International Journal of Biosciences (IJB)
ISSN: 2220-6655 (Print) 2222-5234 (Online)
Vol. 1, No. 4, p. 36-50, 2011
http://www.innspub.net

RESEARCH PAPER OPEN ACCESS

Computational genomic study of LTP pathway in the context of Raf/Ksr homologue in human and chimpanzee

Farzana Ahmed[1], Abdullah-Al-Emran[2*]

[1]*Department of Mathematics and Natural Science, BRAC University, Dhaka, Bangladesh*

[2]*Departmet of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Santosh,Tangail, Bangladesh*

Received: 04 June 2011
Revised: 26 June 2011
Accepted: 27 June 2011

Key words: Comparative genomic study, long term potentiation, Raf/KSR1.

Abstract

LTP (Long Term Potentiation) pathway, which is responsible for long-term memory development is defined as the long-lasting enhancement in communication between two neurons that results from stimulating them simultaneously. One of the most important genes concerning this LTP pathway is Raf which actually encodes for KSR1 (kinase suppressor of ras-1) compartmentalizes Hippocampal Signal Transduction and sub serves synaptic plasticity and Memory formation has been explored through the use of comparative genomics in this work. For this purpose, syntenic regions for KSR1 genes in human, chimpanzee, mouse and other organisms are computationally searched to find about the conserved regions and their function through the course of evolution. A coding region was found on chromosome 16 syntenic region for KSR1 of chimpanzee. Sequence comparisons among members of the Raf family reveal the presence of three blocks of conserved sequences: conserved regions (CR) 1, 2, and 3. In this work conserved regions are found in near the C-terminal and N-terminal region was absent in both human and chimpanzee which suggests C terminal region is very important and N –terminal region is dispensable for raf-1 activity which supports the previous finding. Phylogenetic analysis showed that the KSR1 of human and chimpanzee are evolutionary distant from other organisms and human is most closely related with chimpanzee than any other organism which is also supported by previous evolutionary concept.

---

*Corresponding Author: Abdullah-Al-Emran ✉ emranbge@mbstu.ac.bd

Introduction

Comparative genomics exploits both similarities and differences in the proteins, RNA, and regulatory regions of different organisms to infer how selection has acted upon these elements. Human is different from all other species in memory development process. The important features for what human is very different from all other species include their exclusive property of learning, thinking, memorizing and also utilizing this memory for extraordinary discovery and invention. The memory process in the human brain is very complex. Long-Term Potentiation (LTP) is widely considered as the major cellular mechanisms that underlie learning and memory which is the persistent increase in synaptic strength following high-frequency stimulation of a chemical synapse. Studies of LTP are often carried out in slices of the hippocampus (Fig. 1), an organ of central importance in learning and memory (Cooke and Bliss, 2006).

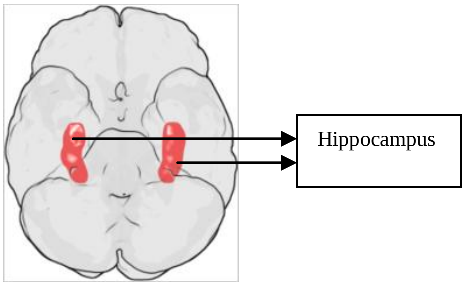


Fig.1. Location of Hippocampus in Human Brain.

Comparative genomics approach used in this study of LTP pathway for identification of genes that are important and play the major role in developing the long term memory and how they have changed through evolution to make the human being more efficient in learning, developing memory and thinking. The genes that play important role in LTP are given in Table 1 (Shahani 2007). These genes are identified on the basis of shared synteny. Shared synteny describes preserved co-localization of genes on chromosomes of related species (Moreno-Hagelsieb *et al.*, 2001). Stronger-than-expected shared synteny can reflect selection for functional relationships between syntenic genes, such as combinations of alleles that are advantageous when inherited together, or shared regulatory mechanisms. The term is sometimes also used to describe preservation of the precise order of genes on a chromosome passed down from a common ancestor (Passarge *et al.,* 1999). This is one of the most reliable criteria for establishing the orthology of genomic regions in different species. (Engström *et al.,* 2007; Heger and Ponting , 2007 ; Poyatos and Hurst, 2007; Dawson *et al.,* 2007). Patterns of shared synteny or synteny breaks can also be used as characters to infer the phylogenetic relationships among several species, and even to infer the genome organization of extinct ancestral species. Within the vertebrate lineage, analyses of different genomes can also highlight similarities and differences in genome structures. Recent analyses in zebrafish have also identified conserved synteny groups between the zebrafish and human genomes (Barbazuk *et al.*, 2000; Postlethwait *et al.*, 2000).

Table 1. List of genes that play very important role in LTP.

Gene	Organism	Location	*Symbol*
FOXP2	Human	7q31	FOXP2
(NG_007491)	Chimp	7	FOXP2
	Mouse	6	Foxp2
CAMK2A	Human	5q33.1	CAMK2A
(NC_000005)	Chimp	5	CAMK2A
	Mouse	18	Camk2a
EGFR	Human	7p12	EGFR
(NG_007726.1)	Chimp	7	LOC463415
	Mouse	11 9.0cM	Egfr
MAPK2	Human	19p13.11	MDS032
(NG_023054)	Chimp	22	MAPK2
	Mouse	8C1	2010315L10Rik
PKC	Human	16p11.2	PRKCB1
(NG_029003.1)	Chimp	16	LOC467927
	Mouse	7	Prkcb1
PKA	Human	19p13.1	PRKACA
(NC_000019)	Chimp	19	PKA C-alpha
	Mouse	8C3	Prkaca
Ras	Human	6p25	RREB1
(NG_016201.1)	Chimp	6	LOC462416
	Mouse	13A3.3	Rreb1
Raf	Human	16p11.2	LOC441755
(NC_000016.9)	Chimp	16	LOC468107
	Mouse	4	LOC637116
GRIN2A	Human	16p13.2	GRIN2A
(NG_011812.1)	Chimp	16	GRIN2A
	Mouse	16.4cM	Grin2a

Kinase Suppressor of Ras1(KSR1) compartmentalizes Hippocampal Signal Transduction and Subserves Synaptic Plasticity and Memory Formation. The ERK/MAP kinase cascade is important for long-term memory formation and synaptic plasticity, with a myriad of upstream signals converging upon ERK activation. (Sara *et al.,* 2006) Despite this convergence of signaling, neurons routinely activate appropriate biological responses to different stimuli. Scaffolding proteins represent a mechanism to achieve compartmentalization of signaling and the appropriate targeting of ERK-dependent processes. Kinase suppressor of Ras (KSR1) functions biochemically in the hippocampus to scaffold the components of the ERK cascade, specifically regulating the cascade when a membrane fraction of ERK is activated via a PKC-dependent pathway but not via a cAMP/PKA-dependent pathway. Specificity of KSR1-dependent signaling also extends to specific downstream targets of ERK. (Sara *et al.***,** 2006) Behaviorally and physiologically, it was found that the absence of KSR1 leads to deficits in associative learning and theta burst stimulation-induced LTP. It was also reported that the KSR1 performs endogenous scaffolding role in controlling kinase activation within the nervous system. The Ras GTPase controls many aspects of normal animal development by stimulating the Raf/MEK/ERK kinase cascade and several other less well characterized signaling pathways. (A.B. Vojtek and C.J. Der, 1998) Kinase Suppressor of Ras (KSR) is a positive regulator of Ras signaling that was identified by genetic studies in *Drosophila* and *C. elegans*, where epistasis analyses suggested that it functions at a step between (or in parallel to) Ras and Raf. (Kornfeld *et al.*, 1995*;* Sundaram, Han, 1995; Therrien *et al.,* 1995; Sieburth *et al.***,** 1999).

Since KSR1 has a great implication in Long Term Potentiation (LTP), therefore the objective of this study was to explore most important genes Raf which is actually KSR1 (kinase suppressor of ras-1) to search for syntenic regions for KSR1 genes in human, Chimpanzee, mouse and other organisms to find about the conserved regions and their function through the course of evolution by using comparative genomics.

Materials and methods

In this study of comparative genomics the full emphasis was given to Raf, which is actually KSR1 (kinase suppressor of ras-1), as KSR1 has a great implication in LTP. (Sara *et al.,* 2006.)

List of materials

Computer

Software tools used:

ClustalW (http://www.ebi.ac.uk/),
FGENESH (http://www.softberry.com/berry.phtml),
GENSCAN (http://genes.mit.edu/GENSCAN.html)
TWINSCAN (http://mblab.wustl.edu/nscan/submit/)
AUGUSTUS (http://augustus.gobics.de/)
PSI-BLAST (http://www.ncbi.nlm.nih.gov/)

Methods

With a target of having homologous KSR1 genes (Gene Accession number: NC_000017) from various species Map Viewer from NCBI (http://www.ncbi.nlm.nih.gov/) was used. The genes were selected on the basis of located on the same chromosome number and same arm in different species. The protein products of selected genes were retrieved from GenPept of NCBI gateway (http://www.ncbi.nlm.nih.gov/). All the proteins are analyzed for homology search by using PSI-BLAST from NCBI (http://www.ncbi.nlm.nih.gov/). For gene prediction and thereby having the protein product of chimpanzee chromosome 16 genomic contig different gene prediction software like FGENESH (http://www.softberry.com/berry.phtml),GENSCAN(http://genes.mit.edu/GENSCAN.html), TWINSCAN (http://genes.mit.edu/GENSCAN.html), AUGUSTUS (http://augustus.gobics.de/) were used. Identity and similarity of each protein with human KSR1 were checked by using bl2seq from NCBI (http://www.ncbi.nlm.nih.gov/) to produce the input file for multiple sequence alignment. Multiple sequence alignment of the protein products was build up by using ClustalW from EMBL-EBI. (http://www.ebi.ac.uk/) Upon the basis of this multiple sequence alignment, phylogenetic tree was constructed to explore the evolutionary distance between the selected organisms using NJ method from EMBL-EBI. (http://www.ebi.ac.uk/).This approach helped to identify the conserved and thus the functional domains that are most important for the comparative genomics study.

Results and discussion

Predicted Protein Sequence for *Pan troglodytes* gi|114665200|ref|NW_001226709.1 Chromosome 16 Genomic Contig by using Gene Prediction Tools was searched and the results are given in Fig. 2. Multiple Sequence Alignment Candidates were Chosen By PSI BLAST (Table 2) and the results of the alignment containing the C and N-terminus region of the predicted protein is given in Fig. 3. Phylogenetic Tree also known as a phylogeny, is a diagram that depicts the lines of evolutionary descent of different species, organisms, or genes from a common ancestor. Phylogenies are useful for organizing knowledge of biological diversity, for structuring classifications, and for providing insight into events that occurred during evolution. The tree was constructed from the result of multiple sequence alignment (Fig. 4).

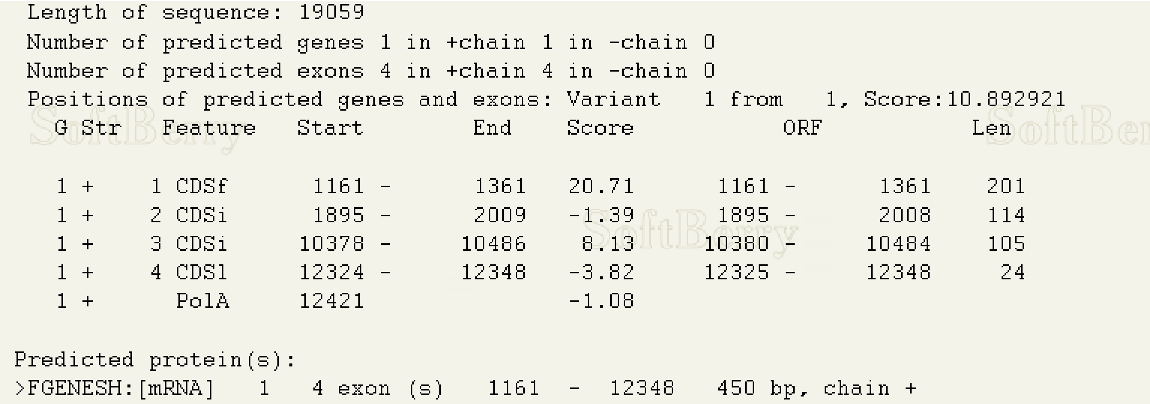
In this work chromosome 16 syntenic region of chimpanzee was analyzed for containing any gene or not by ENSEMBL and UCSC genome browser and the results showed the presence of a coding region within the region. (results not shown in this report). Further analysis with the use of different programs (FGENESH, GENSCAN, AUGUSTAS and TWNSCAN) established the presence of a coding region within the chromosome 16 genomic contig. (result: 3.1.1) According to the AUGUSTAS result (that was used in this report) there is the presence of a coding sequence of 459 base pair that encodes a protein of 152 amino acids. The gene starts from 1161bp-11556bp and it contains 4 exons. (result: 3.1.1.1) According to FGENESH result the contig contains a coding sequence of 450 bp that encodes a protein of 149 amino acids. The gene starts from 1161 bp- 12348 bp and it contains 4 exons. And the poly A tail starts from 12421 bp. (result 3.1.1.2)

Table 2. Multiple sequence alignment candidates.

gi no	Organism	Function
gi|113426068|	*Homo sapiens* (Human)	PREDICTED: Hypothetical protein
>AUGUSTAS predicted protein (gi:114665200)	*Pan troglydates* (Chimpanzee)	choromosome 16 genomic contig
gi|109113709|	*Macaca mulatta* (Rhesus)	PREDICTED: kinase suppressor of ras
gi|73966924|	*Canis lupus familiaris* (Dog)	PREDICTED: Kinase suppressor of ras-1
gi|7305215|	*Mus musculus* (Mouse)	kinase suppressor of ras
gi|109488592|	*Rattus norvegicus* (Rat)	PREDICTED: similar to Kinase suppressor of ras-1
gi|126314007|	*Monodelphis domestica* (Opossum)	PREDICTED: similar to Kinase suppressor of ras-1
gi|118100376|	*Gallus gallus* (Chicken)	PREDICTED: similar to Kinase suppressor of ras-1
gi|125838149|	*Danio rerio* (Zebrafish)	PREDICTED: hypothetical protein
gi|157012738|	*Anopheles gambiae* (Anopheles)	Anopheles gambiae str. PEST
gi|23344920|	*Drosophila melanogaster* (Drosophila)	kinase suppressor of ras
gi|66551614|	*Apis mellifera* (Honey bee)	PREDICTED: similar to kinase suppressor of ras CG2899-PA isoform 1

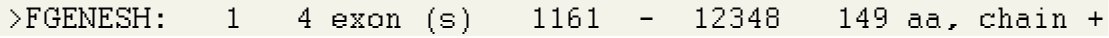
KSR orthologs were subsequently identified in vertebrates and other organisms as well (Therrien *et al.*, 1995). KSR proteins are somewhat similar to Raf in sequence and overall structure in that they have an N-terminal cysteine-rich domain (CA3), a central serine/threonine-rich domain (CA4), and a C-terminal kinase-like domain (CA5). However, unlike Raf, KSR proteins also possess two unique N-terminal conserved areas (CA1 and CA2), lack a Ras binding domain, and may also lack kinase activity (D.K. Morrison, 2001). (Fig. 5: Domain architecture of KSR).

FGENESH Result: (http://www.softberry.com/berry.phtml)



```
  Length of sequence: 19059
  Number of predicted genes 1 in +chain 1 in -chain 0
  Number of predicted exons 4 in +chain 4 in -chain 0
  Positions of predicted genes and exons: Variant   1 from    1, Score:10.892921
    G Str   Feature    Start         End     Score            ORF            Len

    1 +     1 CDSf       1161 -       1361    20.71      1161 -       1361     201
    1 +     2 CDSi       1895 -       2009    -1.39      1895 -       2008     114
    1 +     3 CDSi      10378 -      10486     8.13     10380 -      10484     105
    1 +     4 CDSl      12324 -      12348    -3.82     12325 -      12348      24
    1 +       PolA      12421                 -1.08

 Predicted protein(s):
 >FGENESH:[mRNA]    1    4 exon (s)    1161   -   12348    450 bp, chain +
```

ATGGGAGAGAAGAAGGAGGGCGGTGGTGGGGGTGATGCGGCGGCCACGGAGGGTGGCGCAGGGGCTGCGGCCAGC
CGGGCGCTGCAGCAGTGCGGGCAGCTCCAGAAGCTCATCGTCGTCTTCATTGGCAGCCTGTGCGGGCTGTGCACCAA
GTGCGCTGTGTCCAACGACCTCACCCAGCAGGAGATACAGACCCTGGAGGAAGAAGCAAATGGGTTTGCTTTCCTAG
CTCTGTCCAGTACTTTAGGGACCCTGAGGACTGAAGAGATTCTTGGAGAGCCATCTGGTGTATGTCATGGGTGGGTC
TTTTTTGAAGTATCAAGCTCCCGTGCTGCTAGAGAAAGTCCTCAGCTAGAAACTGGTGCAAATTCTGGAGATGAGGCC
TTGTGGAGTGTTAAGAATGGTTTTCTTCCCAGCAGCTACAGGGAGCATTGCCTATCTCTAGATTAA

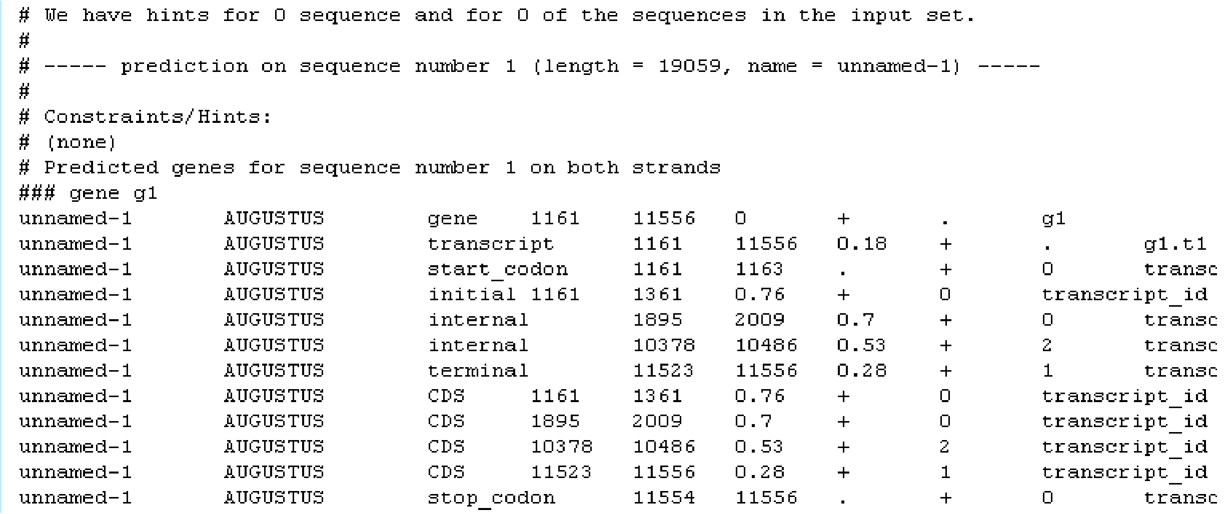


```
>FGENESH:    1    4 exon (s)    1161   -   12348    149 aa, chain +
```

MGEKKEGGGGGDAAATEGGAGAAASRALQQCGQLQKLIVVFIGSLCGLCTKCAVSNDLTQQEIQTLEEEANGFAFLAL

SSTLGTLRTEEILGEPSGVCHGWVFFEVSSSRAARESPQLETGANSGDEALWSVKNGFLPSSYREHCLSLD

AUGUSTAS Result: ((http://augustus.gobics.de/)



```
# We have hints for 0 sequence and for 0 of the sequences in the input set.
#
# ----- prediction on sequence number 1 (length = 19059, name = unnamed-1) -----
#
# Constraints/Hints:
# (none)
# Predicted genes for sequence number 1 on both strands
### gene g1
unnamed-1       AUGUSTUS        gene    1161    11556   0       +       .       g1
unnamed-1       AUGUSTUS        transcript      1161    11556   0.18    +       .       g1.t1
unnamed-1       AUGUSTUS        start_codon     1161    1163    .       +       0       transc
unnamed-1       AUGUSTUS        initial 1161    1361    0.76    +       0       transcript_id
unnamed-1       AUGUSTUS        internal        1895    2009    0.7     +       0       transc
unnamed-1       AUGUSTUS        internal        10378   10486   0.53    +       2       transc
unnamed-1       AUGUSTUS        terminal        11523   11556   0.28    +       1       transc
unnamed-1       AUGUSTUS        CDS     1161    1361    0.76    +       0       transcript_id
unnamed-1       AUGUSTUS        CDS     1895    2009    0.7     +       0       transcript_id
unnamed-1       AUGUSTUS        CDS     10378   10486   0.53    +       2       transcript_id
unnamed-1       AUGUSTUS        CDS     11523   11556   0.28    +       1       transcript_id
unnamed-1       AUGUSTUS        stop_codon      11554   11556   .       +       0       transc
```

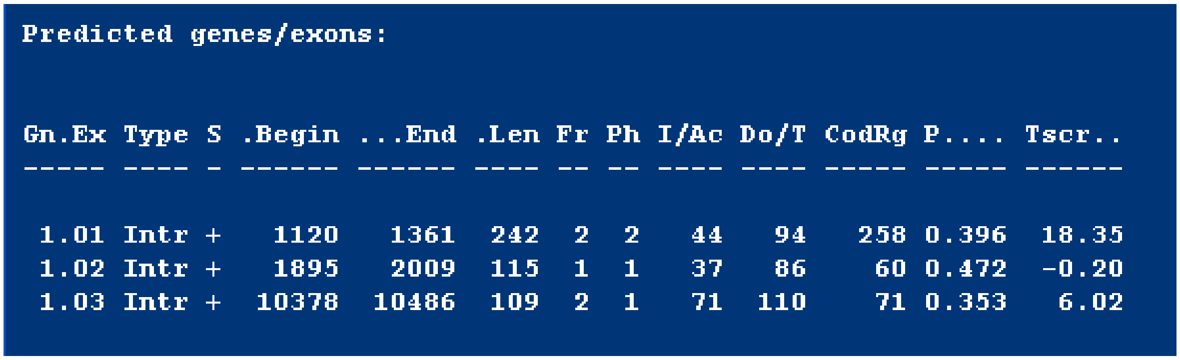
coding sequence

atgggagagaagaaggagggcggtggtgggggtgatgcggcggccacggagggtggcgcaggggctgcggccagccgggcgctgcagcagtgcgggcagctccagaagctca
tcgtcgtcttcattggcagcctgtgcgggctgtgcaccaagtgcgctgtgtccaacgacctcacccagcaggagatacagaccctggaggaagaagcaaatgggtttgctttcctagc
tctgtccagtactttagggaccctgaggactgaagagattcttggagagccatctggtgtatgtcatgggtgggtcttttttgaagtatcaagctcccgtgctgctagagaaagtcctca
gctagaaactggtgcaaattctggagatgaggccttgtggagtgttaagaatggttttcttcccagcagctacagttcagattatcaatttgaagacttactccgctag

protein sequence

MGEKKEGGGGGDAAATEGGAGAAASRALQQCGQLQKLIVVFIGSLCGLCTKCAVSNDLTQQEIQTLEEEANGFAFLAL
SSTLGTLRTEEILGEPSGVCHGWVFFEVSSSRAARESPQLETGANSGDEALWSVKNGFLPSSYSSDYQFEDLLR

GENSCAN Result: (http://genes.mit.edu/GENSCAN.html)

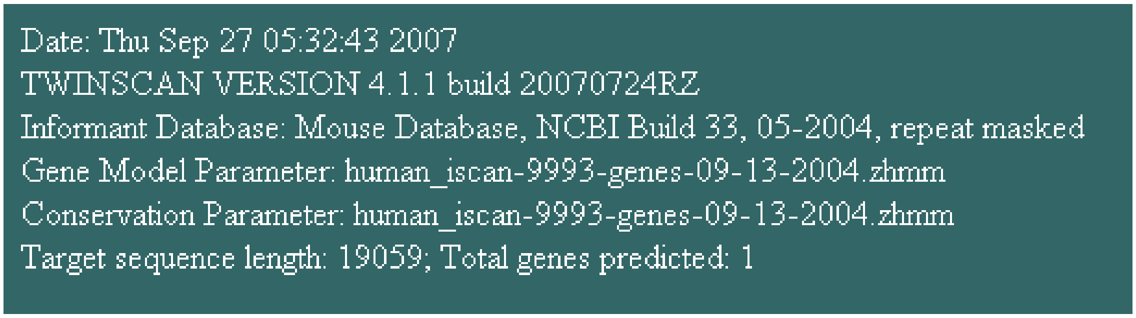


```
Predicted genes/exons:


Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..
----- ---- - ------ ------ ---- -- -- ---- ---- ----- ----- ------

 1.01 Intr +   1120   1361  242  2  2   44   94   258 0.396  18.35
 1.02 Intr +   1895   2009  115  1  1   37   86    60 0.472  -0.20
 1.03 Intr +  10378  10486  109  2  1   71  110    71 0.353   6.02
```

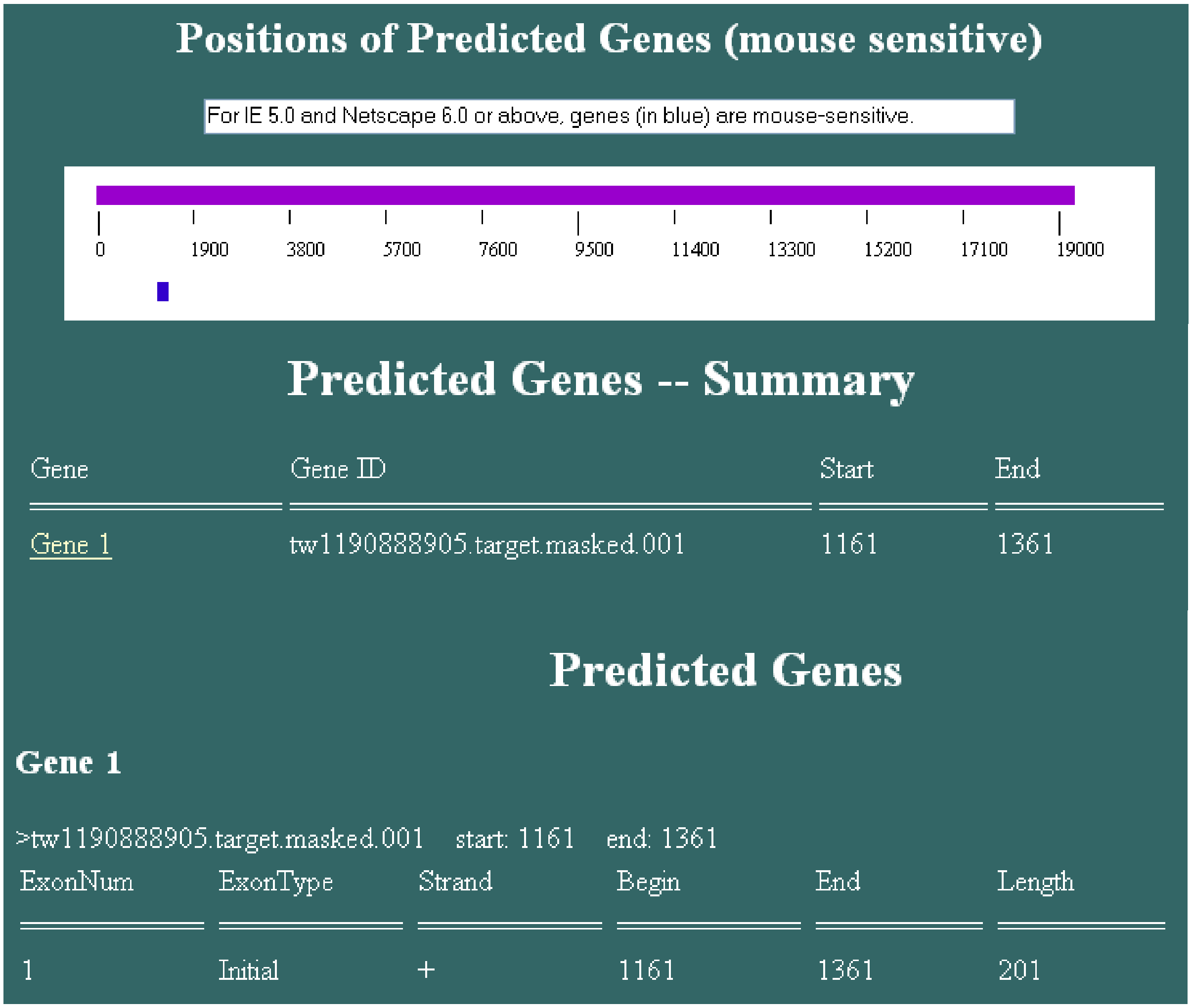
>07:15:03|GENSCAN_predicted_peptide_1|156_aa

XPCGLHRAALPVEAMGEKKEGGGGGDAAATEGGAGAAASRALQQCGQLQKLIVVFIGSLCGLCTKCAVSNDLTQQEI
QTLEEEANGFAFLALSSTLGTLRTEEILGEPSGVCHGWVFFEVSSSRAARESPQLETGANSGDEALWSVKNGFLPSSYX

TWINSCAN Result: (http://mblab.wustl.edu/nscan/submit/)



Date: Thu Sep 27 05:32:43 2007
TWINSCAN VERSION 4.1.1 build 20070724RZ
Informant Database: Mouse Database, NCBI Build 33, 05-2004, repeat masked
Gene Model Parameter: human_iscan-9993-genes-09-13-2004.zhmm
Conservation Parameter: human_iscan-9993-genes-09-13-2004.zhmm
Target sequence length: 19059; Total genes predicted: 1



Positions of Predicted Genes (mouse sensitive)

For IE 5.0 and Netscape 6.0 or above, genes (in blue) are mouse-sensitive.

Predicted Genes -- Summary

Gene	Gene ID	Start	End
Gene 1	tw1190888905.target.masked.001	1161	1361

Predicted Genes

Gene 1

>tw1190888905.target.masked.001 start: 1161 end: 1361

ExonNum	ExonType	Strand	Begin	End	Length
1	Initial	+	1161	1361	201

>tw1190888905.target.masked.001 coding sequences:

ATGGGAGAGAAGAAGGAGGGCGGTGGTGGGGGTGATGCGGCGGCCACGGAGGGTGGCGCAGGGGCTGCGGCCAGCC
GGGCGCTGCAGCAGTGCGGGCAGCTCCAGAAGCTCATCGTCGTCTTCATTGGCAGCCTGTGCGGGCTGTGCACCAAGT
GCGCTGTGTCCAACGACCTCACCCAGCAGGAGATACAGACCCTGGAG

>tw1190888905.target.masked.001 protein sequences:

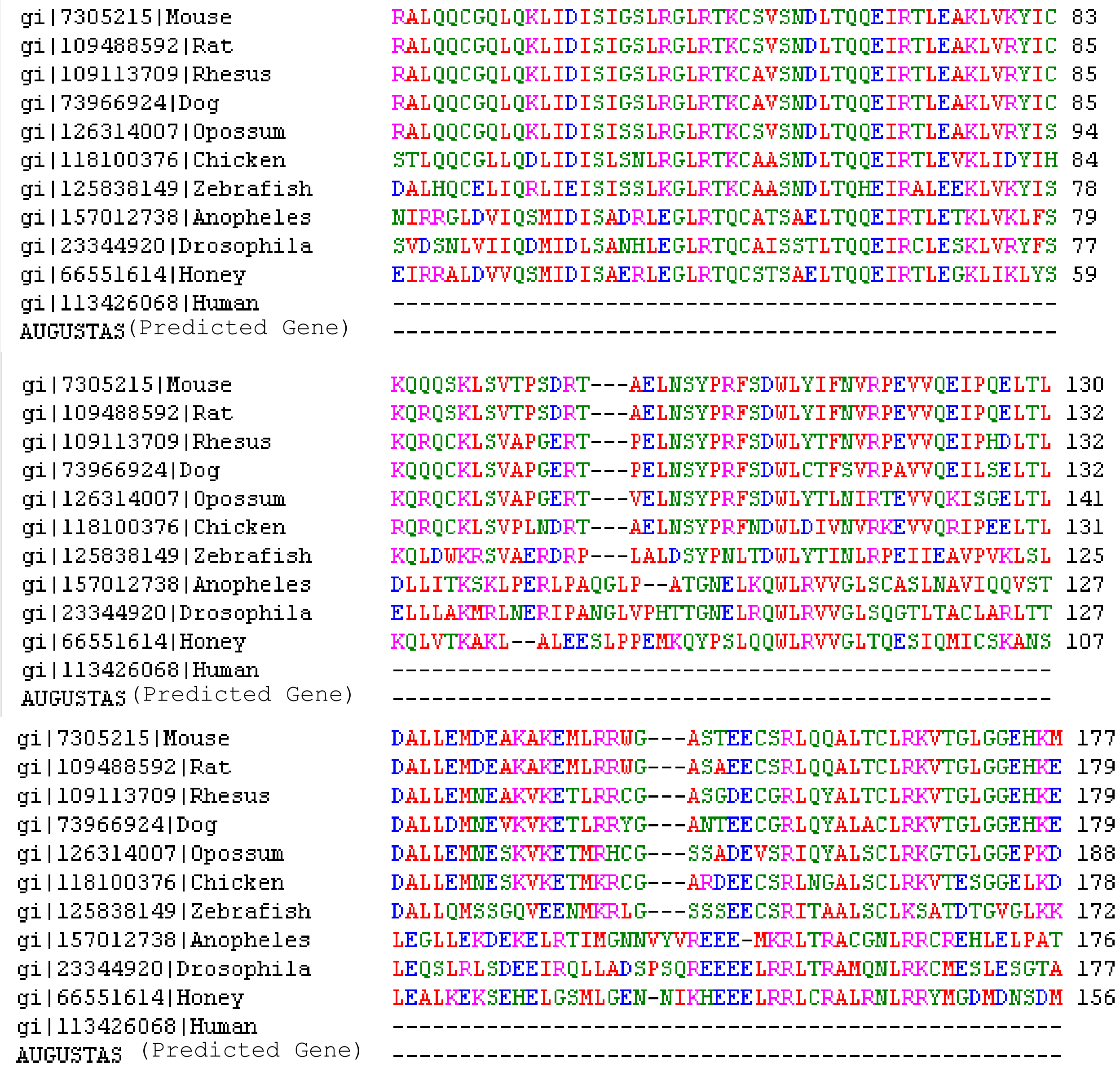
MGEKKEGGGGGDAAATEGGAGAAASRALQQCGQLQKLIVVFIGSLCGLCTKCAVSNDLTQQEIQTLE

Fig. 2. Gene prediction of chimpanzee chromosome 16 genomic contig by using different gene prediction software (FGENESH ,TWINSCAN,AUGUSTUS).

Raf-1 is a serine/threonine kinase poised at a key relay point in mitogenic signal transduction pathways from the cell surface to the nucleus. Sequence comparisons among members of the raf family reveal the presence of three blocks of conserved sequences: conserved regions (CR) 1, 2, and 3. CR1 contains a cysteine-rich "zinc finger" motif, related to a similar structure in protein kinase C, which has been shown to coordinate 2 mol of zinc. (Yu-Hua Chow *et al.*, 1995) CR2 consists of a stretch of 20 amino acids rich in serine and threonine residues; these are targets of phosphorylation by Raf-1 autokinase activity as well as by other serine/threonine-specific protein kinases, including protein kinase C. (Yu-Hua *et al.*, 1995) CR3, which spans the C-terminal half of Raf-1, comprise the catalytic domain.

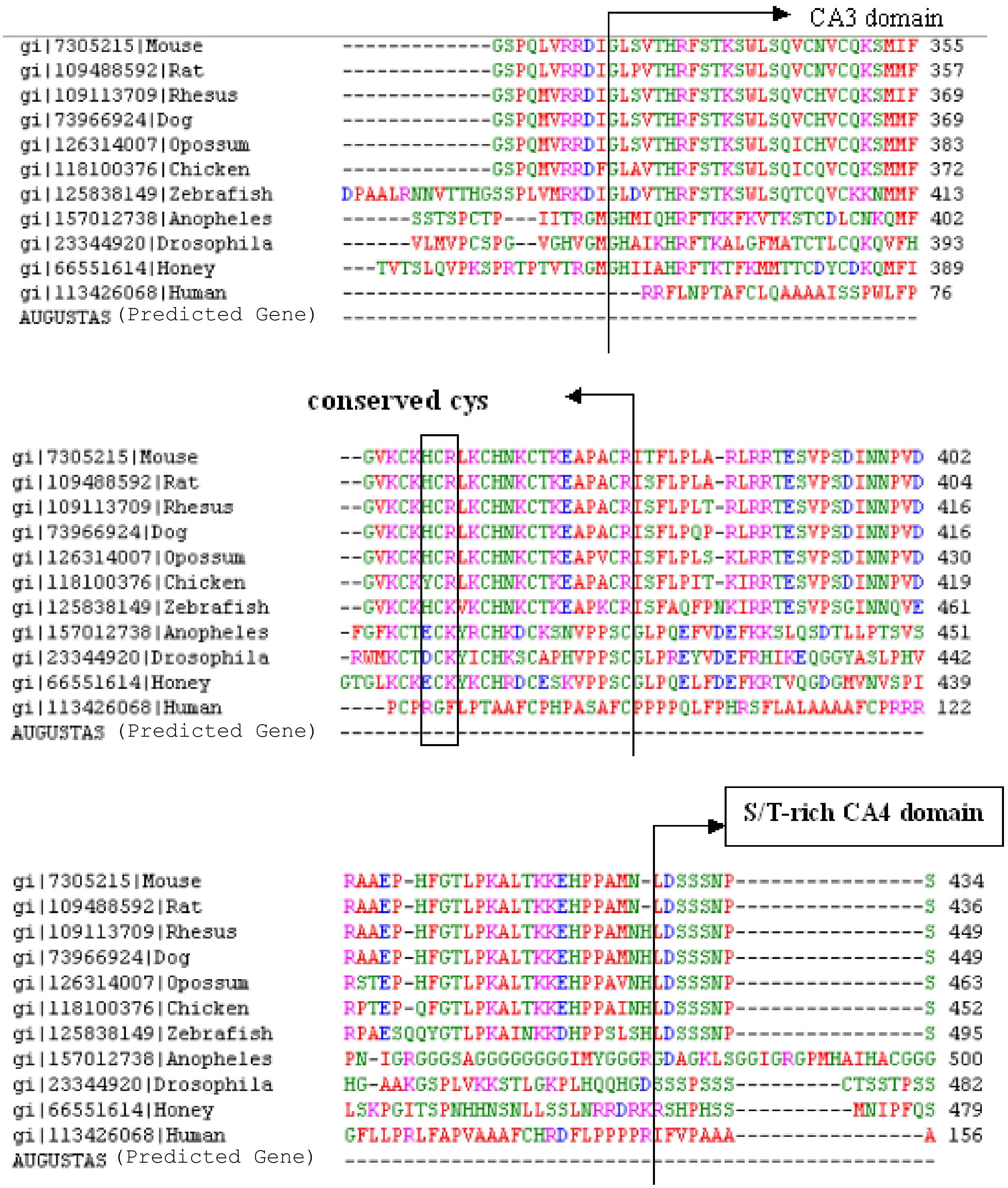
Interspersed among these conserved regions are sequences unique to Raf-1. Current models of Raf-1 regulation postulate that the N-terminal regulatory domain represses the transforming potential and kinase activity of the C-terminal catalytic domain. Activation of the transforming potential of Raf-1 has been associated with N-terminal truncation and/or fusion to other proteins, suggesting that the Raf-1 N-terminal half harbors a negative regulatory domain. Seven internal deletion mutants that together scan the entire N-terminal half of human Raf-1 protein were generated to map functional regions in this regulatory domain. (Yu-Hua *et al.*, 1995).

Amino Terminal Domain (N-Terminus):



```
gi|7305215|Mouse              RALQQCGQLQKLIDISIGSLRGLRTKCSVSNDLTQQEIRTLEAKLVKYIC 83
gi|109488592|Rat              RALQQCGQLQKLIDISIGSLRGLRTKCSVSNDLTQQEIRTLEAKLVRYIC 85
gi|109113709|Rhesus           RALQQCGQLQKLIDISIGSLRGLRTKCAVSNDLTQQEIRTLEAKLVRYIC 85
gi|73966924|Dog               RALQQCGQLQKLIDISIGSLRGLRTKCAVSNDLTQQEIRTLEAKLVRYIC 85
gi|126314007|Opossum          RALQQCGQLQKLIDISISSLRGLRTKCSVSNDLTQQEIRTLEAKLVRYIS 94
gi|118100376|Chicken          STLQQCGLLQDLIDISLSNLRGLRTKCAASNDLTQQEIRTLEVKLIDYIH 84
gi|125838149|Zebrafish        DALHQCELIQRLIEISISSLKGLRTKCAASNDLTQHEIRALEEKLVKYIS 78
gi|157012738|Anopheles        NIRRGLDVIQSMIDISADRLEGLRTQCATSAELTQQEIRTLETKLVKLFS 79
gi|23344920|Drosophila        SVDSNLVIIQDMIDLSANHLEGLRTQCAISSTLTQQEIRCLESKLVRYFS 77
gi|66551614|Honey             EIRRALDVVQSMIDISAERLEGLRTQCSTSAELTQQEIRTLEGKLIKLYS 59
gi|113426068|Human            --------------------------------------------------
AUGUSTAS(Predicted Gene)      --------------------------------------------------

gi|7305215|Mouse              KQQQSKLSVTPSDRT---AELNSYPRFSDWLYIFNVRPEVVQEIPQELTL 130
gi|109488592|Rat              KQRQSKLSVTPSDRT---AELNSYPRFSDWLYIFNVRPEVVQEIPQELTL 132
gi|109113709|Rhesus           KQRQCKLSVAPGERT---PELNSYPRFSDWLYTFNVRPEVVQEIPHDLTL 132
gi|73966924|Dog               KQQQCKLSVAPGERT---PELNSYPRFSDWLCTFSVRPAVVQEILSELTL 132
gi|126314007|Opossum          KQRQCKLSVAPGERT---VELNSYPRFSDWLYTLNIRTEVVQKISGELTL 141
gi|118100376|Chicken          RQRQCKLSVPLNDRT---AELNSYPRFNDWLDIVNVRKEVVQRIPEELTL 131
gi|125838149|Zebrafish        KQLDWKRSVAERDRP---LALDSYPNLTDWLYTINLRPEIIEAVPVKLSL 125
gi|157012738|Anopheles        DLLITKSKLPERLPAQGLP--ATGNELKQWLRVVGLSCASLNAVIQQVST 127
gi|23344920|Drosophila        ELLLAKMRLNERIPANGLVPHTTGNELRQWLRVVGLSQGTLTACLARLTT 127
gi|66551614|Honey             KQLVTKAKL--ALEESLPPEMKQYPSLQQWLRVVGLTQESIQMICSKANS 107
gi|113426068|Human            --------------------------------------------------
AUGUSTAS(Predicted Gene)      --------------------------------------------------

gi|7305215|Mouse              DALLEMDEAKAKEMLRRWG---ASTEECSRLQQALTCLRKVTGLGGEHKM 177
gi|109488592|Rat              DALLEMDEAKAKEMLRRWG---ASAEECSRLQQALTCLRKVTGLGGEHKE 179
gi|109113709|Rhesus           DALLEMNEAKVKETLRRCG---ASGDECGRLQYALTCLRKVTGLGGEHKE 179
gi|73966924|Dog               DALLDMNEVKVKETLRRYG---ANTEECGRLQYALACLRKVTGLGGEHKE 179
gi|126314007|Opossum          DALLEMNESKVKETMRHCG---SSADEVSRIQYALSCLRKGTGLGGEPKD 188
gi|118100376|Chicken          DALLEMNESKVKETMKRCG---ARDEECSRLNGALSCLRKVTESGGELKD 178
gi|125838149|Zebrafish        DALLQMSSGQVEENMKRLG---SSSEECSRITAALSCLKSATDTGVGLKK 172
gi|157012738|Anopheles        LEGLLEKDEKELRTIMGNNVYVREEE-MKRLTRACGNLRRCREHLELPAT 176
gi|23344920|Drosophila        LEQSLRLSDEEIRQLLADSPSQREEEELRRLTRAMQNLRKCMESLESGTA 177
gi|66551614|Honey             LEALKEKSEHELGSMLGEN-NIKHEEELRRLCRALRNLRRYMGDMDNSDM 156
gi|113426068|Human            --------------------------------------------------
AUGUSTAS (Predicted Gene)     --------------------------------------------------
```

```
gi|7305215|Mouse            DSGWSSTDARDSSLG------------PPMDMLSSLGRAGASTQGPRSIS 215
gi|109488592|Rat            DSGWSSAEARESSSG------------PSVDMLSSLGRAGASTQGPRSIS 217
gi|109113709|Rhesus         DSSWSSLDARRESGSGPSTDTLSPASLPWPLGSSQLGRAGNSAQGPRSIS 229
gi|73966924|Dog             DSGWSSLDARRESGSGSSVDTPSPASLPWPPGSAQLGRMGSGAQGPRSVS 229
gi|126314007|Opossum        DPAWSPLEARHESGSGSPSDTLSPAGLTWTPGASQLSKG-ISSQQPRSVS 237
gi|118100376|Chicken        DVLTNLSEVRRENSSTNMTESPCSSAPTWTHSALSLPKG--NSQQARSVS 226
gi|125838149|Zebrafish      EGIPWIADPTEHDSTSADPLSCSMRATSYGLKPTSWPSPSALPRLRSCVS 222
gi|157012738|Anopheles      DGSVSRKGTEPHDLFWDSWDRQPACHRTSPRVNRSLGKCSKQSAVALAAA 226
gi|23344920|Drosophila      S-----SNNDPEQWHWDSWDRPTHIHRG------SVGNIGLGNNSTASPR 216
gi|66551614|Honey           NLYWDSWDRHHLRTGASPRPARSRTTRCSVPSEDSIPYHNNNNNLNSDIL 206
gi|113426068|Human          --------------------------------------------------
AUGUSTAS (Predicted Gene)   --------------------------------------------------


gi|7305215|Mouse            VSALPASDSPVPG------LSEGLSDSCIPLHTSGRLTPRALHSFITPPT 259
gi|109488592|Rat            VSALPASDSPVPG------LSEGLSDTCIPLHTSGRLTPRALHSFITPPT 261
gi|109113709|Rhesus         VSGLPASDSPTPS------FSEGLSDTCIPLHASGRLTPRALHSFITPPT 273
gi|73966924|Dog             VSALPTSDSLPPG------PCEGLSDTCILLHTSGRLTPRALHSFITPPT 273
gi|126314007|Opossum        VSALPTASDALPLGHGLSGYGEGLSDSYVSLHHSGRLTPRTPHSFITPPT 287
gi|118100376|Chicken        VSTIPSSDSLASSHGPSIYADNLLDPFAFPAHSGRLTPRTPHSITVTPPT 276
gi|125838149|Zebrafish      VSTLPPSDFPVVVYTR-------CDGLTDPFSSSPQLNCNVSGLLSTPPV 265
gi|157012738|Anopheles      AASSYHHSSQQRHSNGPGMPGEAMGSTSNTSPQTDDGGGSLSLATTQHVE 276
gi|23344920|Drosophila      THHRQHGVKGKNSALANSTNFKSGRQSPSATEELNSTQGSQLTLTLTPSP 266
gi|66551614|Honey           AQASSITSLSSTSPPSTPLLQRPGRGIKVPTTPPPYKKHQIGLQNTISQP 256
gi|113426068|Human          --------------------------------------------------
AUGUSTAS (Predicted Gene)   --------------------------------------------------


gi|7305215|Mouse            TPQLRRHAKLKPPRTPPPPSRKVFQLLPSFPTLTRSKSHESQLGNRIDDV 309
gi|109488592|Rat            TPQLRRHAKLKPPRTPPPPSRKVFQLLPSFPTLTRSKSHESQLGNRIDEV 311
gi|109113709|Rhesus         TPQLRRHTKLKPPRTPPPPSRKVFQLLPSFPTLTRSKSHESQLGNRIDEV 323
gi|73966924|Dog             TPQLRRHTKLKPPRTPPPPSRKVFQLLPSFPTLTRSKSHESQLGNRIDEV 323
gi|126314007|Opossum        TPQLRRQNKLKPPRTPPPPSRKVFQLLPNFPTLTRSKSHESQLGNRIDEV 337
gi|118100376|Chicken        TPQAKRRHKLKPPRTPPPPCRKVFQLLPSFPTLTRSKSHESQLGNRIDDV 326
gi|125838149|Zebrafish      TPPPKRYGKLKPSRTPPPPSRKLLNLLPSITLTRSK--SQSQLANRIDDL 313
gi|157012738|Anopheles      GAALLNPQHAISPDEASTSGCTPSPSPPNSPSNVQLHNRQNRRGFPTTPP 326
gi|23344920|Drosophila      PNSPFTPSSGLSSSLNGTPQRSRGTPPPARKHQTLLSQSHVQVDGEQLAR 316
gi|66551614|Honey           EVFPLTKSKSHESQLANRLENGDTASSCGNMDPVSSVGRRNRLPTETGSC 306
gi|113426068|Human          -------MAFCPPRLCAPPAPGLLARRGVLTPRRCEFCHRGFLPPAAFCH 43
AUGUSTAS (Predicted Gene)   --------------------------------------------------


gi|7305215|Mouse            TPMKFELPH----------------------------------------- 318
gi|109488592|Rat            APMRFELPH----------------------------------------- 320
gi|109113709|Rhesus         SSMRFDLSH----------------------------------------- 332
gi|73966924|Dog             SPMRFDLPH----------------------------------------- 332
gi|126314007|Opossum        PSIKFDLPH----------------------------------------- 346
gi|118100376|Chicken        PPIKFELSQ----------------------------------------- 335
gi|125838149|Zebrafish      TPNKSGKKNKPLLNVQVNGGANGWEDSPSRSPLLSARSPCLPPNTPSHLT 363
gi|157012738|Anopheles      PRKKHQTALLSFASGGGGGASSSASSVCSPATATV--------------- 361
gi|23344920|Drosophila      NRLPTDPSPDSHSSTSSDIFVDPNTNASSGGSSSN--------------- 351
gi|66551614|Honey           SSNADITGENLLLNSNSPVKSPPYSSAESDDNSFKG-------------- 342
gi|113426068|Human          PPPPRLFAP----------------------------------------- 52
AUGUSTAS (Predicted Gene)   --------------------------------------------------
```

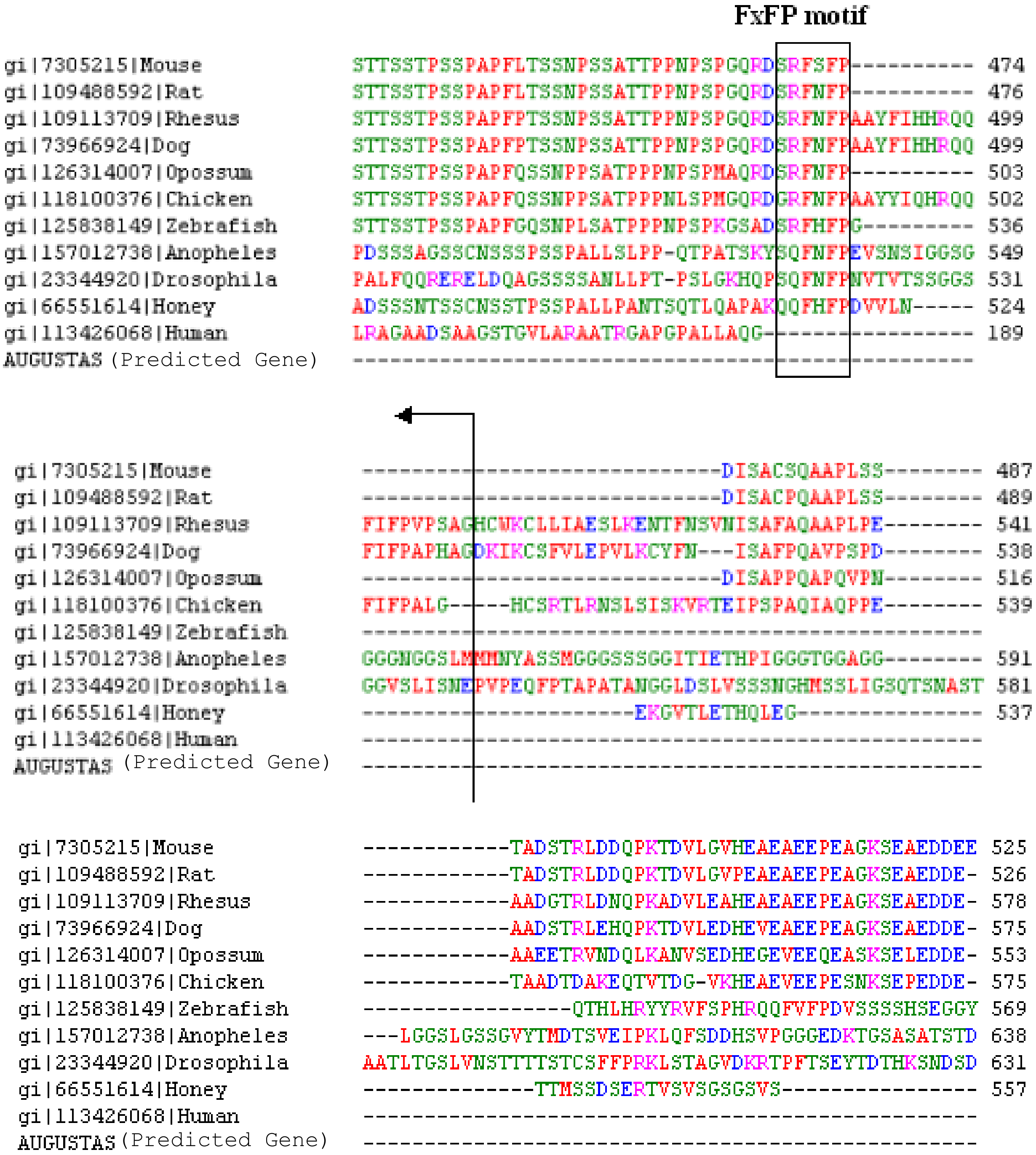


```
gi|7305215|Mouse                -------------GSPQLVRRDIGLSVTHRFSTKSWLSQVCNVCQKSMIF 355
gi|109488592|Rat                -------------GSPQLVRRDIGLPVTHRFSTKSWLSQVCNVCQKSMMF 357
gi|109113709|Rhesus             -------------GSPQMVRRDIGLSVTHRFSTKSWLSQVCHVCQKSMIF 369
gi|73966924|Dog                 -------------GSPQMVRRDIGLSVTHRFSTKSWLSQVCHVCQKSMMF 369
gi|126314007|Opossum            -------------GSPQMVRRDIGLSVTHRFSTKSWLSQICHVCQKSMMF 383
gi|118100376|Chicken            -------------GSPQMVRRDFGLAVTHRFSTKSWLSQICQVCQKSMMF 372
gi|125838149|Zebrafish          DPAALRNNVTTHGSSPLVMRKDIGLDVTHRFSTKSWLSQTCQVCKKNMMF 413
gi|157012738|Anopheles          ------SSTSPCTP---IITRGMGHMIQHRFTKKFKVTKSTCDLCNKQMF 402
gi|23344920|Drosophila          ------VLMVPCSPG--VGHVGMGHAIKHRFTKALGFMATCTLCQKQVFH 393
gi|66551614|Honey               ---TVTSLQVPKSPRTPTVTRGMGHIIAHRFTKTFKMMTTCDYCDKQMFI 389
gi|113426068|Human              --------------------------RRFLNPTAFCLQAAAAISSPWLFP 76
AUGUSTAS (Predicted Gene)       --------------------------------------------------
```

```
gi|7305215|Mouse                --GVKCKHCRLKCHNKCTKEAPACRITFLPLA-RLRRTESVPSDINNPVD 402
gi|109488592|Rat                --GVKCKHCRLKCHNKCTKEAPACRISFLPLA-RLRRTESVPSDINNPVD 404
gi|109113709|Rhesus             --GVKCKHCRLKCHNKCTKEAPACRISFLPLT-RLRRTESVPSDINNPVD 416
gi|73966924|Dog                 --GVKCKHCRLKCHNKCTKEAPACRISFLPQP-RLRRTESVPSDINNPVD 416
gi|126314007|Opossum            --GVKCKHCRLKCHNKCTKEAPVCRISFLPLS-KLRRTESVPSDINNPVD 430
gi|118100376|Chicken            --GVKCKYCRLKCHNKCTKEAPACRISFLPIT-KIRRTESVPSDINNPVD 419
gi|125838149|Zebrafish          --GVKCKHCKVKCHNKCTKEAPKCRISFAQFPNKIRRTESVPSGINNQVE 461
gi|157012738|Anopheles          -FGFKCTECKYRCHKDCKSNVPPSCGLPQEFVDEFKKSLQSDTLLPTSVS 451
gi|23344920|Drosophila          -RWMKCTDCKYICHKSCAPHVPPSCGLPREYVDEFRHIKEQGGYASLPHV 442
gi|66551614|Honey               GTGLKCKECKYKCHRDCESKVPPSCGLPQELFDEFKRTVQGDGMVNVSPI 439
gi|113426068|Human              ----PCPRGFLPTAAFCPHPASAFCPPPPQLFPHRSFLALAAAAFCPRRR 122
AUGUSTAS (Predicted Gene)       --------------------------------------------------
```

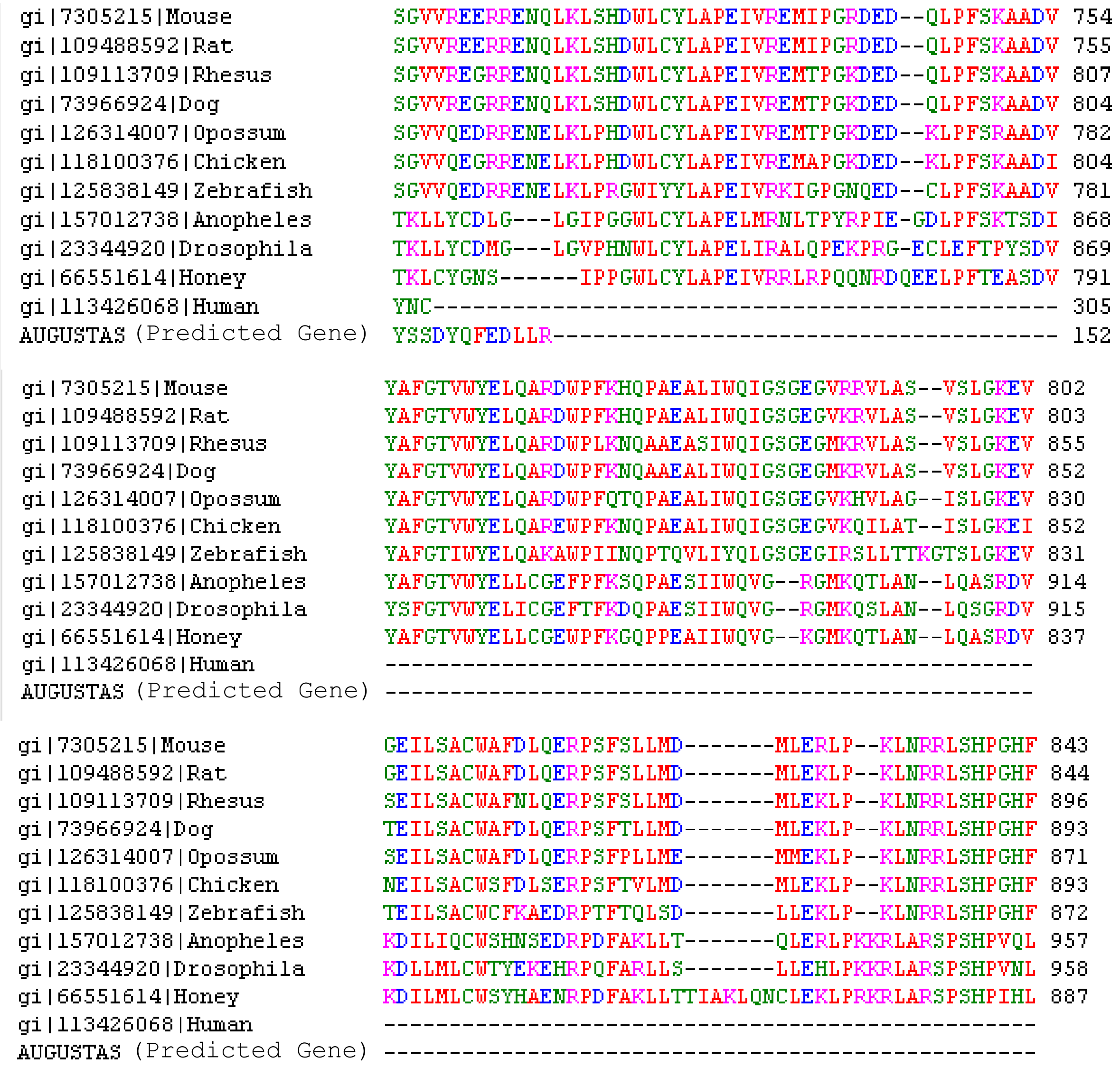
```
gi|7305215|Mouse                RAAEP-HFGTLPKALTKKEHPPAMN-LDSSSNP----------------S 434
gi|109488592|Rat                RAAEP-HFGTLPKALTKKEHPPAMN-LDSSSNP----------------S 436
gi|109113709|Rhesus             RAAEP-HFGTLPKALTKKEHPPAMNHLDSSSNP----------------S 449
gi|73966924|Dog                 RAAEP-HFGTLPKALTKKEHPPAMNHLDSSSNP----------------S 449
gi|126314007|Opossum            RSTEP-HFGTLPKALTKKEHPPAVNHLDSSSNP----------------S 463
gi|118100376|Chicken            RPTEP-QFGTLPKALTKKEHPPAINHLDSSSNP----------------S 452
gi|125838149|Zebrafish          RPAESQQYGTLPKAINKKDHPPSLSHLDSSSNP----------------S 495
gi|157012738|Anopheles          PN-IGRGGGSAGGGGGGGGIMYGGGRGDAGKLSGGIGRGPMHAIHACGGG 500
gi|23344920|Drosophila          HG-AAKGSPLVKKSTLGKPLHQQHGDSSSPSSS---------CTSSTPSS 482
gi|66551614|Honey               LSKPGITSPNHHNSNLLSSLNRRDRKRSHPHSS----------MNIPFQS 479
gi|113426068|Human              GFLLPRLFAPVAAAFCHRDFLPPPPRIFVPAAA----------------A 156
AUGUSTAS (Predicted Gene)       --------------------------------------------------
```

Carboxy Terminal Domain (C-Terminus)



```
gi|7305215|Mouse             STTSSTPSSPAPFLTSSNPSSATTPPNPSPGQRDSRFSFP---------- 474
gi|109488592|Rat             STTSSTPSSPAPFLTSSNPSSATTPPNPSPGQRDSRFNFP---------- 476
gi|109113709|Rhesus          STTSSTPSSPAPFPTSSNPSSATTPPNPSPGQRDSRFNFPAAYFIHHRQQ 499
gi|73966924|Dog              STTSSTPSSPAPFPTSSNPSSATTPPNPSPGQRDSRFNFPAAYFIHHRQQ 499
gi|126314007|Opossum         STTSSTPSSPAPFQSSNPPSATPPPNPSPMAQRDSRFNFP---------- 503
gi|118100376|Chicken         STTSSTPSSPAPFQSSNPPSATPPPNLSPMGQRDGRFNFPAAYYIQHRQQ 502
gi|125838149|Zebrafish       STTSSTPSSPAPFGQSNPLSATPPPNPSPKGSADSRFHFPG--------- 536
gi|157012738|Anopheles       PDSSSAGSSCNSSSPSSPALLSLPP-QTPATSKYSQFNFPEVSNSIGGSG 549
gi|23344920|Drosophila       PALFQQRERELDQAGSSSSANLLPT-PSLGKHQPSQFNFPNVTVTSSGGS 531
gi|66551614|Honey            ADSSSNTSSCNSSTPSSPALLPANTSQTLQAPAKQQFHFPDVVLN----- 524
gi|113426068|Human           LRAGAADSAAGSTGVLARAATRGAPGPALLAQG----------------- 189
AUGUSTAS (Predicted Gene)    --------------------------------------------------

gi|7305215|Mouse             -----------------------------DISACSQAAPLSS-------- 487
gi|109488592|Rat             -----------------------------DISACPQAAPLSS-------- 489
gi|109113709|Rhesus          FIFPVPSAGHCWKCLLIAESLKENTFNSVNISAFAQAAPLPE-------- 541
gi|73966924|Dog              FIFPAPHAGDKIKCSFVLEPVLKCYFN---ISAFPQAVPSPD-------- 538
gi|126314007|Opossum         -----------------------------DISAPPQAPQVPN-------- 516
gi|118100376|Chicken         FIFPALG-----HCSRTLRNSLSISKVRTEIPSPAQIAQPPE-------- 539
gi|125838149|Zebrafish       --------------------------------------------------
gi|157012738|Anopheles       GGGNGGSLMMMNYASSMGGGSSSGGITIETHPIGGGTGGAGG-------- 591
gi|23344920|Drosophila       GGVSLISNEPVPEQFPTAPATANGGLDSLVSSSNGHMSSLIGSQTSNAST 581
gi|66551614|Honey            ----------------------EKGVTLETHQLEG--------------- 537
gi|113426068|Human           --------------------------------------------------
AUGUSTAS (Predicted Gene)    --------------------------------------------------

gi|7305215|Mouse             ------------TADSTRLDDQPKTDVLGVHEAEAEEPEAGKSEAEDDEE 525
gi|109488592|Rat             ------------TADSTRLDDQPKTDVLGVPEAEAEEPEAGKSEAEDDE- 526
gi|109113709|Rhesus          ------------AADGTRLDNQPKADVLEAHEAEAEEPEAGKSEAEDDE- 578
gi|73966924|Dog              ------------AADSTRLEHQPKTDVLEDHEVEAEEPEAGKSEAEDDE- 575
gi|126314007|Opossum         ------------AAEETRVNDQLKANVSEDHEGEVEEQEASKSELEDDE- 553
gi|118100376|Chicken         ------------TAADTDAKEQTVTDG-VKHEAEVEEPESNKSEPEDDE- 575
gi|125838149|Zebrafish       -----------------QTHLHRYYRVFSPHRQQFVFPDVSSSSHSEGGY 569
gi|157012738|Anopheles       ---LGGSLGSSGVYTMDTSVEIPKLQFSDDHSVPGGGEDKTGSASATSTD 638
gi|23344920|Drosophila       AATLTGSLVNSTTTTSTCSFFPRKLSTAGVDKRTPFTSEYTDTHKSNDSD 631
gi|66551614|Honey            --------------TTMSSDSERTVSVSGSGSVS---------------- 557
gi|113426068|Human           --------------------------------------------------
AUGUSTAS (Predicted Gene)    --------------------------------------------------
```

```
gi|7305215|Mouse            ----------DEVDDLPSSRRPWRG-------PISRKASQTSVYLQEWDI 558
gi|109488592|Rat            ----------DEIDDLPSSRRHWRG-------PISRKASQTSVYLQEWDI 559
gi|109113709|Rhesus         ----------DEVDDLPSSRRPWRG-------PISRKASQTSVYLQEWDI 611
gi|73966924|Dog             ----------DEVDDLPSSRRPWRG-------PISRKASQTSVYLQEWDI 608
gi|126314007|Opossum        ----------DELDDLPYSRKPWKG-------MISRKASKTSVYLQEWDI 586
gi|118100376|Chicken        ----------DEVEDLPNRRPHLQG-------MIYRKPSQTSVYLQEWDI 608
gi|125838149|Zebrafish      ----------TGESDQLDGVASQSS-------SGHR-------------- 588
gi|157012738|Anopheles      ----------SSSDRTPIRLDSTEE-----RDHESSWPRQNSLSLKEWDI 673
gi|23344920|Drosophila      KTVSLSGSASTDSDRTPVRVDSTE-------DGDSGQWRQNSISLKEWDI 674
gi|66551614|Honey           ----------TDSERTPVRVDSQDSQVSDGEPGDSRWPRQNSLSMREWDI 597
gi|113426068|Human          -----------FLAEAPAPRAPCLG----------------------PCG 206
AUGUSTAS(Predicted Gene)    --------------------------------------------------

gi|7305215|Mouse            PFEQVELGEPIGQGRWGRVHRGRWHGEVAIRLLEMDGHNQDH--LKLFKK 606
gi|109488592|Rat            PFEQVELGEPIGQGRWGRVHRGRWHGEVAIRLLEMDGHNQDH--LKLFKK 607
gi|109113709|Rhesus         PFEQVELGEPIGQGRWGRVHRGRWHGEVAIRLLEMDGHNQDH--LKLFKK 659
gi|73966924|Dog             PFEQVELGEPIGQGRWGRVHRGRWHGEVAIRLLEMDGHNQDH--LKLFKK 656
gi|126314007|Opossum        PFEQVELGEPIGQGRWGKVHRGRWHGEVAIRLLEIDGNNQDH--LKLFKK 634
gi|118100376|Chicken        PFEQIELGDPIGQGRWGKVHKGKWHGEVAIRLLEIDGNNQDH--LKLFKK 656
gi|125838149|Zebrafish      -EDVLELGELIGKGRWGKVCRGRWHGEVAVRLLEIDGNNQEH--LKLFKK 635
gi|157012738|Anopheles      PYDDLHLKEKIGNGRFGTVHRALWHGDVAVKLLKEDYVADER-TLEAFKL 722
gi|23344920|Drosophila      PYGDLLLLERIGQGRFGTVHRALWHGDVAVKLLNEDYLQDEH-MLETFRS 723
gi|66551614|Honey           PYDELKIGEPIGTGRFGTVYRGYWHGNVAIKVLNMDYYLDDDKILEAFKL 647
gi|113426068|Human          LHRVALPAEAMGENKEGGGGGDAAATEGGTGAAASRALQQCGQLQKLIVI 256
AUGUSTAS (Predicted Gene)   ----------MGEKKEGGGGGDAAATEGGAGAAASRALQQCGQLQKLIVV 40
                                      :*  : *         : .          :     : :

gi|7305215|Mouse            EVMNYRQTRHENVVLFMGACMNPPHLAIITSFCKGRTLHSFVRDPKTSLD 656
gi|109488592|Rat            EVMNYRQTRHENVVLFMGACMNPPHLAIITSFCKGRTLHSFVRDPKTSLD 657
gi|109113709|Rhesus         EVMNYRQTRHENVVLFMGACMNPPHLAIITSFCKGRTLHSFVRDPKTSLD 709
gi|73966924|Dog             EVMNYRQTRHENVVLFMGACMNPPHLAIITSFCKGRTLHSFVRDPKTSLD 706
gi|126314007|Opossum        EVMNYRQTRHENVVLFMGACMNPPHLAIITSFCKGRTLHSFVRDPKTSLD 684
gi|118100376|Chicken        EVMNYRQTRHENVVLFMGACMNPPHLAIITSFCKGRTLYSFVRDPKISLD 706
gi|125838149|Zebrafish      EVMNYRQTRHENVVLFMGACMHPPHLAIITSFCKGRTLYSVVRD--SKLD 683
gi|157012738|Anopheles      EVATFKKTRHENVVLFMGACMKPPRLAIVTSLCKGNTLFTHIHLRKDKFN 772
gi|23344920|Drosophila      EVANFKNTRHENLVLFMGACMNPPYLAIVTSLCKGNTLYTYIHQRREKFA 773
gi|66551614|Honey           EVATFRKTRHENLVLFMGACMKPPRLAIVTSMSKGMTLYTHIHLRKDKFN 697
gi|113426068|Human          FIGSLCGLCTKCAVSNDLTQQEIQTLEILFN------------------- 287
AUGUSTAS (Predicted Gene)   FIGSLCGLCTKCAVSNDLTQQEIQTLEEEANGFAFLALSSTLGTLRTEEI 90
                             : .      :  *    :  .   *    .

gi|7305215|Mouse            INKTRQIAQEIIKGMGYLHAKGIVHKDLKSKNVFYDNGKVVITDFGLFGI 706
gi|109488592|Rat            INKTRQIAQEIIKGMGYLHAKGIVHKDLKSKNVFYDNGKVVITDFGLFGI 707
gi|109113709|Rhesus         INKTRQIAQEIIKGMGYLHAKGIVHKDLKSKNVFYDNGKVVITDFGLFGI 759
gi|73966924|Dog             INKTRQIAQEIIKGMGYLHAKGIVHKDLKSKNVFYDNGKVVITDFGLFGI 756
gi|126314007|Opossum        INKTRQIAQEIIKGMGYLHAKGIVHKDLKSKNIFYDNGKVVITDFGLFGI 734
gi|118100376|Chicken        INKTRQIAQEIIKGMGYLHAKGIVHKDLKSKNVFYDNGKVVITDFGLFGI 756
gi|125838149|Zebrafish      INKIRQIAQEIVKGMGYLHAKGIVHKDLKSKNVFYDSNKVVITDFGLFGM 733
gi|157012738|Anopheles      LNRTTLVAQQISQGMGYLHARGIVHKDLKTKNIFLENGKVIITDFGLFSA 822
gi|23344920|Drosophila      MNRTLLIAQQIAQGMGYLHAREIIHKDLRTKNIFIENGKVIITDFGLFSS 823
gi|66551614|Honey           INKTTIIAQQISQGMGYLHARGIIHKDLKSKNIFLESGKVVITDFGLFSV 747
gi|113426068|Human          -----------------------------------IPHMVEIKNHCLFKI 302
AUGUSTAS (Predicted Gene)   LGEPSGVCHGWVFFEVSSSRAARESPQLETGANSGDEALWSVKNGFLPSS 140
                                                                     :.:  *
```



```
gi|7305215|Mouse          SGVVREERRENQLKLSHDWLCYLAPEIVREMIPGRDED--QLPFSKAADV 754
gi|109488592|Rat          SGVVREERRENQLKLSHDWLCYLAPEIVREMIPGRDED--QLPFSKAADV 755
gi|109113709|Rhesus       SGVVREGRRENQLKLSHDWLCYLAPEIVREMTPGKDED--QLPFSKAADV 807
gi|73966924|Dog           SGVVREGRRENQLKLSHDWLCYLAPEIVREMTPGKDED--QLPFSKAADV 804
gi|126314007|Opossum      SGVVQEDRRENELKLPHDWLCYLAPEIVREMTPGKDED--KLPFSRAADV 782
gi|118100376|Chicken      SGVVQEGRRENELKLPHDWLCYLAPEIVREMAPGKDED--KLPFSKAADI 804
gi|125838149|Zebrafish    SGVVQEDRRENELKLPRGWIYYLAPEIVRKIGPGNQED--CLPFSKAADV 781
gi|157012738|Anopheles    TKLLYCDLG---LGIPGGWLCYLAPELMRNLTPYRPIE-GDLPFSKTSDI 868
gi|23344920|Drosophila    TKLLYCDMG---LGVPHNWLCYLAPELIRALQPEKPRG-ECLEFTPYSDV 869
gi|66551614|Honey         TKLCYGNS------IPPGWLCYLAPEIVRRLRPQQNRDQEELPFTEASDV 791
gi|113426068|Human        YNC----------------------------------------------- 305
AUGUSTAS (Predicted Gene) YSSDYQFEDLLR-------------------------------------- 152

gi|7305215|Mouse          YAFGTVWYELQARDWPFKHQPAEALIWQIGSGEGVRRVLAS--VSLGKEV 802
gi|109488592|Rat          YAFGTVWYELQARDWPFKHQPAEALIWQIGSGEGVKRVLAS--VSLGKEV 803
gi|109113709|Rhesus       YAFGTVWYELQARDWPLKNQAAEASIWQIGSGEGMKRVLAS--VSLGKEV 855
gi|73966924|Dog           YAFGTVWYELQARDWPFKNQAAEALIWQIGSGEGMKRVLAS--VSLGKEV 852
gi|126314007|Opossum      YAFGTVWYELQARDWPFQTQPAEALIWQIGSGEGVKHVLAG--ISLGKEV 830
gi|118100376|Chicken      YAFGTVWYELQAREWPFKNQPAEALIWQIGSGEGVKQILAT--ISLGKEI 852
gi|125838149|Zebrafish    YAFGTIWYELQAKAWPIINQPTQVLIYQLGSGEGIRSLLTTKGTSLGKEV 831
gi|157012738|Anopheles    YAFGTVWYELLCGEFPFKSQPAESIIWQVG--RGMKQTLAN--LQASRDV 914
gi|23344920|Drosophila    YSFGTVWYELICGEFTFKDQPAESIIWQVG--RGMKQSLAN--LQSGRDV 915
gi|66551614|Honey         YAFGTVWYELLCGEWPFKGQPPEAIIWQVG--KGMKQTLAN--LQASRDV 837
gi|113426068|Human        --------------------------------------------------
AUGUSTAS (Predicted Gene) --------------------------------------------------

gi|7305215|Mouse          GEILSACWAFDLQERPSFSLLMD-------MLERLP--KLNRRLSHPGHF 843
gi|109488592|Rat          GEILSACWAFDLQERPSFSLLMD-------MLEKLP--KLNRRLSHPGHF 844
gi|109113709|Rhesus       SEILSACWAFNLQERPSFSLLMD-------MLEKLP--KLNRRLSHPGHF 896
gi|73966924|Dog           TEILSACWAFDLQERPSFTLLMD-------MLEKLP--KLNRRLSHPGHF 893
gi|126314007|Opossum      SEILSACWAFDLQERPSFPLLME-------MMEKLP--KLNRRLSHPGHF 871
gi|118100376|Chicken      NEILSACWSFDLSERPSFTVLMD-------MLEKLP--KLNRRLSHPGHF 893
gi|125838149|Zebrafish    TEILSACWCFKAEDRPTFTQLSD-------LLEKLP--KLNRRLSHPGHF 872
gi|157012738|Anopheles    KDILIQCWSHNSEDRPDFAKLLT-------QLERLPKKRLARSPSHPVQL 957
gi|23344920|Drosophila    KDLLMLCWTYEKEHRPQFARLLS-------LLEHLPKKRLARSPSHPVNL 958
gi|66551614|Honey         KDILMLCWSYHAENRPDFAKLLTTIAKLQNCLEKLPRKRLARSPSHPIHL 887
gi|113426068|Human        --------------------------------------------------
AUGUSTAS (Predicted Gene) --------------------------------------------------
```

Fig. 3. Multiple sequence alignment of the predicted gene with that of human and other species showing N-terminus and C-terminus region.

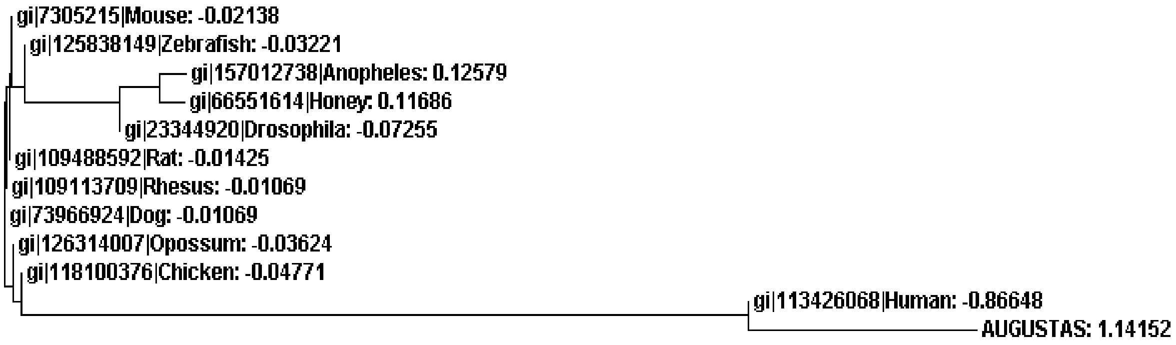


Fig. 4. Phylogenetic tree constructed on the basis of predicted gene and showing humans are most closely related to Chimpanzee.

Effects of the deletion mutations on kinase activity of Raf-1 were evaluated using a baculovirus/insect cell overexpression system and an in vitro kinase assay with the known physiological substrate of Raf-1, mitogen-activated protein kinase kinase. (Yu-Hua *et al.,* 1995) Deletion of amino acids 276-323 in the unique sequence between conserved regions 2 and 3 leads to modest elevation of Raf-1 basal kinase activity, whereas deletion of amino acids 133-180 in conserved region 1 results in diminished kinase activity. Surprisingly, none of the Raf-1 N-terminal deletion mutants, including a truncated version that is transforming in rodent fibroblasts, exhibits greatly increased levels of basal kinase activity (Yu-Hua *et al.,* 1995). In addition, while activation of Raf-1 kinase by Ras requires sequences in conserved region 1, only the C-terminal half containing the kinase domain of Raf-1 is required for activation by Src. These findings demonstrate that N-terminal deletions in Raf-1 do not necessarily result in constitutively elevated basal kinase activity and that the N-terminal regulatory domain is completely dispensable for Raf-1 activation by Src.

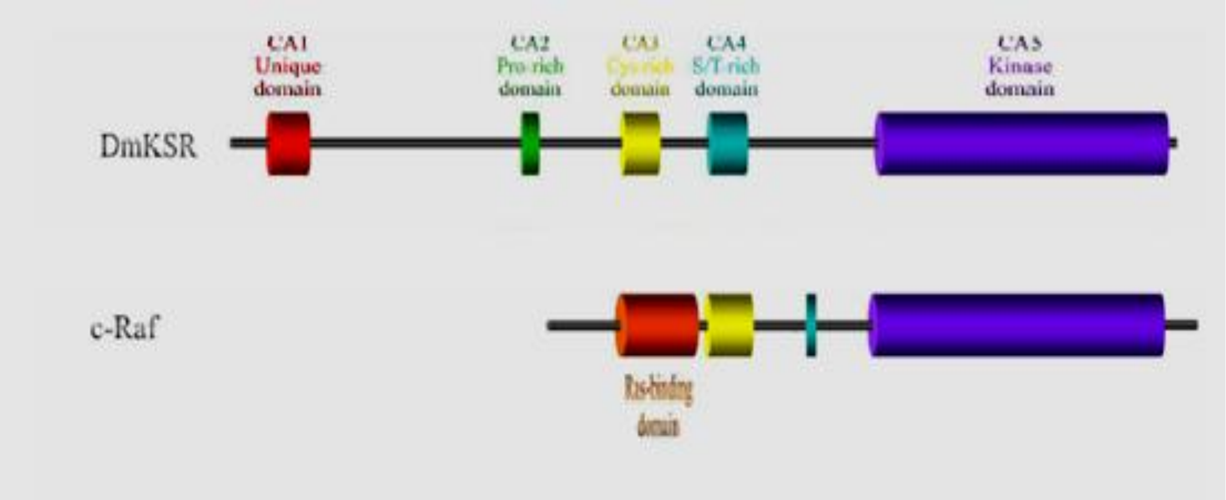


Fig. 5. Domain architecture of KSR (Catherine, 2005).

In our multiple sequence alignment conserved regions are found in near the C-terminal, which suggests the importance of C terminal region for Raf activity. Furthermore it was also found that the N-terminal region is missing in both human and chimpanzee and thus N –terminal region is dispensable for raf-1 activation and support the previous finding. Moreover CA3 domain with conserved cys and the S/T-rich CA4 domain with FxFP motif were also been identified. From phylogenetic tree it was also been elucidated that the KSR1 of human and chimpanzee are evolutionary distant from other organisms and human is most closely related with chimpanzee than any other organism.

References

Vojtek AB, Der CJ. 1998. Increasing complexity of the Ras signaling pathway. J. Biol. Chem. **273**, 19925–19928.

Barbazuk W, Korf I, Kadavi C, Heyen J, Tate S, Wun E, Bedell J, McPherson J, Johnson S. 2000. The syntenic relationship of the zebrafish and human genomes. Genome Research **10,** 1351–1358.

Catherin AK. 2005. Lipid binding and the scaffolding function of the Kinase Suppressor of Ras. PhD thesis, University of Pittsburgh, USA, 15-27.

Cooke SF, Bliss TV. 2006. Plasticity in the human central nervous system. Brain **129**, 1659–73.

Dawson DA, Akesson M, Burke T, Pemberton JM, Slate J, Hansson B. 2007. "Gene order and recombination rate in homologous chromosome regions of the chicken and a passerine bird". Mol. Biol. Evol. **24,** 1537–52.

Sieburth S, Sundaram M, Howard RM, Han M. 1999. A PP2A regulatory subunit positively regulates Ras-mediated signaling during Caenorhabditis elegans vulval induction. Genes Dev. **13**, 2562–2569.

Morrison DK. 2001. Ksr: a mapk scaffold of the ras pathway. J. Cell Sci. **114**, 1609–1612.

Engström PG, Ho Sui SJ, Drivenes O, Becker TS, Lenhard B. 2007. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. Genome Research **17**, 1898–908.

Heger A, Ponting CP. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 Drosophila genomes. Genome Res. **17,** 1837–49.

K Kornfeld, Hom DB, Horvitz HR. 1995. The ksr-1 gene encodes a novel protein kinase involved in ras-mediated signaling in Caenorhabditis elegans. Cell **83**, 903–913.

Sundaram M, Han M. 1995. The C. elegans ksr-1 gene encodes a novel Raf-related kinase involved in Ras-mediated signal transduction. Cell **83,** 889–901.

Therrien M, Michaud NR, Rubin G, Morrison DK. 1996. KSR modulates signal propagation within the MAPK cascade. Genes Dev. **10**, 2684–2695.

Moreno-Hagelsieb G, Treviño V, Pérez-Rueda E, Smith TF, Collado-Vides J. 2001. Transcription unit conservation in the three domains of life: a perspective from *Escherichia coli*. Trends in Genetics **17**, 175–177.

Passarge EB, Horsthemke RA, Farber. 1999. Incorrect use of the term synteny. Nature Genetics **23,** 387-93.

Postlethwait J, Woods I, Ngo-Hazelett P, Yan Y, Kelly P, Chu F, Huang H, Hill-Force A, and Talbot W. 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res. **10,** 1890–1902

Poyatos JF, Hurst LD. 2007. The determinants of gene order conservation in yeasts. Genome Biol. **8**, 233-40.

Shahani N. 2007. A Comparative study of genes related to Long-Term potentiation (LTP) and memory development. 4th year thesis, University of Dhaka, Bangladesh, 10-23.

Sara CS, Caterina MH, Michele KD, Deborah KM, Sweatt JD. 2006. Kinase Suppressor of Ras1 Compartmentalizes Hippocampal Signal Transduction and Subserves Synaptic Plasticity and Memory Formation. Neuron **50**, 7 65-79.

Yu-Hua C, Kevin P, Toni HJ, Paul Dent, Thomas WS, Richard J. 1995. Functional Mapping of the N-terminal Regulatory Domain in the Human Raf-1 Protein Kinase. J Biol Chem. **270**, 14100-14106.