFLLCLPLLLHLLSGWSRAGWVDTHCLCYDFIITPKSRPEPQWCEVQGLVDERP
FLHYDCVNHAKAFASLGGKKNVTKTWEQETETLRDVVDFLKGQLLDIQVENLPIEPLT
LQARMSCEHEAHGHRGSWQFLFNGQKFLFDSDNNRKTWALHPGAKMTEKWEKNRDMVTM
FFQKISLGDCKMWLEEFMLYWEQMLDPTKPPSLAPGTTQPKAMATTLSPWSLLIIFLCFI
LAGR

>hs_ULBP3 Chain=None Position=6:150384286-150390231(-) GeneID=ENSG00000131019
ProteinID=ENSP00000403562 Score=191.3 E-value=2.7e-55 Length=0 Comment=Not detected in genome
search
MAAAAIPAILPRLAILPYLLFDWSGTGRADAHSLWYNFTIIHLPRHQQWCEVQSQVDQK
NFLSYDCGSDKVLMSHGLEEQLYATDAWGQLEMLREVGRRLRELEADTELEDFTPSGPL
TLQVRMSCECEADGYIRGSWQFSDGKRKFLFDSDNNRKTWVHAGARRMKEKWEKDSGLT
TFFKMVSMRDCSKWLRDFLMHRKKRLEPTAPPTMAPGLAQPKAATTLSPWSFLIILCFI
LPGI

>hs_ULBP2 Chain=None Position=6:150263136-150270371(+) GeneID=ENSG00000131015
ProteinID=ENSP00000356320 Score=182.2 E-value=1.3e-52 Length=0 Comment=Not detected in genome
search
MAAAAATKILLCLPLLLLSGWSRAGRADPHSLCYDITVIPKFRPGRWCAVQGVDEKT
FLHYDCGNKTVTPVSPGLKKNVTTAWKAQNPVLEVVLDILTEQLLDIQLENYTPKEPLT
LQARMSCEQAEGHSSGSWQFSDGQIFLLFDSEKRMWTTVHPGARKMKEKWENDKVVAM
SFHYISMGDCIGWLEDFLMGMDSTLEPSAGAPLAMSSGTTQLRATATTLILCCLLIILPC
FILPGI

>hs_RAET1G Chain=None Position=6:150238014-150244257(-) GeneID=ENSG00000203722
ProteinID=ENSP00000356329 Score=167.9 E-value=2.2e-48 Length=0 Comment=Not detected in genome
search
MAAAASPFAFLLRLLPLLLSSWCRTGLADPHSLCYDITVIPKFRPGRWCAVQGVDEKT
FLHYDCGSKTVTPVSPGLKKNVTTAWKAQNPVLEVVLDILTEQLLDIQLENYTPKEPLT
LQARMSCEQAEGHSSGSWQFSDGQIFLLFDSENRMTTWHVPGARKMKEKWENDKDMTM
SFHYISMGDCIGWLEDFLMGMDSTLEPSAGAPPTMSSGTAQPRATATTLILCCLLIMCLL
ICSRHSLTQSHGHHPQSLQPPHPPLHPTWLLRRVWLWSDSYQIAKRPLSGGHVTRVTL
IIGDSSHSLPCPLALYTINNGAARYSEPLQVSI

>hs_PROCR Chain=None Position=20:33759876-33765165(+) GeneID=ENSG00000101000
ProteinID=ENSP00000216968 Score=112.5 E-value=5.5e-32 Length=0 Comment=Not detected in genome
search
MLTTLLPILLSGWAFCSQDASDGLQRLHMLQISYFRDPYHVYQGNASLGGHLTHVLEG
PDTNTTIIQQLPEPESWARTQSGLSYLLQFHGLVRLVHVERTLAFPLTIRCFGLCEL
PPEGSRAHVFEVAVNGSSFVFRPERALWQADTQVTSQVVTTLQQLNAYNRRTYELRE
FLEDTCVQYVQKHISAENTKGSQTSRSYTSVLVGLVGSFIIAGVAVGIFLCTGRRRC

>mm_H2-Q10 Chain=chain01 Position=17:35470366-35474000(+) GeneID=ENSMUSG00000067235
ProteinID=ENSMUSP00000066419 Score=387.9 E-value=4.8e-115 Length=3635 Comment=Overlaps single
annotation & in proteins
MGAMAPRTLLLLLAAALAPTQTQAGSHSMRYFETSVSRLPGLGEPFIIIVGYVDDTQFVRF
DSDAETPRMEFRAPWMEQEGPEYWERETQRAKAGNEQSFHVSLRLLGYYNQSESGSHTIQ
WMYGCKVGSDFRFRGLYQAYDGRDYIALNEDLKTWTAADVAIIITRRKWEQAGAAEYY
RAYLEAECEVWLLRYLELQKETLLRTPPKTHVTHHPGSEGDVTLRCWALGFYPADITLT
WQLNGEELTQDMLVETRPAGDGTQKQWASVVVPLGKEQNYTCHVYHEGLPEPLTRWEP
PPSTDSIMSHIADLLWPSLKLWYWL

>mm_H2-Q7 Chain=chain02 Position=17:35439442-35443208(+) GeneID=ENSMUSG00000060550
ProteinID=ENSMUSP00000071843 Score=377.1 E-value=6.5e-112 Length=3767 Comment=Overlaps single
annotation & in proteins
MALTMLLLLVAALTLIETRAGQHSIQYFHTAVSRPGLGEPWFISVGYVDDTQFVRFSD
AENPRMEPRARWMEQEGPEYWERETQIAKGHEQSFRRSLRTAQSYNQSKGGSHTLQWMI
GCDMGSDGRLRLRGYLQFAYEGRDYIALNEDLKTWTAADMAAQITRRKWEQAGIAEKDQAY
LEGTCMQSLRRYLQKGKTLRLTDPKKAHVTHHPRSYGAVTLRCWALGFYPADITLTLWQL
NGEELTQDMELVETRPAGDGTFOKWASVVVPLGKEQNYTCHVNHGELPEPLTLRWGWP
PPYTVSNMATIAVVVDLGAVAIIIGAVVAFVMNRR

>mm_H2-K1 Chain=chain03 Position=17:33996617-34000013(-) GeneID=ENSMUSG00000061232
ProteinID=ENSMUSP00000025181 Score=370.7 E-value=5.8e-110 Length=3397 Comment=Overlaps single
annotation & in proteins
MVPCTLLLLLAAALAPTQTRAGPHSLRYFVTAVSRPGLGEPYMEVGYVDDTEFVRFSD
AENPRYEPRARWMEQEGPEYWERETQKAKGNEQSFVRDLRTLLGYNQSKGGSHTIQVIS
GCEVGS DGRLLRGYQYAYDGDYIALNEDLKTWTAADMAALITKHKWEQAGEAERLRAY
LEGTCVWELRRYLKNGNATLLRTDSPKAHVTHHSPEDKVTLLRCWALGFYPADITLTLWQL
NGEELIQDMELVETRPAGDGTFOKWASVVVPLGKEQYYTCHVYHQGLPEPLTLRWEP
TVSNMATVAVLVVLAIAVITGAVVAFVMKRRRNTGGKGGDYALAPGSQTSDLSPDCKV
MVHDPHSLA

>mm_H2-Q1 Chain=chain04 Position=17:35320811-35324047(+) GeneID=ENSMUSG00000079507
ProteinID=ENSMUSP00000072942 Score=366.9 E-value=8.9e-109 Length=3237 Comment=Overlaps single
annotation & in proteins
MALGRLLLLLAAALTLTKTGAGSHSLRYFETS VSRPGFGKPRFISVGYVDDTQFVRFSD
AKNPRYEPRAPWMEQEGPEYWERNTRRVKGSEKRFQESLSTLLSYNQSKGGIHTFQKLS
GCDLGS DGRLLRGYQYAYDGLDYIALNEDLETTAADVAAQETRRKWEQAGAAEKHRTY
LEGKCLMHLHRYLELGKEMLLRTDPKKAHVTHHPRSQGDVTLRCWALGFYPADITLTLWQL
NGEELTQDMELVETRPAGDGTFOKWASVVVPLGKEQNYTCHVYHEGLPEPLTLRWEP
TVSNMVI IAVLVVLAIAVIVIGAVVIGVMVSVFMKRRRNTGGKGGEDCALAPSRDQAQSSD
ISLLDCKA

>mm_H2-Q2 Chain=chain05 Position=17:35342590-35345706(+) GeneID=ENSMUSG00000091705
ProteinID=ENSMUSP00000078138 Score=364.5 E-value=3.6e-108 Length=3117 Comment=Overlaps single
annotation & in proteins
MALRRLLLLLVAALKLITRAGSHSLRYFTTAVSRPGLGEPRIIVGYVDDTQFVRFSD
AENPRMEPRAPWMEQEGPEYWERNTQVSKENEQSFVRSLGTALSYNQSKGGSHTLQWL
GCDLGP DGSLLRGYEQSAYDGRDYIALNEDLITWTAADLAALKTRSKLEQAGLAEKRRAY
LEVVDCLTWLRRYLELGKETLLRTDPKKAHVTHHPRSQGDVTLRCWALGFYPADITLTLWQL
NGEDLTQDMELVETRPAGDGTFOKWASVMVFPGEPRYTCHVEHEGLPEPLTLRWEP
TDSYMVII IAVLVVLAIAVIGAVVIGVMVSVFMKRRRNTGGKVRDYAQPGRDPSQSSDISLLEL

>mm_Gm8909 Chain=chain06 Position=17:36165140-36168190(-) GeneID=ENSMUSG00000073402
ProteinID=ENSMUSP00000133663 Score=362.3 E-value=1.4e-107 Length=3051 Comment=Overlaps single
annotation & in proteins
MAPRTLLLLLAAALILITETRAATEGPGRRSGSHREPPPGPHSMRYFTTAVSRPGLGEPRIIVGYVDDTQFVRFSD
FETAVSRPGLGEPWFISVGYVDDTQFVRFSDGDAENPRMEPRAPWMEQEGPEYWERETQIA
KGHEQGFQGSRLNLLHYNQSAGGSHTFQRMYGCDLGS DGRLLRGYQFAYDGSYIALN
QDLKTWTAADLAAQITRRRWEQGGVAETLRLAYLEDPCLEWLRRLYLELGKETLLRTDPK
HVTHHPRSEDDVTLRCWALGFYPADITLTLWQLNGEELTQDMELVETRPAGDGTFOKWAS
VVPLGKEQNYTCHVYHEGLPEPLTLRWEP RPSRDSNMV I IAVLVVLVGVV I IIGAMVPFVL
KSRRKIGGKEENYALAGGSNSVQGSALFLEAFSHP

>mm_H2-D1 Chain=chain07 Position=17:35263369-35266725(+) GeneID=ENSMUSG00000073411
ProteinID=ENSMUSP00000134570 Score=361.5 E-value=2.6e-107 Length=3357 Comment=Overlaps single
annotation & in proteins
MGAMAPRTLLLLLAAALAPTQTRAGPHSMRYFETAVSRPGLGEPRIIVGYVDDNKEFVRF
DSDAENPRYEPRAPWMEQEGPEYWERETQKAKGQEQWFRVSLRNLGYYNQSAGGSHTLQ
QMSGCDLGS DWRLLRGYLQFAYEGRDYIALNEDLKTWTAADMAAQITRRKWEQSGAAEHY
KAYLEGECEVWHLHRYLKNGNATLLRTDSPKAHVTHHPRSKGEVTLRCWALGFYPADITL
WQLNGEELTQDMELVETRPAGDGTFOKWASVVVPLGKEQNYTCRVYHEGLPEPLTLRWEP
PPSTDSYMV I IAVLVVLAIAVIGAVVAFVMKRRRNTGGKGGDYALAPGSQSSEMSLRDC
KA

>mm_Gm10499 Chain=chain08 Position=17:36142494-36145667(-) GeneID=ENSMUSG00000073403
ProteinID=None Score=361.4 E-value=2.7e-107 Length=3174 Comment=Overlaps single annotation; not
in proteins
PPPGSHSMRYFETALSRLGEPRIIVGYVDDTQFVRFSDGDAENPRYEPRAPWMEHEGR
EYWERETQIAKRNEQSFRRSLRTAQRYNQSEGGGLHTFQLLFGCDVGS DGRLLRGYLQF
AYDGRDYIALNEDLNTWTAADLAAALITRRKWEQAGAAEHYKAYLEGECEVSLRRYLQLGK
ETLLRTAHVTHHPRSEGDVTLRCWALGFYPADITLTLWQLNGEELTQDIELVETRPAGDGT
FOKWAASVVVPLGKEQNYTCHVEHEGLPEPLTLTGKGGDYAPAGGR

>mm_H2-Q4 Chain=chain09 Position=17:35379894-35383607(+) GeneID=ENSMUSG00000035929
ProteinID=ENSMUSP00000080159 Score=360.6 E-value=5e-107 Length=3714 Comment=Overlaps single
annotation & in proteins
MASTMLLLLVAVAQTLEIRAAPRPAPLPARSPAGGPRSSGSHRALPPGPHLLSYFYTS

VSRPGLGEPFRFISVGYVDNTEFVRFDSDAENPRYEPRAPWMEQEGPEYWERETQKAKGNE
QIFRVNLRLLSYNQSAGGSHTIQVISGCEVSGDGRLLRGYQQFAYDGRDYIALNEDLK
TWTAAADMAAQITRRKWEQAGATEKSKAYLEGACVQSLRRYLELGKETLLRTPPKAHVTC
HHRSDGDVTLRCWALGFYPANI LLTWQLNGEELTQDMELVETRPAGDGTQKQWASVVVPL
GKEQNYTCHVHHEGLPEPLTLRWEPPSTVSNMANVAVLVVLGAWPSLQLWVLL

>mm_H2-K2 Chain=chain10 Position=17:33975748-33978533(-) GeneID=ENSMUSG00000067203
ProteinID=None Score=359.2 E-value=1.1e-107 Length=2786 Comment=Overlaps single annotation; not
in proteins

PPPGSHSLRYFHTAMSRPGLGEPFRFIAVGYVDDTQFVRFDSDAENPRMEPRAPWMEQEGP
EYWERETQVAKNTEQISRVLNLRTRGYYNQSEGGGSHYQRMSCGEVSGDGRLLRGYQY
AYDQDYIALNEDLKTWTTADIAAQITGQKLEQVGAERERAYLEGACMESLRRYLELGN
ATLLRTHAVTHHPGSEDDVTLRCWALGFYPADITL TWQLNGEELTQDMELVETRPAGDGI
FQKWA AVVPLGKEQNYTCHVHHEGLPEPLTL

>mm_H2-Q6 Chain=chain11 Position=17:35425098-35428612(+) GeneID=ENSMUSG00000073409
ProteinID=ENSMUSP00000109511 Score=357.6 E-value=3.9e-106 Length=3515 Comment=Overlaps single
annotation & in proteins

MALTTLLLVAAALTLIETRAGPHSLRYFHTAVSWPGLVEPRFIIVGYVDNQFVRFDS
AENPRMEPRARWMEQEGPEYWERETQKAKGHEESFRVSLRTRAQRYNQSKGGSHTLQWY
GCDVGSDELRLLRGYQFAYEGRDYIALNEDLKTWTTAAADMAAQITLHKWEQAGIAERDRAY
LEGACVQSLRRYLRKQETLLCTDPPKAHVTHHPRSYGAVTLRCWALGFYPADITL TWQL
NGEELTQDMELVETRPAGDGTQKQWASVVVPLGKEQNYTCHVHHEGLPEPLTLRWEPPPS
TVSNMANVAI LVVLA WPSLELWVIL

>mm_H2-B1 Chain=chain12 Position=17:36080782-36083975(-) GeneID=ENSMUSG00000073406
ProteinID=ENSMUSP00000134155 Score=348.1 E-value=3.1e-103 Length=3194 Comment=Overlaps single
annotation & in proteins

MAQRTLFLLLAAALTMIE TRAGPHSMRYFETAVFRPGLGEPFRFISVGYVDNTEFVRFDS
AENPRSEPRAPWMEQEGPEYWERETQIAKDNEQSFQWGLRNLHYNQSKGGFHTFQRLS
GCDMGLDGRLLRGYQFAYDGRDYITL NEDLKTWMAADLVALITRRKWEQAGAAELYKFY
LEAECVEWLRRLYLRKQETLLRTPPKAHVTHHPRPAGDVT LRCWALGFYPADITL TWQL
NGEELTQDMELVETRPAGDGT* KWA AVVPLGKEQNYTCHVHHEGLPEPLTLRWEPPPS
TGSNMVNI AVLVLGAVII IEAMVAVLKS SRKIA IILPGPAGTKGSSAS

>mm_H2-T23 Chain=chain13 Position=17:36030829-36032432(-) GeneID=ENSMUSG00000067212
ProteinID=ENSMUSP00000099739 Score=337.4 E-value=2.4e-101 Length=1604 Comment=Overlaps single
annotation & in proteins

MLLFAHLLQLLVSA TVPTQSSPHSLRYFTTAVSRPGLGEPFRFIIVGYVDDTQFVRFDS
ENPRMEPRARWIEQEGPEYWERETWKARDMGRNFRVNLRTLLGYYNQSNDESHTLQWY
CDVGPDRLLRGYQEA YDGRDYISL NEDLRSWTANDIASQISKHKSEAVDEAHQRAYL
QGCPVEWLHRYLRIGNETLQSDPPKAHVTHHPRSEDEVTLRCWALGFYPADITL TWQL
GEEELTQDMELVETRPAGDGT* KWA AVVPLGKEQNYTCHVHHEGLPEPLTLRWEPPST
VSNMVI IAVLVLGAVII LGAVVAVFMKRRRHIGVKGYAHVLSKSFQTS DWPQKA

>mm_Gm1127 Chain=chain14 Position=17:36048831-36058120(-) GeneID=ENSMUSG00000079492
ProteinID=ENSMUSP00000109371 Score=334.5 E-value=2.4e-99 Length=9290 Comment=Overlaps single
annotation & in proteins

MGLTSCPLLLLLLATALAPTQAGKSSHSLRYFATAISRPGFGEPRFTAVGYVDDTQFMR
FDSSENPRAEPCKPWVQMEPEYWEQETRKFKETQNFRTCLYNLLHLYNQSDGPHTI
QDMHGCVVGPDPQLHGHYQHAYDGHDIITL NEDLSSWTAADAVAQITQHKWEEAGVAEE
YKAYLEGTCVETLHRLLEEAQQHSDPPNTHVTRHPRPEGDVTLRCWALGFYPADITL TWQ
LNGEELTQDMELVETRPAGDGT* KWA AVVPLGKEQNYTCHVHHEGLPEPLTLRWEPLP
STDSNMVIIAALVVF GAVII IGAVVAFMMKRRRTGGKGVYCWEP RPHSPLQDGADSCVL
SGKQIICACAKGIL

>mm_Gm7030.1 Chain=chain15 Position=17:36126997-36129174(-) GeneID=ENSMUSG00000092243
ProteinID=ENSMUSP00000133734 Score=334.0 E-value=3.6e-99 Length=2178 Comment=Overlaps single
annotation & in proteins

MGLTSCPLLLLLLATALAPTRAGKSSHSLRYFATAMSRPGLGEPFRFTAVGYVDDTQFMR
FDSSENPRAEPCKPWVQMEPEYWEQETRKFKETQNFRTCLYNLLHLYNQSDGPHTI
QDMHGCVVGPDPQLHGHYQHAYDGHDIITL NEDLSSWTAADAVAQITQHKWEEAGVAEE
YKAYLEGTCVETLHRLLEEAQQHSDPPNTHVTRHPRPEGEVTLRCWALGFYPADITL TWQ
LNGEELTQDMELVETRPAGDGT* KWA AVVVPFGEPRYTCHVYH

>mm_Gm7030.2 Chain=chain18 Position=17:36109867-36111436(-) GeneID=ENSMUSG00000092243
ProteinID=ENSMUSP00000133734 Score=328.5 E-value=1.2e-98 Length=1570 Comment=Overlaps multiple
annotations; one in proteins

PPPGSHSLRYFHTVSRPGLGEPFRFIIVGYVDDTQFVRFDSSENPRMEPRARWIEQEGP
EYWERETRKARDMGRNFRVNLRTLLGYYNQSKDGSHTLQWY GCDVGPDRLLRGYQQA
YDQDYISL NEDLRSWTATNLASHISKCKSEAVDEAHQRAYLQGPCVEWLHTYLQLGSE
TLRSADVTRHPRPEGDVTLRCWALGFYPADITL TWQLNGEELTQDMEFVETRPAGDGT
QKQWASVVVPLGKEQNYTCHVHHEGLPEPLTL

>mm_H2-Q5 Chain=chain16 Position=17:35394347-35398010(+) GeneID=ENSMUSG00000055413
ProteinID=None Score=333.5 E-value=6e-99 Length=3664 Comment=Overlaps single annotation; not in
proteins

PPPGPHSLRYFHTAVSRPGLGEPFRFIIVGYVDDTQFVRFDSDAENPRMEPRARWMEQEGP

EYWERETQVAKGHEQSFQGSRLTAQSYYNQSKGGGSHLTQWMYGCDMGSDGRLLRGYLQF
AYEGRDYIALNEDLKTWTAADMEAQITRRKWEQAGIAERDRAYLETHVTHHPISYDAVTL
RCWALGFYVPDITLTWQNGEELTQDTELVEVTRPAGDGTFOKWAAMVVPFGEEQNYTCHV
HHEGLPEPLTLGRDSS*SSDRSLPDGK

>mm_Mrl Chain=chain17 Position=1:155132358-155137784(-) GeneID=ENSMUSG0000026471
ProteinID=ENSMUSP0000027744 Score=331.9 E-value=1.3e-99 Length=5427 Comment=Overlaps single
annotation & in proteins
MMLLLPLLAVFLVKRSHTRTHSLRYFRLAVSDPGFVVPFISVGYVDSHPITTYDSVTRQ
KEPKAPWMAENLAPDHWERYTQLLRGWQOTFKAELRHLQRHYNHSGLHTYQRMIGCELLE
DGSTTGFLQYAYDQDFIIFNKDTLSWLAMDYVAHITKQAWEANLHELQYQKNWLEEECI
AWLKRFLYGRDTERTEHPVVRTTRKETFPGITTFFCRAHGFYPPEISMWTKNGEEIA
QEVYDYGVLPSGDGTQVWLSVNLDPQSNVDVYSCHVEHCGRMVLEAPRESGDILRVSTI
SGTTILIALAGVGLIWRRSQELKEVYMQPTQVNEGSSPS

>mm_H2-M2 Chain=chain19 Position=17:37481479-37483256(-) GeneID=ENSMUSG00000016283
ProteinID=ENSMUSP00000131297 Score=316.6 E-value=3.4e-95 Length=1778 Comment=Overlaps single
annotation & in proteins
METSAFFTLLFTLYLLGASLALAQTFKGSLSLRYFDIAVSRPGLLEETHYMTVGVVDDTE
FVHFDNEAENPRFEPVWMEQMGQKYWDDQTRIAKAAEQQIRVYFQKLRDYYNQSQNSS
HTIQRMTGCIYIGPDGHLHAYRQFGYDQDYLTNEDLSTWTAADAAAIEITRREWEATNV
AEFWRVYLEGPCMVWLFKYLTVGNETLLRTEPPKAYVTHHPREGEVTLRCWALGFYPSD
IIMIWRDGEDQTDQMDVIETRPAGDGTFOKWAAMVVPVSGKEQNYTCHVAHEGLPEPLTL
RWSRPPQSFIIFIIIVAVGLVLLGASVATLVMWKKSSGGERGSL

>mm_H2-M3 Chain=chain20 Position=17:37270512-37272537(+) GeneID=ENSMUSG00000016206
ProteinID=ENSMUSP0000035687 Score=313.8 E-value=2.3e-94 Length=2026 Comment=Overlaps single
annotation & in proteins
MGSSSNRALLHMVVVSLAVTQTSKGSLSLRYFHTAVSRPGRGEPQYISVGYVDDVQFQRC
DSIEEIPRMEPRAPWMEKERPEYWKELKLVKKNIAQSARANLRTLLRYYNQSEGGSHILQ
WMVSCVEGPDMLLGAHYQAAAYDGSYITLNEEDLSSWTAVDMVVSQITKSRLESAGTAEYF
RAYVEGECLELLHRFLRNGKEILQRADPPKAHVAAHHPKGDVTLRCWALGFYPADITLT
WQKDEEDLTQDMELVETRPAGDGTFOKWAAMVVPVSGEEQRYTCYVHHEGLTEPLALKWGR
SSQSSVIMVIVASLVLLGGVITIVVCKRRRAGER

>mm_H2-T3 Chain=chain21 Position=17:36187258-36189880(-) GeneID=ENSMUSG00000054128
ProteinID=ENSMUSP00000099736 Score=304.5 E-value=1.8e-91 Length=2623 Comment=Overlaps single
annotation & in proteins
MRMGTMPVPGTLLILLAASQGQTQTCPSGSHSLRYFYFALSRAISEPWYIAVGYLDDTQFV
RFNSSGETAYFKLAPWVEQEGPEYWARETEIVTNSAOFFRENLOTMLDYYNLSQNGSHT
IQVYMGCEVEFFGSLFRAYEQHGYDGRDYIALNEDLKTWTAADTAAEITRSKWEQAGYTE
LRRTYLEGPKDLSLLRYLENRKKTQECTDPPKTHVTHHPREGEVTLRCWALRFYPADIT
LTWQNGEELIQDTELVEVTRPAGDGTFOKWAAMVVPVPLGKEQYKTCYVHHEGLPEPLTLRW
EPPQTSMPNRTTVRALLGAMIILGFMSSGVMWRRKNNGGNDNTAAYQNEREHLISLSP
RAESEALGVEAGMKDLPSAPPLVS

>mm_H2-M5 Chain=chain22 Position=17:36987662-36989214(-) GeneID=ENSMUSG00000024459
ProteinID=ENSMUSP00000131365 Score=304.5 E-value=1.1e-91 Length=1553 Comment=Overlaps single
annotation & in proteins
MRSPALSTLLSLLLTGALALTLVRAGIHSLQFFATMTQPGLREHSFIFVVFVDDTQFLC
YNNKGNQRMPEPRALWVQMGPEYWEQOTRTRVVKVIEKIALVNLQEAAMDYHNSKDGSHVF
QCVYGCVEGPDGLFLRGHEKHAYDGRDYLTLSPLDHSWVAGDTAAQITLRRWEKSGVSEQ
RQSFLLKGECEVSLRITYLEIGKETLLRTPPKAHVTHHPREGEVTLRCWALGFYPANILL
TWQWDEEDLTQDMDLIETRPAGDGTFOKWAAMVVPVSGEEQRYTCHVQHEGLTQPLVLKWD
PSKHTIPIMGITVGLLLFGVVFAGVAVIWMRKRKGFQIILKTFLLG

>mm_Azgp1 Chain=chain23 Position=5:137985116-137989821(+) GeneID=ENSMUSG00000037053
ProteinID=ENSMUSP00000038559 Score=284.6 E-value=4.3e-85 Length=4706 Comment=Overlaps single
annotation & in proteins
MVPVLLSLPPLLGPVAFQETGSYYLTFYTLGSLRPSKGFPRFQATAFLNDQAFFHNSNS
GKAEPVGPWSQVEGMEDWEKESQLQRAREEIFLVTLKDIMDYKDTGSHTFQGMFGCEI
TNNRSSGAVWRYAYDGEDFIEFNKEIPAWIPLDPAANTKLEAEKVYVQRAKAYLEEE
CPEMLKRYLNYRSRHLDRIDPPTVTITSRVIPGGNRIKFLAYGFYQRI SLHWNKANK
LAFEPERGVFPNGNGTYLSWAEVEVSPQDIDPFFCLIDHRGFSQSLSVQWDRTRKVKDEN
NVVAQPQ

>mm_H2-M10.1 Chain=chain24 Position=17:36323382-36325847(-) GeneID=ENSMUSG00000024448
ProteinID=ENSMUSP00000025322 Score=271.6 E-value=1.1e-80 Length=2466 Comment=Overlaps single
annotation & in proteins
MRNTGPCTLLLLMVALDLNHYCAGSHWLQTFNIVILEPGMLEPRFIQVSYVDSIQYQGF
DSKDPASAGMQPRAAWIELEPPEYWEKETSRLVLELSQVERQVLRMLVKKNGHKMDGYHTLQ
EYVGCNVANDGSFLGGHYRLTYGYDYLTNEDLSSWTAEGKAQAQAKSRWEDAGEAKRR
KTYLQGECEVQRLRYLDLQKETLLCSADAPQTHVTHHVRPEGNVTLRCWALGFYPADITMT
WKRDNNTQDMLPDRPAGDGTFOKWAAMVVPVSGEELRYTCHVHHEGLPEPLTLKWEK
PHTIPIAILIGLVGLTVVGTVVIFLVWRK

>mm_Hfe Chain=chain25 Position=13:23705868-23708301(-) GeneID=ENSMUSG0000006611
ProteinID=ENSMUSP00000089298 Score=270.5 E-value=3.5e-81 Length=2434 Comment=Overlaps single
annotation & in proteins
MLSL SAGLPV RPL LLLLLLWSVAPQALPPRSHSLRYLFMGASEPDLGLPLFEARGYVDDQ
LFVSYNHESRRAEPRAWILEQTSSQLWLHLSQSLKGDWYMFIVDFWTIMGNYNHKSQVTK
LGVVSESHILQVVLGCEVHEDNSTSGFWRVYGDQDHLFCPKTLNWSAAEPGAWATKVE
WDEHKIRAKQNRDYLEKDCPEQLKRLELGRGVLGQVPTLVKVTTRHWASTGTSLRCOAL
DFFPQNI TMRWLKDNQPLDAKDVNPEKVL PNGDETYQGWLTAVAPGDETRFTCQVEHPG
LDQPLTASWEPLQSAMIIGIISGVTVCAIFLVGILFLILRKRKASGGTMGGYVLTDC

>mm_H2-T24 Chain=chain26 Position=17:36014807-36017547(-) GeneID=ENSMUSG00000053835
ProteinID=ENSMUSP00000109389 Score=270.2 E-value=1.6e-81 Length=2741 Comment=Overlaps single
annotation & in proteins
MWALIFWLLSHPQDGGARSHSLHYCYSAVTEPGPGVPSFFASGFLDNQPFIIHYDSRSMKA
EPCADWLRENAQYFTHETEVEFTNRMKIFQLSLRNIRQYYNSSGTSQSRADGFRQQAGPHT
LQFTYGCERYNRRTTGHWQYGYDGSYLTLDLGSQYIAATFIAGYTKRKNWENNEYWLEK
EKTYLEKECILWLQRYLTMGGKNFTRTDPPKTTVTHQFKPKENVTLRCWALGFYPADITL
TWQLNGEELTQDTELVETRPSGDGTFQKWA AVVVP SGEEQRYTCHVQHEALTPQLVWKWE
PLQLTPTTGVYARGSCSPQATLLSVLAFPLFGIVLVFGLTRYKTQVRRKNWPAPSVPEE
ERL

>mm_H2-M10.3 Chain=chain27 Position=17:36365526-36368114(-) GeneID=ENSMUSG00000058124
ProteinID=ENSMUSP00000073236 Score=262.6 E-value=5.3e-78 Length=2589 Comment=Overlaps single
annotation & in proteins
MRNPGCCTLLLLVAMDLNQY CAGSHWLQTFNIAFLEPGMINSRIFHIGYVDSIQYQGF
SKEPMAILKPRAAWMEQELPKYWNVETTKILLLSQIERRILYFMIEKYEHRMNDYHTLQE
VYGCNVATDGRFLHGHFRLLTYGYDYLTLNEDLSSWTAEGKAEYMKNRWENMSEAERWK
TYLRGECVQRLLRYLDLKGKETLLRS DAPRTHVTHHVRPEGNVTLRCWALGFYPADITLW
KRDGNLTDQMELPDTRPAGDGT FQKWA AVVVP SGEELRYTCHVHHEGLPEPLTLKWEPP
QTIPIIAILIGLVGLTTLVVGTVVIFLVWKK

>mm_H2-M10.6 Chain=chain28 Position=17:36812476-36814954(+) GeneID=ENSMUSG00000037130
ProteinID=ENSMUSP00000039908 Score=258.4 E-value=9.1e-77 Length=2479 Comment=Overlaps single
annotation & in proteins
MRKPGPWL LLLLLLMAIDLTYCAGSHWLQIFNIVILEPGMLEARFIQVGYVDSIQFQGF
SKEPNARMQPRAAWMEQEPPEYWEKETAKALSFSLSGRRILKYMMSANKDNNGYHTLQE
VYGCNVANDGSFLEGHYRLTYGYDYLTLNEDLSSWTAEGEVEKFKTYQEQGMTESWR
TYLLGECVERLLRCLDLGKKTLLHSDAPRTHVTHHVRPEGNVTLRCWALGFYPADITMTW
KRDGNHTQDMELPDTRPAGDGT FQKWA AVVVP SGEELRYTCHVNHEGLPGPLTLKWEPP
HTIPIIAILIGLVGLALVVGTVVIFLVWKK

>mm_H2-M1 Chain=chain29 Position=17:36670257-36671903(-) GeneID=ENSMUSG00000037334
ProteinID=ENSMUSP00000040435 Score=245.7 E-value=3e-74 Length=1647 Comment=Overlaps single
annotation & in proteins
MKNFESQTL LLLLLLMTLTAITKHPNGSHTLRYVYVYLLSWPGLPQLIFLGYVDDTQIMGF
NSISENLGVE SRAPWMYTEEFWEKTTDNV VREHYILKEIMRSVLHIYNYSIIGYHTIQK
TYGCQVMHRRYFSHGFFKLAFLNLHDYITL NEDLKTWRGVGKAGEMLKEMWEKIKYANQVK
SFLQITCVNLLHRRFLAFGKKSLLRDTDPKIHMTHKIRPDRKTLRCWAFNFPPEITLW
QRDGSNQTQDMEMIETRPSGDGT FQKWA AVVVP SGEHRYTCHVNHEGLSEPIITRWTKH
EPPEPTIPFLAIVIALVLGALLMGAVMTFLIWKRRTRGKKSWS

>mm_H2-M10.2 Chain=chain30 Position=17:36283644-36286125(-) GeneID=ENSMUSG00000023083
ProteinID=ENSMUSP00000023845 Score=242.2 E-value=5.1e-72 Length=2482 Comment=Overlaps single
annotation & in proteins
MNPGPCN FLLLLLVALDLTYCAGSHWLQTFNTVILKHGTLEPRIIQVGYVDSIQYQGFDS
KEPTARMQPRAAWMEQEPPEYWEKETAQVLRSLTNERLLRYMMIYNEQSNEEYHTLQEV
FGCNVDNDGSFLGGYRLTYGYEFINL NEDLSSWTAQGEAAGFLKTDLVNVGAAEGWRT
YLLGECTERLLRCLDLGKETLLRS DAPRTHVTHHVRPEGNVTLRCWALGFYPADITLWTK
RDGKNYTQDMELPDTRPAGDGT FQKWA AVVVP SGEELRYTCHVHHEGLPEPLTLKWEPPQ
TIPIIAILIGLVGLTFLVGTVVVFLVWKK

>mm_H2-M11 Chain=chain31 Position=17:36547379-36549004(+) GeneID=ENSMUSG00000037537
ProteinID=ENSMUSP00000042522 Score=238.5 E-value=5.3e-72 Length=1626 Comment=Overlaps single
annotation & in proteins
MKTFTVTEALFLLLQVLLAMTSHPDGTHFFGFFQTLFTLFPWMPKQFISVGFVDDIQFERF
NSRRDQRTEHCAPWKDQKPEYWKDNTDLVLSYFQDLTEILQRMLKIYNYSLTGYHTIQ
RRYGYLLPRGYFRNGFFEVVFNHDHSIRLNEDLSTWTVPVGFAEILREEDWSSGFTQNV
KNFLEVECVDLFLTELEYGKEILLRDTIPKIHVIRKVRPDKKITLRCWALKFYPAEITLT
WERDKSNQTL DMEVTETMPTGDGT FQKWA AVVVP SGEHRYKCHVNHEGLPEPITLRWVP
PEPTISFMHIVIVVVLGALLMGAMTLLIWKRRTR

>mm_H2-M10.5 Chain=chain32 Position=17:36773208-36775712(+) GeneID=ENSMUSG00000037246
ProteinID=ENSMUSP00000047766 Score=236.1 E-value=3.5e-70 Length=2505 Comment=Overlaps single
annotation & in proteins
MRNPGCCTLLLLVAMDLIQY CAGSHWLQTFNAVILEPGMVNSRIFVHIGYVDSIEYQRYD
SKEPIAVLLPRVAWMEQVPMYTWSETASVAELS QLDRRILHFMVKNQEQRMDDYHTLQE
VYGCNVANDGSFLGGHFRLLTYGYDDLYLNLNLS SWIAHGNAEYVKS RWDGEGDAERWK
TYLQGVVERLLRYMVLGKEALLRS DAPRTHVTHHVRPEGNVTLRCWALGFYPADITMTW

KRDGNNHTQDMELPDTRPAGDGTFOKWAAVVPSGEELRYTCHVHHKGLPEPLKWKWEPP
HTPIIIAILLIGLVGLTLLVVGTVVIFLVWKK

>mm_H2-M10.4 Chain=chain33 Position=17:36459530-36462025(-) GeneID=ENSMUSG0000048231
ProteinID=ENSMUSP00000130832 Score=227.6 E-value=1.2e-67 Length=2496 Comment=Overlaps single
annotation & in proteins
MRNPGHFACILCLLVITDRPQYCEGSHWLKTFRIVIMEPGILEPRFIQVSYVDSIQYQGF
DSRSETAGMQPRAAWMKQEPPEYWKNETEHAMGASLLARRTLIYMVTENNKKNDYHTLQ
EVFGCNVAHDGSLFGGHYGLTYGYDYIILNEDLNSWTTEGKVGKFNPDRTQGSVTEGW
RTYLKGECTERFLRCLDLGKETLLRSDAPRTHVTHKVTPEGNVTLRCWALGFYPADITLT
WKRDGNNHTQDMELPDTRPAGDGTFOKWAAVVPSGEELRYTCHVHHEGLPGPLTLKWP
PQTPIIIAILLIGLVGLTLLVVGTVVIFLVWRK

>mm_H2-M9 Chain=chain34 Position=17:36640694-36642350(-) GeneID=ENSMUSG0000067201
ProteinID=ENSMUSP0000084411 Score=224.4 E-value=6.9e-68 Length=1657 Comment=Overlaps single
annotation & in proteins
MKNFESQPLLLFFMVTLVIAKHPNGSHTLRFVSTFLSWPRHLEQIFLIYVDETQIMGF
NSISESQRMESRVPWLNELNAEFWELATQDVLKEKSFVTGIMNKLHLYNDSMTGYHIIQ
ETYGCQVKQRTYFSAFMELLDTHDYITLNEQLQWRAVKGAAEIVKEEWKINLVKSS
KSFLLGACVEGLLQYLNFGKYYLLRDTTPKIHMTYKIRPDRKITLRCWAFNFYPPPEITLT
WQRDGSNQTQDMEVIEIRPSGDGTFOKWAAVVPSGEEHRYTCHVNHEGLSEPVTLRWDP
PEPTIPFLPMIIALVLGALLMGSMVTFLIWKRRTTRGKKGWS

>mm_Mill2 Chain=chain35 Position=7:18855930-18858418(+) GeneID=ENSMUSG0000040987
ProteinID=ENSMUSP0000072223 Score=223.6 E-value=4.8e-67 Length=2489 Comment=Overlaps single
annotation & in proteins
MKASSGKPREFRPAVLLLILGLLLRDSRGSSIQGLADVEVHGSSRLTRHTLRYNVRAH
SLEGEKTLQVLVIYVDEELFLKYNDSRETEPLGCWIKGHGNETCARETNLLKVEEK
LRGMAEVIQKKSQEEGLHTLQATLGCCELLSNGSTRGFWHLYGDGQNFLLTFDQKTLTWT
DGPSTQNKMFWKTHAPRADLVKTFLLDDICPAHLQRYLASLRNGLQDTPGPMVTVTCRNY
PVGRVTLTLCRAFNLTYTREATLVWLQDGKPVQKTFRSETILPSGDGTQYQARVSIKRVLPQ
EPQFSCNLRHGNHSIMQTAVSGHAAEDSQDVASSATASAGSALPVVLAVALARAN

>mm_Fcgrt Chain=chain36 Position=7:45095167-45102678(-) GeneID=ENSMUSG0000003420
ProteinID=ENSMUSP0000003512 Score=210.4 E-value=3.4e-63 Length=7512 Comment=Overlaps single
annotation & in proteins
MGMLPWALSLLLVLPLQTWGSETRPPLMYHLTAVSNPSTGLPSFWATGWLGPQQYLYTN
SLRQEADPCGAMWENQVSWYWEKETTDLKSKEQLFLEALKTEKILNGTYTLQGLLGC
LASDNSSVPTAVFALNGEEFMKFNPRIGNWTGEWPETEIVANLWMPQDAARKESEFLN
SCPERLLGHLERGRNLEWKEPPSMRLKARPGNSGSSVLTCAAFSFPPELKFRLRNGL
ASGSGNCSTGPNGDGSHAWSLLEVKRGDEHHYQCQVEHEGLAQLPTVDLDSARSSVPV
VGIVLGLLLVVVAIAGGVLLWGRMRSGLPAPWLSLSDGSDLLPGNLPEAEPOGANA
FPATS

>mm_Mill1 Chain=chain37 Position=7:18262429-18264862(+) GeneID=ENSMUSG0000054005
ProteinID=ENSMUSP0000069083 Score=201.7 E-value=1.7e-60 Length=2434 Comment=Overlaps single
annotation & in proteins
MLLSRNLRALAAIHLWIVYLLLEDDLGTCAEGDNQRLVASAPYQDIEITLTKPRVQAVAE
PHTLRYDLMALESLEVPGLPQFLTLRYFDDEPFLPYKKNSSITDSQEPRIKDHLEAETWGR
ETDDLQEEEEQLKMLAEITAQNGQNTDLHILQATFGCELQRNGSTRGFWKLYDQGNFL
TFDQKTLTWTVDGPNSTQKNTFWKTRAPRADLVKTFLLDDICPAQLQRYLASLRNGLLNTG
FPKIVITFRNYPVGRITLTCRAFRLYTRVATLTLWLYRKPVQKTFGSETILPSGDGTQY
AWVSIKRVLPQESQFSCNLKHGNHNINEPAAEAPVYGARREQPPTSGVGRVQKSLWSA
MTTALVVISWTLSQLMGLLWFCSSGFCFLQCV

>mm_Gm4271 Chain=chain38 Position=17:36732992-36735336(+) GeneID=ENSMUSG0000092521
ProteinID=None Score=183.3 E-value=2.9e-54 Length=2345 Comment=Overlaps single annotation; not
in proteins
GPGMLEVIQVSYVDSIQFQGVNSKEPTARIQLWAA*MEQEPPEYWEKHKERLNGSGAE
YHTLQGMYGCSVANDGSFLLGHY*FTYCYDYIILNETLSSWNAQDKVAQMLKSSWETED
MTESWKTYLQGECEVERLLRCLDIGKETLLHSTHVTHHVRPEGNVTLRCWALGFYLDITM
TWKRDGNNHTQDMELPDTRPAGDGTFOKGAAVVPSGEELRYTCHVHHEAFHSEMG*GVG
ATAS*QVLCYTAIGMDSAQRSDASLPV*

>mm_Cd1d1 Chain=chain39 Position=3:86997345-86998906(-) GeneID=ENSMUSG0000028076
ProteinID=ENSMUSP00000029717 Score=171.9 E-value=4.9e-52 Length=1562 Comment=Overlaps single
annotation & in proteins
MRYLPWLLWAFLOVWQSEAAQKNYFRCLQMSSFANRSWRTDSVVWGLDLQTHRWSN
DSATISFTKPKWSQKLSNQWEKLOHMFQVYRVSTTRDIQELVKMMSPKEDIYPIEIQLSA
GCEMYPGNASEFLHVAFQKGYVVRFWGTSWQTVPGAPSWLDLPIKVLNADQGSATVQM
LLNDTCPLFVRGELLEAGKSDLEKQEKPVAVLSSVSSAHGHRQLVCHVSGFYKPKVVMW
MRGDQEQQTGRGDFLPADETWYLAQTLQDVEAGEEAGLACRVKHSLSGGQDIIYWDAR
QAPVGLIVFIVLIMLVVAVVYIWRRRSAYQDIR

>mm_Znr1 Chain=chain40 Position=17:36950232-36971123(-) GeneID=ENSMUSG0000036315
ProteinID=None Score=170.9 E-value=3.3e-50 Length=20892 Comment=Overlaps multiple annotations;
none in proteins: (Znr1, ENSMUSG0000036315);(H2-M6-ps, ENSMUSG0000083534);(Ppp1r11,
ENSMUSG0000036398);(ZNRD1-AS1_2, ENSMUSG0000099318)

LSAGSHSLRYLVTTTVSGPGLRDSHFVIVGYMNYMQFMRFDSDGTTQRIQARGPWVKQMG
PVYLEMERRKMESYSHRARENLSHTFQCLISCNMGPDRRLLRGHRHAFDGSYYIILNLD
RKTWTEADSTAQITPQRQWKAEGVAEAFSHFLKVHVSHHPRPEGEVTLRCWALDCFPADVM
LTWQRDEEDLTQVMDLVETRPAGDGTQKWAALVVPSSGRSRRKGRK*HSYINATKGD

>mm_Gm8835 Chain=chain41 Position=17:36202377-36206070(-) GeneID=ENSMUSG00000092457
ProteinID=None Score=167.9 E-value=6.2e-49 Length=3694 Comment=Overlaps single annotation; not
in proteins
PPPGSHWMRYFSTAMSRPGHWKPLYMEAGYVDDTQFVRFSDAENPRYEPRAPWMKLGQP
EYWELQTKVAEGHEKSFQVHLRRTLGGYNNQSKGGAYEGRDYIALNDDLKT*LAADRAALI
TPNK*EQADEAEHYKAYQESTCME*LQAKVTHHARPEGVTLRCWALGFYPADITLTWKR
QEEDLTQADRHSSQTSDDLSELP*K

>mm_Gm4256 Chain=chain42 Position=17:36407623-36409118(-) GeneID=ENSMUSG00000092508
ProteinID=None Score=160.1 E-value=1.5e-48 Length=1496 Comment=Overlaps single annotation; not
in proteins
TARMQLWAA*MEQEPPEYWEKEKERLNVSGAGYHTLQGMYGCVANDGSPFLGHNRFYTC
GYDYTILNENLRWNAQDKVAQTLRDSWEAEGITEIWRTYLLGECTERLLRCLDLGKETL
LRSTHVTHHVRPEENVTLRCWALGFYPADSTLTWKRNGKNHTQDMELPDTRPAGDGTQFK
WAAVLVPSGEEELRYTGHVYHERLPEPLTL

>mm_Gm8868 Chain=chain43 Position=17:36558570-36561661(+) GeneID=ENSMUSG00000092332
ProteinID=None Score=147.7 E-value=5.1e-45 Length=3092 Comment=Overlaps single annotation; not
in proteins
LYLVAHFLGFYQTLFT*PGLPEPQFISVGFVDDTGYHTIQKRSGCYVLLQGDGFSHGFFFI
AINDHDYIRLNEDLRTWTTVSKFAEMHKKLWSSSSFKQIVETYLGRYVDLLFTELEYGK
AFLHWSNACET*VRPDKKITLRC*AISFYSAEITLTWETDESNTSDMEVSETMPAADGT
FQKWAAVVPSGEDHRYTCHVNHEGLPEIITL

>mm_Cdld2 Chain=chain44 Position=3:86986987-86988548(+) GeneID=ENSMUSG00000041750
ProteinID=ENSMUSP00000039583 Score=138.1 E-value=1.2e-41 Length=1562 Comment=Overlaps single
annotation & in proteins
MRYLPCLLLWAFQVWQSEVQKQNYTFRCLQTSFANISWRTDSLILLGDLQTHRWSN
DSAIISFTKPKWSQGLSNQWQEKLOHMFQVYRVSFTRDIQELVKMMSPKEDYPIEIQLS
GCEMYPGNASESFFHVAFQKQYAVRFRGTSWQVRLGAPSWLDELPIKVLNADQGTSAVQT
LLNDTWPQFARGLLEAGKSDLEKQEKPVAWLSSVPSSAHGHLQVCHVSGYKPKVWVMM
RGDQEQQGTTHRGDFLPNADETWYLQATLDVEAGEEAGLACRVKHSLLGGQDIIILYWDARQ
APVGLIVFIVLIMLVVVGAVVYIWRSSAYQDIR

>mm_Gm8810 Chain=chain45 Position=17:36065297-36067779(-) GeneID=ENSMUSG00000091373
ProteinID=None Score=112.3 E-value=1.8e-33 Length=2483 Comment=Overlaps single annotation; not
in proteins
VQ*FFYSSRYFYSAVSRPGIGEPFRFRVVSMDSSLDSTPQWLQGCQDVEPNRLLHWSNQ
HAYDGLDLSTWKKDLSSWTAATHTGPQIPHHKIEDYAHVTRHPRPAGDVTLRCAVWGFYP
ADITLTWQLNGEDLTQDMELVETRPAGDGTQKWAAVLPSGKEQKFTCHLQHEGLYQPL
TLAQSSDLNLEA*K

>mm_Gm8877 Chain=chain46 Position=17:36701636-36703209(-) GeneID=ENSMUSG00000092298
ProteinID=None Score=95.4 E-value=9.7e-29 Length=1574 Comment=Overlaps single annotation; not
in proteins
FYQALFTWPGLLLEAQFISVVDTTQFERFNREDV**MEHCAPWKDKKPKNKWDRTYAVL
DCFQQLRVLMNRRLKGYHTTQRRYDCYVLFQGDGFSHKFFEVAFNDHDFIRLNKDLMTWTP
VGKFAEMCKKLWSSGVPGLQTYLLGKRDESNQTSHEVSETMPAGDGSF*KWADVVPV
SGEHRYSCHVNHKGLPELITL

>mm_2410137M14Rik Chain=chain47 Position=17:36978664-36981332(-) GeneID=ENSMUSG00000064308
ProteinID=None Score=68.5 E-value=1.7e-21 Length=2669 Comment=Overlaps single annotation; not
in proteins
GYMDNIQITSFSKDLSEQWVEPWVFFIAQEPHELEELMRLARRILILGQIELWTLFGYHS
QRENGKWLKYLKGEEMLQAHATNVAHSTLEGDVTLRCAWALDFHPAVIILTWRDEKDL
TQDMDLVETRPAAEAFQKWAAVLPSGEEHKYTCVHQPEGLSEPLIL

>mm_Gm20392 Chain=chain48 Position=17:36210909-36216059(-) GeneID=ENSMUSG00000092362
ProteinID=None Score=62.7 E-value=1.8e-19 Length=5151 Comment=Overlaps multiple annotations;
none in proteins: (Gm20392, ENSMUSG00000092362);(Gm5682, ENSMUSG00000092265)
WRMEPLTPWVEQEGLEYWEQTEWQE*GAEYQE*VDLRSLLHYNNQSKGGVWLVWRYLEK
KEMLQRTQRHHTHLRPEGDVTLRSWALGFYPADITVTTWQLNGKELTHDMELVETRPAGD
TFQKWAAVVPSGEEQNIYLACVP*RAV*DGVRVVV

>mm_chain49 Chain=chain49 Position=7:18041285-18043860(-) GeneID=None ProteinID=None Score=11.1
E-value=0.0022 Length=2576 Comment=No overlapping annotations
LPQDPDLLHYELISLY*EGSWGGFLVLEHLDDFFLCYDGEISILGPRIKEHVAGTWRK
EIEKTNLTWKMSVPLACSTNVLWERHHRFDQIKTLLYNIYPDYLIQYFASMVSVIQHSS
TVGCHNLKYSVFNIYLFNLIPDLLWDGEPQKPTFWPRYPLSSGD*SFQTMWSTVILAE
EPRFTCLIRLIRTPWSLQTLMR

>mm_Raet1d Chain=None Position=10:22360552-22374139(+) GeneID=ENSMUSG00000078452
ProteinID=ENSMUSP00000093471 Score=54.3 E-value=4.7e-15 Length=0 Comment=Not detected in genome search
MAKAAVTKRHHFMIQKLLILLSYGYTNGLDDAHSRLRCNLTIKAPTPEVVKCFVGEILL
HLSNINKTMTSGDPGETANATEVGECLTQPLKDLCKQLRDKVSNTKVDTHKNGYPHLLQV
TMIYPSQGGQTPSATWEFNISDSYFFTFYTEIMSWRSANDESGVIMNKWKDDGFEVVKQLK
FLIHGCSQKMFELKQSKKPRSTSRSPSITQLTSTSPLPPTSHSTSKKGFISVGLIFIS
LLFAFAFAM

>mm_H2-T10 Chain=None Position=17:36115876-36121465(-) GeneID=ENSMUSG00000079491
ProteinID=ENSMUSP00000127602 Score=239.9 E-value=1.6e-70 Length=0 Comment=Not detected in genome search
MSWVLRAAVVCALLQLDARPSWTRIPLGSHSLRYFYTAVSRPGLGEPWFIIIVGYVDDMQ
VLRFSKKEETPRMAPWLEQEEADDWEQOTRIVTIQQLSERNLMTLVHFYNKSMDDSHTL
QWLQGCDEVEDRHLCLWYNQLAYDSEDLPTLSENPSSCTVGNSTVQPISQHLEGHCSDLL
QKYLEKGERLLRSDPPKAHVTRHPRPEGDVTLRCWALGFYPADITLTWQKDGEEELTQEV
EFVETRPAGDGTQKWAAVVVPLGKVQSYTCHVDHEGLPEPLTLRWEPAWYQKPIWIVA
TVFSILLICLCVARRPMKKNAGGRGRRDTQEAGRDSPODSSKTVVDEEMGVCFWKIKYC
KTCLGHSPGTSVGESELLSP

>mm_H2-T22 Chain=None Position=17:36037128-36042747(-) GeneID=ENSMUSG00000056116
ProteinID=ENSMUSP00000077111 Score=241.2 E-value=7.8e-71 Length=0 Comment=Not detected in genome search
MSWVLRAAVVCALLQLDARPSWTRIPLGSHSLRYFYTAVSRPGLGEPWFIIIVGYVDDMQ
VLRFSKKEETPRMAPWLEQEEADNWEQOTRIVTIQQLSERNLMTLVHFYNKSMDDSHTL
QWLQGCDEVEDRHLCLWYNQLAYDSEDLPTLSENPSSCTVGNSTVPHISQDLKSHCSDDL
QKYLEKGERLLRSDPPKAHVTRHPRPEGDVTLRCWALGFYPADITLTWQKDGEEELTQDM
ELVETRPAGDGTQKWAAVVVPLGKVQSYTCHVYHEGLPEPLILRWEPAWYQKPIWIVA
MVFILFIICLCVVCICMKKNAGGRGRRDTQEAGRDSPODSSKTVVDEEMGVCFWKIKYC
KTCLGHSPGTSVGESELLSP

>mm_Ulbp1 Chain=None Position=10:7444873-7473477(-) GeneID=ENSMUSG00000079685
ProteinID=ENSMUSP00000136637 Score=72.7 E-value=1.7e-20 Length=0 Comment=Not detected in genome search
MELTASNKVLSCCLLSLLCLSVCLCPRIEETASLNCNIYKVNRSSESGQHSHEVQGLLRNP
LFVYKDKKCHAIGAHNRSMNATKICEKEVDTLKDGIDIFKGLLLHIVQETNTTGKPLTLQ
AEVCGQYEVDKHFTGYAIVSLNGKNI FRVDTSTGNWTQLDHEFEKFIEMCKEDKVLAAFL
KKTTEGDCRTWLDELMLHWHKEHLEPAGSFTLMIILCVIAIAFLGLLFGVSKCLRHLRTRK
KIGLQSSPPPLDSDSLTVPSTSPQSSVCGTMIQCLCPRKLKSPVFMQIDLQSSAPPLDSDS
LTVPETCSVKKEDEFPTASQNSVLLTSDIDGIP

>mm_Raetle Chain=None Position=10:22158569-22374139(+) GeneID=ENSMUSG00000053219
ProteinID=ENSMUSP00000066627 Score=60.3 E-value=8e-17 Length=0 Comment=Not detected in genome search
MAKAAVTKRHHFMIQKLLILLSYGYTNGLDDAHSRLRCNLTIKDPSTADLPWCVDKCSVDE
ITLHLNINKTMTSGDPGKMANATGKCLTQPLNDLCQELRDKVSNTKVDTHKNGYPHL
QVTMIYPSQGGQTPSATWEFNISDSYFFTFYTENMSWRSANDESGVIMNKWKDDGDLVQQ
LKYFIPQCRQKIDEFKQSKKPRSTSRSPSITQLTSTSPLPSPSHSTSKKGFISVGLIF
ISLLFAFAFAM

>mm_Procr Chain=None Position=2:15575117-155755471(+) GeneID=ENSMUSG00000027611
ProteinID=ENSMUSP00000029140 Score=84.7 E-value=4.9e-24 Length=0 Comment=Not detected in genome search
MLTKFLPLLLLLLPGCALCNSDGSQSLHMLQISYFQDNHHRHQGNASLGLLTHLEGP
SQNVITLQLPWQDPESWERTESGLQIYLTQFESLVKLVYRERKENVFFPLTVSCSLGCE
LPEEEEEGSEPHVFFDVAVNGSAFVSRPKTAVVWVSGSQEPSKAANFTLKLQNLNAYNRTRY
ELQEFLODTCVEFLENHITQNMKGSQTGRSYTSLVLGILMGCFIIAGVAVGIFMCTSGR
RC

>cf_DLA-12 Chain=chain01 Position=12:933548-936109(+) GeneID=ENSCAFG00000032222
ProteinID=ENSCAFP00000031424 Score=383.0 E-value=2.1e-110 Length=2562 Comment=Overlaps single annotation & in proteins
VWAMVPGTLALLLSGALAVTLTRAGSHSLRYFYTSVSRPGRGDRPFIAGVYVDDTQFVRF
DSDAATGRMEPRARWMEQEGPEYWDQTRTIKETARTFRVDLDTLRGYYNQSEAGSHTFQ
WMFGCDLPGGRLLRGYSQDAYDGADYIALNEDLRSWTAADAAALITRRKREAAGDAGHL
RNYLETTCVEWLRRYLEMKGKETLLRAEPPSTRVTRHPVSDHEVTLCWALGFYPAEITLT
WQRDGEDQTDTEVVDTTRPAGDGTQKWAAVVVPSGQEQRYTCHVQHEGLAEPVTRRWE
SPLSTIVIVSIAALVLLVAVGIVGAVIWRKQSRGGKPGYSHAARDDSAQGSVSLTAPR

>cf_DLA88.1 Chain=chain02 Position=12:892933-895431(-) GeneID=ENSCAFG00000000487
ProteinID=ENSCAFP00000000718 Score=376.8 E-value=1.6e-108 Length=2499 Comment=Overlaps single annotation & in proteins
MEVMPRALLVLLSAAALAVTLTRAGSHSLRYFYTSVSRPGRGDRPFIAGVYVDDTQFVRF
DSDAATGRMEPRARWMEQEGPEYWDGETRKYKETAQLYRVDLDTLRGYYNQSEAGSHTI
QTMYGCDLPGGRLLRGYSQDAYDGADYIALNEDLRSWTAADAAALITRRKREAAGDAGH
LRNYLETTCVEWLRRYLEMKGKETLLRAEPPSTRVTRHPVSDREVTLRCWALGFYPPEITL
TWQRDGEDQTDTEVVDTTRPAGDGTQKWAAVVVPSGQEQRYTCHVQHEGLAEPVTRRWE
PSPLSTIVIVSIAALVLLVAVGIVGAVIWRKQSRGGKPGYSHAARDDSAQGSVSLTAPR

RGETPETLKWGGLGHWEPAAGSSLCSVSLWKL

>cf_DLA88.2 Chain=chain05 Position=12:1012553-1021445(-) GeneID=ENSCAFG00000000492
ProteinID=ENSCAFP00000000727 Score=316.3 E-value=7.4e-90 Length=8893 Comment=Overlaps single
annotation & in proteins
MRVVMRGTPPPPPLPSGALAVTLTRAGSHSLRFFHTAVSRPGRGDPLYISVGYVDDTQF
VRFDSDAATGRMEPRAPWVEQEGPEYWDGETRNLKETAQVYRVLDLTLRGYYNQSEAAGS
HTIQTMYGCDLPGGRLLRGYSQDAYDGADYIALNEDLRSWTAADAAALITRRKREAAGD
AGHLRNYLETTCVEWLRRLRYLEMGKETLLRAGSPKGTLTHHPFSDCEVTTLTCQDGEDQTQD
TELVETRPAGDGTFOKWAADVVPSPGQEQRYTCHVQHEGLAEPVTRRWESPRLTPIVIGI
VAGLLVFPVSGTMVAGAVLWRK

>cf_DLA-64 Chain=chain03 Position=12:983833-986467(+) GeneID=ENSCAFG00000000500
ProteinID=ENSCAFP000000031410 Score=370.2 E-value=1.5e-106 Length=2635 Comment=Overlaps single
annotation & in proteins
MEVVMRPRALLVLLSAALAVTLTRAGSHSLRFFHTAVSRPGRGDPLYISVGYVDDTQFLRF
NSDAASPKVEPRARWMEQEGPEFWEEQTEIAKVHAQTSRSNLQTALGYYNQSEAGSHTFQ
WTSgcdvpgdgrllrgyEQFAYDGADYIALDEDLRSWTAADAAAQITRRKWEAAGAAQYY
RVYLQGECEVQSLKYLKLERGKETLQRTDPPKIYLRHPISDHEVTLRWCWALGFYPAEITLT
WQRDGEDQTQDTEVVDTTRPAGDGTFOKWAADVVPSPGQEQRYTCHVQHEGLAEPVTRRWEP
SPLSTIVIVSIAALVLLVAVGIVGAVIWRKQSRGKGPYSHAARDDSAQGSVDVSLTAPR
V

>cf_DLA-79 Chain=chain04 Position=18:41142781-41145286(+) GeneID=ENSCAFG00000008234
ProteinID=ENSCAFP00000012096 Score=319.9 E-value=1e-91 Length=2506 Comment=Overlaps single
annotation & in proteins
MPVMPRTLLLLLLLLLSSGTLGLAQTSAGSHSLRFLNTVVSRRPGHGEPRHVAVYVDDT
PFRFNSEREGRRPEPLVRWLEQEGPEYWEERTLNSRTCTQVLRRLNEVSQDYSNQSRST
GSHTFQITGCDVGPDRFLRGYQRHAYDGLDYITLSEDLRSWIVEDPVAQITRRKWDAA
TVAENRKNFLEGRCLWLRRLHENGRETQORADPPKTSVTRRPISEHEVTLCWALGFYP
AEITLTWQRDGEDQTQDTELVEVTRPGDGTFOKWAADVVPSPGQEQRYTCHILHKSPLKPI
TLRWEPPTQIPVSIAGLVLLVITGAMLVGVVIRWKKHSGRKGGSYSQAENKYGGSS
AQDSQMSVMAPNEQASCLVGL

>cf_AZGP1 Chain=chain06 Position=6:9623328-9626753(-) GeneID=ENSCAFG00000014889
ProteinID=ENSCAFP00000021945 Score=296.9 E-value=2.7e-86 Length=3426 Comment=Overlaps single
annotation & in proteins
MDTIVSALLSLLLGLGPAVPQETQGGPYSLSFFYTGLSRPSDGFPSFQATAYLNDQDFH
YDSETGKAIPRYPWSQMEGIEDWEKESKLOKAREDFMVTLKDIMYKDKESHTFQGM
FGCELQNNKNSGAFWRYYDGRNFIENKEIPAWVPQDPAALNTKKKWEAEVYVQRAKA
YLEEECPVMLQRYLEYKTYLDRQEPSPVITSHTGTEPIQTLKCVWSGFYQEIHLHWI
QADDTQETKSGGALLPSGNNTYQAWVMSASPDLASFSYSCLVKHSSLSQPLTVLWDKR
QGAVRAEGSEDTQVQ

>cf_HFE Chain=chain07 Position=35:24033561-24036503(+) GeneID=ENSCAFG00000010925
ProteinID=ENSCAFP00000016096 Score=277.4 E-value=5.1e-79 Length=2943 Comment=Overlaps single
annotation & in proteins
MSPGARRARLLLLLLLLLFLQLPTVAQRRPRRSHSLRYLFMGASVPDLGLPLFEAR
GYVDDQLFVSYSHESRRAEPRAQWVRTGAASQLWLQLSLSKLGWDMHFIVDFWTIMDNHN
HSKVTKLGVSSSHTLQVILGCEVQEDNSTGFWKYGYDGNHLEFCPETLDWRAAEPKA
QATKLEWEVNKIRAKQNRAYLQRDCEQLRQLLELGRGVLDQVPPVVKMTHHVTSVAVT
LRCQALSFYPQNTMKWLKDRQPLDAEDVEPKDVLPNGDGTYQRWVALAVAPGEEQRYTC
QVEHPGLDQPLTASWEAPMSGTLVVGIIISGIAVCIIVLFTGILFRILRKRQASRGAMGDY
VLAE

>cf_FCGRT Chain=chain08 Position=1:106963801-106989648(-) GeneID=ENSCAFG00000003625
ProteinID=ENSCAFP00000005404 Score=220.4 E-value=1.2e-61 Length=25848 Comment=Overlaps multiple
annotations; one in proteins
MGVPRPRSWGFLFLPTLRADSHLSLLYHLTAVSAPPPGTPAFWASGWLGPQQYLSY
NNLRAQAEPYGAWWENQVSWYWEKETDLDLRTKEGLFLEALKALGDGGPYTLQGLGCEL
GPDNTSVPVAKFALNGEDFMTDPKLGWTWGDWPEPETVSKRWMQAGAVSKERTFLLYS
CPQRLGLHLRGRGNLEWKEPPSMRLKARPGSPGFSVLTCSAFSFPPELQLRFLRNLGA
AGSGEGDFGPNGDGFSHAWSSLTVKSDEHHYRCLVQHAGLPQPLTVELESPAKSSVPV
GIVIGLLLTAFAVGGALLWRRMRKGLPAPWMSLRGDDVVALPTPGVPKDADS

>cf_CD1B Chain=chain09 Position=38:23317871-23319502(-) GeneID=ENSCAFG000000024889
ProteinID=ENSCAFP00000016818 Score=190.6 E-value=5.1e-55 Length=1632 Comment=Overlaps single
annotation & in proteins
MRLLLLLLWPTALYPGGGGEAAPADLQGPSTSYRVIQISSFANSSWAQNOGSGWLDVQIH
GWDADAGRAVFLKPKWSKGNFSDEEMVELEEIIQVYLTGFIVLQVQDHAPEFQMQYFPFIEQ
VAGCQHPDRGTESFLRGALGGLDFLSLKNHSCVPAPEGGSRAQRICELILQYEGIRDIA
EKLLFETCPREFLLGLLDAGKAELERRPVRPEAWLSAGSPSPGRLQLVCHVSGFYKPKVWV
TWMRGEQEQQTRRGDVLPHADGTWYLRATLDVAAQEAAGLSRVKHSSLGGQDMVLHWG
NSISIGLISLAVILPFLIFLIGLFFWLWRRRWSYQSIS

>cf_chain10 Chain=chain10 Position=38:23455331-23457040(-) GeneID=ENSCAFG000000031841
ProteinID=ENSCAFP000000040338 Score=184.7 E-value=2.5e-53 Length=1710 Comment=Overlaps single
annotation & in proteins

MLFLQLVLLAVLLPEGDSEDDSQEPVSRVIRTSSLYNRSWTQNGSVWLDVQIHAWDN
KNRTFVFRWPAQGDLSNERMEADQLFYSNYIFYNLVYHDHVSQWQLECFRYLREGGK
EVADVCSVLGSLFQLQERGSIGF

>cf_CD1A6 Chain=chain11 Position=38:23380178-23384680(-) GeneID=ENSCAFG00000023677
ProteinID=ENSCAFP00000016817 Score=179.6 E-value=8.8e-50 Length=4503 Comment=Overlaps multiple
annotations; one in proteins

MLFLQLVLLVLLPEGDSEDDSQEPVSRVIRTSSLYNRSWTQNGSAWLDDELQTHSWDS
DSGTFLFLWPWAKGKLSKEELIERERTFHTFSIRFPLIFQDSVSDWQLEYFPFQVQMAEGC
GLYFGKPSVGMQIAYQGSDDLVSFQNKSWWSPKGGRRQQVCKLLNQYHVVNLRHSHI
SDFCPHYLLGLLDAGKADLQRQVRPEAWLSAGSPSPRSDHLRLVCHVSGFYKPKVWVTWMR
GEQEQQGQSWRGDVLPHADGTWYLVQVSLDVKAKEVAGLSCRVRHSSSLGGQDMVLHWRPHS
MGLVFLVVIPLVLLAGLAWLWKRKWKTHNRPOCTDFPLK

>cf_chain12 Chain=chain12 Position=38:23347787-23349491(-) GeneID=ENSCAFG00000023735
ProteinID=ENSCAFP00000035027 Score=175.9 E-value=1.2e-50 Length=1705 Comment=Overlaps single
annotation & in proteins

MLFLQLVLLAVLLGGDSEDDSQEPISFRILTTSFYSSSTQNGSAWLDDELQTHGWND
KTGAFRYLQSWSKGNFSNEELLEQAQNLFYTYTIRSPSTFHNHIRDQWLEYFPFQIQLVLGC
DSHFGEASAGFLQLAYQGSDDLVSFQNTSWRPSPEGGSSRAQKICSLFIEYHVVFSEIVHKLK
FDSCPRFLLGLLDAGKAYLQRKVRPEAWLSAGSSPGGHLRLVCHVSGFYKPKVWVWMMR
GEQEQQGTLYSDILPNTDGTWYLVQVSLDVKSKEAAGLSCRVRHSSSLGGQDMVLHWEQPHS
MGLVFLVVIPLVLLAGLAWLWKRKKAHWRPQCTDFPSEQEPSSPGSSTYLNPAQH

>cf_CD1A8 Chain=chain13 Position=JH374180.1:2244-3945(+) GeneID=ENSCAFG00000025281
ProteinID=ENSCAFP00000016822 Score=175.1 E-value=2.1e-50 Length=1702 Comment=Overlaps single
annotation & in proteins

MLFLQLVLLAVLLGGDSEDDSQEPISFRILTTSFYSSSTQNGSAWLDDELQTHGWNN
KTGAFRYLQSWSKGNFSNEELLEVAQNLFYTYTIRSPSTFHNHIRDWHLEYFPFQIQLVLGC
DSHFGEASAGFLQLAYQGSDDLVSFQNTSWRPSPEGGSSRAQKVCSLFNQDHSVHEIVRKLK
NEICPRILLSLLDAGKVDLQRQVRPEAWLSTGSPSGHLRLVCHVSGFYKPKVWVWMMR
GEQEQTGSRGDVLPHADGTWYLVQVSLDVKAKEAAGLSCRVRHSSSLGGQDMVLHWEQPPS
MGLVFLVVIPLVLLAGLAWLWKRKKAHWRPQCTDFPSEQEPSSPGSSTYLNPAQH

>cf_CD1E Chain=chain14 Position=38:23295855-23297490(-) GeneID=ENSCAFG00000011437
ProteinID=ENSCAFP00000016826 Score=174.1 E-value=4.9e-50 Length=1636 Comment=Overlaps single
annotation & in proteins

MPLLLLLLFGGLVQRGASTGASQAGPPHATEEPPSFRVLQTSFANYSWAYTQGGGWL
GELQTHGWDNVRDTRIFLWPWSRGNFSAVELKNLQSLFALYFHGFAIEVQAFARYFQFEY
PFELQMSAGCRLHTGKASESFLNGAYQGSDFLSFQGNISWYSPGAGSRARKVCEMLNRYQ
DIKEIKVSLIGYICPQFLAGILEAGKAEGRQVRPEAWLSADSPGPGRLRLVCHVSGFY
KPKVVRVTWMRGEQEQRGTRRGDFLPHADGTWYLRATLDVAAREAAAGLSCRVKHSSSLGGQD
MVLHWGGGNSALLTSLGLAAVVTLALPVVHTCCKKRSSNRKAPAPSPDSMGNTTPKPR
TSGHQLYTPQESWVKNRFLEKLNKASLNRLWRR

>cf_CD1D.1 Chain=chain15 Position=38:23328749-23330959(-) GeneID=ENSCAFG00000024030
ProteinID=ENSCAFP00000016820 Score=164.1 E-value=4.1e-47 Length=2211 Comment=Overlaps single
annotation & in proteins

MLPLECILLALLLQGDNAQAIQEHVFTIIQISSFVNQSWVQHRGSGWLQDMQTHGWDT
DSGTIIFLHTWSKGNFSDEELLDLELLFRVYLIGLTREAAQYVSQLHFTYFPFEIQVVGGC
ELRSSDFSKGFLRSAYEGSDVFTFQNMVSLVSPGADSKAQSVCYLINQYEGIKIIVYRLI
TNTCPRFVGLGLFDAAKVDYKQVRPEAWLSTGPPAPRGRRLVCHVSGFHPKPKVQVTWMR
GEQEQQGTQRGDILPHADGTWYLRVTLDAAREAAAGLSCRVKHSSSLGGQDIVLVWGHHPV
MYLILLAVTVPLELLLDLGLWFRKRCSYQDIP

>cf_CD1D.2 Chain=chain18 Position=38:23490486-23491711(-) GeneID=ENSCAFG00000030554
ProteinID=ENSCAFP00000016814 Score=138.2 E-value=2.1e-39 Length=1226 Comment=Overlaps single
annotation & in proteins

IGRALILLFRLLSGLPATIWSVSQLHSVTLVQHNQVSVFANGSWARTDQAWLGELQ
AFGWSNGSDAVRCLRPWARGAWGAGRWRALQEQLGLYRRSFTDRVRELVKMLQLRYPLEI
QVSAGCEVRPNTSEDFHAAVQGEELISFQGSWVPAQAPRWVHRATKELNKDQGTTR
TVRELLRDTCPFFVRGLLEAGRSELEKQERPEAWLSAGPAPGPRRLVCHVSGFHPKPKV
RVTWMRGEQEQRGTRRGDVLPHADGTWYLRVTLDAAREAAAGLSCRVKHSSSLGGQDMVLH
WGRSRSSAWLVAVALGSLLVIGCLGCSAAWCRARRSYQDIL

>cf_chain16 Chain=chain16 Position=38:23437375-23439070(-) GeneID=ENSCAFG00000023735
ProteinID=ENSCAFP00000035027 Score=160.7 E-value=3.9e-46 Length=1696 Comment=Overlaps multiple
annotations; one in proteins

AFQGPVSRVQVILTTSSLYNRSWTQNGSAWLDDELQTHGWDSKMGAFIFLWPWSTGNLSNEE
LDPFQVQVTLGCELHFGPEPSVGFVRLAYQGEDLISFQNKSWWSPKGGRRQQVCKLLNQYHVVNLRHSHI
LDHDDTELTHELITGHCPRFVLSLLDAGKADLQRQGCPLAPARSGHLRLVCHVSGFYKPK
VRVTWMRGEQEQRGTRRGDVLPHADGTWYLVQVSLDVKAKEAAGLSCRVRHSSSLGGQDMVLH
HW

>cf_chain17 Chain=chain17 Position=38:23363410-23365089(-) GeneID=ENSCAFG00000023735
ProteinID=ENSCAFP00000035027 Score=157.6 E-value=6.3e-45 Length=1680 Comment=Overlaps single
annotation & in proteins

YFQEPISFQIILTTSFYRSRWTQNVGSAWLDDELQTYGWDSDSGTFTCMWPWSRGNFSNKE

LMEEQSFHTFSIRFPLIFQDHFSEWQSPDPFVQVAAKGCCELHIGETA VGF M* IAYQGED
LVSFQNKSGWPSPPNGGRRRA*QVSRLFNQYHVINVRLQTHISDSCPPFLGLLDAGKTHLQ
RQWLSAGPSPGPRQLQLVCHVSGFYKPVVWVTWMRSEQEQQGTQLPHADGMWYRQVSLDV
KVKETAGLSCWVTHSSLLGGHGPPLG*

>cf_IER5 Chain=chain19 Position=7:14523431-14528827(+) GeneID=ENSCAFG00000029718 ProteinID=None
Score=96.9 E-value=7.9e-27 Length=5397 Comment=Overlaps single annotation; not in proteins
SLPGTHSLTYFFLGI SDPSHGLTEFFSHPII IYTCCWQKELWALWVENSMPGHWERYTQ
GFHTYQGMVG*GSLEDRSTIFLQHAYDQDFIILNKDLFFWMATNNMARITKWAQEA NWH
EL*YQKNWLEEHSITSLKFLLELRHGME*ITFPRIITPPCRIHGFYPPGICMIWLKNGEEI
VQEIGKGDILPKGDGSYWTWGSIELDPQSGDLYSC

>cf_chain20 Chain=chain20 Position=38:23424398-23426095(-) GeneID=ENSCAFG00000023735
ProteinID=ENSCAFP00000035027 Score=80.5 E-value=3.2e-22 Length=1698 Comment=Overlaps single
annotation & in proteins
GFQGPVSFQIILTTFSFYNRSS TQNQGS AWLDELQTPGRNPKTGSFFYRQPWSKGNFSNKE
LMKAENL FHKLLNEICPPI LLSLLDAGKADLQRQWLSAGPSPGPRQLWL VCHVSSFYKPKP
VQVTWMRGEQEQQGTRRGEVLPHADGTWYL*VYLYVA AWEVAGLSCQVRHSS LGGQDMVF
HW

>cf_chain21 Chain=chain21 Position=38:23403891-23405377(-) GeneID=ENSCAFG00000023735
ProteinID=ENSCAFP00000035027 Score=42.9 E-value=1.9e-10 Length=1487 Comment=Overlaps single
annotation & in proteins
SAWLDELQTHGGDPKTGAISFLQAWSNGNFSNEELEHEHW RPSPEGGRRAQKFCSLFNQDH
VSHEIVRKLLENEICPPI LLSLLDAGKADLQRQWLSAGPRPRSGHLWLACHVSSFYKPKPVR
VTWMRGEQEQQ

>cf_PROCR Chain=None Position=24:24184440-24187910(+) GeneID=ENSCAFG00000007945
ProteinID=ENSCAFP00000011762 Score=99.7 E-value=8.3e-29 Length=0 Comment=Not detected in genome
search
MLTTLPLLLCLLLPSWAFCSQEASDGPRNLHMLQISYFRDPYQVWHRGNASLGLTTHV
LEGQGNVTTLQQLPLEEPESWARRE RRLKTYLDEFNLLVRLVHQERRVNFPLTIRCSLG
CELPPEGSPA HVFVEAVNESAFVSFQPEKALWVAGPQVPSRVVYTLHLQ LNSYNRTRYE
LREFLQDTCVQYVQEHNASKNSKGSQTGRSYTSLVLGVLVGSFIIIVGAVGIFLCTGGRR
C

>cf_ENSCAFG00000000399 Chain=None Position=1:40635694-40637301(+) GeneID=ENSCAFG00000000399
ProteinID=ENSCAFP00000000574 Score=92.9 E-value=8.7e-27 Length=0 Comment=Not detected in genome
search
FTGATSLFCEFSVKPQSSAGR PWCEVRGYVKNRFLSYNCSKELTLVGP MTRMNDTGF
RETQKETLNDLMEELQKKVRDIKAEIYPKNGSLDLHGEMKCVRGASRYTSASWKFANQO
ITHLFDS ENKWTVLQAPGRQFQGTLDSDRDTNFLT KVSNGDCKSWMEYIMNHWDEKLE
TTVPQSRPQDMAPSGASGAPSTWIFLLFLYCAILLGI

>cf_RAET1e Chain=None Position=1:40602291-40604598(-) GeneID=ENSCAFG00000024105
ProteinID=ENSCAFP00000032793 Score=164.3 E-value=6.2e-48 Length=0 Comment=Not detected in
genome search
AGMETPFSARVVLGQLPFSFTDAHSLCLNLT VKSQARPEHPWYEVQGSVDKPKPFLQYDSD
SSKVRPLGLL GKKVNATKAWTEL TQMLADVGOELRMILPDIKLENNMTKGPPTLHKL LSC
QCEAEWHFNISRQMTLFFDAMKISWTVVNPGARGIKEEWESKGMADYFRIRISMGD CNWL
QEFLEHWKMLEPSILAPSITVPVANQSSSILKRLWII LVVIVSSVLITSAYQLFHYQKFL
LCTMRELQRKTNPWQKLTLMTI

>bt_BOLA.1 Chain=chain01 Position=23:28502997-28505915(-) GeneID=ENSBTAG0000002069
ProteinID=ENSBTAP00000031126 Score=408.8 E-value=7e-118 Length=2919 Comment=Overlaps single
annotation & in proteins
QMRVMRPR TLLLLSSGLVLTETLAGSHSLRYFYTGVS RPLGEP RPIAVGYVDDTQFV
RFDSDAADPRTEPRVPWMEQEGPEYWDRETRNLKDAAQTFRVNLNTLLGYYNQSEAGSHT
LQWMYGCDVEPDGRFLRGRYQDAYDGRDYIALNQDLRSWTAADTAAQITKRKWEAAGEAE
RRRNYLEGRVCVEWLRRLYLENGKDTLLRADPPKAHVTRHPSSDREVTLRCWALGFYPEEIS
LTWQHNGEDQTQDMELVETRPSGDGTFQKWAALVVPSGEEQRYTCRVQHEGLREPLTLRW
EPPQTSFLTMGIIVGLVVLVAVVAVGAVIWRKKHSGEKGRITYTQAASSDSAQGS DVSLTV
PKV

>bt_BOLA.2 Chain=chain04 Position=23:27863473-27865992(+) GeneID=ENSBTAG00000022590
ProteinID=ENSBTAP00000035741 Score=370.4 E-value=1.4e-106 Length=2520 Comment=Overlaps single
annotation & in proteins
RVMGPR TLLLLSSGLVLTETWAGSHSLSYFNTGVYQPLGEP RFFAVGYVDDTQFARFD
SDAPNPRMEPRAPWMEQEGPEYWEEMTRDAKESQQKSRCLYNLRGYYNQSEAESHI LQV
MFGCEVGPDRLLRGFWQKAYDGRDYIALNEDLRSWTAADTVAQITKRKWDVSGQAKIQR
NYLEVKCVQWLLRHLETKGDTLLRADPPKTHVAHHRISDREVT LRCWALGFYKPEITLW
QQDGEDLTQDMELVETRPSGDGTFQKWAALVVPSGEEQRYTCRVQHEGLREPLTLRWEP
QLSVLIKGIIVGLILLMVTGAVVTGAVIWRKKHSGEKGRGYTQTASSDSAQGS DVSLTV
KG

>bt_BOLA.3 Chain=chain06 Position=23:27647460-27651819(+) GeneID=ENSBTAG00000005182
ProteinID=ENSBTAP00000031125 Score=368.1 E-value=6.2e-106 Length=4360 Comment=Overlaps single
annotation & in proteins

MRVVGQRTLLLLLLDALILLETETRAGSHSLRYFYTAVSRPGLGEPFRFISVGYVDDTQFVR
FDSDADDPRIEPTARWVEQEGPEYWHQETQRTKDTAQFFRVYLNLTGRGYYNQSEAGSHTV
QEMYGCDVGPDGLLRGYDQFAYDGRDYIALNEDLRSWTAADTAAQVTKHNAEAAAGDAARV
R IYLEGKCVWELRRYLVTKDGLLRADPPKTHVAHHPISDREVTLRCWALGFYPPEEISLT
WQRNGEDQTDQDMELVETRPSGDGTFQKWAALVVPVSGEEQRYTCHVQHEGLQEPLTLRWEP
PQPSIPIMGIIVGLVHLMVTGAMVTGAVIWRKKHSGEKGRIYSQASSSDSQGSDVPLTV
PKV

>bt_JSP.1 Chain=chain02 Position=23:28470224-28473004(-) GeneID=ENSBTAG00000020116
ProteinID=ENSBTAP00000011795 Score=401.8 E-value=8.6e-116 Length=2781 Comment=Overlaps single
annotation & in proteins

MRVMGPRTLLLLLSGVLVLTETRAGSHSLRYFYTAVSRPGLGEPFRFIIIVGYVDDTQFVFR
DSDAGDPRVEPRARWVEQEGPEYWDQETRAKADHAQFFRLGLNLTGRGYYNQSEAGSHTLQ
WMYGCDDVGPDRLLRGFWQFQYEGRDYIALNEDLRSWTAADTAAQISKRKWEAADVAERQ
RNYLEGTCVWELPRYLENGKDTLLRADPPKAHVTHHPISGREVTLRCWALGFYPPEEISLT
WQRNGEDQTDQDMELVETRPSGDGTFQKWAALVVPVSGEEQRYTCSVQHEGLQEPLTLRWEP
PQTSFSLTMGIIVGLVLLVTVGAVVAGFVIWMKKRSGEKGNYIQASRSDSAQGSVSLTV
KV

>bt_BOLA-NC Chain=chain03 Position=23:28355165-28357959(-) GeneID=ENSBTAG00000019386
ProteinID=ENSBTAP00000021451 Score=384.4 E-value=1.1e-110 Length=2795 Comment=Overlaps single
annotation & in proteins

MGPRTLLLLLGLVLPDRTRAGPHSMRYFLTAVSRPGLGEPFRFITVGYVDDTQFVFRDSD
RPDPRMEPRARWVEDEQEGPEYWDQETRIQKENTQTFRANLNTLGRGYYNQSEAGSHTIQWMH
GCGVSDGRLLRGYDQFAYDGRDYIALNEDLRSWTAADTAAQITKRKWEAAGEAERFRNY
LEGEVCLRRHLENGKDTLLRADPPKAHVTHHPISGREVTLRCWALGFYPPEEISLTWQR
NGEDQTDQDMELVETRPSGDGTFQKWAALVVPVSGEEQRYTCRVQHEGLQGLTLRWGKGE
PPQPSVPMIGIIVGLVLLVTVGAVVAVVARAVIWSKKRTGEKGRITY

>bt_chain05 Chain=chain05 Position=23:27689770-27692225(+) GeneID=ENSBTAG00000012208
ProteinID=ENSBTAP00000016194 Score=370.2 E-value=1.7e-106 Length=2456 Comment=Overlaps single
annotation & in proteins

QTRRIRVMGPRTLLLLLSEVLVLTETWAGSHSLRYFYCTVSRPGLGEPFRFAVGYVDDTQ
FARFDSADPNPRMEPRAPWMEQEGPEYWEAMTRDAKKAQQLRTGLNLTGRGYYNQSEAGS
HTLQWVFGCDVGPDRLLRGYDQFAYDGRDYIALNEDLRSWTAANTVAQITKRKWEWTSGE
AEFQRNYLEVKCVQWLLRHLETGKDTLLRADPPKTHVAHHPISDREVTLRCWALGFYPPEE
ISLTWQDGEDQTDQDMELVETRPSGDGTFQKWAALVVPVSGEEQRYTCHVQHEGLQEPLSL
RWKPPQPSIPIMGIIVGLVLLVTVGAVVTVGAVIWRKGTGRGYTQAASSDSAQGSVSLM
VPKV

>bt_MR1 Chain=chain07 Position=16:63652896-63655404(+) GeneID=ENSBTAG00000009924
ProteinID=ENSBTAP00000013095 Score=325.4 E-value=3.3e-95 Length=2509 Comment=Overlaps single
annotation & in proteins

MMLLLPLIIVLMMKLSDARTHSLRYFRLGISEPGYGIPEFISAGYVDSHPITMYNSVSQL
KEPRALWMEENLAPDHWEYRYTQLLRGWQAFKVELKQLQHYNHSGFHTYQRMIGCELE
DGSITGFLQYAYDQDFLIFNKDLSWMMAMNDVADIIRRVWEANRHELQYQKNWLEEECI
AWLKRFLYEGKDALQRTEPPKVRVNHKETFFGITTLYCRAYGFYPPEEISINWMKNGEEIF
QDTDYGGILPVGDTYQVWVVELDPQNGDIYSCHVEHGGVHMLQGFQESSETILLVVKAA
VGFIVLAIALAGVILAWRKRPRGKKNVICLSTPEH

>bt_AZGP1 Chain=chain08 Position=25:36982591-36984540(-) GeneID=ENSBTAG00000026236
ProteinID=ENSBTAP00000037042 Score=304.5 E-value=1.1e-88 Length=1950 Comment=Overlaps single
annotation & in proteins

MVPVLLALLLLGPAVSEETQAGNYSLSFLYTGLSKPREGFPFQAVAYLNDQPFPHYNS
EGRRAEPLAPWSQVEGMEDWEKESALQRAREDIFMETLSDIMDYKDRSHTFQGAFC
ELRNNESSGAFWGYAYDQDFIKFDKEIPAVVPLDPAQNTKRKWEAEAVVQRAKAYLE
ECCPGMLRRYLPYSRTHLDRQESPSVSVTGHAAAPHKRTLKCLAYDFYPRSIGLHWTRAG
DAQEAESGGDVLPSGNGTYQSWVVVGVPPEDQAPYSCHVEHRSLTRPLTVWPDPRQQA

>bt_chain09 Chain=chain09 Position=23:27842654-27852650(+) GeneID=ENSBTAG00000007075
ProteinID=ENSBTAP00000009299 Score=298.3 E-value=2.9e-85 Length=9997 Comment=Overlaps single
annotation & in proteins

MRVMGSRTLVLVLLPGALVLTETWAGSHSLRYFYTAVSRPGLGEPFRFIIIVGYVDDMQFVFR
DSDADDPRIEPTARWVEQERPEYWDQATQRTKDTKQTFRANLNNLTGRGYYNQSEAGSHTLQ
EMYGCDVGSDDGFLRGMQFAYDGRDYIALNEDLRSWTAADPAAQVSKRKEQSGAAGER
NYLNRECVGELRRYLETGKDTLLRTPPKTHVTCHPISGREVTLRCWALGFYPPEEILLT
QRDGEDQTDQDMELVETRPSGDGTFQKWAALVVPVSGEEQRYTCHVQHEGLQEPLTLRWEP
ETSFSLTMGIIVGLVLLVTVGAVVAVVTVGAVIWRKKRSGERGHYTAESKCRGDTIPGTLRLV

>bt_chain10 Chain=chain10 Position=23:28302417-28306062(+) GeneID=ENSBTAG00000038619
ProteinID=ENSBTAP00000048068 Score=293.8 E-value=5.3e-86 Length=3646 Comment=Overlaps single
annotation & in proteins

MLWGPVLLYFFLAYLPETQTRSHSLQYFYSVVSEPGVPSFMAFGFVDNQPFIYRVDSEE
MKAKSCVHWLREEPSYFDDETKIFTSRMKIFHLNLRNVQYYNQTKEDGLNRLAQAQKQTS
PHTLQFTYGCLELDEGRITWWHQYGYDGEDYLSLHMDPLQYTAATFVAQYTKQKWEAGGN
FIERDKNYLEKECILWLWRYLTFGGESLNRTEPPKTHMTHHRISDREVTLRCWALGFYPA
KISLTWQRNGEDQTDQDMELVETRPSGDGTFQKWAALVVPVSGEEQRYTCHVQHEGLQEPLT
LRWEPLKTSVPIITGIIVGLVLLVTVGAVIWRKKSGYHSTNGKNRITYFQTAI

>bt_chain11 Chain=chain11 Position=7:41789508-41797877(-) GeneID=ENSBTAG00000004815
ProteinID=ENSBTAP00000006320 Score=289.5 E-value=1.6e-82 Length=8370 Comment=Overlaps single
annotation & in proteins
MAELANLKLKNSWPRGTHSLRYHYLDLSEPDPSLPKFQAVGVDDQFFIRYDSRVDKAE
PQAPWIMPMNAKYWEKETFKQRKWAELHQVETWRMMGYHNHSSVLTGMHSTHRMFGCEIQ
EDGHSNSFWQFGYDQGDHVSMDMETLSWVSANPVALLTKRRMETEHCYAEYNKAYLEGPC
LASLRRYLELGGQRFTRREPPVWVTKHSAQDGGTTLRCWALGFYQDITLSWWLGEKKL
NSKLEYVEARPSGDGTYTWMMAVWVPAGDETYTCHVQHCSLNHTLTVSWEMPSHPGLTA
MVISLILILLVFGGGVSLTKCLQVLGGGKGG

>bt_chain12 Chain=chain12 Position=23:28330913-28333698(-) GeneID=ENSBTAG00000005146
ProteinID=ENSBTAP00000025832 Score=285.1 E-value=2.7e-80 Length=2786 Comment=Overlaps single
annotation & in proteins
MRVMGPRTLQLLSGILVLIETLAFHSLRYFYTAVSRPGLGEPFIIIVGYVDDTEFVRF
DSDAPDRMEPRARWVEQEGPEYWDQETRAKAKDAAQTFRANLNSLRGYNQSEAGSHTLQ
LMYGCVDVPGDGLRGYEQYGDGRDYIALNEDLRSWTAADTAAQISKRKVEAAGDAARV
RIYLGRCVESLRRYLETGKETLLRADPPKAHVTHHPISGREVTLRCWALGFYPPEIILLTW
QHNGEDQTQDMELVETRPGDGNFQKWAALVVPSEGEQRYTCHVQHEGLQEPLTLRWEPPQ
TPFLTMGIIVGLVLLVAVVAGAVIWRKKRSGEKGRTYTQASSSDSAQGSVSLTVPKV

>bt_BT.105339 Chain=chain13 Position=23:27715371-27718199(+) GeneID=ENSBTAG00000001476
ProteinID=ENSBTAP00000026794 Score=284.4 E-value=3.7e-81 Length=2829 Comment=Overlaps single
annotation & in proteins
LLLLSGVLVLTETWPGSHLSYFCTCVSRPGLGEPFRFAVGVVDRTQFARFSDAPNPRM
EPRAPWMEQEGPEYWEEMTRDAKKAQQRMLALNLRGYNQSEAGSHTLQFVFGCHVGL
DGRLLRGIWRTAYDGADYLLNEDLRSWTAANTVAQITKRKWEKTSGEAEFQRNYLEVKCVQ
WLLRHLEKGDITLLRADPPKTNVAHHRISDHDITLRCWALGFYPPEIISLTWQRDREDQVQ
DMEFVETRPSGDGTFFQKWAALVVPSEGEQRYTCCVQHEGLQKPLTLRWEPPQPSIPIMGI
IVGVLLMVTGAVVTEAVIWRKKHSGEVGRGYTQASNDGAQGSVPLTVLKG

>bt_HFE Chain=chain14 Position=23:31617675-31635117(+) GeneID=ENSBTAG00000013065
ProteinID=ENSBTAP00000017371 Score=278.0 E-value=5e-79 Length=17443 Comment=Overlaps single
annotation & in proteins
MGPRARFALLLILLRTAATQGRPPRSHSLRFLFMGASKPDLGLPLFEALGYVDDQLFVS
YDHESRAERRAPWLWGRATSQLWLQLSQNLKGDHMFIVDFWTIMDNHNQSKVTKLQAL
PESHTLQVILGCEIQEDNSTRGFWKYGYDGDHLEFRPETLDWRAAEPRAVTQKLEWEVN
KIRAKQNRAYLDRDCPEQLLHLELGRGPLEQQVPLVKTTHHVTSLSLTLRCLRNFPY
QNITIRWLKDKQFLDAKEIKPEDVLPNGDGTQAWVALAMLPGEQRYSCQVEHPGLDQP
LTATWEPSLSGTLVTGILSGIAVCVIFLIGILFRILKRRQSSRGAAVNYALAECE

>bt_FCGRT Chain=chain15 Position=18:56416586-56430488(+) GeneID=ENSBTAG00000013926
ProteinID=ENSBTAP00000018500 Score=207.5 E-value=9.4e-58 Length=13903 Comment=Overlaps multiple
annotations; one in proteins
MRLPRQPWGLGLLVLLPGALSAENYRSLQYHFTAVSAPAAGTPAFWVSGWLGPQOYLS
YNNLRAQAEYAGAWWESQVSWYWEKETMDLRNQETLFLQALQALGEGPPTMQGLLGCCEL
GPDNVSVPAKFAALNGEEMFDPKLGIDWDGWPESTRVSIKWTQOPEAVNKEKTFLLYS
CPHRLGHLEGRGNLEWKEPPSMRLKARPGSPGFSVLTCSAFSFPPELKLRLFRNLGLA
IGSGEIDMGPNQDGSFYAWSSLTVKSDEHHYRCVVQHAGLAQPLTVELESPARTSVPVV
GIVIGLFLLLTVAAGGALLWRRMRKGLPAPWISFRGEDVGLLPTPLGSKDGES

>bt_CD1B5 Chain=chain16 Position=3:11753955-11755658(-) GeneID=ENSBTAG00000038502
ProteinID=ENSBTAP00000012398 Score=198.5 E-value=2.8e-57 Length=1704 Comment=Overlaps single
annotation & in proteins
MLLLPLLLLAVIVPGGDNEAFQGPSTFHLIQISTFANSTWTQNGSGWLLDQLIHGWDS
DSGTAIFLKPWSKGNFSDEEVAEMEELFRVYFIGFTLEVQDIVSEFQFEYFPFVIQGIAGC
ELHSGKVIQSLFRAGFGGLDFMSIKNRSCVPEPEGSDAQWFCVFITQYQILRIIDTLL
SETCPRYLLGVLDAGKAEQKQVQKPEAWLSSGPTPRPGRLLLVCHVSGFYKPKVPRVMMWR
GEQEQPQTQGNIMPADWTWYLRVTLDDVAGEAAGLSCRVKHSSLDGQDIIILYWGHPKY
IGLISVAIIVPSLILICLALWLWRRWSYQTIL

>bt_CD1E Chain=chain17 Position=3:11725705-11727347(-) GeneID=ENSBTAG00000009421
ProteinID=ENSBTAP00000038206 Score=190.8 E-value=6.7e-55 Length=1643 Comment=Overlaps single
annotation & in proteins
MLLLPPLLPLLFKGLLCHGASIVGLQVPGPHVAEESPSFRLIQISSFANHSWTKTQGS
GWLGDQLQTHGWDSVLASIRFLRPWAQGNFSKEELKNIQAFQLYLHSPFREVQAYASQFQ
FEYFPFLQVSGFCLHSGKALETFLLNGAYQLDFLSFQENSWKSSPEAGSRAENVCKVNL
HYRVIKEIVQRLLSDTCRFLAGVLEAGKAELEKQVQKPEAWVSKGSPGPRLLLVCHVS
GFHPKPIWVMMWRKQKHPGTRRQDVLVADGTWYLRVTLDDVAGEAAGLSCRVKHSSLDGQDIIILYWGHPKY
GQDIIHGGYSVFWILICLIVIVTLVMLVIVDLSLFFKKQSSNQDVFSSISHPAPFPAKGTQ
DPRNSGHQLCLAQESWIKNKLKWKARLNQL

>bt_CD1B3 Chain=chain18 Position=3:11849341-11851113(-) GeneID=ENSBTAG00000022893
ProteinID=ENSBTAP00000013155 Score=189.8 E-value=1.1e-54 Length=1773 Comment=Overlaps single
annotation & in proteins
MLLLPPLLGLVILPGGDNEVDFQGPSTFHLMQISTFVNSTWAQNGSGWLLDQLIHGWES
DSGTAIFLKPWSKGNFSDEEVELVDLFRAYFIGFTREVQDRVNEFQLEYFPFVIQVTAGC
ELHSGEAIESSLRGALGGLDFVSIQNHSCVPAPDSGSRGQKFCALTYQYQGISDIERLL

SETCPRYLLGVLDAGKAEALQRQVKPEAWLSSGPTPGPGRLLLLVCHVSGFYKPKVVRVMWMMR
GEQEQPQTQQGDLMPNADWTWYLRVTLNVAAGEAAGLNCRVKHSLSLGDQDIILYWGHPST
IGLILVAIIIPSLILLICLALWFRRRWSYQNIL

>bt_chain19 Chain=chain19 Position=3:11903102-11906623(-) GeneID=ENSBTAG00000026180
ProteinID=ENSBTAP00000001274 Score=186.7 E-value=1.1e-53 Length=3522 Comment=Overlaps single
annotation & in proteins
LGVILPGGDNEDVFGQPTSFHLIQISTFVNNTWAQNGSGWLDLQIHGWESDSGTAIFL
KPWSKGNFSDDEVTELVDFRAYFIGFTREVQDRVNEFQLECEYSPLIWKDSQ

>bt_chain20 Chain=chain20 Position=3:11784601-11786208(-) GeneID=ENSBTAG00000000963
ProteinID=ENSBTAP00000001272 Score=185.3 E-value=2.8e-53 Length=1608 Comment=Overlaps single
annotation & in proteins
LSLFSFYPPDFVIQGIACCEMHTGKAIGSFLKRAFRGLDFVSVKNDSCAPAPEGGSKAQR
FCALIIQYQAICDTIAKLLLETCPQYLLSVLDAGKAEALQRQVKPEAWLSSGPTPGPGRLL
LVCHVSGFYKPKVVRVMWMMRGEQEQPQTQQGDLMPNADWTWYVRITLDDVAAGEAVGLSCR
RHSSSLGDQDIILYWGHPSTGLIFVAIIVPSLILLICLALWFRRRWSYQTIM

>bt_CD1A.1 Chain=chain21 Position=3:11994653-11996562(-) GeneID=ENSBTAG000000038263
ProteinID=ENSBTAP00000009739 Score=183.2 E-value=7.8e-53 Length=1910 Comment=Overlaps single
annotation & in proteins
MLFLQPLLLLALLVGGDNEQNKNSVPEGFQEPISFEVICVLSFHNSSWVQSLGSGWLGEL
QTHGWKSNSGTFIYLWVWSKGNFSNEELMELQNYLHTNFVRFVQAFYSHARKWQFEYPPFE
VQIAEGCELHAGEVVPVGMRIAYQGSDFLSFQNKSWVSSPEGGKRAQVLRRLFNLFRTGAQ
EIIHKLLSDTCPRLLDAGKAYLQRQVRPEAWLSLGPSPGPGQLTLVCHISGFYKPK
IWMWMMRGEQEQQGTQRSDVLPNADGTWYLRVSLDVEASEASGLSCRVRHSSSLGGQDIIL
YRDHSSMGWIALAVIVTLVLMAGLAFWLKHWTHRESPPSSVLPLE

>bt_CD1A.2 Chain=chain22 Position=3:11537780-11539323(-) GeneID=ENSBTAG000000039366
ProteinID=ENSBTAP000000050624 Score=177.5 E-value=4.2e-51 Length=1544 Comment=Overlaps single
annotation & in proteins
MLFLQIALLLSLLPCGDSSENDPQEPITFKIIRISSFYSPFFAQNLSAWLDELQTHAWDN
NSDRVILYLRPWSKGNFSNEELMDVENLHTFFIRLGOVLHNHASQWQLQYPPFELQIAGGC
EMHIRDASVGFVRIGYQGSDFLSFQKDVVWVSPPEGGISAQFVCTLFNLYRGTQEIHKLL
SDTCPRLFLSLLDAGKAYLQRQVRPEAWLSLGPSPGPGQLMLVCHVSGFYKPKIWMWMMR
GEQEQQGTQRSDVLPNADGTWYLRVSLDVEASEASGLSCRVRHSSSLGGQDIILYWDHSS
TAWIILAMIVPLVLLAGLAFWLKRSWSTYMSDA

>bt_chain23 Chain=chain23 Position=23:27600454-27602353(+) GeneID=ENSBTAG00000008959
ProteinID=ENSBTAP000000035745 Score=176.9 E-value=2.2e-50 Length=1900 Comment=Overlaps single
annotation & in proteins
MWVTELRLLLLLSESLALTKTWAGHIHLGHGDRPKERRPGWQPLPVPVGHSLRYFHIL
VSRPGLGRDLYQSVGYLDDTQFVRYNSDTANPRMEPRAPWMEQEGPQYWDRTQKTIAREHS
QAESNLQVIVGNHNHSHRESHSFLLSGCDVGPPEGCIVRGYEHFAYDGAGYITLNEDEMR
SWSAAYTQIIRRKWEVEGAAEQYAYVETECVDWLRKYLEKKGKETLRAVPPKTHVTRHPIS
DREVTLCWALGFYWAESLWQDGEDQTDQDMEVVETRASGDACHVQHEGLQKPLTLRW
EPPQTPFLTMGIIVGLVLLVAVVAGAVIWRKRSDDGGGSSYTQTAREDR

>bt_chain24 Chain=chain24 Position=3:12187739-12189538(-) GeneID=ENSBTAG000000039993
ProteinID=ENSBTAP000000053071 Score=174.5 E-value=3.9e-50 Length=1800 Comment=Overlaps single
annotation & in proteins
RKWGCWLFLLFWGLPQIWGSLGDPQTPFSFHGLQISSLANSSWTRTDCLGWLGLQPYTW
RNESDTRIRFLKWPVSRGTFSDQWEQLQHTFQVYRSSFYKVLWVFKRLHAEFLSSVPLEI
QGSAGCELLQGTSESFLRAAFQGRDVLVSFGMSWVWVAPDAPPVQVEVCKVINLDQGTKE
TVHLLLDHICPELVRLGTLQTKSELEKQVKPEAWLSSGPPGPHLLLVCHVSGFYKPKV
RVMWMMRGEQEPEGTQGDVMPNADSTWYLRVTLNVAAGEAAGLSCRVKHSSSLGDQDIILY
WDGNRVSRGLVILVLLVFLVGLVFWGGLVFWFRKHRRYQDIP

>bt_chain25 Chain=chain25 Position=3:11881620-11883417(-) GeneID=ENSBTAG000000038426
ProteinID=ENSBTAP000000031077 Score=170.9 E-value=4.4e-49 Length=1798 Comment=Overlaps multiple
annotations; one in proteins
PFVIQLTAGCELHSGEAGISSLRGALGGLDFVSIQNHSCVPAPDSGFRGQKFCALMTQNA
VISYIERLLSETCPQYLLGVLDAGKAEALQRQVKPEAWLSSGPTPGPGRLLLLVCHVSGFYK
PKVVRVMWMMRGEQEPEGTQGDIMPADWTWYLRVTLNVAAGEAAGLNCRVKHSLSLGDQ

>bt_chain26 Chain=chain26 Position=23:27725380-27727018(-) GeneID=ENSBTAG000000010166
ProteinID=ENSBTAP000000013413 Score=168.6 E-value=3.3e-48 Length=1639 Comment=Overlaps single
annotation & in proteins
MGLSHVWVFLAHAFAFFVLLGNAAGSHSLLYNMTVLSRDGFVQSRFFAEGYLDROAFLHYD
HKKGRAEPWRWPEKLAETWETETMDLNEVSWKELRKLAEIISLQEEKGLHLSQETVQ
CDINEDSHPGQFRLLYFNGELLLSCYPEPHGCTLPQSSARTLAMEMELSKHYQAHVQVQEL
CRRLRSYLESWPGFTERTPEPPAVNVTCSDSEGVMVHLTGKTLVHQRSRWIVSIPVAVVFI
IGFCVYCYIKRKRKTASATGRPEPIRLQDLQFQTEPTDHNGLTHPEFQSLCHTPAPSV

>bt_chain27 Chain=chain27 Position=19:19951498-19966398(+) GeneID=ENSBTAG00000007238
ProteinID=ENSBTAP000000009522 Score=165.7 E-value=1.2e-46 Length=14901 Comment=Overlaps single
annotation & in proteins
ARRQRVRYTPGYLLLVTLKLGQMSLAQTTDAHSLCLDLTVKSQSRPGQVQVQGSVD

TKPFLWYSDSDSNKVKPLGFLGKEVNDTKAWTEISQTLVAAGRELRMLVLPVVKLDRKGTGP
PTLVQVKKCCQREAEQCSCASLNFLSLDQQTALLLDTMSITWTVIDPGATGIKEEWENNOEL
AEYFRITISTGDCSYWLWFLKHWEKMLVPEPT

>bt_chain28 Chain=chain28 Position=3:12171328-12173150(-) GeneID=ENSBTAG00000047776
ProteinID=ENSBTAP00000012703 Score=163.1 E-value=8.5e-47 Length=1823 Comment=Overlaps single
annotation & in proteins
VGCWLFLLFFWGLPQIWRSFETPHLSFFFRGLQISSFANRSWTRTDGLAWLGELQPYTLRN
ESNTIGFLKPKWSRGTFFSDQQWEQLQHTFLVYRSSFTRDIWEFVKMLPSDYPLEIQVSGGC
ELLPRNISESFLRAAFQGGKDVLSFQGMSSWVAPDAPPLIEEVIKVLNQDQGTKEVHLL
HDICPELVRGLLQTKSELEKQVKPEAWLSSGSPGPGRLLLVCHVSGFYKPKVVMWMMR
GEQEEPGTRQGDVMPNADSTWYLRVTLVDAAGEAAGLSCQVKHSSLGDQDIIILYWDGNRV
SRGLIVLVLVLFVLLFVGGVLFVFRKHRRYQDIP

>bt_CD1d Chain=chain29 Position=3:12142237-12144049(-) GeneID=ENSBTAG00000024960
ProteinID=ENSBTAP00000033859 Score=162.3 E-value=1.6e-46 Length=1813 Comment=Overlaps single
annotation & in proteins
DSRSESARKWAGCFCSSGGFRRLGGVLRMSGTGLRAERGSRAGKGGDARGGRAETDAP
RKAGSLIHQENLLELLPAFSCCLAPGDRARFRASRHHPPFNLTVCFFLFSFLLPVSLSLPC
PRLRSVFSFAPHMSFPFQFLQISSFNRSWTRTDGLAWLGELQPYTWRNESNTIGFLKPK
WSRGTFFSDQQWEQLQHTLLVYRSSFTRDIWEFVEKLVHVEYPLEIQIATGCELLPRNISES
FFRAAFQGGKDVLSFQGMSSWVAPDAPPFIQEVCKRLNQDQGTKEVHLLHDIWPELVRG
LLQTKSELEKQVKPEAWLSSGSPGPGRLLLVCHVSGFYKPKVVMWMMKGEQEEPGTRQ
GDVMPNVDSTWYLRVTLVDAAGEAAGLSCQVKHSSLGDQDIIILYWDGNHVSRLIVLVLV
LVFVLLFVGGVLFVFRKHRRYQDIP

>bt_chain30 Chain=chain30 Position=3:12111049-12112873(-) GeneID=ENSBTAG00000047988
ProteinID=ENSBTAP00000051276 Score=162.2 E-value=1.5e-46 Length=1825 Comment=Overlaps single
annotation & in proteins
VGCLLFLLFFWGLPQIWRSFETPQTPFSFQGLQISSFANRSWTRTDGLAWLGELQPYTWRN
ESDTIRFLKPKWSRGTFFSDQQWEQLQHTLLVYRSSFTRDIWEFVEKLVHVEYPLEIQIATGC
ELLPRNISESFLRAAFQGGKDVLSFQGMSSWVAPDAPPFIQEVIKVLNQDQGTKEVHLLH
DIWPELVRGVLTQTKSELEKQVKPEAWLSSGSPGPGRLLLVCHVSGFYKPKVVMWMMR
GEQEEPGTRQGDVMPNADSTWYLRVTLVDAAGEAAGLSCQVKHSSLGDQDIIILYWDGNHVS
SRGLIVLVLVLFVLLFVGGVLFVFRKHRRYQDIP

>bt_chain31 Chain=chain31 Position=3:12036739-12038669(-) GeneID=None ProteinID=None
Score=158.2 E-value=2.7e-45 Length=1931 Comment=No overlapping annotations
DFQEPVSVFEVICVLSFNHSSWVQSLGSGWLGELQTHCWKSNPGTIIYLWPKSGNFSNEE
LMEQLQDPPFVQIAKGCCLQAGEAPVGFMRVAYQGSDFLSPQNSWSSPEGGRAQVLR
RLFNLFQQTQLIIHKLLSDTCHFLGLLDARKAYLQROGKLTLVCHVSGFYKPKPI*VMW
MRGEQEQGTQRSVLPNSDGTSYLRVSLDVEAIEASGLSCRVRHSSLGGQHTILY

>bt_chain32 Chain=chain32 Position=3:11853516-11859752(-) GeneID=None ProteinID=None
Score=131.9 E-value=1.8e-35 Length=6237 Comment=No overlapping annotations
VFHWPTSYHIIHIWTFANSSWAQNGSGWLDLQIQGWSDLDGTTIFLKPWSKGDPPFVIQ
DIAGCELHSGEAISSLRGALGELDFVRIQNHSCVPAPDSGSRGQKFCALTTQYPGISDI
IERLLSETCLRYLLGVLDAGKAELOKQWLLSSGPTPGPGCLLLVCHVSGFYKPKVVMWMMR
GEQEQPGTQQGDIMPAD*TWYLRVTLDVATGGAAGLSC*MKHSSLGDQDIIILYWDGNHVS
LLHCR

>bt_chain33 Chain=chain33 Position=23:28393697-28398497(-) GeneID=None ProteinID=None
Score=125.2 E-value=1.2e-33 Length=4801 Comment=No overlapping annotations
VDYTPQLFQFSDDDPNLRMEARALWMEQEGPEYWDWNMQGIKDTAQTRFRVNLNSLWGYNQ
SKVGTAAADTAAQITKPK*EASGDAQCYRNYLEGTCVEQFHRFQENRIHVYTHPISDLEVT
LRFWALGFYPIELLTWQHDGEDLTQDMELVETRPSGDGTFQKVVVVVPSGEEQRYTCRV
QHEGLQEPLTLPSSDQAQGCDDVTFHQ*

>bt_chain34 Chain=chain34 Position=23:27661963-27663563(-) GeneID=ENSBTAG00000010166
ProteinID=ENSBTAP00000013413 Score=124.6 E-value=1.9e-35 Length=1601 Comment=Overlaps single
annotation & in proteins
PSPGSHSL*YINITVLSQDGSVQASSFAEGFLDRQTFLLHYDHKGAGLHSLQETVGCNINED
SHPQGFRLLYFNGLLELLSCYPEPHGCTLPQSSARTLAMEMELSKHYQAYVQGELCWRLRS
YLESWTGFTEERTVNVTHSQDSEGMVHLTCKAFGFPRISISVVWFRNEEPMWVNDQESGDV
LPDNGNTYYTWETVKIPQGEQVVKCIVEHSGNHSSHLAP

>bt_chain35 Chain=chain35 Position=3:11834222-11835843(-) GeneID=None ProteinID=None
Score=105.8 E-value=1.2e-29 Length=1622 Comment=No overlapping annotations
SNQGLGWLDLQHLGWESDSTGDIIFLKPWSKSNFSDEEMTKLEDLF*VYFIFGIEVQNH
INEDPFVIOVIESCGLHSGKAIGSSLRGAIGGLDFVRIQNHSCVAPDSGSRGQKFCARM
TQYPGISNIIERLMSSETCPQYWLSSGPIPGGRLLLVCHVSGFYKPKVVMWVVRGEQEP
GTQQGDIMLNADWTWYLLVTLVDAAGEAAGLSC*VKHSSLGDQDIIIL

>bt_chain36 Chain=chain36 Position=3:11885588-11891650(-) GeneID=ENSBTAG00000039189
ProteinID=ENSBTAP00000051860 Score=102.6 E-value=2e-26 Length=6063 Comment=Overlaps single
annotation & in proteins
VLVFLYTDPPFVIOVIAQCELHSGEAISSLRGALGGLDFVRIQNHSCVPTRDSGSRGQK
CLMTQYQGISDIMERLHSETCPRYLLGVNLARKAELOKQVKPEAWLSSGPTPGPGCLLLV

CHVSGFYPKPVRVMWMRGEQEOPGTQQGDIMPNDWTWYLRVTLQDGVAGVKHSSLGDQDI
 ILYWGEKQAIGRYFVAIIVPSLILSIGLALWFWRCW

>bt_chain37 Chain=chain37 Position=3:12025126-12026824(-) GeneID=None ProteinID=None Score=98.0
 E-value=1.9e-27 Length=1699 Comment=No overlapping annotations
 GFQKPI SF*YIWIL*FYNHPWVQNLGSA*LGDLQTHGWKSSSHTVIYLWPWSKGNFNNEE
 LMELQSVSHRRFTIVLQVFNHNNANQWQTQEIHKLLNDICPCFLLGLLEAGKAYLQR*GHL
 MLVCHASGFYPKPIWVIWMRGEQE*GTQKSDVLAADGTWYLRVSLDVEASEASGLSCR
 VRHSSLGGQDIIL

>bt_chain38 Chain=chain38 Position=3:11981092-11982777(-) GeneID=None ProteinID=None Score=96.0
 E-value=9.2e-27 Length=1686 Comment=No overlapping annotations
 GFPEPISF*YIWIL*FYNHPWVQNLGSA*LGDLQTHGWKSSSHTVIYLWPWSKGNFNNEE
 LMELQSVSHRRFTIVLQVFNHNNANQWQTQEIHKLLNDICPCFLLGLLEAGKAYLQR*GHL
 GSELLSFQTHGHLVCHASGFYPKPLWVMWIVGQE*GTQKSDVLAADGTWYLRVSLD
 VEASEESGLCCQVRHSSLVQDIIL

>bt_chain39 Chain=chain39 Position=3:11794354-11796050(-) GeneID=None ProteinID=None Score=92.8
 E-value=8e-26 Length=1697 Comment=No overlapping annotations
 GFQEPISF*YIWIL*FYNHPWVQNLGSA*LGDLQTHGWKSSSHTVI*YLPWSKGNFNNEE
 LMELQSVSHRRFTITQEIHKLLNDICPCFLLGLLEAGKAYLQR*GHLMLVCHTSGFYPKP
 IWVMMWGEQE*GTQKSDVLAADGTWYLRVSLDVEASEASGLSCWVRHSSLGGQNIIL

>bt_ENSBTAG00000039444 Chain=None Position=9:85813503-85814535(+) GeneID=ENSBTAG00000039444
 ProteinID=ENSBTAP00000048357 Score=169.1 E-value=2.1e-49 Length=0 Comment=Not detected in
 genome search
 LLSQNWEKRPMLCLLPTSPSLQFAFTDAHSLSFNLTVDPPQPRGQPCCEVQGVQDGEVLL
 SYDCCHAKIIFTSPLEGEAEWAQIEILRDIRDQLRDFTEKHTITDPLTLQARMTCCRED
 DRHVSRSWFLGNGQMRHLFDSENGHWVVDHPGGRWMEKKEWENDRAVTDFLKVKVSMGDCR
 AWLQDFMACWEKMLKTTECDPSEVSCPRAMTRGCSVILLPFLPPDI

>bt_ENSBTAG00000047902 Chain=None Position=9:88273909-88299080(-) GeneID=ENSBTAG00000047902
 ProteinID=ENSBTAP00000053867 Score=199.0 E-value=2.9e-58 Length=0 Comment=Not detected in
 genome search
 MDRKAGPGARLGFQAQVLLVALRFCTARGGPPCAVSQRRTLCNAHSLCYNFTVYVPHSPG
 EPWCVVQGVQDGNVFLSYHCGGTKIQSTSPLEGEVKTNTTWETQTELRDIGNFLKQQLP
 DI IPEKHTARDPLTLQGRMTCRCEEDGHISGSWQFGFNGEMCLRFDSENGHWITDHSRGR
 QI KEKWENDRAVTDFFKVKVSMGDCRAWLQDFMVCWEKMLKTSASPTTGPPVQMPAPPIK
 SKPWILPVVLTSTFIITFPRLSPFQEQVMLLPSPNRCVGLRTOQLLGLP

>bt_ENSBTAG00000048114 Chain=None Position=9:88632916-88639753(+) GeneID=ENSBTAG00000048114
 ProteinID=ENSBTAP00000054263 Score=168.1 E-value=4.2e-49 Length=0 Comment=Not detected in
 genome search
 HSLCLDLTVKSSQSRPGQPWCQVQGSVDTKPFLQYDSASNKVKPLGFLGKEVNDTKAWTEI
 SQTLEAGKELRMVLPVTKLDENETRGPPTLQVLCQCREAQCSGASLHFSNLGRNTALL
 LDTMSITWTVIDPGATGIKEEWENQELAEYFRTISTGDCSYWLEFLKHWEKMLVPEPT
 ESLIMAADISQSASIRLDSIILLIITQLVLIASS

>bt_ENSBTAG00000024751 Chain=None Position=9:88249142-88250182(+) GeneID=ENSBTAG00000024751
 ProteinID=ENSBTAP00000016352 Score=179.3 E-value=2e-52 Length=0 Comment=Not detected in genome
 search
 VVLGARGAFLLYQPSAALLGSHISGGRDTHSLFYNTVDRHPSGPEPWCVVQGVQIDGN
 FFLSYDCGGTKIQSTSPLEGEVKTNTTWETQTEPLRNIGDFLKGQLPDI IPEKHARVES
 SPLTLQARMTQCCEEDGHISGSWQFGFNGEMCLRFDSENGHWTEVHSGRRMKEKWEKDR
 AVTDFKKNVSMGDCQAWLQDFIVCWEKML

>bt_ENSBTAG00000038366 Chain=None Position=3:11883112-11883708(-) GeneID=ENSBTAG00000038366
 ProteinID=ENSBTAP00000052888 Score=57.8 E-value=1.9e-16 Length=0 Comment=Not detected in genome
 search
 LGVILPGGDNADVFGQPTSFHLMQISTFVNNTCAQNQGGWLLDQLIHGLESDLGTAIFL
 KPWSKGNFSDDEVTELDLFRAYLIGFTCGMQDRVNEFQLECEYSPLIWKDSQ

>bt_ULBP27 Chain=None Position=19:19833725-19875946(+) GeneID=ENSBTAG00000024490
 ProteinID=ENSBTAP00000033925 Score=179.2 E-value=2.1e-52 Length=0 Comment=Not detected in
 genome search
 MGGSKSSLDLFLVLLIVLFSGTSSDAHSLSYNVTIDPRPRDGPWCVEVQGEVDQKVFLSY
 DCGRAKIKSISPLGEEVKSMAWETQINTLKDGTGDLLEKQMPDVTPEKHIDKGPLTLQAR
 MTCWREDNGHTSASWEFGFNGQLCLLFDSENGYWTMVHSGRRMKEKWEKENDRAVMDFFKK
 VSMGDCQRWYHTFLVRWEKMLKTAASPTAFLRVNSTATATKHVTCILPVLLSSFIITVF
 LG

>bt_ENSBTAG00000038891 Chain=None Position=9:88244953-88245726(-) GeneID=ENSBTAG00000038891
 ProteinID=ENSBTAP00000048618 Score=197.5 E-value=7.9e-58 Length=0 Comment=Not detected in
 genome search
 TDAHSLCYNFTVDSQPRPGQPWCVVQGVQDGVKFLSYDCGRAKIQFTSPLEGEVKTTKAW
 ETQTELRDIGNLREQLPVDVTSEKRTVTDPLTLQGRMTCRCEEDGHISGSWQFGFSGQM
 CLLFDSENGHWTEVHSGRRMREKWEKENDRAVNTFFKVKVSMGDCRVWLQDFLVHWEEMLKT

>bt_PROCR Chain=None Position=13:65052810-65106553(+) GeneID=ENSBTAG00000008291
ProteinID=ENSBTAP00000010911 Score=90.5 E-value=3.8e-26 Length=0 Comment=Not detected in genome search
MLTTLLPPLLPLLLPGWALCSQEASDGPWDLHMTQVSYFRNPSQVWHRGNATLGGVLTHTVL
EGPGRNVSIQQLQPLQEPDSWALTKIYLNRYLEEFVGLVQVVHQERGVTPLIIRCFGLG
ELPPEGSEARVFFVAVNGSSFNFPKKTASWVAEPHAPSRVVITYTDQLNKYNRTRVEL
REFLQDTCVQYIQKHITNNLKGSTGRSYTSLVLGVLVGCFTVTVGAVGIFLCTGRRR
C

>bt_ULBP3 Chain=None Position=9:85948859-85961158(-) GeneID=ENSBTAG00000038112
ProteinID=ENSBTAP00000050046 Score=185.3 E-value=3.4e-54 Length=0 Comment=Not detected in genome search
MGGSKTSLGLFVLLPVLVFSRTSSDAHSLSYNVTIDPRPRDGPWCEVQGEVDQKVFLSY
DCGRAKIMYMSPLGKVKSMNAWETQDTLRLDIGDLLKEQMPDVTPEKHIDEVSGLLTL
QARMACWREDNGHTSASWQFGFNGQTCLLFDLENGHWTMVHSGRRMKEKENDRAVTEF
FKKVSMDGCQRWFQDFLLRWEKMLKTASPTTGPSTMQPMAPDSSYIAWIATGVLGAFVM
TIIIIACIRYKRRKLCSEAPDRCSVGLRTQSLGFCFSPAVYFRAKRSDLRNPKSVYQL

>bt_ENSBTAG00000039875 Chain=None Position=9:88307676-88310265(-) GeneID=ENSBTAG00000039875
ProteinID=ENSBTAP00000045991 Score=151.3 E-value=3.8e-44 Length=0 Comment=Not detected in genome search
QGRMPLAQTGHLILLILLIEARKTPGNAHSLCQDLTVKSQSRPGLHMCQVQGSVDTKP
FLWYDSDSNKVKPLGFLGKEVNDTKAWTEISQTLVAAGRELRMVLPVIKLDKRGTPPTL
QVKLCCQREAEQCSGASHFSLNGQTALLDTSITWTVIDPGATGKEEWENNELAEYF
RTISTGDCSYLWELFKHWEKMLLPEPTEPIIMAPDISQSASIRLVSCIILLIITQLVLI

>bt_ENSBTAG00000001836 Chain=None Position=19:19937602-19947884(-) GeneID=ENSBTAG00000001836
ProteinID=ENSBTAP00000002400 Score=185.5 E-value=2.9e-54 Length=0 Comment=Not detected in genome search
MGGSESSLGLFVLLLVIVLFSGTSSDAHSLSYNVTIDPRPRDGPWCEVQGEVDQKVFLSY
DCGRAKIKYMSPLGEEVKSMNAWETQDTLTKDIGDLLKEQMPGVTPEKHIDKVGSGPLTL
QARMTCWREDNGHTSASWQFGFNGQLCLLFDSENGHWTMVHSGRRMKEKENDRDVDF
FKKVSMDGCQHWYQDFLLSWKMLKTASLTTPSTMQPMAPDSSYIAWIATGVLGAFVI
IIVLACIHYKNR

>bt_ULBP3.02 Chain=None Position=17:39881230-39884057(+) GeneID=ENSBTAG00000026437
ProteinID=ENSBTAP00000037395 Score=167.3 E-value=7.1e-49 Length=0 Comment=Not detected in genome search
TDAHSLSYNITIDPRPRDGPWCEVQGEVDQKVFLSYDCGRAKIKYMSPLGEEVKSMNAW
EAQTDLRLDIGDLLKEQMTDVTPEKHLKDVSGSPLTLQARMSCWQEDNGLISASWQFGFN
GQLCLLFDSESGYWTMVHSGRWMKEKENDRAVTEFFKKVSRADCQRWYQDFLLHWEKML
LKTASPTTGPSTMQPMAPDSSYSTWIATGVLGAFVITIIIIACIHYKRRRLCSQEARDR
CSVGLRTQSLGSCCPAVHLRAKRSDLRNPKSVYQ

>bt_ENSBTAG00000023970 Chain=None Position=19:19919500-19921026(+) GeneID=ENSBTAG00000023970
ProteinID=ENSBTAP00000011321 Score=179.3 E-value=2e-52 Length=0 Comment=Not detected in genome search
LGEASGSFSPSTPSPLCFAFTDAHSLSYNVTIDPRPRDGPWCEVQGEVDQKVFLSYDCG
RAKIKSISPLGEEVKSMNAWETQINTLKDIDGDLLEQMPDVTPEKHIDKGPLTLQARMT
WREDNGHTSASWQFGFNGQLCLLFDSENGYWTMVHSGRWMKEKENDRAVTEFFKKVSM
GDCQRWYHTFLLSWEKMLKTAGKASPTAPLRVNSTATATKHVTCILPVL

>bt_ENSBTAG00000036061 Chain=None Position=9:88262868-88266693(+) GeneID=ENSBTAG00000036061
ProteinID=ENSBTAP00000025021 Score=171.3 E-value=4.5e-50 Length=0 Comment=Not detected in genome search
GSPCAMNQHCTLCNAHSLSYNVTIDPRPRDGPWCEVQGEVDQKVFLSYDCGRAKIMYMS
PLGEEVKSMNSWETQDTLRLDIGDLLKEQMPDVTPEKHIDEGLTLQARMTCWREDNGHT
SASWQFGFNGQLCLLFDSENGYWTMVHSGRRMKEKENDRAVMDFFKKVSMGDCQRWYH
TSLVRWEKMLKTAASPTAPLRVNSTATATKHVTCILPVLSSFIIIVFLG

>bt_ENSBTAG00000046595 Chain=None Position=19:19885026-19898292(-) GeneID=ENSBTAG00000046595
ProteinID=ENSBTAP00000056117 Score=190.5 E-value=9.8e-56 Length=0 Comment=Not detected in genome search
DTHSLFYNFVDPHPSPGEPWCVVQGVQVDFNFFLSYDCGGTKIQSISPLGEEVKTNAWE
TQTELRDIRDFLKGQLPDIPEKHTARGPLTLQARMTQCEEDGHISGSWQFGFNGEMC
LRFDSNGHWTVEVHSGRRMKEKENDRAVDFPKVSMGDCQTCQLDLIVCWEKTLKTS
ASPTTVPHTVQPTAFTSNHITKIIILGVLGAFIIISIVVWIIHKNSRRRCSQEAPDRCSV
LRTQSLGFCFSPFTLEPRDQTLGIPSLSTSYDDTVAAPSRVSCHI

>bt_RAET1G Chain=None Position=9:88232044-88402262(-) GeneID=ENSBTAG00000039329
ProteinID=ENSBTAP00000052134 Score=176.2 E-value=1.7e-51 Length=0 Comment=Not detected in genome search
MAKGERGGPTRLGFLDLLIVWFSGTPGDAHSLSFDFTVDPQPRPGHPWCEIQSQVDGK
VFLSYDCGHAKIIIPSVLREEVKTINAWETQIETLRDIRDWIKDHMHDFTEKHMPPDR
TLQARMTCRCEDDRHVSGSQFGLNGVMILHFDSENGHWRVDPGGRRMKEKENDRAV
DFLKKVSMGDCRGWLQDFMVRWKEILKTTASPTTVPPTVQPTAPPISHVTWIAPGVLVSF
VIMGIVAWIYKRRRLCSQEAPDRCSVGLRTQCLLGCFCSPAFTLEPRDQTLGVSLSLST
YDDTVAAPSRVSCQI

>gg_Bf1 Chain=chain01 Position=16:74245-75385(-) GeneID=ENSGALG0000000178
ProteinID=ENSGALP0000000233 Score=334.1 E-value=8.3e-96 Length=1141 Comment=Overlaps single
annotation & in proteins
MRPCGAVLGLLRLGLLLAAVCGAAELHSLRYIHTAMTDPGPGQPWYVDVGYVDGELFV
HYNSTARRYVPRTWEAAKADQQYWDGQTOIGQRNERSVKVSLDLTQERYNQGGSHTVQ
WMFGCDILEDTIRGYRQVAYDGKDFIAFDKDMKTFATAVPEAVPTKRKWEEGGVAEGWK
SYLEETCVCWEVLRVVEYGAELGRRERPEVRVWGKEADGILTLSCRAHGFYPRPIVVSWL
KDGAVRQDAQSGGIMPNGDGTYYHTWVTIDAQPGDGDKYQCRVEHASLPQGLYSWEPQ
PNLVPIVAGVAVAIAIVVGVGFIIYRRHAGKKGKGYNIAPDREGSSSSSTGSNPSI

>gg_BfIV21 Chain=chain02 Position=16:59825-60964(+) GeneID=ENSGALG00000024372
ProteinID=ENSGALP00000042920 Score=333.6 E-value=1.2e-95 Length=1140 Comment=Overlaps single
annotation & in proteins
MGSCGALGLGLLLAAVCGAAELHSLRYIRTAMTDPGPGPLPWFVDVGYVDGELFMHYNST
ARRAVPRTEWIAANTDQYWDRETQIVQGSEQINRENLDILRRRYNQGGSHTVQWMSGC
DILEDGTIRGYHQAAAYDGRDFVAFDGTMTLTAAVPEAVPTKRKWEEGGYAEGLKQYLEE
TCVWEVLRVVEYGAELGRRERPEVRVWGKEADGILTLSCRAHGFYPRPIVVSWLKDGAV
RGQDAQSGGIVPNDGTYYHTWVTIDAQPGDGDKYQCRVEHASLPQGLYSWEPQPNLVPI
IVAGVAVAIAIAIVVGVGFIIYRRHAGKKGKGYNIAPDREGSSSSSTGSNPSI

>gg_chain03 Chain=chain03 Position=JH375207.1:7801-8953(-) GeneID=ENSGALG00000024340
ProteinID=ENSGALP00000009984 Score=330.2 E-value=1.3e-94 Length=1153 Comment=Overlaps single
annotation & in proteins
MGPSEVVVLLGLLGLGAAVCGSLHSLRYFLTGMDTPGPGMPFRFVIVGYVDDKIFGIYDYSK
SRTAQPIVEMLPQEDREHWAQTQKAQGGGERDFDRGLRGLPERYKSKGSHTLQMMFGCD
ILEDGSIIRGYDQYAFDGRDYIAFDMDTMTFTAADPVAEITKRKWEETEGTYAERWKHELGT
VCVQNLRRYLEHGKAALKRRVQPEVRVWRKEANGILTLSCAHYGFYPRPIAISWMDKGMV
RDQETHWGGVVPNSDGTYYHTWVTIDAQPGDGDKYQCRVEHASLPQGLYSWEPQPNLPI
VAVAVVAIAVAIAAVVGLVVKSKSGMEKKDYEAAPGHDEVSSSSATGSELSI

>gg_chain04 Chain=chain04 Position=AADN03017514.1:1388-2540(-) GeneID=ENSGALG00000026292
ProteinID=ENSGALP00000023473 Score=329.5 E-value=2.1e-94 Length=1153 Comment=Overlaps single
annotation & in proteins
MGPSEVVVLLGLLGLLQLLVGLGAAVCGSHSLRYFLTGMDTPGPGMPFRFVIVGYVDDKIF
GIYDYSKSRTAQPIVEMLPQEDQEHWAQTQKAQGGGERDFDWFLSRLPERYKSKGSHTMQ
MMIGCDILEDGSIIRGYDQYAFDGRDFLAFDMDTMTFTAADPVAEITKRKWEETEGTYAERW
KHELGTVCVQNLRRYLEHGKAALKRRVQPEVRVWRKEANGILTLSCAHYGFYPRPIAISW
MKDSMVQDQETHWGGVVPNSDGTYYHTSAIDVLPEDGDKYRRCRVEHASLPQGLFSWEPQ
PNLPIEAWLVVPLVVLVVALIALLVWKFLSG

>gg_chain05 Chain=chain05 Position=JH375207.1:19904-21056(-) GeneID=ENSGALG00000027445
ProteinID=ENSGALP00000000235 Score=326.5 E-value=1.5e-93 Length=1153 Comment=Overlaps single
annotation & in proteins
MGPSEVVVLLGLLGLLQLLVGLGAAVCGSHSLRYFLTGMDTPGPGMPFRFVIVGYVDDKIF
GIYDYSKSRTAQPIVEMLPQEDQEHWAQTQKTQGGGERDFDWFLSRLPERYKSKGSHTMQ
MMIGCDILEDGSIIRGYDQYAFDGRDFLAFDMDTMTFTAADPVAEITKRKWEETEGTYAERW
KHELGTVCVQNLRRYLEHGKAALKRRVQPEVRVWRKEANGILTLSCAHYGFYPRPIAISW
MKDSMVQDQETHWGGVVPNSDGTYYHTSAIDVLPEDGDKYRRCRVEHASLPQGLFSWEPQ
PNLPIEAWLVVPLVVLVVALIALLVWKFLSG

>gg_chain06 Chain=chain06 Position=JH375196.1:7436-8588(-) GeneID=ENSGALG00000029093
ProteinID=ENSGALP00000000459 Score=325.6 E-value=3.4e-93 Length=1153 Comment=Overlaps single
annotation & in proteins
MGRGGVWGARDTVRCSAVGCRAMGPSEAVVLLGLLGAECGSHSLRYFMTGMDTPGPGM
PQFVIVGYVDGELFGKYDSKSRWVHPIVEKLPQEDQEHWDQTLKAREGELEFSEGLHRL
QVRYNRSRGGSHTLQKMPFGCDILEDGSIIRGYDQYAFDGRDYIAFDMDTMTFTAADPVAEIT
KRKWEETEGTYAERWKHELGTVCVQNLRRYLEHGKAALKRRVLPPEVRVWGKEANGILTLFC
RAYGFYWPISLSWMDKGMVRDQETHWGGVVPNSDGTYYHTSAIDVLPEDGDKYRCHVEH
ASLPQGLFLWEPQPNLIPSGAGAVIAIVAVIAAVVGLVVKSKSDKRELGGRRKGLRS
STRPRRWIQRGLHRYSVELGLRRGPCGWSGSSSLRSPVGPAGAILGPALPGEPPGW

>gg_chain07 Chain=chain07 Position=AADN03013240.1:123-1275(-) GeneID=ENSGALG00000027293
ProteinID=ENSGALP00000021123 Score=325.6 E-value=3.4e-93 Length=1153 Comment=Overlaps single
annotation & in proteins
MTRRCVGSAAAGVVPRTPRPLPPEFLISILTALASEIWLAPGDVTRAHCRHWPWGAE
PMGTRGGVWGARDTVRCSAVGCRAMGPSEAVVLLGLLGAECGSHSLRYFMTGMDTPGPG
MPQFVIVGYVDGELFGKYDSKSRWVHPIVEKLPQEDQEHWDQTLKAREGELEFSEGLHR
LQVRYNRSRGGSHTLQKMPFGCDILEDGSIIRGYDQYAFDGRDYIAFDMDTMTFTAADPVAEIT
KRKWEETEGTYAERWKHELGTVCVQNLRRYLEHGKAALKRRVLPPEVRVWGKEANGILTLF
CRAYGFYWPISLSWMDKGMVRDQETHWGGVVPNSDGTYYHTSAIDVLPEDGDKYRCHVE
HASLPQGLFLWEPGSGVGRMRLEGWAAPSPDMPPLSSRATAQ

>gg_chain08 Chain=chain08 Position=16:393075-394227(-) GeneID=ENSGALG00000027955
ProteinID=ENSGALP00000040364 Score=324.4 E-value=6.3e-93 Length=1153 Comment=Overlaps single
annotation & in proteins
MPRFVIVGYVDDKIFGIYDYSKSRTAQPIVEMLPQEDQEHWAQTQKAQGGGERDFDWFLSR
LPERYKSKGSHTMQMMIGCDILEDGSIIRGYDQYAFDGRDFLAFDMDTMTFTAADPVAEIT

TKRRWEEEGTYAERWKHELGTVCVQNLRRYLEHGKAAVKRRERPKVRVWGKEADGILTLS
CHAHGFYPRPIAISWMKDMVRDQETRWGGIVPNRDGTYHASAAIDVLPEDRDNRYRCRVE
HASLPQPLFSWEPPLNLIIPGGAGAVVTIVFVIAAVVGLVWKRKSGKEKKDYEAAGHD
GESSGSATALVWNGSGGVPVVGAVPVCAPPLGPRPGQYWAQPSLENPQGEGSGCGCGP
T

>gg_chain09 Chain=chain09 Position=16:366610-367762(+) GeneID=ENSGALG00000024348
ProteinID=ENSGALP00000021998 Score=323.2 E-value=1.6e-92 Length=1153 Comment=Overlaps single
annotation & in proteins
GSHSLRYFLTGMTDPGPGMPRFVAVGYVDDKIFGKYDSKSRVWHPIVEMLPQEDQEHWDI
QTQMAREGELEFSEGLHRLQVRYNRSGGSHTLQKMFCDILEDGSIRGYDQYAFDGRDFL
AFDMYTMTFTAADPVAEITKRRWETGGTYAERWKHELGTVCVQNLRRYLEHGKAAKRRV
QPEVRVWGKEADGILTLSCHAHGFYPRPIAISWMKDMVRDQETHWGGVVPNSDGTYHTS
AAIDVLPEDGDKYCRVEHASLPQPLFSWEPQPNLIPIVAGAVVAIVAVIAAVVGLVWV
KRKSGKEKKGYEAAGHDVSSGSATGSEPSI

>gg_chain10 Chain=chain10 Position=AADN03010737.1:934-2072(-) GeneID=ENSGALG00000028268
ProteinID=ENSGALP00000017922 Score=322.4 E-value=2.8e-92 Length=1139 Comment=Overlaps single
annotation & in proteins
MGPSKVVVGLLLGALGAEACWPHSLRYFVTGMTDPGPGMPRFVIVGYVDGDLFGKYDSK
IKSAQPIVEMLPQEDQEHWAQTQKARGGERDFDWFSLRPERYNKSGGSHTLQRMIGCD
ILADGSIKRGDKYAFDGRDYIAFDMDTMTFTAADPVAEITKRRWETEGTYAERWKHELGT
VCVQNLRRYLEHGKAAKRRRERPEVRVWGKEANGNLTLSCAHGFYPRPIAISWMKDMV
GDQETHCGGVVPNSDGTYHASAVINVLPKDGDYCRVEHASLPQPSLFLWEPQPNLIPI
VAGAVVAIVAVIAVVVGLVWVWKSKEKKGYEAAPGHGESSISATGSEPSI

>gg_chain11 Chain=chain11 Position=JH375206.1:4911-6063(-) GeneID=None ProteinID=None
Score=321.2 E-value=5.5e-92 Length=1153 Comment=No overlapping annotations
VPAGSHSLRYFLTGMTDPGPGMPRFVIVGYVDGDLFGKYDSKSRVWHPIVEKLPQEDREH
WAAQTEKARGGERDFDRGLGSHTLQKMFCDILEDGSIRGYDQYAFDGRDFLAFDMDTMM
FTAADPVAEITKRWETEGTYAERWKHELGTVCVQNLRRYLEHGKAAKRRKVRVWGKEA
DGILTLSCAHGFYPRPIAISWMKDMVRDQETQWGGVVPNSDGTYHASAAIDVLPEDGD
KYCRVEHASLPQPLF

>gg_chain12 Chain=chain12 Position=AADN03013018.1:2694-3846(-) GeneID=ENSGALG00000027944
ProteinID=ENSGALP00000042327 Score=321.2 E-value=5.5e-92 Length=1153 Comment=Overlaps single
annotation & in proteins
MTRGGARGVRCGAVRCGAVRCGAVPCGAVPCRAVPCHAMGPCKAVLLGLLGALGA
AACGSHSLRYFLTGMTDPGPGMPRFVIVGYVDGDLFGKYDSKSRVWHPIVEKLPQEDREH
WAAQTEKARGGERDFDRGLGSLPERYNKSGGSHTLQKMFCDILEDGSIRGYDQYAFDGR
DFLAFDMDTMMFTAADPVAEITKRWETEGTYAERWKHELGTVCVQNLRRYLEHGKAAK
RRGEDGRAAWGWAGCGAGAQCGLVSPAHNVTNLQSGPRCECGRRPTGS

>gg_chain13 Chain=chain13 Position=JH375203.1:5472-6634(-) GeneID=ENSGALG00000029079
ProteinID=ENSGALP00000041779 Score=314.5 E-value=5.5e-90 Length=1163 Comment=Overlaps single
annotation & in proteins
MGPSEAVVGLLLGSLGAAVCGSHSLHYFLTGMTDPGPGMPQFVIVGCVDGELLWNYSSL
GRTVRPIMGWLPQEDQEHWAETTKKARDVELDFYEFGLRQLQVHYNKSGGSHTLQKMFCD
ILEDGSIRGYDQYAFDGRDFLAFDMDTMTFTAADPVAEITKRRWEEEGTYAERWKHELGT
VCVQNLRRYLEHGKAAKRRGEDGRGTWGWAGYEAGAQCGLLSPVHNVTHLQSGPRCEG
GRRPTGS

>gg_chain14 Chain=chain14 Position=JH375199.1:13104-14256(-) GeneID=ENSGALG00000026145
ProteinID=ENSGALP00000042857 Score=313.3 E-value=1.5e-89 Length=1153 Comment=Overlaps single
annotation & in proteins
MTRGGVWARGAAASSAMGPSEAVVGLLLGALGAAECGSHSLRYFMTGMTDPGPGMPR
FVIVGYVDGELFGKYDSKSRVWHPIVEKLPQEDREHWDGQSQKARDGELEFSEGLHRLQV
RYNRSGEHCRSHSAMELGQELCGAGASA

>gg_chain15 Chain=chain15 Position=AADN03010643.1:2948-4100(+) GeneID=ENSGALG0000002614
ProteinID=ENSGALP0000004106 Score=308.3 E-value=4e-88 Length=1153 Comment=Overlaps single
annotation & in proteins
MTRMTDPGPGMPQFVIVGYVDGELFGKYDSKSRVWHPIVEKLPQEDREHWAQTQKAREG
ELEFSEGLHRLQVRYNRSGGSHTLQKMFCDILEDGSIRGYDQAFDGRDYIAFDMDTMT
FTVADPVAEISKRRREEEGTYAERWKHELGTVCVQNLRRYLEHGKAAKRRMQPEVRVWR
KEADGILTLSCAHGFYPRPIAISWIKDMVRDQETHWGGVVPNSDGTYHASAAIDVLP
DVKYCRVEHTSLPQPLFLWEPQPNLIPIVAGAVITIVAVIAAVIGLVVWKSKEKSR
ELGGKAVVAGSPLPQ

>gg_chain16 Chain=chain16 Position=JH375199.1:26158-27320(-) GeneID=ENSGALG00000027693
ProteinID=ENSGALP00000040357 Score=295.0 E-value=3.4e-84 Length=1163 Comment=Overlaps single
annotation & in proteins
MTDPPGPGMPQFVIVGYVDGELLWNYNSLCRTVVRPIMGWLPQEDQEHWAETTKKARDVELD
FYEFGLRQLQVHYNKSGGSHTLQKMFCDILEDGSIRGYDQYAFDGRDFLAFDMDKMTFTA
ADAVAEISKRRWEEEGTYAERWKHELGTICVQNLRRYLEHGKAVLKRREPEVRVWGKEA
NGILTLSCRAYGFYPRPIAISWMKDMVRDQETHWGGIVPNSDGTYHASAAIDVLPEDGD
KYRRCRVEHASLPQPLFSWEPQPNLIPIVAVVAVVAVIASVIGLVVWKSKEKGPDL
RIGPDMSPMQTLILSAGKEKKKSYEAAGHDRESSVSAMVIV

>gg_chain17 Chain=chain17 Position=16:406868-408117(+) GeneID=ENSGALG00000026466
ProteinID=ENSGALP00000040545 Score=272.3 E-value=1.9e-77 Length=1250 Comment=Overlaps single
annotation & in proteins
MGPSEAVLLGALLGALAAACGSHSLRYFKTRMTDPGPGMPQFVIVGCVGELLWNYNSL
SRTVVRPIMGWLPPIILSPPTSPQPHSTPGAPQPMKPLTPHPCCASGSHTLQKMGICDILE
DGSIRGYVQYAFDGRDFLAFDMDTMTFTAADPVAEITKRRWEEEGIYTEGCKHELGTICV
QNLRRYLEHGKAVLKRREPERVVRWGKEANGILTLSCRAYGFYPRPIAISWMKDMVRDQ
ETHWGGIVPNSDGYHASAAIDVLPKDVDKYWCRVEHTSLPQGLFSWELQPNLTPSVAG
AVGAIVAVIAAVVGVVVVWKS KSGKEKKGYEAAAGHDGESSV

>gg_chain18 Chain=chain18 Position=JH375197.1:5985-7051(+) GeneID=ENSGALG00000004617
ProteinID=ENSGALP00000007338 Score=268.2 E-value=1.3e-75 Length=1067 Comment=Overlaps single
annotation & in proteins
MGPSEAVLLGALLGALAAACGSHSLRYFKTRMTDPGPGMPQFVIVGCVGELLWNYNSL
GRTVVRPIMGWLPQEDQEHWD AETQKARDVELHFYEFGLGRLQVHYNKSGGSHTLQKMGICD
ILEDGSIIRGYVQYAFDGRDFLAFDMDTMTFTAADPVAEITKRRWEEEGIYTEGCKHELGT
ICVQNLRRYLEHGKAALKRRREPERVVRWGKEADGILTLSCRAYGFYPRPIAISWMKDMV
RDQETHWGGM

>gg_chain19 Chain=chain19 Position=AADN03010745.1:20-1052(-) GeneID=None ProteinID=None
Score=246.9 E-value=6.2e-70 Length=1033 Comment=No overlapping annotations
YDSKSRVWHLSEMLPQEDREHWD TQKARDVELVFCGFLGRLQVHYNKSGGSHTLQKMG
FCDDILEDGSIRGYVQYAFDGRDFLAFDMDTMTFTAADPVAEITKRRWEEEGIYAEERWKH
ELGTVCVQNLRRYLEHGKVALKRRKRVVRWGKEADGILTLSCRAYGFYPRPIAISWMKDMG
VRDQETHWGGIMPNRDGYASAVINVLPEGDGKYQCRVEHASLPQGLF

>gg_chain20 Chain=chain20 Position=AADN03014659.1:36-1021(+) GeneID=ENSGALG00000027123
ProteinID=ENSGALP00000042458 Score=242.3 E-value=2.5e-67 Length=986 Comment=Overlaps single
annotation & in proteins
MKPLTPHPCCASGSHTLQKMGICDILEDGSIIRGYVQYAFDGRDFLAFDMDKMTFTAADAV
AEISKRRWEEEGIYTEGCKHELGTICVQNLRRYLEHGKAVLKRREPERVVRWGKEANGIL
TLSCRAY

>gg_CD1A1 Chain=chain21 Position=16:2390-4170(+) GeneID=ENSGALG00000012494
ProteinID=ENSGALP00000028603 Score=163.7 E-value=5.8e-45 Length=1781 Comment=Overlaps single
annotation & in proteins
MQPHAILLLFFPPTTAEPEALFSPAGSHMLKLLHFATFQNSTSVLVGGLGLLGDVVKMG
SLDSRTGNIRYRPLRPLPKGDWDVIESSIKSYVRDFSRVLVQMYTTPYPPVVFQSSIG
CELOSNGTIRTFPDIAYEGQNLFRNLNLAGTWDQMOHNLQSAKAEHLMANASTLNEVIQV
LLNDTCVDILRLFIQAGKADLERQVPPMAVVFARTAGQAQLLLVCRVTSFYPRPIAVTWL
RDGREVPSPALSTGTVLPNADLT YQLRSTLLVSPQDGHGACRVQHCSLGDRLSLVPWE
DSKWGLSAGLALLLAAAVAVVVLVRRYRKRQRVDEVRSIPLAEHRGTARDGTAAAGQYG
GCDRETPDEGRGHI

>gg_CD1.2 Chain=None Position=16:5769-8509(+) GeneID=ENSGALG00000027865
ProteinID=ENSGALP00000042318 Score=89.1 E-value=1.9e-26 Length=0 Comment=Not detected in genome
search
MWPCHLFLFLLHKTWAEAEETSCPPPEESQFFQLFYTLGNSVSTELTGMALLADVPIM
VLDPHWNLNINCRPWVQETAEVTEVKKILSFSMVGIRNTRIRFMHEMTAKAGLDYPRVFI
HTGCKLYTNGTRWSFVNIGEGGRDLVYELSRERWVQVQRSTLLAKVMSNTLTDLRAVSGF
LEHVFSSSPFNILMLHEEGRTDLERRVPPMAVVFARTAGQVQLLLVCRVTSFYPRPIAV
TWLRDGREVPSPALSTGTVLPNADLT YQLRSTLLVSPQDGHGACRVQHCSLGDRLSLV
PWENPSASSTVGITITILLAAITGGIWWRRRKHAGSGTDFRTFLI

>gg_ENSGALG00000027431 Chain=None Position=AADN03010681.1:349-1273(+) GeneID=ENSGALG00000027431
ProteinID=ENSGALP00000043158 Score=146.9 E-value=5.8e-43 Length=0 Comment=Not detected in
genome search
MGPSEVVVLGALLGALAAACWPHSLRYFLTGMDTPGPGMPRFVIVGYVDGDLFGKYDSK
IKSAQPIVEKLPQEDQEHWD TQKARDGELDFCGFLGSLPEQYKNSGGSHTLQMMFGCD
ILEDGSIIRGYDQY

>gg_ENSGALG00000004772 Chain=None Position=16:380198-384541(-) GeneID=ENSGALG00000004772
ProteinID=ENSGALP000000028184 Score=230.5 E-value=1.3e-68 Length=0 Comment=Not detected in
genome search
LGSLPEQYKNSGGSHTLQMMFGCDILEDGSIIRGYDQYAFDGRDHIAFDMDTMMFTAADP
VAEITKRRWEEEGEYAEERWKHELGNVCVQNLRRYLEHGKAALKRRVQVVRVVRWGKEADGIL
TLSCAHAGFYWPPIAISWMKDMVRDQETHWGGVVPNRDGYHASAAIDVLPEDGDKYQC
RVEHASLPQGLFSWEPQPNLIPVAGAVVAIVAVIAAVVGLVVVWKS KSGKEKKGYEAAAP
GRDGESSISVTGSELSI

>gg_ENSGALG00000027045 Chain=None Position=AADN03016313.1:424-1169(+) GeneID=ENSGALG00000027045
ProteinID=ENSGALP00000042263 Score=117.2 E-value=3.7e-34 Length=0 Comment=Not detected in
genome search
MTDPGPGMPQFVIVGCVGELLWNYNSLGRVVRPIMGWLPQEDQEHWD AETQKARDVELH
FYEFGLGRLQVHYNKSGGSHTLQKMGICDILEDGSIIRGYVQYAFDGRVTSPLIWR

>gg_PROCR Chain=None Position=9:3879408-3881194(-) GeneID=ENSGALG00000006547
ProteinID=ENSGALP00000010561 Score=60.8 E-value=1.8e-17 Length=0 Comment=Not detected in genome search
MLRLLLLCAALGCGAGGDAPLFTFTMLQWTRVSNNGYAFWGNATLGGRLSHLLEDNRNVTV
LPLEPPAGWARQDMVANYLSYFSGIVQVFSKERPLNYTQNLHCRGCLFPNGTTCSEFY
EVS LNGTAF LTFHVPNATWKLWRPRKDPVATFAQQEELMKYSETTHHLQHFLNNTTCVDILW
AQSPQTGKHRGRSHAPLVGLLILGVS AVVGMAGVIFLCTGGSC

>tg_chain01 Chain=chain01 Position=22_random:1550-3070(+) GeneID=ENSTGUG00000017273
ProteinID=ENSTGUP00000017552 Score=298.1 E-value=7e-91 Length=1521 Comment=Overlaps single annotation & in proteins
VLHSLHYLHVAVSEPSGVPQFTSIGFVDGIPFVRYDSEGRAEPLTQWMKDGAEPGYWD
EETQISVRNQHV DARNLET LQERYNQSRGLHTALWAYGCDLLSDG SVRGSRRYGYDGRDF
ISFDLGS GKF LAADSAAEITRRRWEQEDMAERLKNYLKHKCPEWLKRYVGYGQKELERKE
PPDVHVSGKEEYGTLLLSCHAYGFYPTITVSWMKENATLDQEMEWGGIVPNSDGTFFHTW
ARIEALPEEWEQYRCKVDHPGMLEPGLFAWEPTSGRNLTMVAVSVIAAILLITVVLIGFG
VWKLQSGRRDRSGYNVAAGKDVGMNGSTAGTEE

>tg_chain02 Chain=chain02 Position=12:31611-32691(+) GeneID=ENSTGUG00000003538
ProteinID=ENSTGUP00000003643 Score=170.3 E-value=7.9e-53 Length=1081 Comment=Overlaps single annotation & in proteins
GFTTIRLLQTTTFQNTSFVDMEGGLLEDIELGSLDKHTWSIHFCQPWVHPALPRADWDT
IENLIKIYLLQGFNHLINEGAVQREVPYFPVVOCTAGCVLPNRTSQAFGYVGYGQDFLS
FDTKNITWTLSDTELSRYVQSFLQNYTAFTELEILFNDCVDDMEVLLHYGRAALERQ
ELPLATVFARTPSLDQLLVCHVTGFYPRPISVAWLRDGOEVPVPPALNTSTILPNADLT
YQLRSVLAVAPRDGHSYVCRVRHRSLSLIPWGNSEVVLTGLMAGLLAAMAVAAML
V

>tg_chain03 Chain=chain03 Position=Un:70832059-70834629(+) GeneID=ENSTGUG00000015195
ProteinID=ENSTGUP00000015546 Score=43.2 E-value=4e-15 Length=2571 Comment=Overlaps single annotation & in proteins
VLHSLQFLEVALSEPGPRVPSLIVGDVDTGTPCERCGSERGWMEAQTPGMGAGAELGYWD
RQRPCSRGTGTPWPTVTWRRGRYNQSRGGLHTVQGVSSCDLLSNSVHGSHQYGYEGWDFI
SFQLGSGSFAVSN TAVWISQR CWESRGIMVEQMKHYLGHTYVEGPPKYIRYGREALEHKE
PPDAHVYRKVEHGTFRALSCHMYGFPCRTIRMSWMKGEEMQDQEIELGGIIPNSNGTFHTW
ARAEALPGEREQHWCRVEHPRMLELGLIIV

>tg_ENSTGUG00000014179 Chain=None Position=Un:90577703-90578180(-) GeneID=ENSTGUG00000014179
ProteinID=ENSTGUP00000014563 Score=109.8 E-value=6.4e-32 Length=0 Comment=Not detected in genome search
RGRYNQSRGGLHTVQGVSSCDLLSNSVHGSHQYGYEGWDFIFFQLGSR SFAVSN GAVWIS
QRCWEHKGIMVEQMKHYLGHTCVERPPKYIRYGREALEHK

>tg_PROCR Chain=None Position=9:5279190-5280817(+) GeneID=ENSTGUG00000006783
ProteinID=ENSTGUP00000006970 Score=56.6 E-value=3.6e-16 Length=0 Comment=Not detected in genome search
MLRLLLLGGALGCAAEPAAPLAFTMLQLTRVYMGNSIFRGNASLNGQLSHLLEGNNVTQV
LPLEPPDAWARRQDEVIAYLRNFRLLVMMFNKERPINFTHHLCCHLGCRLYPNGTAQSFY
EVTLNRTAF LSFHVP SATWERRWP GELPVAAFAQEQLMKYPIITQDLQYFLNNTCVSILQ
AQSVRTGKVSGRSRAPLVGLVLGSLGLLMAGIFLCTGGSC

>tg_ENSTGUG00000016646 Chain=None Position=Un:51086513-51086992(+) GeneID=ENSTGUG00000016646
ProteinID=ENSTGUP00000016940 Score=115.1 E-value=1.6e-33 Length=0 Comment=Not detected in genome search
RGRYNQSRGSLHTVQGVSSCDLLSNSVHGSHQYGYEGWDFISFQLGSGSFAVSN TAVWI
SQR CWEHKGIMVEQMKHYLGHTCVERPPKYIRYDREALEHK

>tg_ENSTGUG00000015564 Chain=None Position=Un:104347401-104347872(+) GeneID=ENSTGUG00000015564
ProteinID=ENSTGUP00000015895 Score=101.0 E-value=2.6e-29 Length=0 Comment=Not detected in genome search
RYNQSRGSLHTVQGVSSCDLLSNSVHGSHQYGYEGWDFISFLLGSR SFAVSSGATWITQR
CWESRGIMVEQMKHYLGHTCVERPPKYIRYGWAEALEHK

>tg_ENSTGUG00000014985 Chain=None Position=Un:124963299-124964275(-) GeneID=ENSTGUG00000014985
ProteinID=ENSTGUP00000015335 Score=135.3 E-value=1.8e-39 Length=0 Comment=Not detected in genome search
VLHSLQFLEVALSEPGVPQSLIVGDVDTGTPCERCGSERGRMEAQTPGMGAGAELGYWD
SQSWDIGTARAGTLGDLES DRLFSPRSPGLHTVQGVSSCDLLSNSVHGSHQYGYEGWDFI
SFQLGSELAVSNSAVWISQR CWESRGIMVEQMKHYLGHTCVERPPKYIRYGREALEHK

>tg_ENSTGUG00000015443 Chain=None Position=Un:42292316-42292793(+) GeneID=ENSTGUG00000015443
ProteinID=ENSTGUP00000015776 Score=98.4 E-value=1.4e-28 Length=0 Comment=Not detected in genome search
RGRYNQSRGSLHTVQGVSSCDLLSNSVHGSHQYGYEGWDFISFLLGSGSFAVSSGALWIS
QRCWESRGIMVEQMKHYLGHTCVERPPKYIRYGWAEALEHK

>tg_ENSTGUG00000014981 Chain=None Position=Un:124954309-124954785(+) GeneID=ENSTGUG00000014981
ProteinID=ENSTGUP00000015332 Score=103.5 E-value=4.5e-30 Length=0 Comment=Not detected in
genome search
RGRYNQSRGSLHTVQGVSSCDLLSNSIHGSHQYGYEGWDFISFLLGSRSFVAVSNGAVWIS
QRCWEHKGIMVEQMKHYLGHTCVERPPKYIRYGREALEHK

>tg_ENSTGUG00000015460 Chain=None Position=Un:73084048-73084525(-) GeneID=ENSTGUG00000015460
ProteinID=ENSTGUP00000015793 Score=107.3 E-value=3.5e-31 Length=0 Comment=Not detected in
genome search
RGRYNQSRGSLHTVQGVSSCDLLSNSVHGSHQYGYEGWDFISFQLGSGSFVAVSNTAVWIS
QRCWESRGIMVEQMKHYLGHTCVERPPKYIRYGWGALEHK

>mg_chain01 Chain=chain01 Position=GL425503.1:2182-3320(+) GeneID=ENSMGAG0000005937
ProteinID=ENSMGAP00000005893 Score=286.9 E-value=5e-83 Length=1139 Comment=Overlaps single
annotation & in proteins
RYFLTGMTDPGPGMPQFTAVGYVDGKAFGNYDSERRTVRPIVDWLPREDREHWDAAETKKA
RSGELDFYVGLGWLREHYNKSQSGSHTLQKMFSGDILEDGVSVRGYDQYAFDGRDYIAFEMD
TMTFTAANAVAQSTKRKWESEGTYAERWKHELGTVCVENLRRYRGARKDRAERRERPEVR
VRGKEADGILTLSCRAHSFYPRPITISWMDKGVVQDQETHWGGIVPNSDGTYHASATIDV
PPGDWDKYQCRVEHASLLQSSFFSWELQPNLILIVAGAIITVVAVVAVVAGLVVWRKSG
KDEKCYRAELGHDGESSMATGTVWDR

>mg_chain02 Chain=chain02 Position=GL425629.1:399-1386(-) GeneID=ENSMGAG0000006286
ProteinID=ENSMGAP00000006312 Score=198.3 E-value=1.4e-55 Length=988 Comment=Overlaps single
annotation & in proteins
LVPSEAVVLGLLGSLSTVCSHSLRYFVTAMTDPGPGMPQFVIIIGYVDGKAFGNYDSK
SWRAQPIVEILPQEDWEHWDAAQTQRAQDGEVDLYWFLGRLQKLYNKSQSGSHTLQKMFSGD
ILEDGGIQGYNQYTFDRRDYIAFDMTMTFTAAGAVAQITKRRLEEEBGTCAVRFLCVENL
RRYLEHGKAVLKRERPEVQVRGKEADGILTLSCRAHGFYPPWPIAWRDKGKVRDQETHW
GRIVPNSDSTCHASATIDVLPDGDKYRCRVEHASLPQPSLFSWKPQPNLIPIAAGAVIAI
MAVIAIMAVIAIIVGLVVRASHVKAEGRG

>mg_chain03 Chain=chain03 Position=GL429735.1:864-2485(+) GeneID=ENSMGAG0000009608
ProteinID=ENSMGAP00000018632 Score=140.2 E-value=1.7e-38 Length=1622 Comment=Overlaps single
annotation & in proteins
MQPHAILLFFFPGTWAEPEALFPPSPAGSHMLKLLHFATFQNSTSVLVGGVGLLGDVEM
GSLDSRTGNLHYLPPWMPRLPKGNETTRARFSGILQHHSFVSELFQRREGTEDEPFFV
QTSVGCLELYSNETIRTFEDIAYNGQIFLRFRLDTGTWDMQMNQLAAIAEHLMLVNSTLN
EVIQVLLGDTCEVLRIFIQAGKADLERQVPPIAVVFARTAGPAQLLLVCRVTSFYPRPI
TVTWLRDQKELPPSPALSTGTVL

>mg_ENSMGAG0000009766 Chain=None Position=GL424372.1:74-1058(+) GeneID=ENSMGAG0000009766
ProteinID=ENSMGAP00000010090 Score=227.0 E-value=1.2e-67 Length=0 Comment=Not detected in
genome search
VSGSHTVQRKIGCDILEDGTTTRGYHQVAYDGRDFIAFDKDKTKFTATVPEAVLTKRRWED
DGSVAERFKHYLEETCWEWLRTYMERGKAELEERPEPVRVWGGKADGILTLSCRAHGFY
PRPIDVSWMKDQVAGNQDTSQGGIAPNSDGTHTLVTTIDARPEWWDKYQCRVEHASLQPP
GPFYSWEPSQPNLPIVVGIVIVAIAIAIVGGVGFIIYRSHAGKKGKGYNIAPG

>mg_PROCR Chain=None Position=11:3905742-3914577(-) GeneID=ENSMGAG0000002693
ProteinID=ENSMGAP00000002298 Score=67.3 E-value=2.1e-19 Length=0 Comment=Not detected in genome
search
AVTLGCGSGGAAPLAFIMLQWTQVSKGSSVFWGNATLGGRLSHILEDHNNVTOVLPLEPPA
VWAQQRDMVANYLSYFSGIVQVFSKERPLNYTQILHCLRGCCFLPNGTAHSFYEVSLNGT
AFLTFHVPTATWELRWRPRDPVATFARRELMKYSETTRNLQHFINTCVDILRAQSPQTG
KQRGRSHAPLVGLLILGVSAVVGMVAVGIFLCTGGSC

>mg_ENSMGAG00000017223 Chain=None Position=GL429694.1:2-782(+) GeneID=ENSMGAG00000017223
ProteinID=ENSMGAP00000018230 Score=53.8 E-value=6e-16 Length=0 Comment=Not detected in genome
search
RWSFANIGEGGRDLVTVYDLNRRERWLPRRSTPLAELMSNTLTDLRAVSGLLEHIFSTSLPN
YIIMLHKEGKADLERRVPPIAVVFARTAGPAQLLLVCRVTSFYPRPITVTVWLRDQKELPP
SPALSTGTVLPNADLTYQLHSTLLVSPRDGHSYACRVQHCSLGDRLVLPVENPSSASTV
ITIAILLTIVIVGAVVWRRHR

>mg_ENSMGAG00000016965 Chain=None Position=GL424516.1:24657-24833(+) GeneID=ENSMGAG00000016965
ProteinID=ENSMGAP00000017795 Score=96.8 E-value=3.9e-28 Length=0 Comment=Not detected in genome
search
VSGSHTVQRMYGCDILGNSTTRGYNQYAYDGRDFIAFDKDKTKFTAAVPEAVLTKRKWE

>ac_chain01 Chain=chain01 Position=GL344125.1:19677-25330(+) GeneID=ENSACAG00000006478
ProteinID=ENSACAP00000006366 Score=348.6 E-value=2.8e-101 Length=5654 Comment=Overlaps single
annotation & in proteins
MGLSRKEPFLLLLVSAALLLGGSSSHSLRYFATIVSEPQGEVPPQFFSVGYVDDQQQFV
SYDAKARRRPAVWPWIRKDDDPQYWERNSQIVWNTQVFRGSLVNNARYNQSGSHT
WQQMYGCELRDGGSKGGYRQFAYDGRDYISFDKETLTYTAADVPAQNTKRWAEPAIAQ
GRKVYLEEICEWLQRYLDYKESLLRTDKPEVKVTRKEDYDGMETLICRVGGFPYKPID
ITWTKDGEVWMDQVFHGLVSPNSDGTYYTWRVSMVDPKESERYKCHVEHDGLPNPVDVAW

EEPDSNLGIIIGCVVGVLLLIIGVMAGIYFKKRQDGYKAAAASDQGSNSSGTGTAVPI

>ac_chain02 Chain=chain02 Position=GL344125.1:34918-47343(+) GeneID=ENSACAG00000006478
ProteinID=ENSACAP00000006366 Score=346.7 E-value=9.9e-101 Length=12426 Comment=Overlaps single
annotation & in proteins
AGSSSHSMRYVYTSVSEPGQEQPQFFHVAVVDDQEFVSYDAQARRVLPTAPWIRKVDKDD
POYWERNISQYVWVWQNFVRVLDGLTSLGYYNQSGGEPGSHTLQKMYGCELRLDGSKGGYRQ
FAYDGRDYISFDKETLTYTAADVPAQNTKRWKWEADLATIQYRKAYLEEICIEWLQRYLDY
GKESLLRTVKVTRKEDYDDMETLICRVGGFYPKEIDITWTKDGEVWTQDVFHGLVSPNSD
GTYTWRSVTVDPKERERYKCHVEHDGLPNPVDV

>ac_chain03 Chain=chain03 Position=2:199513343-199515738(+) GeneID=ENSACAG00000017226
ProteinID=ENSACAP00000016954 Score=344.6 E-value=3.9e-100 Length=2396 Comment=Overlaps single
annotation & in proteins
MPLSRGAPAVLPFLLAFLVGRGAGSSSHSMRYFVTSVSEPGQVQPFQFSYVGYVDDQEFVS
YNASTRRYLKVPWISKVEKNDPDYWERNTLYAQGHERSFRDHLATLAEYYNQSGGLHTF
QWYMGCELRRDWSKGGYQYAYDGRDYISLDKDTLTWMAADVPAQNTKRWKWEADFRDNEY
KTTYLEETCIEWLQRYLNYGKETLLRTEKPEVKVTRKEDYDGMETLICRVGGFYPKDIDI
DWTRDGEVWLQDVFHGLVSPNSDGTYYTWRSVTVDPKERERYKCHVEHDGLPNPVDVAWE
EPDSNLGLIIGCVVGVLLLIIVMAGIAYVFKKSQDGYKAAAASDQGSNSSGTGTAVPI

>ac_chain04 Chain=chain04 Position=GL344036.1:83552-89335(-) GeneID=ENSACAG00000006674
ProteinID=ENSACAP00000006547 Score=344.1 E-value=5.8e-100 Length=5784 Comment=Overlaps single
annotation & in proteins
SGTSKLPILGLPLEEDPLGGPSSSSSHSMCYFVTSISEPGQVQPFQFVYVGYVDDQEFVSY
NTGTRRYLAKVPWISQVEKNDPKYWERNTLYAQGHEQSFRDHLATLAEYYNQSGGLHTFQ
WYMGCELRRDWSKGGYQYAYDGRDYISLDKDTLTWMAADVPAQNTKRWKWEADFRDNEY
KIYLEETCIEWLQRYLNYGKETLLRTEKPEVKVTRKEDYDGMETLICRVGGFYPKDIDI
WTRDGEVWLQDVFHGLVSPNSDGTYYTWRSVTVDPKERERYKCHVEHDGLPNPVDVAWE
PASNLGLIIGCVVGVLLLIIVMAGIAYVFKKSQDGYKAAAASDQGSNSSGTGTAVPI
QT

>ac_chain05 Chain=chain05 Position=2:199546590-199563181(+) GeneID=ENSACAG00000011070
ProteinID=ENSACAP00000010940 Score=337.9 E-value=4e-98 Length=16592 Comment=Overlaps multiple
annotations; one in proteins
MGLSQEDPFLLLFGVSALLLGGSSGSSSHSLRYVYTSVSEPGQEQPQFFSVGYVDEQEFV
SYDSKAKRRFPVAVPWIRKVEEEDPEYWEQGSQVLSAEQVFRMDLENVARYYNQSGGAHI
LQYMGCELRRDWSKGGYQYAYDGRDYISFDKETLTYTAADVPAQNTKRWKWEADFRDNEY
RNKAYLEDICIEWLQRYLDYGKETLLRTEKPEVKVTRKEDYDGMETLICRVGGFYPKEID
ITWTKDGEVWTQDVFHGLVSPNSDGTYYTWRSVTVDPKESERYKCHVEHDGLPNPVDVAW
EEPASASNLGLIIGVWVGLLIIAVLAGIAYVFRNSQDGYRAAAGNDQGYSSSGAGKWD
WAHL

>ac_chain06 Chain=chain06 Position=GL344134.1:50311-56905(-) GeneID=ENSACAG00000024528
ProteinID=ENSACAP00000019090 Score=337.1 E-value=9e-98 Length=6595 Comment=Overlaps single
annotation & in proteins
LSRDGPFLLGLVGSALLLGGSSAGSSSHSLRYFTTIVSEPGQEQPQFFSVGYVDDQEVTSY
DAKARRDLKAPWPKVWIKWERNISQVANNWEQIFRVGLTLLGYYNQSGGAHTLQWYMG
EMRKDRSKVGYFQYGYDGRDYLSLDKETLTWTAANVPAQNTKRWKWEANHAFLQYKDYLE
KDCIEWLQRYLDYGKETLLRTEKPEVKVTRKEDYDGMETLTCHVGGFYPKDIDIDGEVWM
QDVFHGLVSPNSDGTYYTWRSIKVDPKESHEHYKCHMEHDGLPNPVDVAWEPPASNLGLII
GCVLGIIFLIAVMAGIAYV

>ac_chain07 Chain=chain07 Position=GL344164.1:25218-31763(-) GeneID=ENSACAG00000023667
ProteinID=ENSACAP00000018603 Score=334.1 E-value=4.8e-97 Length=6546 Comment=Overlaps single
annotation & in proteins
VGFPPFAEPRHLINIICAFLLTSLAISHSLRYFYTAASEDNKWLSSFIIVGYVDDQEQICQ
YDSIMREKRCVPMWMSKVEEEDPSYWEETEEMSRNSEQFFKESLVTAKNRYNQSGGFHTL
QQMYGCDLRKDGSKAGYQYAYDGRDYISFDKETLTWTAADVEAQFTKRWKWEADFRDNEY
LKAYLEEEICILWLKYLEYKGTLLRTEAPEVKVTRKEDYDGMETLICRVGGFYPKEIDI
EWTRDGEVWMQDVFHGLVSPNSDGTYYTWRSVTVDPKERERYKCHVEHDGLPSPVDMAWK
EPDSASKSVLIGCSVGIPLPITGLGVYFQKHQDDYKTVSGSKKFNSSKEGKSIQFVRQS

>ac_chain08 Chain=chain08 Position=GL344462.1:1558-6374(-) GeneID=ENSACAG00000000562
ProteinID=ENSACAP00000000544 Score=329.2 E-value=1.5e-95 Length=4817 Comment=Overlaps single
annotation & in proteins
MGSLLGLGFIVLLAGSCSGCSSHSMYFVTSVSEPGQVQPFQFSYVGFVNDQEFVFNAST
KRYLKPVPWISQVEKDDPDYWERNTLYAQGHERSFRDHLATLAKYYNQSGGLHTFQWYMG
CELRRNWSKRGYQYAYDGRDYISLDKDTLTWTAADVEAQFTKRWKWEADFRDNEYKFKFYL
EETCIKWLQRYLNYGKETLLRTEKPEVKVTRKEDYDGMETLICRVGGFYPKEIDITWTRD
GEVWMQDVFHGLVSPNSDGTYYTWRVWVTVDPKERERYKCHVEHDGLLNPVDVAWEPPASK
SVLIGCAVGILLPIAGLAVYF

>ac_chain09 Chain=chain09 Position=GL343748.1:166921-168031(+) GeneID=ENSACAG00000011968
ProteinID=ENSACAP00000018536 Score=324.6 E-value=4.7e-94 Length=1111 Comment=Overlaps single
annotation & in proteins
MALPLWLLGPAKSLFLLDGSLVLAALPKGTAEPEMAPGASFHFLRFYFTVTESSPGVPH
FVIVGYVDDQVFFQYDAQTRKAVPRVPMWKEKVEADPQYVWEQTVQAQNAEVGFRVDLAN

LRERYNQTTGGFHTWQWMMYGCELRKDRSKGGYYQYAYDGRGSLSFDDKTLTWTAAADVPAQI
TKRRWEREKIQARFNHYLEGEICDWLQNYLDYQKQTLRTERPVGVRVSRRTASDQGEAL
ICQAHGFHPREIDATWRKGEENMDHETLRRNIAPNSDGTYYHTWLSIEIDPKEMDLLRCHL
EHASLSTPMVLAYEPEGVNI

>ac_chain10 Chain=chain10 Position=GL344798.1:12321-17723(-) GeneID=ENSACAG00000027871
ProteinID=ENSACAP00000019987 Score=324.0 E-value=4.9e-94 Length=5403 Comment=Overlaps single
annotation & in proteins
FNLHKSLSLFLHILSPVFMPSFSSSSSSSHSLRYFYTAASEDNKWLSSFIIVGYVDDQRIC
QYDSIVREKRKPCVPWMSKVEEEDPSYWAEEETVSRNSELFFKVSLVTANNRYNQSGGFHT
LQOMYGCDLRKDGSKAGYRQYAYDGKDFISFDKETLTWTAADVKAQITKRRTWDAQSTRNQ
YLKAYLEEEECIKWLQKYLEYGNETLLRTEKPEVKVTRKVDYDGMETLICRVGGFYPKDID
IEWTRDGEVVMQDTFHGLVSPNSDGTYYTWRISITVDPKERERYKCHVEHDGLPSPVDVAW
EDPASKSVLIGCAVGILLPITGLAVYFPEKHQDDYKTVSASDQKFNISEVKNVEVPL

>ac_chain11 Chain=chain11 Position=GL344714.1:6300-11691(-) GeneID=ENSACAG00000022634
ProteinID=ENSACAP00000007706 Score=323.3 E-value=7.9e-94 Length=5392 Comment=Overlaps single
annotation & in proteins
KDVPIIKLCFILENAALAEELRRCFSSSSSSSHSLRYFYTAASEDNKWLSSFIIVGYVDDQRI
CQYDSIVREKRACVPWMSKVEEEDPSYWAEEETVSRNAEVFFKVSLVTANNRYNQSGGFH
TLQOMYGCDLRKDESKGGYRQYAYDGKDFISFDKETLTWTAADVEAQITKRRWDAQSTRN
QYLKAYLEEEECIKWLQKYLEYGNETLLRTEKPEVKVTRKVDYDGMETLICRVGGFYPKDI
DIEWTRDGEVVMQDTFHGLVSPNSDGTYYTWRISITVDPKERERYKCHVEHDGLPSPVDVA
WEDPVSKSVLIGCAVGILLPITGLAVYFQKHQDDYKTVSGSDKKSNTSKEGKFTAERVE

>ac_chain12 Chain=chain12 Position=GL343965.1:57526-62923(-) GeneID=ENSACAG00000022657
ProteinID=ENSACAP00000019604 Score=319.7 E-value=9.3e-93 Length=5398 Comment=Overlaps single
annotation & in proteins
MARLLLLRVVALLLGHGPIALLNPAPGVSSSSSSSHSLRYFYTAASEDNKMHSSFIIVGYVD
DQRICQYDSIVREKRKPSVPWMSKVEEEDPSYWKDESEESRNAEVFFKESLVIARNRYNQSG
GGFHTLQOMYGCDLRKDGSKAGYRQYAYDGKDFISFDKETLTWTAADVEAQITKRRKWDQAQ
SMRNQYLKAYLEEEECIKWLQKYLEYGNETLLRTEKPEVKVTRKVDYDGMETLICRVGGFY
PKEIDITWTRDGEVVMQDTFHGLVSPNSDGTYYTWNVTVDPKERERYKCHVEHDGLLNP
VDVAWEPEPDSKFGILLIGCIVGALLAIAAIAGLAVYFRE

>ac_chain13 Chain=chain13 Position=GL343520.1:556834-564191(-) GeneID=ENSACAG00000006498
ProteinID=ENSACAP00000006420 Score=296.7 E-value=7.7e-85 Length=7358 Comment=Overlaps single
annotation & in proteins
MSPVSGLFPPLLGAIVSPFGCSAHSRSHRLHYFFTVGSEPGGGLPHFSIVGYLDDEQFV
RYDSETRRRDLPRSSWIRKMEHDAQYWDWQTLRSRSELYFRVSLWTRFNHRSRHHHTWQW
MHGCEVSADGRKGGLSQFGYDGRDFSLDKETLTWTAASNPKAQVTKRKEADRVWVQGRK
TFLEEDCIDWLHKYLNKYGKETLLRKEKSPVVKVTRKAGYGDLETLKRAHGFYKPEIDITW
RMEGNVMEQETRRGGIAPNADGTYYHTWLSIEVDTKDRDLRCHVEHDSLPEPLDAAWEEA
GVLWPLVAGISGAVAVVILLGVILVLYLKRREHEGYRATASS

>ac_chain14 Chain=chain14 Position=GL343520.1:493814-503767(-) GeneID=ENSACAG00000029516
ProteinID=ENSACAP000000022668 Score=280.6 E-value=3.6e-81 Length=9954 Comment=Overlaps single
annotation & in proteins
MGQPWLLGAAALAFLLASITGAPSHSLSYLFVVSEPGPLPQFMVSGHFDGQLFASYS
ERRSGVPRTEWMEEMVELDPGYWDWLNRNARNSEQIFKWQIGTVMRHYNHSGGLHTMQWV
FSCTLSEEDGRKGGSNKFSYDGEALSFDKETLRTWASDAVAEGYKKEWAEAGRTQRSKV
YLEDKICISMLHKKHLEVRKGVPRKRVAPVVKVTHKAGIGGLETLVCHIHGFYKPKDIDATWR
KDEEVREQETLRGNITPNSDGTYYNTWISIEVDEKDRDKYQCHVEHDSLPEGAGSGLHFPV
VECGAIPLPWIIVGSALVILIVTLIVIIILKKRGSQNPATNAEMQEMNPPQRRRRCIS
VAE

>ac_chain15 Chain=chain15 Position=GL343520.1:527815-534009(-) GeneID=ENSACAG00000024638
ProteinID=ENSACAP00000018398 Score=260.8 E-value=2e-75 Length=6195 Comment=Overlaps single
annotation & in proteins
MGAFRGLFFFFLFFPGTLAGSFSSMRGFYTATWDPSSQAQPQVAIVIYVDDQLATQYNSS
TRQALPQVYWMKMEDEHIPSFWEVNLTGQRLELRLKKTLENLKERNSQSTGFLTFQLMT
GCEIRKNGHKEGYHQYSYNGRDFLSFDPKTQWTAADLEAQITQRQWDGVDVSRHVMYY
VEEECPEDLQRFVEYKESLLRRETPVVTISRKAYDGLLETLLCRVHGFFYKPEIDVNWKR
DGEVWVQDTFSGGILPNADGTYYHTWISIEVEPKERDRYRCHVEHDSLEELVLAWEEDRGS
VWPLVGGVVGAVVGLLVAAGIILYKKGQEAETEPEQVEDPLMNG

>ac_chain16 Chain=chain16 Position=GL344250.1:24152-28586(+) GeneID=ENSACAG00000006471
ProteinID=ENSACAP00000006333 Score=239.0 E-value=9.9e-69 Length=4435 Comment=Overlaps single
annotation & in proteins
MAVVQVDGVTIGHYDSQVRRAI PRVPWITRVEEAYPMFWDWVTDRAQKAEWFGWDMKKL
QKLNHNSGGIYTLQSMYCEMDGDGDRIQGYGLYNGEDFLSFDIETRIWTVSDVRAET
LKKRWDSNVDRSQIKDYLERNCIEYLQKFLDYGKEVLTIKEPPTVKVSYDGLKTLTCRV
DGFYKPEIDAVWKKDSEVWEQETLRGGVVPNSDGTFFHTWLRLEVDPKDRDLRCHVDHAA
LEEALDTAWEPEASIPIIIIIIIAVLSVLVAHALLLWYKRNELPKIISKPFERQPLNT
GGTKGQSQPDS

>ac_chain17 Chain=chain17 Position=GL343777.1:59265-67362(-) GeneID=ENSACAG00000000468
ProteinID=ENSACAP00000000429 Score=194.3 E-value=2.4e-55 Length=8098 Comment=Overlaps single
annotation & in proteins
LVTHYRLSGRWETVWFSSTHNLYYSLTMWEHTGDPVHFVFLHNYLDYESLYYDNNIREVI
PHATWINKAEKEYPGYKQNLGAASLERGLRADLLYLKNNINQTMSTLTWQMSFGCMLH
PDNTTSSYYKTAYNGKDFLAFDTVKGSWTALSVEGQNNKLEWDARNVTNKSFKTFLELGC
IAWLKKYLEYLEGVSLKKNPPKARVTIRVKPDGEEVLTCTRAYGFSPKEISVNWVNDRDVI
TAGVVYGNVTFNFDGTFYTWLRLITVDPKERSNFQCRMEHEMLKPKIKVFLRKPCKGAVVG
RV

>ac_chain18 Chain=chain18 Position=GL343619.1:393831-405607(+) GeneID=ENSACAG000000003922
ProteinID=ENSACAP000000003807 Score=186.2 E-value=4.6e-53 Length=11777 Comment=Overlaps single
annotation & in proteins
ASSHSLYYFRTYMSEAIIPGAPRYVEIGYLDGQPIERYDSKSGKMHPLTPWISHSVSLMEN
YWEVQSSLAQLNEEGFEHFTVILERYNGYDVLHTLQCVDCGCELSDDGTEKLYHREAYD
GKELTDWDIYVWDIYRHRKGFWGQTCIGWLRRLHLEFQKKNLPTDAEAPDVKLIHRPINGSL
ETLICKVNGFFPKDINAFWLKDGKIVDDENIHVKVVPNADGTYQARLTIQVDPTEAQHYE
CHIDHASLNEPLHKGWRWSADSAWPIVGSVAVAAGLVVVVSLGIFYTSRVWRGAGASRS
LSMGRIQRKHL

>ac_chain19 Chain=chain19 Position=2:189407192-189410738(+) GeneID=ENSACAG000000011395
ProteinID=ENSACAP000000011131 Score=157.3 E-value=3.2e-44 Length=3547 Comment=Overlaps multiple
annotations; one in proteins
MRMACRYFLVFFYVIGAAFFHLPVLPVWPFRLQTIISFQNTSATEIMGTIAFLGDVETHS
LDHTWIKIFLQPTWQSAFTPLKWEMLGQLFRASFDIDFKKAINNMVAASNYSPFVIQSF
FFCEIGTDGTRKGFYKGAANGDDVLGYSTDNATVWVQKDTPLAVAVQDFLNRNKGTTANM
RSLLLNECIDILESSLKGTNETLHRQEKPVAVVFAQEPATTDSSLVLCVQVTFYPLHIN
VSWLQDEVALPSSRINSTTILHNYDLTYQIRSSLAIKSMETSHSYVCRIQHSGLDGKSLV
ILWERKHRYRVTIVVVVVASILVVVAGLVFLYQKRRQYEDVNVQAIKSTARQ

>ac_ENSACAG000000003088 Chain=None Position=AAWZ02037434:7354-13029(-) GeneID=ENSACAG000000003088
ProteinID=ENSACAP000000003009 Score=229.7 E-value=2.4e-68 Length=0 Comment=Not detected in
genome search
HFNKLTHTNRRNSSVGMGFHTLQOMYGCDLRKDGSKAGYRQYAYDGKDFISFDKETLTV
TAADVEAQTFRKRWDAQYTRNQYLKAYLEAECIGWLQKYLEYGKETLLRTEPTEVMVTRK
VDHDGMETLICRVGGFYPKKEIDIEWTRDGEVWTQDVFHGLVSPNSDGYTYTWRISITVDPK
ERERYQCHVEHDGLLKPVDVAWEEPASNLRRRIIGCVVVALVAIAAAGLAVYFKKCGDC
KATSGGRRGLQNSNKKKNIKM

>ac_ENSACAG000000025539 Chain=None Position=GL344035.1:52225-52899(+) GeneID=ENSACAG000000025539
ProteinID=ENSACAP000000018273 Score=78.5 E-value=1.2e-22 Length=0 Comment=Not detected in genome
search
INLPRLGRNFLISYFFLQTCFSFSSHSLKYFYTAISEDNELLSFYVVGMDQORISQYD
STVRLELPCVSWMKKVEEVDPSYWKDETEVSRNSEQFFKENLVIARNRYNQSRG

>ac_ENSACAG000000024309 Chain=None Position=AAWZ02038573:7544-9660(-) GeneID=ENSACAG000000024309
ProteinID=ENSACAP000000018317 Score=229.2 E-value=3.4e-68 Length=0 Comment=Not detected in
genome search
PSLNLHHLICISLSGFHTLQOMYGCDLRKDGSKAGYRQYAYDGKDFISFDKETLTVTAADV
KAQFTKRWDAQYTRNQYLKAYLEEEICIGWLQKYLEYGKETLLRTEPTEVMVTRKVDHDG
METLICRVGGFYPKKEIDIEWTRDGEVWTQDVFHGLVSPNSDGYTYTWRISITVDPKERERY
KCHVEHDGLLKPVDVAWEEPGM

>ac_ENSACAG000000024939 Chain=None Position=AAWZ02038988:2977-5236(-) GeneID=ENSACAG000000024939
ProteinID=ENSACAP000000019917 Score=99.4 E-value=7.9e-29 Length=0 Comment=Not detected in genome
search
FYLLALRDVLDLICTAVEITTSGCSSHMLYFVTSVSEPGQVQPSYVGFVNDQEFVY
NASTKRYLPKVPWISQVEKDDPDYWERNTLYAQGHERSFRDHLATLAKYYNQSGG

>ac_ENSACAG000000025947 Chain=None Position=AAWZ02040156:3915-5955(-) GeneID=ENSACAG000000025947
ProteinID=ENSACAP000000021228 Score=230.2 E-value=1.9e-68 Length=0 Comment=Not detected in
genome search
PSLNLHHLICISLSGFHTLQOMYGCDLRKDGSKAGYRQYAYDGKDFISFDKETLTVTAADV
EAQITKRWDAQSMRNQYLKAYLEEEICIGWLQKYLEYGKETLLRTEPTEVMVTRKVDHDG
METLICRVGGFYPKKEIDIEWTRDGEVWTQDVFHGLVSPNSDGYTYTWRISITVDPKERERY
QCHVEHDGL

>ac_ENSACAG000000024081 Chain=None Position=AAWZ02039289:774-2823(+) GeneID=ENSACAG000000024081
ProteinID=ENSACAP000000020986 Score=233.6 E-value=2e-69 Length=0 Comment=Not detected in genome
search
PSLNLHHLICISLSGFHTLQOMYGCDLRKDGSKAGYRQYAYDGKDFISFDKETLTVTAADV
EAQITKRWDAQSMRNQYLKAYLEEEICIEWLKYLEYGKETLLRTEPTEVMVTRKVDHDG
METLICRVGGFYPKKEIDIEWTRDGEVWTQDVFHGLVSPNSDGYTYTWRISITVDPKERERY
QCHVEHNGL

>ac_PROCR Chain=None Position=3:33612803-33622411(-) GeneID=ENSACAG000000009038
ProteinID=ENSACAP000000008831 Score=69.1 E-value=7.6e-20 Length=0 Comment=Not detected in genome
search

MLFFNMFLLSLALCHWAYGAESHNFMTVQYSYFPNKTLVESVGYATLDGTRHTHLESHNG
HVNVSQQLPLETSDRWEMRRTOQLHDYLSRLKALVYLLTLERNISYPLEVHCTATCQLSEN
GTSFFFEVLLNKVEFLIFRGAKNWEPLQETSEAIYASTNLNKYNETRGHVEFFLQETCI
NFIREHTDVKEALTKGREGRPHPTPLVLGIILGALAIMGFAVCLFLCTGGR

>xt_mhcla Chain=chain01 Position=GL173531.1:22383-54316(-) GeneID=ENSXETG0000017065
ProteinID=ENSXETP0000037199 Score=318.5 E-value=5.6e-91 Length=31934 Comment=Overlaps single
annotation & in proteins
VYCLLPDLYTFRMDVRLVPLLLTLGVS AVYSGSHSLRYYYTAVSDRAFGLPEFSSVGVYVD
ETQINRYSSDNGRNEPAAQWMKQNEGPEYWRDQTQNSKGM EAVFKHNKVKGMDFRNQTSG
THSFQNMYGCELREDGSTRGYDQYGDGREFMALDTRWVYVPSVREAQISTQKWN SPEV
NEPERERNYQENICIESLKRYLGYGQAELEERRVRPHVKISDHQSDDITELRCQAYGFYPR
EIDVKWRNRRGRDDVHDEAKEILPNPDGSYQLRATAVVTPEKGD SYACHVEHSSLEEKLI
VWVPGPNKDGKNNV I IACVAVAVLMITAAAVGVFLYKQRQESKGYTETASNESPASSVA
RA

>xt_chain02 Chain=chain02 Position=GL172781.1:692190-695424(-) GeneID=ENSXETG0000017938
ProteinID=ENSXETP0000038903 Score=276.9 E-value=1.3e-78 Length=3235 Comment=Overlaps single
annotation & in proteins
MYMQIIISVLACASCKYKYAH IIVAGSHSYSYFYTGNSDGGQGHPHFTAAAYLD DFFPIGVY
TSETKTFQIKTKWIEEKADPALWKNYTNIFQRWETDFRQGVKMM EHFNQTEGFHSLQWM
YGCLEDDDDGGIRAYDL DGYDGKEFMALDLD RVIH ISSRPETE AIAHKWNTETIEAEKMT
YLEVQC IETLKKYISYGKEHLERKVTPOVKVFHQQSAGFIK LHCQVYGFPPQAIKVKWLR
SEEEKVQPFVQLLPNDFTYQTRVITITVAPEEMDTYFCHVEHSSLEKPHIVVWVPIEP
SRNNLLYMSLGASSV LLLILTFVMGLWLGRRRPGKKQKMFCSVVISKDRHSPSTTPI SQ

>xt_chain03 Chain=chain03 Position=GL173193.1:381302-386177(-) GeneID=ENSXETG0000009073
ProteinID=ENSXETP0000019930 Score=264.1 E-value=6.6e-75 Length=4876 Comment=Overlaps single
annotation & in proteins
LLPFMFVLLGSH TLLYHITIVSAPT TGV PQYSIIMYVDGVQYGRYNSDIRQAETSIPSLR
ELSKHIEKQTQYVQY EYMSQRHKLKFLIGFYNNTKSDRDTHIYQSRLACELHDDGTIGGY
QEIAPFGKALLIFDKENYTYVPV TQEA EIV AQLWNKH YDSTNTKLYMENDCIEHLKLYLP
YISTDLERKVP PNVKVSNSESESGTKLHCRVYGFYPRDVEVKWIKNGRDEIHSEEAQ IIL
PNPDGTYQIRVSVGVTP EGGATYSCHIDHSSLEKPLVVKFDSTEPNNGRYLIIISVIT
LLLLLLPFALGLCIYRKRKTKGVA AQT VATCRD SHSRNIGQR

>xt_chain04 Chain=chain04 Position=GL173193.1:354444-357204(-) GeneID=ENSXETG0000009076
ProteinID=ENSXETP0000007870 Score=257.4 E-value=7.3e-73 Length=2761 Comment=Overlaps single
annotation & in proteins
VVSAPT PGPVQYSIIMYVDGVQYGRYNSDIRQAQFS PSLNHLSEHLEMQTKFAQTSEVW
QRHKL NFMGVFNKTKSNGNIHIYQRKACELHDDGTIGGYQEIAPFGKDFLVFDKEKVV
YVPVTHEAVMVSHLWKNRYDPTSSK K YMETDCIQHMEMYLPYISTDLEKVPVKVQISSS
ESENGTELHCRVYGFYPRDVEVKWIKNGRDEIHSEEAQ IILPNPDGTYQIRVSVGVTP EGG
GATYSCHIDHSSLEKALVVPFARHTVAALRFLVIDV LLLLLLLLLL IILGIYICRKRKDSR
GSYQSGSKESEV

>xt_chain05 Chain=chain05 Position=GL173063.1:966777-971978(-) GeneID=ENSXETG00000031111
ProteinID=ENSXETP0000035926 Score=257.1 E-value=7.6e-73 Length=5202 Comment=Overlaps single
annotation & in proteins
VVLLFSLGVC AVHCGSH TLEYHIALVSSPAPGVPQSTITVYIDGLKYGKYDSETRRRARFL
TPSLSSLTEHLD MQTKYAQRFEVIQKHMAFLTG YLNRTYGN GDFH VYQRRFACELHEDG
TVSGYEEIAYNGKEVMMFDKERVVYVPV TQEVLTMTQ QWNQHYNDAKINKIYMENECIHY
MKMYLPYISTDLEKRVHPEVKVSSSES SVSGTKLHCRVYGFYPRDVEVKWIKNERDEIYSE
EAAQ IILPNPDGTYQIRVSVGVTP EEGATYSCHVDHSSLAKPLAVCFGSWKSQDCLVPCNA
GLSPLIWI FFLNFCVRRRVPDPQC KCCT

>xt_chain06 Chain=chain06 Position=GL173063.1:979645-982170(-) GeneID=ENSXETG00000031629
ProteinID=ENSXETP0000035924 Score=256.9 E-value=7.9e-73 Length=2526 Comment=Overlaps single
annotation & in proteins
TSTLRIPILQYTCSVCSCCPAGSHTLEYHIALVSSPGVGPQSTITAYIDGLKYGKYDSD
TSRARFLTPSLSSLTEHLD MQTKYAQRFEVIQKHMEFLMG YLNKTYVNGGFNIYQRKFA
CELHEDGTVSGYEEIAYNGKEVMMFDKERVVYVPV TQEVLTMTQ QWNQHYNDAKINKIYM
ENECIQHMKMYLPYISTDLEKRVCPVKVSSSES DGEQKLHCRVYGFYPRDVEVKWIKNG
RDEIHSEEAQ IILPNPDGTYQIRVSVGVTP EEGATYSCHVDHSSLEKPLVVPVFKPSSN
IIVPIVAALALALVAL

>xt_chain07 Chain=chain07 Position=GL173193.1:1938-6154(-) GeneID=ENSXETG0000009064
ProteinID=ENSXETP0000019874 Score=252.8 E-value=1.6e-71 Length=4217 Comment=Overlaps single
annotation & in proteins
MNIILSLHNSHC IAGPHILO YRLTLISTPTPGSPQCTITAYMDGLNYGRYDSDTGHGQA
LVP SLSIPSLAEHREMQRHARGVELIQ RDKIKFLMGFLNK TNGNGDFH VYQRFACELH
EDGTVRGEYEEIAPFGKEFISVDKERVVYVPV TQEALGITELW NKRYDLQDNKIYLENECI
QHMKMYLPYISTDLEKRVPPKVKVSSSES ESGTKLHCRVYGFYPRDVEVKWIKNGRDEIH
SEEAQ IILPNPDGTYQIRVSVGVTP EGGATYSCHIDHSSLEKPLVVPVFKPSSN

>xt_chain08 Chain=chain08 Position=GL173193.1:87977-91470(-) GeneID=ENSXETG0000009068
ProteinID=ENSXETP0000019887 Score=248.3 E-value=4e-70 Length=3494 Comment=Overlaps single
annotation & in proteins

MACIVFLLFGLGIAAVSCRSHLLYHITVVSAPTGPVPOYSIIMYVDGAQYGRYNSDIRQ
AQFFSPSLSPLEHLEMQTRYAQAFVWQRHRLNFLMGVFNKTKGGDRDTHIYQRKAACE
LRDDGTSGGQEIADFQKFLAYDKEKVVYIPSTQEAVMVSHLWNKHYSSTNSKMFMEID
CIQHMKMYLPYISTDLEKKVPPKVKVSCSSESQTKLHCRVYGFYPRDVEVKWIKNGRDEI
HSEEAQILPNPDGTYQIRVSVGVTPPEGATYSCHIDHSSLEKAITVPFEPNKRISILYII
PVLASLVLLIALIALGVFICRKKKDPGSGSHQVSAEDKE

>xt_chain09 Chain=chain09 Position=GL173193.1:259437-268566(-) GeneID=ENSXETG00000033941
ProteinID=ENSXETP00000053836 Score=244.7 E-value=4e-69 Length=9130 Comment=Overlaps single
annotation & in proteins
QNKYLPFYSSYTPVSVSIGSHLHYDVTIVSAPTGPVPRYLVSVMYVDGVQYQYGRYNSDIR
HAEALYPSLRALSEHIERQTYAQEHEMRQRHRLNFKGYLNTKESNGNIYQRRLACELY
EDGTIGGYEIEAFDQKFFSFDKERVVYVPTQEAAMLSYLWNIRYDPTDTKIYMENDCI
EHIKLYLPYISTDLEKKVPPKVKVSSSESQTKLHCRVYGFYPRDVEVKWIKNGRDEIH
SEEAQILPNPDGTYQIRVSVGVTLDKNSNYSCLVHDSSLETPIVNFVGDACNKESEMPP
SLLVLSVCPVIGQVAVSLHVLGIFIRKRMVYVTVG

>xt_XB-GENE-5894693 Chain=chain10 Position=GL173193.1:425428-427883(-)
GeneID=ENSXETG00000016466 ProteinID=ENSXETP00000019932 Score=243.3 E-value=1.1e-68 Length=2456
Comment=Overlaps single annotation & in proteins
MACYISLVFLLGFPVAVSGSHLHYIISMLSAPEAVPQYSIVVYADGLPIGRYNSDLHR
AQFLIPSLNVLTEHLELQTKFAQRWQVYQERKMAFLMGFLNRTSDRGTIHIYQRKFSCEL
EEDGTIGGYQEFALDQREVITFDRENEVFVPATQEAVTMLPRWNYLGNAGKNGMYMENE
CIEHIKLYLPIMMDLEKKVPPKVKVSSSESQTKLHCRVYGFYPRDVEVKWIKNGRDE
IHSEEAQILPNPDGTYQIRVSVGVTPPEGATYSCHIDHSSLEKALVVPFEGETKGYVIY
IVIGVVAALLALCLQDSLTKRPFLIFLGFKGRQSVSTDEQD

>xt_chain11 Chain=chain11 Position=GL173193.1:298049-301990(-) GeneID=ENSXETG0000009077
ProteinID=ENSXETP00000053837 Score=241.7 E-value=3.3e-68 Length=3942 Comment=Overlaps single
annotation & in proteins
HTLHYDVTIVSAPTGPVPRYLVSVMYVDGVQYGRYNSDTCHAEALYPSLALSEHIEKQTO
YAQGYEMRQRRLKSLNGLNKRTRGNGNIYQRRLACELYEDGTIGGYEIEAFDQELLVY
DKERVVYVPTQEAEMLSFIWNIRCGPTDTKIYMENDCIEHIKLYLPYISTDLERKVP
VNISSESSESQTKLHCRVYGFYPRDVEVKWIKNGRDEIHSEEAQILPNPDGTYQIRVSV
GVTPDKNSNYSCLVHDNSLETPIIVTFVPPNNGRSHLYFIIPVSVIFLLLLLILGIFIRK
Q

>xt_chain12 Chain=chain12 Position=GL173193.1:58340-69297(-) GeneID=None ProteinID=None
Score=237.9 E-value=3.7e-67 Length=10958 Comment=No overlapping annotations
APLGSHTLQYHISLVSDPGYGVPOFSVVMYIDGLQYGRYNSDTQCAQALAPSLNALSEHI
ENQTKYAEYEVWRHRLNFLMGFFNKSQKGDTHIYQRKACELHEDGTIGGYEIEAFDQK
EFFIFDKERVVYVPTQEAAMVSHLWNIRYDPTDQSKVYVENDCIENLKLKLYLPYISTDLEK
KVKVSSSESSESQTKLHCRVYGFYPRDVEVKWIKNGRDEIHSEEAQILPNPDGTYQIRV
SVGVTPPEGATYSCHIDHSSLETPLV

>xt_chain13 Chain=chain13 Position=GL173193.1:16498-36539(-) GeneID=None ProteinID=None
Score=235.8 E-value=1.8e-66 Length=20042 Comment=No overlapping annotations
APLGSHTLQYHISLVSDPGYGVPOFSVVMYIDGLQYGRYNSDTQCAQALAPSLNALSEHI
ENQIKFAQYEVWRHRLNFLMGFFNKTGTGDIHVYQRKACACERHDDGTVDGYQEIADF
GKELLVFDKERAEPVPTQEAEMVAHLWNHIESADSKVYVENDCIENLKLKLYLPYISTDL
EKKKVKVSSSESSESQTKLHCRVYGFYPRDVEVKWIKNGRDEIHSEEAQILPNPDGTYQI
RVRVGVTPPEGATYSCHIDHSSLEKPLV

>xt_XB-GENE-5873924 Chain=chain14 Position=GL176203.1:8516-10992(-) GeneID=ENSXETG00000015263
ProteinID=ENSXETP0000003387 Score=235.6 E-value=3e-66 Length=2477 Comment=Overlaps single
annotation & in proteins
PAAHTLHYAFTLVNSAQGSYFVAVFVDDAQIGRYSSDTQKVEPLDWMKERLEPQHWET
MTSWGKHYQNCHTKDMGFFIDRFNNTYGDRGPFYQVNFACEVHEDGGVSGHEAFANNGN
EFMFFDKGNMRFVVRREAQVWVNGQFSAIKHKNHMEQDCVQWMLYLRVYKGRLE
KKVQPTVKVFSRESGDATKLYCWWYGFYPRDVEDVVRVNRNGTEEDSEEHKEILPNPDGTYQ
IRITLTMQPADVDSYCHVDHSSVNGTQIYKLEPDKRGHVIACVAVVVFIMAVAVVVA
AVLKFRRR

>xt_chain15 Chain=chain15 Position=GL173193.1:280451-286573(-) GeneID=ENSXETG00000032875
ProteinID=ENSXETP00000019918 Score=234.4 E-value=5.4e-66 Length=6123 Comment=Overlaps single
annotation & in proteins
GSHTLHYHVTIVSAPTGPVPRYLVSVMYVDGVQYGRYNSDIRHAEALYPSLALSEHIEKQ
TQYQGYEMRQRRLKFLNGLNKRTRGNGNIYQRRLACELYEDGTIGGYEIEAFDQELL
VYDKERVVYVPTQEAEMLSFIWNIYCGPTDTKLYMENDCIEHIKLYLPYISTDLERKAV
PGNHVTLFSESESQTKLHCRVYGFYPRDVEVKWIKNGRDEIHSEEAQILPNPDGTYQIR
VSVGVTLDKNSNYSCLVHDSSLETPI

>xt_chain16 Chain=chain16 Position=GL173193.1:549443-557399(-) GeneID=ENSXETG00000033834
ProteinID=ENSXETP00000019884 Score=228.4 E-value=2.6e-64 Length=7957 Comment=Overlaps single
annotation & in proteins
ISSAAYPLQYSIIAYVDDVQYGRYNSDTRRCEPLIQSLMVFSEHLENQTKLAQQTET
QKAFMNLITDLLNKNYDNNINIVFQTKAACVLHEDGTVDVYKEVALDQKELIVFDKERQE
FVPATQEAVALLAQLLNQANAMEDKIFMENECTOHLTLYLPYIKNYLDKVKVPRVKSSE

ESGKTLHCRVYGFYPRDVEVKWIKNGRDEIHSEEAQILPNPDGTYQIRVSVGVTPEEG
 ATYSCHIDHSSLEKTLVVPFETETKNSHITYILIAVGIVLFLAFGLGFFIQRKMK

>xt_chain17 Chain=chain17 Position=GL174797.1:5412-7994(-) GeneID=ENSXETG00000025015
 ProteinID=ENSXETP00000053705 Score=228.1 E-value=3.6e-64 Length=2583 Comment=Overlaps single
 annotation & in proteins
 DNITMSYIVVLLFSLGVCVVHCGRHTLQYDISLVSSYSHSVPHYTTITAYIDGLQYGRYDS
 ATRRGQVLIPSMVIPSRSTHLKHMLLQIEFAQKQEVAEKQKMESLGMFLNKTHGNGEFHV
 FQRKFACELHKDGTISGYDEIAFDGKELMTFDREVRVHVVPVQEAALAMTQLWNRNYDYAK
 INKLYMENDCIENMKMYLLYVSTDLERKVLPPKVVSSSES DSGADLHCSVYGFYPRDVEV
 KWIKNGRDEIHSEEAQILPNPDGTYQIRVSVGVTPEEGANYSCHIDHSSLENPMVVPFV
 SNRRSISQIKIPIMAAFFLIFLALGMLTRMRNRKGSSESSQESLSSEENPE

>xt_chain18 Chain=chain18 Position=GL173193.1:198714-203900(-) GeneID=ENSXETG00000033980
 ProteinID=ENSXETP00000019901 Score=224.4 E-value=4.2e-63 Length=5187 Comment=Overlaps single
 annotation & in proteins
 MHMAYVDDLRFARYNNEINRCEFWIPALGTYERMTMQKNYMHNFESHQKTKLLTDLY
 NKTSGKDAIHVYQTKAACEMHEDGTITAYQEVAFDQKELIAYDYQTTETFIPTTPEAQTV
 QIWNENYVLEKVFLENYCTHRLSTYLPALASDLEKLVPPKVVSSSESESGTKLHCRVY
 GFYPRDVEVKWIKNGRDEIHSEEAQILPNPDGTYQIRVSVGVTPEGGATYSCHIDHSSL
 EKTLLVVPFPNGSYQSGSTETDIEN

>xt_chain19 Chain=chain19 Position=GL173193.1:181934-186043(-) GeneID=ENSXETG00000030073
 ProteinID=ENSXETP00000063559 Score=219.2 E-value=1.4e-61 Length=4110 Comment=Overlaps single
 annotation & in proteins
 GNMAGILPLVFLIFSLGTSVVYSGSHTLKYITKLTAPYPGLPDYLMHAYVDDLRYARYN
 KEINRCEYWIIPALGAKWEPMTMQKNYMHNFESHQKTKLLTDLYNKTSGKDAIHVYQTK
 SACEMHEDGTITAYQEVAFDQKELIAYDYQTTETFIPTTPEAQIVAQIWNKNYAKIEKIFL
 EKFTNRLSKYLNALASDLEKLVPPKVVSSSESESGTKLHCRVYGFYPRDVEVKWIKN
 GRDEIHSEEAQILPNPDGTYQIRVSVGVTPEGGATYSCHIDHSSLEKTLVVPFPEYNNRS
 ILYIIIPICAAALLLMIGLGLFLFKKGGKSHSNGAAYSYTWTIDIGR

>xt_XB-GENE-5757686 Chain=chain20 Position=GL173193.1:168702-175190(-)
 GeneID=ENSXETG00000014445 ProteinID=ENSXETP00000060129 Score=219.0 E-value=1.7e-61 Length=6489
 Comment=Overlaps multiple annotations; one in proteins
 GNMAGILPLVCLIFSLGTSVVYSGSHTLRFHVTKLTAPYPGFPDYMLLAYVDDLRFARYN
 NELNRCFEWIPALGALNERITKQKNYTHNFKISHQKTKLLTDLYNKTSGKDAIHVYQTK
 SACVMHEDGTISAYQEVAFNGKELIAYDYQMGTFIPTTPEAQVAAQVWNEYAKIEKVF
 ENFCTHRLSWYLPALASHLEKLVPPKVVSISES DSGAELHCRVYGFYPRDVEVKWIKNG
 RDEIHSEEAQILPNPDGTYQIRVSVGVTPEGGATYSCHVEHSSLEKTLVVPFGYRLRIS
 LVAIHVAVMLLLLNILVTAKFLIKAGQRHNPNGSYQSGSTETDIEN

>xt_chain21 Chain=chain21 Position=GL173193.1:130442-134120(-) GeneID=ENSXETG0000009089
 ProteinID=ENSXETP00000019892 Score=216.8 E-value=7.2e-61 Length=3679 Comment=Overlaps single
 annotation & in proteins
 MHMAYVDDLRYARYNNEINRCEYWIIPALSTLYERMTMQKNYTHNFKISHQKTKLLTHLY
 NKTSGKDAIHVYQTKSACEMHEDGTISAYQEVAFDQKELIAYDYQMETFIPTTPEAQVVA
 QIWNEDYAKREKIYLNKLCYRFSKYLNALASDLEKLVPPKVVSISES DSGAELHCRVY
 GFYPRDVEVKWIKNGRDEIHSEEAQILPNPDGTYQIRVSVGVTPEGGATYSCHVDHSSL
 EKALVVPF

>xt_chain22 Chain=chain22 Position=GL173063.1:992422-996364(-) GeneID=ENSXETG00000030689
 ProteinID=ENSXETP00000035918 Score=215.5 E-value=2e-60 Length=3943 Comment=Overlaps single
 annotation & in proteins
 MGIVGFLLLGLGAHIASCSSHILHYCITAVVSDPAYGLPQYSIVGYVDNLLIGRYSSQTR
 RPEFLIQSLRDLFEHTEELTRIIHKKENADTMIMKVILSSLNKTGDGDFHIFQIKYACEL
 LEDSSISGKEELALNAKTFITYNTENPEYIPVPAALTAQAQWTELYAKLEKDYMEHECV
 SHLKLKLYLKKDLEKLVPPKVVSSSESESGTKLHCRVYGFYPRDVEVKWIKNGRDEIH
 SEEAQILPNPDGTYQIRVSVGVTPEGGATYSCHTDHSSLEKTLVVPFAEANNMGLLYIL
 IALAVTVILLAVLLGVFIHRKRSRQPVNTEETD

>xt_chain23 Chain=chain23 Position=GL173193.1:444157-447998(-) GeneID=ENSXETG00000034352
 ProteinID=ENSXETP00000058545 Score=207.6 E-value=3.8e-58 Length=3842 Comment=Overlaps single
 annotation & in proteins
 MACYISLVFLGFPVAVYSGDHTLQFYVAVSSTAYPGLLQYSIIAYVDDVRYGRYNSDTRR
 CESLIQSPMVFSEHLDGQTNMAQOTEITQKVLNMLIIDLLNKTHTDINIVYQTKAACVLE
 DGTVDVYEEVALDGKELVDFDKERVEFVPATQKAVLLAQWLNKQAFAMEDKIFMNECTQ
 HLTLYLPYIKSYLDEKVPVKLSSSESESGTELHCRVYGFYPRDVEVKWIKNGRDEIHSE
 EAAQILPNPDGTYQIRVSVGVTPEEGATYSCHIDHSSLENPLVVPFETETKNSHITYILI
 AVGVIVLFLAFGLGFFIQRKMKGLKGRSQSVSSEEQD

>xt_chain24 Chain=chain24 Position=GL173193.1:512159-517175(-) GeneID=ENSXETG00000033504
 ProteinID=ENSXETP00000058304 Score=206.1 E-value=1e-57 Length=5017 Comment=Overlaps single
 annotation & in proteins
 MACYISLVFLGFPVAVYSGSHTLKYIITMLSTAYPGLPQYSVIAYVDDVRYGRYNSDTRH
 AEFLIQSLMVFSEHSEGLNKQAQLIETMQAKIMNFIMDSSNMTYDINIVYQIAACVLE
 DGTVDVYEEGALDGKQLVDFDKETVEFVPTREAVLLAQWVNKQAFAKKYKILMNRCTQ
 HLTLYLPYIKSYLDEKVPVKLSSSESESGTKLHCRVYGFYPRDVEVKWIKNGRDEIHSE

EAAQILPNPDGTYQIRVSVGVTPPEGATYSCHIDHSSLENPLVLFETETKNSNITYILI
AVGVIVLFLAFGLGFFIQRKMKGFKDSQQSASTNEQE

>xt_chain25 Chain=chain25 Position=GL173193.1:585862-595649(-) GeneID=ENSXETG00000030785
ProteinID=ENSXETP00000053844 Score=198.9 E-value=1.9e-55 Length=9788 Comment=Overlaps single
annotation & in proteins
MVSVAAPGLPQYSITAFVDDVMIGRYSSDTRDIETFFVPVPEILSWHVRRKMEYAFHLEEE
EKKGLEKVMVFSNMTYGDADIHVYQKKFICELDDDGSVGGYQEAAPFDGESFATPDKEKGV
FVPLTQKASDVIQQCANDLYAKSDKFYMENYCIHDLKLYLPYLKMLDKKVCVKVSISS
LESNNTKLHCCVYGFHPREVDVKWIKNGRDEIHSEEAQILPNPDGTYQIRVSVGVTPPE
GGATYSCHIDHSSLENLTVVPPF

>xt_chain26 Chain=chain26 Position=GL173193.1:530734-535327(-) GeneID=ENSXETG00000032895
ProteinID=ENSXETP00000061810 Score=195.7 E-value=2.4e-54 Length=4594 Comment=Overlaps single
annotation & in proteins
MACYISLVFLLGFPVAVYSGSYTLKYYITMLSTAYPGLPQYSVIAYVDDVRYGRYNSDTRH
AEFLIQSLMVLSEHLDGNLQKQALIEITQKALMNFIMDSSNKTYDINHVFQTKAACVLHE
DGTVDVYKEEALDGKQLIVFDKETMEIVPATGEEAVALQWLNKQAYPKYKISMENECTQ
HPLPLYLLYIKNYLDKVPNVKLSGSESENGTELHCRVYGFYPRDVEVKWIKNGRDEIHSE
EAAQILPNPDGTYQISVGVKPEGATYSCHIDHSSLEKALVVPFEKETKNSHITYILIAV
GVIVLFLAFRLGFFYKKKVK

>xt_chain27 Chain=chain27 Position=GL173193.1:490297-495395(-) GeneID=None ProteinID=None
Score=191.1 E-value=4.1e-53 Length=5099 Comment=No overlapping annotations
CFLGSPFLQYITMLSTAYPGLPQFSIIAFVDDVRYGRYNSDTRRAEPLIQSLMVFSEQL
ESLTKLAQQTELTKLFMTLIMDLLNKTYGNVFPQIAACVLHEDGTVDICKDTALDGKQ
LLVFDKERVEIVPVTQEAVALQWVNKQGYAKKYKIFMENECTQHLPSPFLCISYLDKK
VPHVKLSGSESGTKLHCRVYGFYPRDVEVKWIKNGRDEIHSEEAQILPNPDGTYQIRVS
VGVTPPEGATYSCHIDHSSLENPLV

>xt_chain28 Chain=chain28 Position=GL173193.1:238026-245746(-) GeneID=None ProteinID=None
Score=190.2 E-value=4.1e-53 Length=7721 Comment=No overlapping annotations
LLSGSHLLQYHITRLTAPSPGLPELIKIGYVDGLRYARCTSETHRCEFLIPALETVYERM
TLKEKCINYEFSQFERLQLLTDIYNKTSGFHIIYQMKFGCVQDDGTGTGIYQEVAFDSKE
LITYDKHTGTFTPTSQEARIVAQLWNKNYAKVEKIYMENCTPRLALFPPAIAADLEKKK
VKVSSSESDGEQKHLHCRVYGFYPRDVEVKWIKNGRDEIHSEEAQILPNPDGTYQIRVSV
GVTPEEGATYSCHIDHSSLENPLV

>xt_chain29 Chain=chain29 Position=GL173193.1:411576-415530(-) GeneID=None ProteinID=None
Score=183.3 E-value=9.4e-51 Length=3955 Comment=No overlapping annotations
LPPGKHVLEYMTMTSAPIPGISTFSAITAHVDGVQHGRYTSETGRAEPLIPSLSMLTEHL
DMQTHFARHWVHVYQRKYICELQDSGKVLGYHAFVFNKGEVIAFDREVEVFVPMPEAEL
LIPSWNKYLDVVKHKKAYLKNVCIEHLKLYLSYMLPELDRKKVKTSISRSDSGKRLCCHV
YGFYPPKVVHIKVMKNGTEEVSPPEFKHILPNPDGTYQIRVSVGVTPPEGATYSCHVDHSS
LNSTMVI

>xt_chain30 Chain=chain30 Position=GL173193.1:614050-621836(-) GeneID=ENSXETG00000030515
ProteinID=ENSXETP00000035927 Score=156.4 E-value=6.8e-43 Length=7787 Comment=Overlaps single
annotation & in proteins
DINHVTQTKAACVLHEDGTVDVYQVEALDGKQLIVFDKDLVEFVPATQEAVALLTQLWTKH
YANAKDKLFLENCTQHLLKLYIPYIKSYLDKKAVPRVKMSNSESGTKLHCRVYGFYPRD
VVVKWIKNGRDEIHSEEAQILPNPDGTYQIRVSVGVTPPEGATYSCHIDHSSLEKTLVV
PFGDNDLPFSIHSVETETKN

>xt_chain31 Chain=chain31 Position=GL173193.1:110268-114833(-) GeneID=None ProteinID=None
Score=146.2 E-value=6.8e-40 Length=4566 Comment=No overlapping annotations
GLLGSHTLQYDVTLVSAPTPGVPQFSIIIVYVDGVPYGRYNSDIGRAEWNERYDSTYTEIF
MEIDCIEHMKMLPYISTDLEKKNVQISSSESESGTNLHCSGYGFYPRDVEVKWIKNGRD
EIHSEEAQILPNPDGTYQIRVSVGVTPPEGATYSCHIDHSSLENPLV

>xt_ENSXETG00000002125 Chain=None Position=GL173193.1:152859-156665(-)
GeneID=ENSXETG00000002125 ProteinID=ENSXETP00000004543 Score=152.6 E-value=2.3e-45 Length=0
Comment=Not detected in genome search
AGKDAIHVYQTKSACVMHEDGTISAYQEVAFNGKELIAYDYQMGTFIPTTPEAQVAAQVW
NEDYAKIEKVFLFNCTHRLSWYLPALASHLEKLVPPKVKVSISESDSGAELHCRVYGFY
PRDVEVKWIKNGRDEIHSEEAQILPNPDGTYQIRVSVGVTPPEGATYSCHVEHSSLEKT
LVVPPFEYNNRSILYIIIPICGALLLMIGLGLFLFKKGGKDPNGSYQSGSTETDIEN

>tn_chain01 Chain=chain01 Position=Un_random:91605827-91610885(-) GeneID=ENSTNIG00000001016
ProteinID=ENSTNIP00000002167 Score=242.6 E-value=1.4e-75 Length=5059 Comment=Overlaps single
annotation & in proteins
VTVSSGIPNLPEYVVVGLDDVQTDHYDSITGRAEPRQEWKVKVAEHPEDWEGQTQIAA
GIHQNFKATIEILKQRFNQTGAGVHIFQLMYGCEWDETERNGFHWQGYDGEDFISFKL
KEEIIYAAKREAAYSARRWNQDRAEIEQQKNYYQTCIKWLKLYLDYGRSSLMRTEPPPLS
VPSRKEPSSPVTCVATGFYPPGATLSWRRGEEELHEVEPEGILPNPDGTFQMSAALLD
SSVPPEDWSSYKCVFQLSGGQEVPTSLDKNPSITTAIVGVGAVLVVLVGLAIAGFIMYRQ
KKGESWR

>tn_chain02 Chain=chain02 Position=Un_random:91642436-91643710(-) GeneID=ENSTNIG00000001085 ProteinID=ENSTNIP00000002941 Score=236.3 E-value=1.1e-73 Length=1275 Comment=Overlaps single annotation & in proteins
MKIAKFFVVLAAALAVYRSRAVIHTLKYFYRSGVVPNPFPEFSGAALLDGAIEIVHYDSDR
RAEPRQEWEMENLSAGDPQYWERETAKFLGAQQTYNARIEILKPRFNQTAGVHVIFQNMYG
CEWDDTDERNGFDQWGYDGEDLMSFNLTEEIIWVAAKREAEITHTWNOQDRAEMEYWKNY
LTRECVSSLKTYLDYGRSSLMRTELPSVSLLOKEPSSPVTCMATGFYPPGGATLSWRRGEE
ELHEEVEPGEILPNPDGTFQMSAALDLSSVPPEDWSSYKCVFQLSGGQEVPTSLDRKRIR
TNWENPNTSTIVRVVISVLLVVLGTIAGFLVYRQKKGKCPGRHHSAGEKKTNMQL

>tn_chain03 Chain=chain03 Position=Un_random:91633011-91634212(-) GeneID=ENSTNIG00000005106 ProteinID=ENSTNIP00000000273 Score=233.9 E-value=8.8e-73 Length=1202 Comment=Overlaps single annotation & in proteins
MNSLVFLLLALAKSVNCFVAVTHTLNYFHTVSSGVPNPFVAVVMVDDVHISHYDSNTR
RTQPRQEWEMENVNADDPQYWEGETGNFMGAHQAYKARIEILKPRFNQTAGVHVYQSTYGC
EWDDETDDEINGFNQGYDAEDFISFKLKEIIVAAKREAEITKRKCDQNRANTDYLSYY
TQICIKWLQKFDYGRSSLMRTELPSVSLLOKEPSSPVTCMATGFYPPGGATLSWRRGEE
LHEEVEPGEILPNPDGTFQMSAALDLSSVPPEDWSSYKCVFQLSGGQEVPTSLDRKRIR
NRGKTGGGGDAENPNTSITTAIVAVAVLVLVLAIGFLVYRQKVLGKCPRAEGHPE
VKEPLKPAPT

>tn_chain04 Chain=chain04 Position=Un_random:41327814-41340383(+) GeneID=ENSTNIG00000003024 ProteinID=ENSTNIP00000005594 Score=228.2 E-value=2.6e-71 Length=12570 Comment=Overlaps single annotation & in proteins
SLPVTHTLKYFYRSGVVPNPFPEFLAVVMVDDVQISHCDSNTRRAEPRQEWMEKVTADDP
QYWERETAKFLDAQQTYNARIEILKPSFNQTAGVHVYQSTYGCWDDDETKKINGFSQGG
YDGEDLLRFKLEAICAAKPEAEILKRSKDEKKAEMESLKYLLTHECVWLKYLQYDGR
SSLMRTELPSVSLLOKEPSSPVTCMATGFYPPGGATLSWRRGEEELHEEVEPGEILPNPDG
TFQMSAALDLSS

>tn_mhc1zea Chain=chain05 Position=Un_random:59040704-59042474(+) GeneID=ENSTNIG00000003449 ProteinID=ENSTNIP00000006036 Score=212.9 E-value=9.1e-67 Length=1771 Comment=Overlaps single annotation & in proteins
KSFMVLLVLLGSLVVDGGLSEIHSLHYIYALSRLPVRNRPGEHFTAMGMLDDHMIDYFD
SDKQVKVPKQTMKTELEENYWEKGTQSRKSKQWFKVNIKILKDRFRQNDSDIHVLQWV
HGCNIDTSGNETEFLHGLIDMYSYDGSFLLTFDEANENWVAPNDALQTKRWDLLQVLKE
YTKVYLKKECVTWLERFLKYQKENTAPVKPAVYALSREANVQANIVLTCMATGFSSINTI
VQIKRDLGLVTKDDGDVRFNGDGTQYKRDQVEIPKSDSKYTCVEVHESGLHEIRVWDT
DEGNTAVVVGAVVVVVVVVVLVIGIGLYMAVKK

>tn_chain06 Chain=chain06 Position=Un_random:91647125-91648450(-) GeneID=ENSTNIG00000000474 ProteinID=ENSTNIP00000003437 Score=200.0 E-value=1.7e-62 Length=1326 Comment=Overlaps single annotation & in proteins
EGGGVGRIRASLGEDGAHLSTGESAGTSLCFPHLCLCLSGVHVLQRMYGCEWDDETNE
SLSGFYQYGDGEDFISFNLTEEAWVAARREAEVTKRRWEKHEGVKHWKTYLTHRCISW
VQEVYVYRQSSLVTELPSPVSLLOKEPSSPVTCMATGFYPPGGATLSWRRGEEELHEEVEP
GEILPNPDGTFQMSAALDLSSVPPEDWSSYKCVFQLSGGQEVPTSLDRKRIRTNWGTGTG
GGDDGEKNGLVLLWEQRLLLFSS

>tn_chain07 Chain=chain07 Position=7:4474373-4475478(-) GeneID=ENSTNIG00000009992 ProteinID=ENSTNIP000000012887 Score=192.6 E-value=1e-60 Length=1106 Comment=Overlaps single annotation & in proteins
MFYTCIFICLFSVTDHFLFAEKHYLYLTAASSGVNTQPEFVAQLVVDNLPGGYCDNSNGKI
QPRRDWAIKMIHEDPKLELYENECVRYQHTYRGIKNIKQFNQSGGIIHFQRIHGCE
WDDDETGEVTAFNQFGYDGEDFISLDPVMTWVAAPQALITKYKWDSEIYIYLIKMFNS
HLCPRDLKMYLQFLKRFQKQKQPSVFLLOKTPSSPVSVCHVTGFHPENGVVFWRKDGQKL
QEEVEYTEILPNHDGTFQTRVDELELSSVSPPEEWSRYECVFHPPGDQEDITLKLKDKVIRT
NWGKTGSACCVICQSETSLKCV

>tn_chain08 Chain=chain08 Position=Un_random:91618580-91625876(-) GeneID=ENSTNIG00000005106 ProteinID=ENSTNIP00000000273 Score=191.5 E-value=3e-60 Length=7297 Comment=Overlaps single annotation & in proteins
YLSVNHHTLKYLYTGSSGIPNLPEYVIVGLVDDVQTDHYDSITGRAEPRQEWRRKVKAEHP
EDWEGQTQIAAGIHQNFKATIEILKQRFNQTGGGYDGEDFISFKLTEEIIWVAARREAEII
KHKWDQERAWIEQLKTYFTQICIDWLKSSVSLLOKEPSSPVTCMATGFYPPGGATLSWRRG
EEELHEEVEPGEILPNPDGTFQMSAALDLSSVPPEDWSSYKCVFQLCGGQEVPTSLDRTO
TSDLSLINCR

>tn_chain09 Chain=chain09 Position=Un_random:83418349-83419239(-) GeneID=None ProteinID=None Score=141.6 E-value=4e-45 Length=891 Comment=No overlapping annotations
LLAVKHGLRYFITLSSGVTEFPPELVLMVLDNFQGGWCDGKQIY*Q*FKSFVI*QLANVC
LSGIHIFQRIHGCDWDDTGEVTFGNQFGYDGEDFISLDPVAMTWWAAKQALITKYKWD
SDTVHLQGRKIFNTHICPEKLMYLOYLHRFQPSVFLLOKTPSSAVSVCHVTGFHPENGVV
FWRKDGKQLQEEVEYTEILPNH

>tn_chain10 Chain=chain10 Position=7:4476900-4477892(-) GeneID=None ProteinID=None Score=102.9 E-value=3.2e-34 Length=993 Comment=No overlapping annotations
LLAVKHSLRYFITSTSGVTEFPDFVAQTVVDNLFQGGWCENNAQPRNEWAKKALDPVTMT

WVAAKPQALITKHKTRSDTVHLQEEKIFNTHICPERLKKYLQFLKRFQKKQPSVFLLOK
TPSSAVSCHVTGFHPENGVVFWRKDGKQLQEEVEYTEILPNHDGTFQTRADLELSSVSPE
EWSRYECVFPHPGDQEDITL

>tn_ENSTNIG0000004925 Chain=None Position=Un_random:74651122-74651415(+)
GeneID=ENSTNIG00000004925 ProteinID=ENSTNIP00000007590 Score=84.5 E-value=2.5e-24 Length=0
Comment=Not detected in genome search
THTLKYFHTVSSGVPNPFVAVVMVDDVHISHYDSDTRRAQPRQEWEMNVS AEHPQYWE
RETGNFMGAHQAYKAWIEILKPSFNQTGGEMSSPLHC

>tn_ENSTNIG00000004145 Chain=None Position=Un_random:66895563-66896295(+)
GeneID=ENSTNIG00000004145 ProteinID=ENSTNIP00000006764 Score=171.7 E-value=2e-50 Length=0
Comment=Not detected in genome search
GVHTFQWMVGCWEWDEIDEINGFDQYGDGDDFLSFKLKDGTVVAAKREAEITKRKWDYD
GAGIDYWKYYHQNQCIKWLKYLGRSSLMRTELPHSVSLLQKEPSSPVICMATVFPYD
GATLSWRRGEEELHEVEPEPEILPNPDGTFQMSAALDLSVPPEDWSSYKCVFQLSGGQE
VHTSLDRKQIRTNWKGTTGGGDDGE

>tn_ENSTNIG00000001973 Chain=None Position=Un_random:46457180-46457895(+)
GeneID=ENSTNIG00000001973 ProteinID=ENSTNIP00000004370 Score=145.6 E-value=1.1e-42 Length=0
Comment=Not detected in genome search
GVHIFQWMYGCEWDEIDERNWFCQWGYDGEDFISFKLKEEYVAAKREAEIITHRWQON
RALIEQQKYYQICIKWLQKFVDYGRSSLMRTELPSVLSLLQKEPSSPVTTCMATGFYPGGA
TLSWRRGEEELHEVEPEPEILPNPDGTFQMSAALDLSVPPEDWSSYKCVFQLSGGQEV
TSLDRKRITTNWEG

>dr_mhcluba Chain=chain01 Position=19:7696486-7721711(+) GeneID=ENSDARG00000075963
ProteinID=ENSDARP00000020667 Score=272.5 E-value=1.1e-80 Length=25226 Comment=Overlaps single
annotation & in proteins
MQSLIGLLLVCLQYASGATHSWKAYTGTGLTEFPEFVALNLIDDQLMGYFDSKTNRF
KSQFQWMDNLGKEYDEQQTNILGYPEVFNKNIKVMMERFNQTQGVHTFQFMYGCEMDD
DGNKQVHWQIGYDGEDFISLDKKTLTWTAANSQAMTTKVWDSTGAEANYWKYLENECI
EWWQYVGYGKDTLTKVSPQVSLQKSSSPVCHVTGFYPSGLKISWQRNGQDHEDEV
ELGELIPNEDGTYQRTSFLNVKPEEWKDKFSCVVEHQSKTINSILTEDEIRTNNTAPL
GIIIGIVVAALLVAIAVAGFVYRRHKGFKVPVQNTSDGGSNSRT

>dr_mhclzba Chain=chain02 Position=1:47656107-47658002(-) GeneID=ENSDARG00000036588
ProteinID=ENSDARP00000053152 Score=236.6 E-value=7.6e-70 Length=1896 Comment=Overlaps single
annotation & in proteins
MGSAVFLSFAVMLVAVVPAWTEKHSLYIYTALSRPVNLPGIYEFTAMGLLDDRQIDYNN
SEQQKIPKQHWMEKMQEDYWEKGTQSRKSKEQWFNVNINILMRMRHNKSDVHVLQWR
HGCEIESQGNVDRFRSGIDEYSYDGRNFLSFDDAESQWVAPVEEALPTKRKWDNVPILNQ
YTKGYLEKEKCVDWLNKRFREYGDKELEGGSSPEVHVFAKRIINGKIKLKLTLATGFYPKD
VILNIRKYRITLPDNEVESTGVRPNEDGTFQLRKSINIYEDEKAEYDCYVSHHTLKEPII
KKWDGCELDGPEGSPGIIAGAIIVLVVLAAGAVYFLRKRSGNNVVKPSSVPTTISG
NKDEKCSMLPGSDSGGSSDGSKSSPTNSQEKMDIV

>dr_mhclzaa.1 Chain=chain03 Position=1:47627093-47631877(-) GeneID=ENSDARG00000086877
ProteinID=None Score=236.0 E-value=1.2e-69 Length=4785 Comment=Overlaps single annotation; not
in proteins
FNTEKHSLYIYTALSRPVNLPGIHEFTAMGLLDDRQIDYNNSDQKQKIPKQHWMEKMQ
EDYWEKGTQSRKSKEQWFNVNINIDLHVLQWRHGCEIDSQDDVDRFSKGIDEYSYDGENF
LSFDDADSQWVAPVDAALPTKRKWDNVPILNQYTKGYLEKEKCVDWLNKRFREYGNELREG
VFARSIINGKAKLKLTLATGFYPKDVYLTIRKYRMALLDKDLESGVVRPNHDGTFQLRM
STYIYEDEKAEYDCYVSHHTLKEPII

>dr_mhclzaa.2 Chain=chain04 Position=3:1079142-1080155(-) GeneID=ENSDARG00000092162
ProteinID=ENSDARP00000118368 Score=234.8 E-value=2.3e-69 Length=1014 Comment=Overlaps single
annotation & in proteins
XHRPESSLGKSPSAIFRCFPVVLFLRRKAEPGKHTHFLSVMMSSGGVASGICALLCVFLL
CGDLPSAQGEKHSLYIYTGLSKPLDLPGIYEFSAMGLLDDRQIDSYNSEEQRKIPKQW
MKEKMQEDYWEKGTQSRKSKEQWFNVNINILMDRMRHNMSDLHILQWRTGCEVEIQGSEV
KFSKIDEYGYDGENFLSFDEAESQWVAPVEEALPTKRKWDNVPILNQYTKGYLEKEKCV
WLNKRFREYADEELRNGSSPDVYKLTCKSTKDETKLKLTLATGFYDKDVMNLNRRNCLPE
DETESTGVRPNHDGTFQLRKSVEIKEDQIDEYDCHLTHRTLKNPIIVTQVNFPPVVE

>dr_mhclzaa.3 Chain=chain10 Position=3:1114702-1122867(-) GeneID=ENSDARG00000074765
ProteinID=ENSDARP00000096436 Score=222.9 E-value=7.5e-66 Length=8166 Comment=Overlaps single
annotation & in proteins
MTSFDRSFTLLYLCLFHVIMSSFRAEKHSLYFIYTGLSRPVDLPGIYEFSAAGLLDDRQI
DSYNSEEQRNIPKQWMEKMQEDYWENRTQSRKEKQLWFYDNVHLLIDRNQSTSVLHV
LQWRHGCEVEKQGENHSPNKSIDEYGYDGEDFLYDDEAESRWVAPVEEALPTKRKWDNLP
ILNQYTKDYLEKEKCVDWLNKRFREYADEELRNASPPDVFKLTKRSTKDETKLKLTLATGF
YPNVMDLNIIRYRHYLPEQETISTGVRPNHDGTFQLSKSVEIKEDQIDNDCSVSHPTFK
QPIIKWDGTDLDGIYTGHPPEVTPVIVIGSVLIFLAILGVVFLVMKCGEQDSSSDTTE
CTSILLGFIADHLNKLTEGENW

>dr_mhclzaa.4 Chain=chain11 Position=3:1089527-1090540(-) GeneID=ENSDARG00000092162
ProteinID=ENSDARP00000118368 Score=221.0 E-value=2.3e-65 Length=1014 Comment=Overlaps single
annotation & in proteins
SSSEKHSLYIYITSLSKPLDLPGIYEFSSAMGLLDDRQIDSYNSEEQQRKIPKQQWMEKMQ
EDYWEKGTWRSRKSKEQWFNVNVDLHLQWRHGCEVEIQGSEVFKSFGVSEYGYDGENLLY
FDDAESRWVAPVEEALPTKRKNWNIPIILNQYTKGYLEKECVDWLNKRFREYADEELRKAKL
TKKSTKDETKLKLTLCLATGFYPNDVMLNIRRNCLPEDETESTGVRPNHDQTFQLRKSVEI
KEDQIDEYDCHLTHRTLKNPVTA

>dr_mhclzfa Chain=chain05 Position=3:508250-511459(-) GeneID=ENSDARG00000088022
ProteinID=ENSDARP00000095791 Score=231.8 E-value=2.2e-68 Length=3210 Comment=Overlaps single
annotation & in proteins
MIMDVKVLTTLLCVFLLYGALPLCRAEKHSLYIYTGSRPVDLPGIYEFSSAMGLLDDRQI
DSYNSREQRKIPKQQWMEKMQEDYWEKGTQSRKSKEQWFNVNVIILDRMRNRNLDLHV
LQWRHGCEVEIKGSEVFKSGKIDEGYDGENFLAFDEAESQWVAPVEEALPTKRKNWDSVP
ILNQYTKGYLEKECVDWLNKRFREYANEELRNGSPPEVHVFAVKSISDKTKLKLTLCLATGF
YSKDTMLVIRRNRLPEKKTSTGVRPNHDQTFQLRKSVEIEQDETDEYDCYMTHTLTKGP
VIARWDGKCKDCLPNLNIWVAVAGVLMGLVVALLLVLLKKKIIDLRLQRLSGSQSPLYQI
QADSESEDYHN

>dr_mhcluca Chain=chain06 Position=19:7669183-7671181(+) GeneID=ENSDARG00000092731
ProteinID=ENSDARP00000117273 Score=231.6 E-value=1.5e-68 Length=1999 Comment=Overlaps single
annotation & in proteins
MADRITQMQKMEYILLVCLPYVDGATHWNAYTATGLSQPFPEFVAQNLIIDGQLMGY
FDSKNTCFKSGFQWMEKMLGTADEQQTNILQRHTAKFKNNIKVAIERFNQTOGVHTLQE
IYGCEWDDQTNLNAFRQYGYNGEDFLTLDFMEMRWITPVQQAMITTOQKWNDRGFIESD
MNYFRSECIEWLQKYLEYKSSLMKTVSPQVSVLQKYFSPVVCATGFYPSGIKISWQK
NGQDHDEDELGELLPNADGTFQKRSTLNVKPEEWKNSKFCVVEHQGERILTEDEIRTN
NPSATTGIIIGVLSAVLLELFAVAGFVVYWKKGFKVPANPVNSDNDSCGCGSGSDSVS
HKSSDSTSDISSVCSNDSSTTALLRNRQKYDKN

>dr_DARE-ZE*0601 Chain=chain07 Position=Zv9_NA257:92605-93624(+) GeneID=ENSDARG00000076734
ProteinID=ENSDARP00000103264 Score=229.8 E-value=6.3e-68 Length=1020 Comment=Overlaps single
annotation & in proteins
MMSGGVASGICALLCVFLLCGDLPSAQGEKHSLYIYITGLSKPLDLPGIYEFSSAMGLLDD
RQIDSYNSEEQQRKIPKQQWMEKMQEDYWEKGTQSRKSKEQWFNVNVIILDRMRHNKSD
LHVLRHGCEVEIQGSEVFKSGVSEYGYDGENLLSFDETESQWVAPVKEALPTKRKNW
NVPILNQYTKGYLEKECVDWLNKRFREYADAELRNGSPDPVYKFTKKSTKDETKLKLTLCL
ATGFYPSDVMLNIRRNRYLPEQETISTGVRPNHDQTFQLSKEVEIKEDQIDEYDCSVSHR
TFKQPIIIKWDGTDLDGKTYGPPETVPVVGSVLILLAILAVVFLVMNYDGECKGLF
IHFTLLKDCGLLLP

>dr_mhclzea Chain=chain08 Position=1:47617360-47618591(-) GeneID=ENSDARG0000001470
ProteinID=ENSDARP00000091917 Score=229.3 E-value=1.5e-67 Length=1232 Comment=Overlaps single
annotation & in proteins
MAVFAALFVSVMLFTVVPVPAWTEKHSLYIYITALSRLPVHLPGIYEFTAMGLLDDRQIDYNN
SQEQKIPKQHWMEKMQEDYWEKGTQSRKSKEQWFNVNLIKILMERMRHNNSDVHVLQW
RHGCEIDSQGNDRVFSKIDEGSYDGRNFLAFDDADSQWVAPVEEALPTKRKNWNVPIINQ
YTKGYLEKECVDWLNKRFREYGDQELREGSPDPVHVFAKKIISGKAKLKLTCMVTGFYPK
DVILITRKYRTALSDNEVESSGVRPNPDGTFQLRKNSTNIYEKAEYDCYVAHRTLKEPIIK
KWDGECQDCSSGTPIGTIFGALIGVLLVLAIVIGGAVYFLANTRMGWRNAL

>dr_mhclzca Chain=chain09 Position=1:47636615-47642920(-) GeneID=ENSDARG00000069471
ProteinID=ENSDARP00000063528 Score=223.4 E-value=2.1e-64 Length=6306 Comment=Overlaps single
annotation & in proteins
MAVLAVLFSVAVMLLSVSVRTEKHSLYIYITALSRLPVDQSGIYQFTAMGLLDDRQIDYNN
NIDQKIPKQHWMEKMQEDYWEKGTQSRKSKEQWFNVNVIILDRMRHNNSDLHVLQWR
HGCEIEIQGNEHRFSKIDEGSYDGRNFLSFDDAESQWVAPVDAALPTKRKNWNVPIINQ
YTKGYLEKECVDWLNKRFREYGDQELRDGSPDPVVFARRIISGKIKLKLTCMATGFYPKD
MILTIRKYRTLPNDLDSGVRPNQDGTFLRKNSTNIYEKAEYDCYVNHRTLKEPIIK
VRWNGEYLSEPPAIIAIIIGVLLVLAIVGVTVWILKKNIIIGNKDEKRSMPNGSANYGR
GSSA

>dr_CABZ01031753.1 Chain=chain12 Position=22:5245353-5261410(+) GeneID=ENSDARG00000059039
ProteinID=ENSDARP00000076471 Score=210.2 E-value=1.5e-60 Length=16058 Comment=Overlaps single
annotation & in proteins
MKIIVVFCIPFVCSGYHSLITTYTGIRGASETPIFIAVAILDGEQIDYDYSVTNKLIPK
QDWMKEYASEEIKWEDCKIRGDVHQIYRSNINVLQRFNQSNGVHVYQRMYGCDWDEETE
ESHGFPDQYGDGEDYVMLDMREIRYITPVQGEITVQKWNNSREQLKTLQHYKYECVYWR
LKQFLGLRKADLEIRDPEVFLQKNPSSPVECFATGFYPSGVIITWLNKGQDHGENVELR
ELKPNEDGTFQRSSSLHVSQEDWKKSLYTCVVEHQRMITQKSGDEIKSNSSSEVFITPER
VLAAILLAALVICLCICYLVYPKPKLPYQPVNQNEEMYEYIVKNE

>dr_CABZ01031754.1 Chain=chain13 Position=22:5239594-5241973(-) GeneID=ENSDARG00000039164
ProteinID=ENSDARP00000057223 Score=144.8 E-value=1.7e-42 Length=2380 Comment=Overlaps single
annotation & in proteins
MSIMKFIIFFNLPFVYSELHTFVVITYTAMGNQTFFAAATLDGQQIGHYDKNTKTLVPKQ
DWNEQIKSKTPWKVDTLTIETVQRTYNNMHDLIEQFPQSNAGHYQGRHGCSWEDKTNHY

DEFHDYAYDGDFFITLVDVKKIKYRASVSKAKYIVKKNNDREQNKTLKEYYKLGCIYWLI
EFLVFTKNAFGGTAPKMFLLQKSPNSDVKCHVTGFYSNNIVILWMKNGQEVSNALLKSG
EILPNEDGTFQRTVTLRVSLQYWRKEQYTCVVRMEKTIIRRNLTETEEIKSNYSEMSITAS
DSSTSILVPIIVVVLVSLILWWSCHQNCRLGQCFHLENAN

>dr_113060 Chain=chain14 Position=8:47112669-47115321(-) GeneID=ENSDARG00000016056
ProteinID=ENSDARP00000123111 Score=137.3 E-value=3e-40 Length=2653 Comment=Overlaps single
annotation & in proteins
MFLRMFCFVCLLPFIGVNVNAGSHSLMALATYIVGQTPFPEFSVVMMLDDLQLAYDSIGW
KTIYRSGSDSKYYDEEQSDAGIVFRDMFYDMKDRAFYLDKDHQNTDGVHVVHQRVLVGCCELL
NNEKTGPHLYWDAFGGQNMEEFIFDTEKHTIQVKMPWVITWDQLKRLHENFMYDNYHPI
CIKTLRRYLNMEKKNVMRKVKPRVRLMQLKLSDSQGLQISCLATGFYPRHINLTLFRDAE
LVDDDQITGGEILPNGDGTQMRKSLIVSEELQKGHKYNCTANYLNLDNKMDIVFDVAE
SDPGSFSVSVVMGVLVFGVLSLISITLIMRRKRRTGVSQYVVAQTSVQDAT

>dr_ch211-147g22.1 Chain=chain15 Position=25:11323396-11327494(+) GeneID=ENSDARG00000023203
ProteinID=ENSDARP00000067873 Score=130.2 E-value=2.6e-38 Length=4099 Comment=Overlaps single
annotation & in proteins
RTMPVILILSLRLIKTGSLSLWMSHYIKGQTPFPEFSVFLMLDDVVMYNGETKKNVFP
RGNATNEDDVFDLNVLRTISDKVHSSFEEKWVATQDVNKTEVVLALQRIVGCCLRHGDK
PGQMIITRDAARGSTDELLYVDKNFYQGTNLNPAFVLMNHLKISMWNHEHLYHPFCIKT
LKGYLEKRRKNQVNRKVKPKVRLLLKLSSEFRVSCLATGFYPRHINLTLRDLGQPVSDHD
VTGGDLLPNGDGTQMRKSLQIRAEEREKHKYTCSAKHLSDLNKLDVNLGERFEPVHSKV
IVACVGLVLLTIAGTII ECRKRKQSGNKCN

>dr_CR339041.1 Chain=chain16 Position=25:11300253-11301251(+) GeneID=ENSDARG00000051711
ProteinID=ENSDARP00000045181 Score=128.1 E-value=1e-37 Length=999 Comment=Overlaps single
annotation & in proteins
IQISILQIIRCNILCAGPHSLMLLATYIKGQTPFPEFSVFLMLDDVSTLYNGETKTYVP
RGNTTAEEDDVFETNVLKLSFMKQVVGTTNNINKTHVNFHFSLEGVLVQLRVVCDLKD
GEPGKMITRDAFRGSTTDELQYFDKFTYQGTNLNVDNLLKIHLEASKRNHELYQPYCI
KTLKGYLKKRTNQVNRKVKPKVRLIQKASDLGWVVSCLATGFYPRHINLTLRDLGQPVSD
DHELTGGDLLPNGDGTQMRKSLQIRAEEREKHKYTCVSKHLSLDNKLVDVFLVADFD
HGEPFKSLIPSVLVVSTLMLVFGVAAAVIARKRRCSTT

>dr_ch211-147g22.4 Chain=chain17 Position=25:11216200-11217255(+) GeneID=ENSDARG00000097275
ProteinID=ENSDARP00000129196 Score=127.2 E-value=2.3e-37 Length=1056 Comment=Overlaps single
annotation & in proteins
MCELKDDGEPGQMITQEAFFEGFTIDTLLYVDHNFTHINLSIPQEQIEQTEMLVKNRHK
LYQPPCYKTLKGYLEKRRDQVNRKVKPEVRLFYEANSDSGDFRVSCLATGFYPRHINLTL
LRDQPVSDYEVVTGGDLLPNGDGTQMRKSLQIRAEERQKHKYTCIKHLGKEYHIDLAE
PQRNHIIAVAVSVLLVCAIIVGLAMLWIKRYQAAG

>dr_BX572619.1 Chain=chain18 Position=25:11082577-11085035(+) GeneID=ENSDARG00000046057
ProteinID=ENSDARP00000067714 Score=123.0 E-value=3.6e-36 Length=2459 Comment=Overlaps multiple
annotations; one in proteins
SHSLWMLVMIYKGTAFPEFSVVMMLDDVSVLYYNGDTNSLIARGNTKADDDVFNALNP
IHDHLQSSFEDRWTLATTHLNKTDGVFALQQLVACELQDDGEPGRMILRNAAFSGSTTDQ
LLFVDKFTYHDSFNVSTHVLHAHHDYKYLCEKLLQPFQTLKGYLVKRRNQINRKVK
PEVRLIQKANSDSGGFRVSCLATGFYPRHINLTLRDLGQPVSDHEVTGGDLLPNGDGTQ
MRKSLQIRAEEREKHKYTCVSKHLSLDNKLIVTDFDPEEFPPLVILSLVSV
FITGVII

>dr_ch211-147g22.3 Chain=chain19 Position=25:11281133-11284387(+) GeneID=ENSDARG00000096830
ProteinID=ENSDARP00000128887 Score=120.9 E-value=1.3e-35 Length=3255 Comment=Overlaps single
annotation & in proteins
MSYVGNMIIILFLCVLLAPTALTGSHSLWMLATYMKGEPYSEISITFMLDDITVGHYNS
KTERYIARHNTTNEDELWNTFYVARHDLHTFFVRKLFKNHTESPQLYQMMSHCELQGN
NKPQMVSKIAFSGSTTDEMSEYENKFTYQSQKEMPSLALELKWRRHESVSYPTCISTLR
NYLKMRTQVNRKVKPKVRLQKANSDSGGFRVNLATGFYPRHINLTLRDLGQPVSENE
VTGGDLLPNGDGTQMRKSLQIRAEERQKHKYTCVSKHLSLDNKLIVTDFDPEEFPPLV
IPSVLIILSLVLFITGVVVIYCRKRRRAVSSKRDIYASASTSEESTGSTVTQPQEQAT

>dr_dkey-52p2.7 Chain=chain20 Position=25:11378000-11380505(+) GeneID=ENSDARG00000097766
ProteinID=ENSDARP00000128245 Score=119.3 E-value=5e-35 Length=2506 Comment=Overlaps single
annotation & in proteins
MILRNAAFSGSTTDQLLFVDKFTYHDSFNVSTHVLHAHHDYKYLCEKLLQPFQTLK
YLVKSRNQINRKVKPKVRLIQKANSDSGGFRVSCLATGFYPRHINLTLRDLGQPVSDHEV
TGGDLLPNGDGTQMRKSLQIRAEEREKHKYTCVSKHLSLDNKLIVTDFDPEEFPPLV
ILLSLVSVFITGVIIYCRKRRRAVVIYLLSKCIKKITS

>dr_CU571323.1 Chain=chain21 Position=8:47095167-47096314(+) GeneID=ENSDARG00000016227
ProteinID=ENSDARP00000022240 Score=115.1 E-value=1.2e-33 Length=1148 Comment=Overlaps single
annotation & in proteins
MKTMLFLLYLLSCLTFTDAGFHSLLVATYVDGQTPPELSVVMMLDDVQIYYDSDTWR
VFHRSPSDSKYYDEEDQSDADAVFHDYDEMKYRVLHLKNLQNHDTGITVLRIVGCELFN
DKPGIYHLWDAHDGKTEKFTFNINHEFQLKNQWFRWDQVMIQKRIVHENIYYPVICI
KVLRRYLNVKNSVMRKVKPRVRLMKKLLPDSQGLQISCLATGFYPRHINLTLFRDAEFPV

DDDQIIIGGEILPNGDGMQMRKSLIVSKEELDEGHEYCTMKHLNLDNKLDIVFDVSGTV
PGCFSVSVVISVLVFMCVSVFIITKLMRRRQDTGRGSEKCDYSPTISSSQDEI

>dr_dkey-52p2.4 Chain=chain22 Position=25:11342613-11350193(+) GeneID=ENSDARG00000096977
ProteinID=ENSDARP00000128999 Score=109.3 E-value=3.8e-32 Length=7581 Comment=Overlaps single
annotation & in proteins
XPTATLYTTSNQFNFSRIACASENMTGKVLVFLFLLSAPTTLVTDHSHLWLLGTIYKGET
QFPKLSFTVMLDDLRVGFYNSETKFIPRDNTTNEDEAVSDVLTVIQNTLEPFMMSTLTF
KNGTESPQLFQITWHCELLDNDKPGQIIFKIAFGSSTDEVSFYNDNFCTQCQNNMATVP
LESFQMHYETVCYPCNIIATIRDYLLKRRQTVQNGKVQPKVRLIKKVHPKSGGFRLSCLATE
FYPHSINLTLRDLGQPVSDHEVTGGNLLPNG

>dr_ch211-147g22.2 Chain=chain23 Position=25:11315123-11316131(+) GeneID=ENSDARG00000051712
ProteinID=ENSDARP00000067872 Score=107.1 E-value=1.5e-31 Length=1009 Comment=Overlaps single
annotation & in proteins
HTHRGGAGTCFNVKCGFTESIRTCNMDKILLFLFLPTSPAPKGSLSLCLLATYIKGSPSP
FPFLSGVMMLDDIPLLYNGDTKTFMRGNTAEDNVFDANAFLSIIGHIQSSVDRWGL
ASRDLNKTDRIFTLQQLVLCESDEGEHGKMSRDAVEGTTTDELQHVHDKFTYKHTLNV
SAYLIDFYLELTKSLHKTFLFQPTCFKTLGSLYLIQRRNQINRKKVKKVRLFKKELSSGFIV
SCLATGFYPRHINLTLRDLGQPVSDHVDVTGGDLLPNGDGTQMRKSLAIRAEERQKHKYS
CSVKHLSLDNKLHVDLDFDHSKPFQSVIPSVLTVLALLLVFGVAAVIWKRCRDSVKCY
SAASTSVENMETT

>dr_CR339041.2 Chain=chain24 Position=25:11257981-11263809(-) GeneID=ENSDARG00000051710
ProteinID=ENSDARP00000084819 Score=105.0 E-value=1e-30 Length=5829 Comment=Overlaps single
annotation & in proteins
MFATYIKGKTPFPPELSSGVMLDDIRVLYYNGVTSFLARGNTAEDDVFDPDDLLENIKGF
IKSEFSFLNAQNTFTKTDGIFVVSQILAMCELKDDGEPGQMGIGQDAFERLTIFTVLYADNK
CTIDINLNIHQEQKEKIEGVKNYRNLIQPFCKYTKLVYLLKRRKQVNRKVEPKVRIHF
KANLDFGGFRVSCLATGFYPRHINLTLRDLGQPVSDHEFTGGDLLPNGDGTQMRKSLIR
RAEDSEKHKYTCSEFKHLDEWHIDLAEPHRNTIWIIVSVLLVCAIIVGLAMLWKRYQTA
RQRENEPNNHQTTHMNAPEVDVLDMDHSRVE

>dr_BX572619.2 Chain=chain25 Position=25:11063812-11066177(-) GeneID=ENSDARG00000086127
ProteinID=ENSDARP00000067869 Score=101.0 E-value=1.1e-29 Length=2366 Comment=Overlaps single
annotation & in proteins
FLCFLCLLSLPTAVVTDHSHLWFLVTVIEGETQFPFASVVYMLDDITVGYYNSETKTCVP
RGNTTNEDEVDLIGYLLNEYYPPIVEKFI GLFKNKTKGEYNRYSTCPVHQLLTVCELP
DKDGVGQMITKIAFEGSSTDEMFIIDTLTFQGMKVKLHLEMFWRWFHAMAYKRGKLL
EKYLKRRATQGRVVKPVRLIQKASDSGFCVSCLATGFYPRHINLTLRDLGQPVSDHEV
TGGDLLPNGDGTQMRKSLIRAEERQKHKYSCSAKHLSDNKLIAATLEFDHGEPYKSVI
PSVLAVLMLLVFGAAAIVWKRRTGTMRFKINL

>dr_dkey-52p2.6 Chain=chain26 Position=25:11363689-11372081(+) GeneID=ENSDARG00000051713
ProteinID=ENSDARP00000067874 Score=98.1 E-value=1e-28 Length=8393 Comment=Overlaps multiple
annotations; one in proteins
WVVATYIKGQTPPEFSYVLMDDITVMYNSDTKTFPPRGNTAEDDLYGSKDHLIIHD
FMRSTFKDKWGLATFKLNKTDGVFALQQLVVLDRADSEPGQIISQNTFEGCITDEVRYV
DKKLTQGTNLVSAFVHHIHHEYVKYLCELIHPFYFKTLKGYLIKRRNQINRKKVKKV
LILKANSDSGGFRVSCLATGFYPRHINLTLRDLGQPVSDHGVTTGGDLLPNGDGTQMRK
LEIRGEEREKHKYTCATHLSLDDKLDITLGLVAEFDGELFKSVIPAVLIVLSLVLVFI
TGVVI

>dr_BX682558.1 Chain=chain27 Position=3:23763318-23765350(-) GeneID=ENSDARG00000055813
ProteinID=ENSDARP00000072675 Score=95.1 E-value=5.1e-28 Length=2033 Comment=Overlaps single
annotation & in proteins
MMYLVLAVCILSASVFSQSGSHSLWVVFATFLTGDIVSQFPEFSAVVMLDDIVIGHYNADD
RSFVPSVTVQESVNIETQITISTVCKGHEGMKTKAYLIDYLNHTRGLHVQORLVGCELL
QNEPGQMMTLEAFNGESGFERRYDIQGDQQTHWKVPVKSRAQLEYDAWLYAHFYRPLCI
SQLRKYLKKEKRRVMARVPRVRIQRTCSKTGKIOMTCLATGFYPRHINLTLRDLGQPV
NEERVMGGELLPNADGTQMRKSVLSAEERQRRRTYCTVNHLSLDNKLIDISIEPGLDP
VIIFPSVLLLLCVFVGLAGFLMWRKKYKQPEQVYTPTSATDQTTEQSQL

>dr_BX572619.3 Chain=chain28 Position=25:11071627-11074602(-) GeneID=ENSDARG00000087161
ProteinID=ENSDARP00000110996 Score=84.6 E-value=7.5e-25 Length=2976 Comment=Overlaps single
annotation & in proteins
SLLSTYIKGETQFPKLSGTAIILNDFIVGYNNNTYIPRGNTTNEDDVIDADQTKTLADYVY
GSFLKRSSLLSQNNQNTSKGLDVYQKLVVCELLSDTPGKMFRRDAARGSTIDEILYTN
TFMYTVIGNVSQELIKPHLEAFKLDFTLFPVCIKTLKTYLKKRSQVNRREMPNSRLF
RKASDSGGFRVSCLATGFYPRHINLTLRDLGQPVSDHEVTGGDLLPNGDGTQMRKSLIR
KAEEREKHKYSCSVKHMGEFHDIL

>dr_chain29 Chain=chain29 Position=8:47125558-47128530(-) GeneID=None ProteinID=None Score=83.9
E-value=1.7e-24 Length=2973 Comment=No overlapping annotations
YFKGSHSLMAFATNIIVHTPFPEFSAVVLLDDLQVMYNSVTRQAVHRNVVHPICIKILQ
GYLQSKNNVMRKRMLKLLSDCQGL*ISCLATGFYPRHINLTLFRDAEPVDDDDQITGGE
ILPNRDGTQMRKSLIIEEELQKGHQYNCVTKHLNLDNKLDI

>dr_chain30 Chain=chain30 Position=25:11228370-11229374(+) GeneID=None ProteinID=None
Score=70.7 E-value=1.1e-20 Length=1005 Comment=No overlapping annotations
ISGSSLSFWLFFSTYIKGKTQFPPEFSYTTTLDDITVGYNSETYIPRGNTTNEDDVIHSDY
IISISHMHESFMRRSRLGQDNKNSKEEFKMKFEQFFYPVCIDLKKNVVKREIEINR
KVRLIQKASDSGGFVVVSLATGFYPRHINLTLRDLRGGQPVSDHEFTGGDLLPNGDGTYQMR
KSLEISAEEREKHNITCSAKHMGEYHIDL

>dr_ighd Chain=chain31 Position=3:34038903-34043015(-) GeneID=ENSDARG00000096248 ProteinID=None
Score=21.7 E-value=1.1e-05 Length=4113 Comment=Overlaps single annotation; not in proteins
PLTGHH*LLTDITSLSQVTTFSQTSLVYMEFTCPHHLRGQAFVLRRLRRPSDSDLGGKIS
SLVCFISDFFPSDILVEWKLNGKQLSRSQFSNSPLVALSSGGFSMHSALILPGEQQKDI
YSCEVSHESQKPIN

>dr_arpc1a Chain=None Position=25:11354468-11368396(+) GeneID=ENSDARG00000096940
ProteinID=ENSDARP00000128878 Score=113.4 E-value=1.4e-32 Length=0 Comment=Not detected in
genome search
MARTSENMIGKVLVFLFLLSGPSTSVLTFHSLQILGTYIKGDTLFPQISFTFMLDDLTVG
LYITERDEFIPRDNTTNEDETVDMLDKIKNELQPLLKGTLIFQNSTESPQLFQIAWHCE
LLDNNKTGQMILKIAFSGSTTDEVSPFYGNFTQCQNNMATVPLEYFQMFKSIYYPVCM
TTLGGYLVKRQVQVNRKVQPKVRLIKKVHPKSGGFRLSCLASEFYYPFINLTLRDLRGGQPV
SDHEVTGGDLLPNGDGTYQMRKSLEIRADEREKHKYTCSAKHLDKNLDVDFEFDPSPYPVK
IVIPVVLLSLVLVLTAVLIHKCMNKQAASPEQEQEPIEMGGRLRK

>dr_CABZ01077449.1 Chain=None Position=13:48786469-48790819(+) GeneID=ENSDARG00000090839
ProteinID=ENSDARP00000107098 Score=55.9 E-value=1.5e-15 Length=0 Comment=Not detected in genome
search
MLAFLLIVSFHEVILGADEHAFQQFIECAFNSQGVDRSWRYGDKDVMHVDLATETVV
GTSEPGKRLAEERKSTEYIRRKEEKLKIVCSAVKTVFLKSNNTLSRAAKPTVLLSNGGQ
GQEYLLKCVVRGFYPNVIRVWRTQKGPPIFFGVSTGTLPHDTGTFQMTSYLSLGNMTAHD
VTCEIEHLSIDGKLRKNYGDNPWILSQITVAVVAFILGFVCTIAVIFVWKRHQTTKSHDD
ETNGASDESEASLSLNVMNISQET

Supplementary Table 4. Opossum *UT* exon 2 and 3 amplicon sequences.

>UT2.cDNA
Not detected

>UT3.cDNA
Not detected

>UT4.exon2.cDNA
AGGCATGTGATTCAGTTCACCACAGTGGGTATAGACCACTCTATCTTGGAACTAAGTGAATTAGCCTCATAGATGACATTGAAGTGGCATCTTA
TAACAAAGGACAGAAGCAAATTCATCAAGATACCTTGCATCTCTGAAGTACTAGGAGTCAATTTACTTACCCAGTGGCATAATTCACCTGGTGA
AACACGAGCAAGATTCCTCCGGCCAATCGAATTCCTTAGCACGGAATGACATTAACCGCCATAG

>UT4.exon3.cDNA
AAACCACACAGCCAGCTCCAGGCAGAATGTGAAATGGACAATGACATAACAGTAAAGAGTCACATCTATTTAATTTGGGATGGGAGGAATGT
ACCAAATAGATGAAGAGTTGGGCACTGGGAGAACATGAATCCCGAATTCAGCAATACCAATACATCTGGAGAGTTCTCTCAGGACCAGTCTA
AGAAAACTACATGAATCTATATTGTTGACTTGATGAAGAAAATGTTGGATATTCAAGCAAGGGCGA

>UT4.aminoacid.cDNA
RHVIQFTTVGIDHSILELSAISLIDDIIEVASYNKGQKQIAIKIPCISEVLGVNYLTQWHNSLVKHEQDFLRPIEFLARNDINRHNHTAQLQAECE
MDNDITVKSHIYLIWDGEECYQIDEEVGHWENMNPFFKQYQYILESSLRSLRKHVMNLYCVDLMKKIVGYSS

>UT5.exon2.cDNA
TTCAGTCCACCACAGGCATATGGGACAGTTGACTGGAGTTGGCACAGCCCACTCTCTTGGAACTAAGTGAATTAGCTTATTAGATGACGTTG
AAGTGATATCTTACAACAAAGCACACAAGCAAATFGATTTAAAGATACCTGGATCTCTAAAGCACTGGGAGTTAAGTATCTTACAGAGCTGCAT
GATTTATTGGTGAAGCAAGCAATATTTTCAGATGGACTATCGAAATCGTAGAACGGAATGACACAAGCCATAATAGTT

>UT5.exon3.cDNA
GTTCCCCACAGGAAACCACAGCCAGCTCCTGGCAGAATGTGAAATAGATAATAACATCACAGTAAAGAGTCACATGCATTTAATTTGGGAT
GGGAGGAATGTTTTCAGAAATAGATGAAGAGTCTGGGCACTGGGAGAACATAAACTCCTGAATTCAGGAATACCAACAGACTCTGGAGAGACCAT
CTGGACTTCTTAAGGAAACAATATATGAAGCTCTATTGTTGACGGGATCAGGAAAAATA

>UT5.aminoacid.cDNA
SVHHRHMQQLTGVGTAHSLLELSAISLLDDVEVISYNKAHKQIDLKIPWISKALGVKYLTELHLLVKQEYFRWTIEIVERNDTSHNSSPTGNH
TAQLLAECEIDNNITVKSHMHLIWDGEECFRIDEESGHWENINPEFKYQQTLEPFWTSLRKQYMKLYCVDGIRKIKG

>UT6.exon2.a.cDNA
TTCAGTCCACCACAGACATGAGCTTCAGTTCCACCACAGTGGGTACAGACCACTCAGTCTTGGAACTAAGTGAATTAGCCTCATAGACGATGTTA
AAGTGGCATCTTATAAGAAAGGACAGAAGCAAATTCATCAAGATACCTGGATCTCTGAAGTTCCTGGGAGTCAATTTATGTCACACAGCTGCAT
GATTTATTGGTGAACCATGAGCAAGATTTCTCTGGCCAATCCAGTACTTAGCACGGAATGACATTAACAAACATAG

UT6.exon2.b.cDNA
TTCAGTCCACCACAGACATGAGCTTCAGTTCCACCACAGTGGGTACACACCACTCAGTCTTGGAACTAAGTGAATTAGCCTCATAGATGACGTTG
AAGTGGCATCTTATAAGAAAGGACAGAAGCAAATTCATCAAGATACCTGGATCTCTCAAAGTATTTGGGAGTCAATTTATCTTACCAGTGGCAT
AATTCAGTGGTGAACCATGAGCAAGATTTCTCTGGCCAATCCAGTACTTAGCACGGAATAACATTAACAAACATAG

UT6.exon2.c.cDNA
TTCAGTCCACCACAGACATGAGCTTCAGTTCCACCACAGTGGGTACACACCACTCAGTCTTGGAACTAAGTGAATTAGCCTCATAGATGACGTTG
AAGTGGCATCTTATAAGAAAGGACAGAAGCAAATTCATCAAGATACCTGGATCTCTCAAAGTATTTGGGAGTCAATTTATCTTACCAGTGGCAT
AATTCAGTGGTGAACCATGAGCAAGATTTCTCTGGCCAATCCAGTACTTAGTATGGAATGACATTAACAAACATAG

>UT6.exon3.a.cDNA
GAAACTGCACAGTCCAGCTCCTGGCAGAATGTGAAATAGACAATGATATAACAGTAAAGAGTCACATCTATTTAATTTGGAATGGGAAGGAAAGT
TACCGGATAGATGAAGAGTTGGGCACTGGGAGAACATAAAAGCCTGAAACCAAGCAATACCAATACATTTCTGGAGAGTCCCCCTCTGGACCAGTCT
AAGGAAACACTACATGAAGTTATAGTGTGCTGACTTGATGAAGAAAATCA

>UT6.exon3.b.cDNA
GAAACTGCACAGTCCAGCTCCTGGCAGAATGTGAAATAGACAATGATATAACAGTAAAGAGTCACATCTATTTAATTTGGAATGGGAAGGAAAGT
TACCGGATAGATGAAGAGTTGGGCACTGGGAGAACATAAAAGCCTGAAACCAAGCAATACCAATACATTTCTGGAGAGTCCCCCTCTGGACCAGTCT
AAGGAAACACTACATGAAGTTATAGTGTGCTGACTTGATGAAGAAAATCA

>UT6.exon3.c.cDNA
GAAACCACACAGCCAGCTCCTGGGAGAATGTGAAATAGACAATGACATCACAGTAAAGAGTCACATCCATTTAATTTGGAATGGGAAGGAAAGT
TACCGGATAGATGAAGAGTTGGGCACTGGGAGAACATAAAAGCCTGAAAGTTAAGCAATACCCATAACATTTCTGGAGAGTCCCCCTCTGGACCAGTCT
AAGGAAAGCTACATGAAGCTATATTGTGCTGACTTGATGAAGAAAATCA

>UT6.aminoacid.a.cDNA
SVHHRHELQFTTVGTDHVSLELSVISLIDDVKVASYKKGQKQITIKIPWISEVLGVNYVTQLHDLNHNHEQDFLWPIQYLARNDINKHGNCTVQL
LAECEIDNDITVKSHIYLIWNGKESYRIDEVGHWENIKPETKQYQYILESPWLSLRKHVMKL.CADLMKKIKG

>UT6.aminoacid.b.cDNA
SVHHRHELQFTTVGTDHVSLELSVISLIDDVKVASYKKGQKQITIKIPWISEVLGVNYVTQLHDLNHNHEQDFLWPIQYLARNDINKHGNCTVQL
LAECEIDNDITVKSHIYLIWNGKESYRIDEVGHWENIKPETKQYQYILESPWLSLRKHVMKL.YADLMKKIKG

>UT6.aminoacid.c.cDNA

FSVHHRHELQFTTVGTHHSVLELSAISLIDDVEVASYKKGQKQITIKIPWISKVLGVNYLTQWHNSLVNHEQDFLWPIQYLVWNDINKHRNHTAQ
LLGECEIDNDITVRSIHILWNGKESYRIDEEVGHWENINPEVKQYPYILESPLWTSRLRKSVMKLYCADLMKKIKG

>mdUT8.exon2.cDNA

ACATGATACAGTTCACAATGGTGGGCACAGGAAGCTCCCTCTTGAATACTTTATTTGTGGACTTTCTGGATGATGTACAGGTGTTCTCCTTTAAC
AAATATAATCATCAGCTCATAGCCAAGGAGGCCCTGGATCTCACAGTCTCGGGGCCAAGTTCATAGAGGACACTTGGCAGAAGTTAGTGGACCA
CGAGAAGAGTTTCTATGGTCTGAGGAAGCTGATGCAGAACAGCACCAAGAGTGACA

>mdUT8.exon3.cDNA

TGAACCACACAGTGAACCTTTCATACTTTGTGAAATTTGAGAGGAACAACCAGCTAGGAAATGAAATCCAGATTGCTGTGAATGGTCAGGATTTT
TGCTGGCTGGATGAGAAGTATGGCGATTGGTTTATCATGGTGCCTGAGGCCGAGCCCTCAAAACCCATTCTGACAAGTAACCTGTGGACCACTCA
GAGGAAGCACTACAT

>mdUT8.aminoacid.cDNA

MIQFTMVGTTGSSLLEYFIVDFLDDVQVFSFNKYNHQLIAKEAWISQVLGAKFIEDTWQKLVDEKSFVWFLRKLQNSKSDMNHTVQLFILCEI
ERNNLGNEIQIIVNGQDFCWLDEKYGDFIMVPEAEFFKPIILTSNLWTTQRKHV

>UT9.exon2.cDNA

TTAAGTTCAGTGCAGTGCAGACAGCCGCTCTCTCTTGGACTTCACTGGTATTGCGAGTATTGATGACATTCAGTGGATTTCTATTACAAACTG
AACCATCAGATTGCAGTAAAGTAGATTGGATCTCTGATGTCTGGGACCGACTTCATGAAGAGATGTGCATTTGATAAAGAGTCATGAGA

>UT9.exon3.cDNA

GGGATGCTGTCTCCTTGATCCATTTTTTGGACTGGAAATGACACCAAGACTGAGAGAAACAACAATGCAGTACATATTAGGCTGTGAATTAGAT
GATGACATCAAGTTGAATAGCAAAATACGAATTGGCTTGGATGGAGAGACTATGATTGAGTTTGAATGAACATCAAAAGACACTGGGTGGTCCAGAG
CCCTAGGGCTGAAAACCTTGAACCCATTGTGAAAAGCCTCTCTGGACTAATTTGAGAAAACGTT

>UT9.aminoacid.cDNA

KPTAVSTDRSLDFTGIAVIDDIQWISYYKLNHQIAVKVDWISDVLGTDPIEEMCHLIKSHERDAVSLIHFLTGNDDTKERNNTMQYILGCELDD
DIKLNKIRIGLDGETMIEFDEHQHRHWVQSPRAENLKPIVESLFWNLNRKR

>UT10.exon2.cDNA

CAGTGGGTACAAGACATCTCTCTTCAATTTACCATAATCCAGTATTGATGATGTGCTAGTGTGCACCTATGATAAAGGGAATAAGCAAAT
GTGTTCAATGGAGCATGGATCACCAGGCCACTAGGAAAGACTTCACTGAGAATAAGAAGGAATTAATTTAAGTGTGAGACGTATTTCCTGCTG
GCTCCTCAGCTTCTTTATAAAGAAAGACACCAAGAGTAACA

>UT10.exon3.cDNA

AAAATCATACAATACAGCTCTATGGAGAGTGTGAACCTGGATAATGACATTTACATTTGGCAGCCAAATCCAGTTTGTGTTGGATGGTGAAGATTTT
ATCAGGACAAATGATAAAGATGGAGCACTGGATTGTCACAAAACCTGAAGCAGAGCACTTCAGAACCATAGCAGAAAGACCTTGGGGGTCAACCT
GAGAAAGAAAGTTATGAAGCATTATTTGTGC

>UT10.aminoacid.cDNA

VGTRHSLFNFTIIHVIDDLVCTYDKGNKQVFNNGAWITRALGKDFTENKKEFNLSAETYFRWLLSFFIKKDTKSNKNHTIQLYGECELDNDIYI
GSQIQFAVDGEDFIRTNKMEHWIVTKPEAEHFRITIAESTLGVNLRKVMKHYC

>UT11.cDNA

Not detected

>UT13.cDNA

Not expressed

>UT15.exon2.cDNA

CACAGGCATGACTTCTATTTACCACCACTGGCAGCTCAGGGTCATTTGGACTTCATTTGTGAACAGTGTGTGGATGATGTTACAGTGTGCTC
CTACCATAAAGAGAAAGCATCAGCTTGTGATGAAGGAAGACTGGATGTACCAAGCCCTGGGAGCCAGTGAATTCACAGAAACAGCAGAAATTC
TGGACCTTGAAAAAATTTCCAATGGGCTTTCAAGAACCTGATAAAGAATGAGGCCAAGAGTGAA

>UT15.exon3.cDNA

AGAACCACACTCTACAGGTGTTTGTAAATGCGAAGTGGACAAGGACATCTGGTGTGAGCAGCCACTTCCAGTTTGTCTTTGGATGGTGAAGATTTT
TGCCAGATGGATGAACAACAGAGCATTTGGATAGCCATGAAGCCTGAAGCCAGCACTTCAAGTTCCTCTGGGACAGCGCCTTCTGGAAAAGGCG
GGTGTGAGACTATATACAGGAAGATTGTATTGATGGCAT

>UT15.aminoacid.cDNA

HRHDFYFTATGSSGSFLDFIVNSVVDVQLCSYHKEKHQLVMKEDWYQALGASEIAQKQQLLDLEKNFQWAFKNLIKNEAKSGHTLQVFNCE
VDKDILVSSHQFALDGEDFCQMQDEQPEHWIAMKPEAQHFKFLWDSAFWKRVRHYIQEDCIDG

>UT17.exon2.cDNA

CATCACAGCCATGAGATGTTTTTACTGCAGTGGGAACGACAAAGACTCTCTTGGACTTCACCCCTGGTCAGCTTCTTGGATGATGTAGAGATAGT
CTTCTATAACAAAAGCAACAGCAGTTTGTGATCAAAGAGGCTGGTTTCCATGCCCTGGGAGCTGACTTTCATTGAAGATATACAGCAGAAGT
TAAAGTACAATGAGATTGACTATCTTTGGGCCCTACAGAACTGGGTAAAGAATGATAAGAGAGGGGAAA

>UT17.exon3.cDNA

AAAGAACCATACAGTCCAGATTTGGCATAACTGTACCTGGACCAAGACATCCATGTGAGCAGTTCGAGTCTGGTGTGCTGTGGATGGCAGAAGAT
TCTGCCAGATGGATGAGCAGATAAGTCACTGGGTAGCCAAGAGCCTGAGGCTGAAAAAATCGTCTCTCTCAGAAAAGATCTACTGCTCCAAG
AAGGTGCAACGCTACCTGGAGGACTATTGTGTCCAGCCAATGA

>UT17.aminoacid.cDNA

HHSHEMFFTAAGTTKLLDFTLVSLDDVEIVFYNNKQQQFVIKEAWVSHALGADFIEDIQQLKYNEIDYLWALQNWVKNDRGENHTVQIWHN
CHLDQDIHVSRRVWCAVDGEDFCQMQDEQISHWVAKKPEAEKIVLSEKIYCSKKVQRYLEDYCVQP