In silico sequence analysis of Gibberellin-20 Oxidase-2, the semidwarfing protein of rice through the application of BLAST, FASTA and WU-Blast 2 programs

ASTHA DUHAN, MANOJ KUMAR SHARMA AND TAVINDERJEET KAUR

Accepted : September, 2009

SUMMARY

Computational prediction of evolutionary relationship for genes and proteins remains a fundamental problem in biology. Multiple sequence alignment plays a very important role in the evolutionary relationship prediction. The quality measurement is the most crucial issue in the multiple sequence alignment. An appropriate scoring function, high residue identity and degree of divergence lead to more accurate and reliable multiple sequence alignments. Using two scoring matrices, Blosum and PAM series of substitution and similarity matrices, sequence analysis tools for identifying evolutionary relationships to GA200x2, the semi-dwarf protein of rice have been selected. The quality of alignment produced by various scoring matrices features the parameters that are important for the selection of right matrix for any alignment tool. We also show that the quality of any alignment and the homologous sequences produced by BLAST has more predictive power in identifying evolutionary conserved residues for the semi-dwarf protein. This was made evident by the conserved residue pattern observed for homologous sequences scanned by BLAST (Blosum80) resulted in 18 conserved residues by multiple sequence analysis in comparison with FASTA and WU-Blast2 that showed 11 and 12 conserved residues, respectively using the scoring matrix PAM 120.

Key words : Rice, GA₂₀, Semidwarf-1, sd-1, Blosum, Pam, FASTA, BLAST, Clustal, Semidwarfing gene

The semidwarfing gene in rice (sd-1) is one of the most important genes deployed in modern rice breeding. Its recessive character results in a shortened culm with improved lodging resistance and a greater harvest index, allowing for the increased use of nitrogen fertilizers. The sd-1 gene was first identified in the Chinese variety Dee-geo-woo-gen (DGWG), and was crossed in the early 1960s with Peta (tall) to develop the semidwarf cultivar IR8, which produced record yields throughout Asia and formed the basis for the development of new highyielding, semidwarf plant types (IRRI, 1967). Plant architecture refers to the collection of all important agronomic characters which determines the grain production. In rice it is mainly affected by plant height, tillering and panicle morphology (Wang et al., 2005). In rice, short stature plants were developed by altering the plant architecture. Short stature in rice was developed by a recessive semidwarfgene sd-1, responsible for high yielding varieties (Nagano et al., 2005). The sd-1 gene codes for the oxidase enzyme involved in the biosynthesis of gibberellins, Gibberellin-20 oxidase-2(GA20ox2).

Correspondence to:

GA20ox2 is tightly linked to the sd-1 locus of rice chromosome 1. Semidwarfing gene inhibits the elongation of lower internodes which makes rice resistant oto lodging (Ogi et al., 1993). The short stature of IR8 is due to a mutation in the plant's sd1 gene, encoding an oxidase enzyme involved in the biosynthesis of gibberellin, a plant growth hormone (Sasaki et al., 2002).Ga20ox2 catalyses the conversion of GA₅₃ to GA₂₀. DGWG sd-1 mutants have 383 bp deletion in genomic sequence which encode a non-functional protein in an *indica* semidwarf IR8 (Monna et al., 2002). However, a substitution of Leu-266 that is highly conserved residue resulted in the loss of function in japonica semidwarf. Rice represents enormous gene pool for improvement of rice cultivars as well as other crops which show significant similarity to rice genome. By making the use of sequence similarity/ homology genetic relationships among cereal crops can be established (Ishii et al., 1996). Comparative genomic analysis revealed that the sdw1/denso gene that controls plant height, yield and quality has located in the syntenic region of the rice semidwarf gene sd1 on chromosome 1 (Jia et al., 2009).

Protein sequences contain valuable information to predict structure and function of gene product. The comparison of two sets is a fundamental task that infers about how two sequences may be related functionally and genetically. The simultaneous alignment of amino acid sequences is now a major tool in molecular biology. The advent of large genome projects led to an explosion of

ASTHA DUHAN, Department of Bioinformatics, Mata Gujri College, FATEHGARH SAHIB (PUNJAB) INDIA **Authors' affiliations:**

MANOJ KUMAR SHARMA, Department of Bioinformatics, Janta Vedi College, Baraut, BAGHPAT (U.P.) INDIA

TAVINDERJEET KAUR, Department of Bioinformatics, Mata Gujri College, FATEHGARH SAHIB (PUNJAB) INDIA

sequence data in public databases. Modern genome annotation and analysis tools rely heavily on accurate multiple alignments. The role of multiple sequence alignments in such systems has changed simply transferring annotation from one sequence to another to a genome wide perspective (Timo *et al.*, 2002).

Multiple alignments are used to find diagnostic pattern to characterize protein families to detect homology between new sequences and existing families of sequences to help and predict the function of new sequences and an essential preclude to molecular evolutionary analysis (Altschul, 1993). Therefore, all methods capable of handling larger problems in practical timescales make use of heuristics. Currently, the most widely used approach is to exploit the fact that homologous sequences are evolutionary related. One can build up a multiple alignment progressively by a series of pairwise alignments following the branching order in a phylogenetic tree (Feng et al., 1987). Pairwise alignment of very closely related sequences can be obtained accurately by using a wide range of parameter values like gap penalities and weight matrix and the sequences of different degree of divergenece (Barton et al., 1987).

Numerous sequence alignment tools has been developed such as FASTA (Pearson *et al.*, 1988), WU-Blast2 (Altschul *et al.*, 1990) and BLAST (Altschul *et al.*, 1997) to identify the sequential pattern of similar and identical residues over a family of proteins to identify homologues or distant evolutionary relationships. Considering this fact we represent the selection of sequence analysis tools for GA200x2 protein sequence taken from SWISS-PROT (http://expasy.org) sequence database (ID P0C5H5.1) to find homologous and ancestral sequences that have similar or related functions/ sequences.

MATERIALS AND METHODS

Application of sequence analysis tools to evaluate the scoring matrices that were employed in BLAST (Blastp), FASTA (fasta3), WU-Blast2 (WU-Blast2 protein) and clustal (Higgins *et al.*, 1994) are highlighted. FASTA is a two-step algorithm (Pearson *et al.*, 1988). The first step is a word search with a specific word size which finds regions in a two-dimensional table that are likely to correspond to highly similar segments of two sequences. The second step is a Smith-Waterman alignment (Smith *et al.*, 1981) centered on the diagonals that correspond to the alignment of highly similar sequence segments. FASTA is a heuristic approach to Smith-Waterman algorithm. The heuristics used by FASTA allow it to run much faster than the Smith-Waterman algorithm (Pearson, 1991). WU-Blast2 stands for Washington University Basic Local Alignment Search Tool Version 2.0 (http://www.ebi.ac.uk/blast2). The emphasis of this tool is to find regions of sequence similarity or homology at a faster speed, with minimum loss of sensitivity. This will yield functional and evolutionary clues about structure and function of query sequence. WU-Blast2 is a gapped version of BLASt allowing for gapped alignments and statistics. The BLAST (Altschul et al., 1997) uses a word-based heuristic similar to FASTAthrough Maximal Segment Pairs algorithm. Blast do not allow gaps and have a valuable property that their statistics is more significant (Karlin et al., 1990). POC5H5.1, a Gibberellin-20 oxidase-2 sequence from Oryza sativa indica from SWISS-PROT database was selected. Programs FASTA and WU-Blast2 from European Bioinformatics Institute (http://www.ebi.ac.uk), BLAST from National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov) and Clustal from EBI were used as sequence analysis tools.

The scoring matrices for all the three sequence analysis tools were compared and the common matrices among them were identified and selected. Five matrices that are common among the three tools like Blosum 50, 62, 80 and PAM 120, 250 matrices were selected to study FASTA and WU-Blast2 whereas Blosum 45, 62, 80 and PAM 30, 70 were selected for NCBI BLAST. P0C5H5.1 (389 amino acid length) sequence was scanned against Uni-Prot Knowledge-Base database for similar and identical residues. The methodology is divided into three parts. First, the raw sequence scanned against FASTA with 5 scoring matrices and the per cent identity residues were tabulated. The same procedure was followed for WU-Blast2 and BLAST programs. Second, the parameters like % identity, scores (E-score, bit score, % positives) and amino acid overlaps were compared with respect to the highest and lowest hits in FASTA, WU-Blast2 and BLAST. Third, of all the 5 scoring matrices, the best parameters is evaluated from each tool and highest score from highest hits and lowest hits are identified and subjected to multiple sequence alignment through ClustalX₂. For all scoring matrices and tools default options are retained.

RESULTS AND DISCUSSION

A search against a protein database yielded several alignments using five scoring matrices and the scores along with amino acid overlaps accompanying these alignments are used to distinguish sequences related to degree of divergence. Five similarity matrices are used to identify the degree of evolutionary divergence. The output of each

[Internat. J. Plant Sci., Jan. - June, 2010, 5 (1)]

scoring matrix for FASTA, WU-Blast2 and BLAST were given in Table 1, 2 and 3, respectively.

Maximum per cent identity using five scoring matrices for FASTA was 100% with GAOX2_ORYSJ(Gibberellin-20 oxidase2 protein from (Oryza sativa japonica) for all matrices and a minimum of 22.8 identity observed with ASCL1_RAT (Achaetescute homolog 1 from rat)for Blosum50. Comparatively 40.5% is having highest identity from lowest hits observed for IRAK1_BOVIN (Interleukin-1 receptor associated protein) when Blosum80 and PAM120 is employed as a scoring matrix. On an average, about 104 protein sequences have shown identity varying from a maximum of 100% to a minimum of 22.8% with five scoring matrices. A maximum of 389 residue overlap and a minimum of 42 residue overlap observed for five scoring matrices using FASTA and the % similarity and the Evalueranges for highest and lowest hits for FASTA are presented in Table 1.

Per cent identity for WU-Blast2 showed GAOX2_ORYSJ for major % identity with P0C5H5.1 with all scoring matrices. SEPS_METM7 (O-phosphoseryl-tRNA(Cys) synthetase) showed the lowest % identity for PAM250 matrix with 118% identity. Interestingly the number of similar proteins varied with different matrices. A highest of 249-250 similar aligned protein sequences was resulted when Blosum 50 80 and PAM 250 were observed in contrast with 103 and 99

similar sequences with Blosum62 and PAM 120, respectively. The %positives, scores and E-value ranges for highest and lowest hits for WU-Blast2 using five scoring matrices have been shown in Table 2.

NCBI blastp results obtained with different scoring matrices are given in Table 3. A maximum of 100% identity observed for all matrices where PAM 30 and 70 replaced against PAM 120 and 250 in above two cases. PAM 30 showed a 100% identity for GAOX2_ORYSJ with lowest 47% limit for VP13B_MOUSE (Vacuolar protein sorting-associated protein 13B). An average of about 95 hits observed when NCBI Blast is used for sd-1 sequences analysis. A maximum of 389 residue overlap and a minimum of 18 residue overlap observed for five scoring matrices. E-value and bit score ranges for highest and lowest hits are presented in Table 3.

The result from PAM 120 matrix with both FASTA and WU-Blast2 are comparable as the maximum per cent identity for both the tools resulted in 100% and 91%, respectively with GAOX2_ORYSJ as a maximum hit and the minimum hit resulted in the 40.5% identity in FASTA and 30% identity in WU-Blast2. Whereas using BLAST, 18 residues is the lowest residue overlap resulted from Blosum 80 matrix. Therefore based on the statistics given in Table 1, 2 and 3, PAM 120 matrix for both FASTA and WU-Blast2 and Blosum 80 matrix from BLAST results were selected to study multiple sequence alignments by ClustalX2.

Tab	le 1 : Comp	arison of	f scoring	matrices s	canned a	against FA	ASTA showing pe	rcent ide	ntity with o	orrespor	iding p	arameters <i>i.e.</i> %
	identi	ty, % sin	nilarity, o	overlap res	idues, E	-value and	l homolog protein	id in obs	erved high	est and lo	west h	ts
Sr		No. of		FA	ASTA (H	ighest hits)		FAS	STA (Lov	vest hits)
No.	Matrix	hits	% identity	% similarity	overlap	E- value	Swiss-Prot id	% identity	% similarity	overlap	E- value	Swiss-Prot id
1.	Blosum50	112	100.0	100.0	389	3.5e-161	GAOX2_ORYSJ	22.8	54.0	189	7.6	ASCL1_RAT
2.	Blosum62	106	100.0	100.0	389	4.8e-201	GAOX2_ORYSJ	36.8	61.4	57	9.5	TATB_PSEF5
3.	Blosum80	109	100.0	100.0	389	0	GAOX2_ORYSJ	40.5	54.8	42	9.6	IRAK1_BOVIN
4.	PAM120	91	100.0	100.0	389	0	GAOX2_ORYSJ	40.5	66.7	42	8	IRAK1_BOVIN
5.	PAM250	106	100.0	100.0	389	1.4e-117	GAOX2_ORYSJ	24.1	63.0	108	10	COBS_CLOAB

 Table 2 : Comparison of scoring matrices scanned against WU-Blast2 showing percent identity with corresponding parameters *i.e.*

 %identity, %positives, Score, E-value and homolog protein id in observed Highest and Lowest hits

		No		WU	BLAST	-2 (Highest	hits)		WUBI	LAST-2	2 (Lowest	hits)
Sr. No.	Matrix	of hits	% identity	% positives	Score	E-value	Swiss-Prot id	% identity	% positives	Score	E-value	Swiss-Prot id
1.	Blosum50	249	88	88	2247	3.0e-221	GAOX2_ORYSJ	20	37	116	0.026	FAT2_SCHPO
2.	Blosum62	103	91	91	1743	2.0e-179	GAOX2_ORYSJ	29	50	56	8.9	RPON_METKA
3.	Blosum80	250	88	88	2843	5.4e-272	GAOX2_ORYSJ	27	42	138	0.0026	KIF1A_AEDAE
4.	PAM120	99	91	91	1745	5.9e-210	GAOX2_ORYSJ	30	63	70	6.2	COBS_CLOAB
5.	PAM250	250	88	88	1701	6.6e-161	GAOX2_ORYSJ	18	48	131	0.0012	SEPS_METM7

[Internat. J. Plant Sci., Jan. - June, 2010, 5 (1)]

•HIND AGRICULTURAL RESEARCH AND TRAINING INSTITUTE

Table	3 : Compa paramo hits in	eters <i>i.e.</i> each mat	scoring %identity rix	matrice , Bit So	es scann core, ove	ed agains rlap resid	st BLAST (Blast lues, E-value and l	o) show homolog j	ing per protein i	cent ider id in obse	tity with erved Hig	h corresponding hest and Lowest
S.		No. of]	BLAST (Highest hi	ts)		E	BLAST (L	owest hit	s)
No.	Matrix	hits	% identity	Bit score	overlap	E-value	Swiss-Prot id	% identity	Bit Score	overlap	E-value	Swiss-Prot id
1.	Blosum4 5	102	100.0	829	389	0.0	GAOX2_ORYSJ	22	31.8	19	6.4	TRA2_ECOLX
2.	Blosum62	101	100.0	793	389	0.0	GAOX2_ORYSJ	52	33.5	18	2.2	DEOC_RHORT
3.	Blosum80	102	100.0	828	389	0.0	GAOX2_ORYSJ	54	34.0	18	2.5	DEOC_RHORT
4.	PAM30	80	100.0	819	389	0.0	GAOX2_ORYSJ	47	32.5	18	9.4	VP13B_MOUSE
5.	PAM70	94	100.0	799	389	0.0	GAOX2_ORYSJ	51	32.0	18	9.6	DEOC_RHORT

ClustalX2 default matrix is Gonnet250scoring matrix that uses an iterative classical distance measures to estimate an alignment of proteins (Gonnet *et al.*, 1992). Out of 91 hits obtained from FASTA with PAM 120, 99 hits obtained from WU-Blast2 with Pam 120 scoring matrix and 102 hits from BLAST with Blosum 80 scoring matrix, we screened the hits by selecting those sequences with % identity greater than or equal to 30. Then in the second round of screening only the crop species were kept. After the screening process only 44, 41 and 21 sequences remained in the hits for FASTA (PAM120), WU-Blast2 (PAM120) and BLAST (Blosum80) programs.

A multiple sequence alignment was done with ClustalX2 for the 21 hits resulted from BLAST program (Blosum80). All sequences were downloaded from SWISS PROT in fasta format and subjected to multiple sequence alignment along with the query sequence. The program was run with the default parameters like a value of 10 for gap open penalty and a value of 0.2 as gap extension penalty. Gaps are represented by "-" symbol. Consensus symbol were given underneath the alignments. Any alignment denoting the symbol "*" means that the residue in that column is identical in all sequences and the symbol ":" means that the conserved substitutions have been observed while "." means that the semi-conserved substitutions have been observed (http://www.ebi.ac.uk/ clustalw). From the output result given in Fig. 1, there are a total of 18 residues with "*" which represented identical residues from all the aligned sequences and a total of 22 residues with ":" which represented conserved substitutions and 12 residues with "." which represented semi-conserved substitutions. This feature is taken as the criteria to evaluate the different methods applied to align GA20ox2 protein using five scoring matrices against FASTA, WU-Blast2 and BLAST tools.

Similarly multiple sequence alignment was done for 44 hits resulted from FASTA with PAM120 matrix. The parameters were set as stated earlier. From the output, there are only 11 residues that represented identical feature of aligned sequences, which show a poor evolutionary relationship. About 12 residues represented conserved substitutions and 9 residues represented semiconserved substitutions. The same procedure applied for aligning the sequences of 41 hits from WU-Blast2 with Pam 120 matrix. The result showed 12 conserved residues for aligned sequences and 16 conserved substitution residues with 6 semi-conserved substitution residues. The relevant data was presented in Table 4.

Therefore, from the above given data, it is evident that applying different scoring matrices for 3 sequence alignment tools resulted in evaluation of one scoring matrix that showed a better multiple sequence alignment with ClustalX2.

Conclusion:

In order to use any sequence alignment tool with different scoring matrices, one must have to quantify scoring matrices that may likely to conserve the physical and chemical properties necessary to maintain the structure and function of the protein. Multiple sequence

Table 4 :	Comparison of alignm GONNET250matrix. (N represents aligned iden conserved substitutions)	ent toolsagainst scoring matrix usin (o. of hits represent the no. of protein atical residues: ':' represents conserve	ng CluatlX2 multipl ns subjected to multi ed substitutions resid	le sequence iple sequen lues and '.	e analysis ce alignm 'represent	s using ent; '*' s semi-
		ClustalX2				
Sr. No.	Alignment tool	Scoring matrix	No. of hits	•*'	•••	:
1.	FASTA	PAM 120	44	11	12	9
2.	WU-Blast2	PAM 120	41	12	16	6
3.	BLAST	Blosum 80	21	18	22	12

[Internat. J. Plant Sci., Jan. - June, 2010, 5 (1)]

•HIND AGRICULTURAL RESEARCH AND TRAINING INSTITUTE

		TDR	68	-
	4 sp 084m83.1 ACCH1 -MESSIPOVAALDESTLIKAFBETKTGVKGLIDAGITEIPSIFRAPPATLTSPKPPSSSDFSIPTIDLKG	GGT	72	
	6 SD 039110.2 GAOX1 -MAVSFVTTSPEE - EDKPKLGLGNIOT-PLIFNPSMLNLOANIPNOFIWPDDEKP-SINVLELDVPL-IDLONL	LS-	71	
	7 sp Q39111.1 GAOX2 -MAILCTTTSPAEK-EHEPKQLLEKDQTSPLIFNPSLLNLQSQIPNQFIWPDEEKP-SIDIPELNVPF-IDLS		69	
	8 sp Q39112.1 GAOX3 -MATECIATVPQIFSENKTKETSSIFDAKLLNQHSHH-IPQQFVWPDHEKP-STDVQPLQVPL-IDLAGF	LSG	69	
	1 sp 004705.1 GAOLDWQPVFDAAVLSGRADIPSOFIWPEGESFTPDAAEELHVPL-IDIGGM	LSG	50	
	6 sp 004707.1 GAOIA WVRP	LSG	50	
pr pr<	5 SP 004/00.1 GAOLEWVQP	LSG	50	
		PDC	74	
si pp (96776.1, 20000)	5 sp 039103.2 G30X1 -MPAMLTDVFRGHP	PLI	60	
p: p:<	5 sp 092T84.2 G30X2 -MSETLSDVFRSHP	PLI	53	
	0 sp 09C971.1 G30X4 -MPSLAEEICIGNLGSLQTLPESFTWKLTAADSLLRPSSAVSFD-AVEESI	PVI	52	
1 1 0.2144.0.1.1.0302	6 sp Q3I411.1 G3O21 -MPTPSHLSKDPRGGPDAV	PVV	52	
I = D (201400 - 1 (2012) HERAM HER	7 sp Q3I410.1 G3022 -MPTPAHLSKDPRGGPDAV	PVV	52	
17/17/2010/2013/10.114/	8 sp Q3I409.1 G3023 -MPTPAHLSKDPHGEPDAV	PVV	52	
The 100030.1 (First) The second sec	7 SP Q95818.1 FL3H A - MAPGT	IDD	48	
I = [1] I = [1] <t< td=""><td>SP P28038.1 FL3M HOK - MAPYSNETFL</td><td>DDE</td><td>53</td><td></td></t<>	SP P28038.1 FL3M HOK - MAPYSNETFL	DDE	53	
9 9 9 9 0 9 0 9 0 9 0 9 0 9 0 9 0	S6 gn 1272 W9 1 2CC1	PDE	21	
pp p10967.1 Accil 60 First True True (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	9 sp 92614.1 G207 -MASOPPFKINFCSIFGSSFPNSTSESNINTSTIOTSGIKLPVIDLSHI	TSG	51	
14 1		ai - 101		
app 10667.1.1 Accc3. by 114.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1				
The 16 (2007) 11.1.1.2.CO. 11.1.2.CO. 11.1.1.1.2.CO. 11.1.1.1.1.2.CO.			140	
<pre>1 do (3911.1.2 choil _ post_tost is accretisty view dots is a space with a provide a constraint of the post is a space with a provide a constraint of the post is a space with a provide a constraint of the post is a space with a provide a constraint of the post is a space with a post is a post is a space with a post is a space with a post is a space with a post is post is a post is post is a post is a post is a post is a post is post</pre>	SP P1096/1.1 ACCH3 SO TPIKHKEIVEXVERASEKWEPFOVVNHGIPTSVLDRTLOGTROPFEQDNEVKKOYYTRDTAKKVVYTSNLDLMKE	SVP	140	
<pre>1 pi 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</pre>	4 SP Q84MB3.1 ACCHI ESITARSUVERIGUARAMAGET VINEGIPM VIERMILGIREFIEQUIEVARGET SKUPAS KAVISSAFULES	₽	148	
a) (a) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	7 sp 0.39111.1 GAOX2 SODSTLEADFVIAEACTKHGFFLVVNHGVSSLITADAHELMESFEDMELAGKVACHCHGRGCGVASSFTGEFS		144	
1 004705.1 0xx101 PAAAA VTELVERACTERGEFEVENTIALADAREVONDETTALIADARESUNDETTALIADA	8 SD 039112.1 GAOX3 BSCLASEATELVSKAATKHGFFLITNHGVDESLLSRAYLHMDSFFKAPACEKOKAORKWGESSGYASSFVGRFS		144	
6 mj 004707.1 004107.1 004	1 sp 004705.1 GA01D EPAAAAEVTRLVGEACERHGFFOVVNHGIDAELLADAHRCVDNFFTMPLPEKORALRHPGESCGYASSFTGRFAS		125	
Si pp (04706.1 (a).1 BFRATA TYTLIO ACT BIOP GY UNHOLD ALLA MARCY TYTLIO CALA BARGET 027AB (FTREPA 127) Si pp (13755.2 (a).2 ALBERSYNCOVEN CALE BIOP CONTOUR DATA ALLA MARCHAPPTTPLICE CALA BARGET 027AB (FTREPA 127) Si pp (13755.2 (a).2 ALBERSYNCOVEN CALE BIOP CONTOUR DATA ALLA ALLA ALLA ALLA ALLA ALLA ALLA	6 sp 004707.1 GA01A TAAAAAEVTRLVGEACERHGFFQVVNHGIDAELLADAHRCVDNFFTMPLPEKQRALRRPGES CGYASSFTGRFAS		125	
j pp j2377.1.2 (acc)	5 sp 004706.1 GA018 PRATAEVTRLVGEACERHGFFQVVNHGIDAELLADAHRCVDAFFTMPLPEKORALERPGES CGYASSFTGRFAS		125	
by 19 UVCEDS 1.1 GAXZ LARGULERAAQUAAAACATUPTUPUTUPUTUPUTULAINA REARPHILT 144 by 19 Q230.2 (3 GAXZ LARGULERAAQUAAAACATUPTUPUTUPUTUUTUUTUUTUUTUUTUUTUUTUAAASUUTUUTUUTUUTUUTUUTUUTUUTUUTUUTUUTUUTUUT	3 Sp P93771.2 GAOX1 AERSSVVRQVCEACERHGFFLVVNHGIEAALLERAHRCMDAFFTLPLGEKQRAQERAGES CGYASSFTGRFAS		127	
bit bp [0.9520.4.2 (300.4.2) [0.0510.4.2 (300.4.2) [0.0511.4.1.1 (300.2.1) bit bp [0.0520.4.2 (300.4.2) [0.0511.4.1.1 (300.2.1) [0.0511.4.1.1 (300.2.1) bit bp [0.0520.4.2) [0.0511.4.1.1 (300.2.1) [0.0511.4.1.1 (300.2.1) bit bp [0.0520.4.2) [0.0510.4.1 (300.2.1) [0.0500.4.1 (300.2.1) bit bp [0.0510.4.1 (300.2.1) [0.0500.4.1 (300.2.1) [0.0500.4.1 (300.2.1) bit bp [0.0500.4.1 (300.2.1) [0.0500.4.1 (300.2.1) [0.0500.4	87 SD PUCSH5.1 GAOX2 DAEGLERAAAQVAAACATHGFFQVSEHGVDAALARAALDGASDFFRLPLAEKRRAFRVPGTVSGYTSAHADRFAS		149	
<pre>style="border: 4.6 border: 4.6 border</pre>	5 SD Q39103.2 G30X1 DLHPD-ATNOIGHACRTWGAFQISNHGVPIGLLQDIEFLTGSLFGLPVQRKLKSARSETGVSGYGVARIASFPN		134	
<pre>t b () CTL (11 1 b) CTL (11 1 b) CTL (11 A) CTL (</pre>	5 SP (92784.2 G30X2 LSDLH-VATLVGHACTTWGAFGITNHGVPSRLLDDIEFLTGSLFRLPVQRKLKAARSENGVSGYGVARIASFFN		127	
7 jp (31410.1 jois) 7 jp (31410.1 jois) 8 pp (3140.1 jois) 9 pp (38918.1 jr.18 a.K. VDEKEGICKO, VERATLEVER, VVERATLEVER, JSC, MERAVBER, JC, CORS, DSC, SPFF, L. 256 9 pp (38918.1 jr.18 a.K. VDEKEGICKO, VERATLEVER, VVERATLEVER, JSC, MERAVBER, JC, CORS, DSC, SFFF, L. 256 126 jp (28018.1 jr.18 a.K. VDEKEGICKO, VERATLEVER, VVERATLEVER, JSC, MERAVBER, JC, CORS, DSC, SFFF, L. 256 126 jp (28018.1 jr.18 a.K. VDEKEGICKO, VERATLEVER, VVERATLEVER, JSC, MERAVBER, JC, CORS, DSC, SFFF, L. 256 127 jp (28018.1 jr.18 a.K. VDEKEGICKO, VERATLEVER, VVERATLEVER, JSC, MERAVBER, JC, CORS, JSC, SFFF, L. 257 126 jp (28018.1 jr.18 a.K. VERAVVKAB, SWELP, VVERATLEVER, VVERATLEVER, JSC, MERAVBER, JSC, SFFF, L. 257 127 jp (28018.1 jr.16 a.K. JSC, VERAVVKAB, SWELP, VVERATLEVER, VVERATLEVER, JSC, VERAVBER, JSC, SFFF, L. 257 129 jp (28018.1 jr.16 a.K. JSC, VERAVVKAB, SWELP, VVERATLEVER, VVERATLEVER, JSC, VERAVBER, JSC, SFFF, L. 205 129 jp (28018.1 jr.16 a.K. JSC, VERAVVKAB, SFFF, JSC, VERAVE, JSC, VERAVBER, JSC, JSC, JSC, JSC, JSC, JSC, JSC, JSC	v sp grooti.1 G30A4 = BUSKEL-VILLEGRASINGAFGLANGISUKLELBELSSISKILFDMPSKKLKAASSDKGVSGYGEPRISPFF 6 sp 0.17411 1 G3021 = NWPDPC-LEEVILLEGRAUUTETLEGRUPETELSUUPLIGEVENDIDEDGR		126	
bit S0 031409:1 10027 DWDRRW LCVUW ACT WALL AND THAT ALL AND ALL A	o py strill 03021 _ ERREC-ARADALANDWGALLANGWGALLANGWGALLANGKALGARANANKAUKTGD - CGIGSPISSFI 7 sp 037410.1 (3022 _ ERREC-ARADALANGWGALLANGWGALLANGKAUKAUKAUKAUKAUKAUKAUKAUKAUKAUKAUKAUKAUK		126	
7 pp [299638.1 F13H A TOPERED ICROTTED THAD INVOITED VIEND THE LARDY FALP PER LEPTING CRT GOTT VER LOGBA 123 pp [26303.1 F13H AB A	8 SD 031409.1 G3023 MRDPF-AAEAVGLAAODWGAFLLVGHGVPLDLLVRVEAATAGMFALPASEKMEAVRPGDSCGYGSPPISEFFS		126	
pp [282038.1] FL3H HOR A - REAGING FLOW AARCENGTY TENDAORLIAMT FLAREFERALPARTELETIMESORK GOT VS. ENGLAGEA 126 126 fe [g [282138] L [GLACCOL] Be	7 SD 095818.1 FL3H A VDGKRGEICROIVEACENWGIFOVVDHGVDTNLVADMTRLARDFFALPPEDKLRFDMSGGKKGGFIVSSHLOGEA		123	
ip [96330.1]FLGI AR Bern-VRARVYKAB ENGLFYVYHOTTILLERLGVYGRKFPRLFSSTERSVAFPEDSKETGOVGKTLÖUDIKABK	sp 228038.1 FL3H HOR A RRAQIRDRVAAACEDWGIFQVIDHGVDADLIADMTRLAREFFALPAEDKLRYDMSGGKKGGFIVSSHLQGEA		126	
66 Bp A221W9.1 ACCO1 AMB QLDBAC HWGPF LINH GJT LLDB WELMTKLEYK DQP SVW VERPSOLSK NGUNNDSAEK 88 91 pp 09C614.1 GOCT EVKRKACVK GWVAAAKEWGPF LINH GJT LLDB WELMTKLEYK DQP SVW VERPSOLSK NGUNNDSAEK 101 91 pp 09C614.1 GOCT EVKRKACVK GWVAAAKEWGPF LINH GJT LLDB VERMTLEKK LJD QP SVW VERPSOLSK NGUNNDSAEK 101 91 pp 09C614.1 GOCT ADWINDT LGV TRADPPER	sp Q96330.1 FLS1_AR ESVRRAVVKASEEWGLFQVVNHGIPTELIRRLQDVGRKFFELPSSEKESVAKPEDSKDIEGYGTKLQKDPEG		125	
9 pp [09C6T41.1 [02CX7] EVKRERCYKONVAAAKENGPFQIVHHOIPKUVERMLLEEKKUPDOPPSVVVEREPSDLEKN80KHMPPAG5EA 127 127 pp [010967.1 [ACCH] SO AASHKDTIFCYMADNPPSU	66 sp A221W9.1 ACCO1 AAMEQLDDACENWGFFEILNHGISTELMDEVEKMTKDHYKRVREQFFLEFASKTLKEGCDDVNKAEK		88	
TD P10967.1 ACCH3 SO ASSNETTICYMANNPPEL	9 sp Q9C6I4.1 G2OX7 _ HEVKRKRCVKQMVAAAKEWGFFQIVNHGIPKDVFEMMLLEEKKLFDQPFSVKVRERFSDLSKNSYRWGNPSATSF	A	127	
pp 910967.1 ACCH3 SO ASSMEDTIFCTWAPNDPSE				
gp [0.967.1] ACCH3_SO ASKRDTTFCTMANNPPT				
4 mg 0	SP P10967.1 ACCH3 SO AASWRDTIFCYMAPNPPSL	SYL	205	
6 mp (29110.2 (mox1. KLPNRTHESPREY.DDMSR-SKYCOTYCDALGHGF-OPFGKVYQDYCRAMSLSLINHLLGHSLGVKRUFF 214 8 mp (29112.1 (mox2. KLPNRTHESPGFS-NDMSG-SKYCOTYCDTGARGECGY-EDFGKVYQDYCRAMSLSLINHLLGHSLGVK	4 sp Q84MB3.1 ACCH1 AANWRDTLGCYTAPDPPRP	NHL	207	
7 sp Q39111.1 (AOX2 LIPKKTL&FC9FS-NDNSG-SRVGTYFDTLCGFF-EPGRYVQFYCBYCBAMSLSLKIMLLGALGUYNRDYF 214 8 sp Q39112.1 (AOX2 KIPKKTL&FRSGPSDPALVWTYVATLGGEH-RRLGVVAFYCSMSRLSLKIMLLGALGUYRDYF 215 1 sp 04705.1 (AOLD KIPKKTL&FRSGPSDPALVWTYVATLGGEH-RRLGVVAFYCSMSRLSLEIMVLGESLGVG	6 sp Q39110.2 GAOX1 KLPWKETLSFRFC-DDMSR-SKSVQDYFCDALGHGF-QPFGKVYQEYCEAMSSLSLKIMELLGLSLGVKF	DYF	216	
8 gp Q39112.1 QAOX3 LDPKKTLGFFFSDEKKIG-SQTVKGFVSKKKGCGY-EDFGKVQFATEAMNTLGLKIMLLGMBLGVERXFF 215 1 gp Q4707.1 QAOLA LDPKKTLGFFSDEKKIG-SQTVKGFVSKKKGCGY-EDFGKVQAFKCSMSKLGLKMVLGEBLGGYG	7 sp Q39111.1 GAOX2_ KLPWKETLSFQFS-NDNSG-SRTVQDYFSDTLGQEF-EQFGKVYQDYCEAMSSLSLKIMELLGLSLGVNF	DYF	214	
<pre>1 sp 004705.1 (ACUL_ LEPKETLEFESCEDPALVUTIVATLGETH - RELGEVIARCSEMSELSLEIMVIGEBLGVGRAHK 193 5 sp 004706.1 (ACUL_ LEPKETLEFESCED</pre>	8 sp Q39112.1 GAOX3_ KLPWKETLSFKFSPEEKIH-SQTVKDFVSKKMGDGY-EDFGKVYQEYAEAMNTLSLKIMELLGMSLGVEF	RYF	215	
6 BD 004707.1 [AC01A LADWAR LISENSOPEDPALVUTIVATIGEL - KRUSENSALSLEIM VLGSLEVATCARE 193 7 BD 004707.1 [AC01A LADWAR LISENSOPED	1 sp 004705.1 GAOLD KLPWKETLSFRSCPSDPALVVDYIVATLGEDH-RRLGEVYARYCSEMSRLSLEITMEVLGESLGVGF	AHY	193	
<pre>3 By 004700.1 MACH_ NAMES_ BUTNETLS RYSSADDEBEEGY TYURIGAL ARE SANDELED TARE SANDELED TO SALD THE AREA TO SANDELED TO SAN</pre>	5 SP 004/07.1 GAOLA KLPWKETLSKKSCPSDPALVVIYIVATLGELH-KRLGEVYARYCSEMSKLSLKIMEVLGESLGVG	AHY	193	
87 sp p0c5%51 GAOXZ LDPMRETLSPERTEDAARD VVANTSSTLGD F.AMREVYOK CEMHENELSTTMLLESLGTE 210 5 sp Q3103.2 GGOXZ LDPMRETLSPERTEDASTED RELMP CHELNYCOVCK CEMHENELSTTMLLESLGTE		HVP	203	
5 ap 039103.2 closel close close closel	87 sp P0C5H5.1 GAOX2 KLPWKTLSFGFHDRAADVVALYSSTLGPDF-APMGRVYOKYCREMKELSLTIMELELSLGVE	GYY	218	
<pre>5 sp 09T84.2 d30x2 KrwsS dFTVTG-SPLEUPRKLWPSHLKTCEITEKEEHMQKLAAKLWWFALGELGVGEKTDTUMAGP 194 0 sp 09C671.1 d30x4 KrwsS dFTIADSYRHFNTWPSHLKTCEITEKEEHMQKLAAKLWWFALGELGVGEKTDTUMAGP 194 0 sp 037411.1 d3021 KCMNS dTTFSPANLRSDLRKLWFRAGHDYRHFCAVMEFFRENNRALADKLLFLVALGELGVGVANVES-E 198 19 g37410.1 d3022 KCMNS dTTFSPANLRSDLRKLWFRAGHDYRHFCAVMEFFRENNRALADKLLFLVALGETGVQAAVES-E 198 19 g37410.1 d3023 KCMNS dTTFSPANLRSDLRKLWFRAGHDYRHFCAVMEFFRENNRALADKLLFLVALGETGVQAAVES-E 198 19 g374410.1 d3023 KCMNS dTTFSPANLRSDLRKLWFRAGHDYRHFCAVMEFFRENNRALADKLLFLVALGTGVQAAVES-E 198 19 g3043411.FL3H A VQDRRIVTFSTVRRD</pre>	5 sp 039103.2 G30X1 KOMWSEGFTITG-SPLNDFRKLWPOHELNYCDIVEEYEEHMKKLASKLMWLALNSLGVSEEDIEWAS	L	201	
0 sp 09C971.1 30X4 KKMNS GYTIADDSYRNHFRIWPFAGHDYHFY CGIIGEYUDEMEKLASRLLYCILGSLGVTVEDIMAHKLE 196 sp 031410.1 3022 KCMNS GYTISDANLSDLRIWPFAGHDYHFCAVMEHFERMRALADKLLLFUVALGITGEQUAAVES-E 198 sp 031410.1 3022 KCMNS GYTISDANLSDLRIWPFAGHDYHFCAVMEHFERMRALADKLLLFUVALGITGEQUAAVES-E 198 sp 031410.1 3022 KCMNS GYTISDANLSDLRIWPFAGHDYHFCAVMEHFERMRALADKLLLFUVALGITGEQUAAVES-E 198 sp 031410.1 3022 KCMNS GYTISDANLSDL	5 sp 092T84.2 G30X2 KKMWSEGFTVIG-SPLHDFRKLWPSHHLKYCEIIEEYEEHMOKLAAKLMWFALGSLGVEEKDIOWAG	P	194	
6 sp Q31411.1 (3021 KCMNS GYT SPANLSDIRILW FRAGED THEFCAVM EFF REMRALADCLL LFLVALGL TCEQVAAVES = 198 8 sp Q31409.1 (3022 KCMNS GYT SPANLSDIRILW FRAGED THEFCAVM EFF REMRALADCLL LFLVALGL TCEQVAAVES = 198 7 sp Q31409.1 (3023 KCMNS GYT SPANLSDI	0 sp Q9C971.1 G30X4 KKMWSEGFTIADDSYRNHFNTLWPHDHTKYCGIIQEYVDEMEKLASRLLYCILGSLGVTVEDIEWAH	KLE	196	
<pre>/ isp g1410.1 g5022 KIMWSE GTFFSPANLRSDLRLWPFAGEDXRBFCAVMEFFREMRALADKLELFLVALGIT GEQVAAVES-E 198 sp g31410.1 g5022 KIMSE GTFFSPANLRSDL</pre>	6 Sp Q31411.1 G3021_KCMWE GYTFSPANLRSDLRKLWPKAGHDYRHFCAVMEEFHREMRALADKLLELFLVALGLTGEQVAAVE	S-E	198	
8 B) Q11409.1 [G3023 KGWWSGUTTSPANLKSDL	7 sp Q3I410.1 G3022 KCMWSEGYTFSPANLRSDLRKLWPFAGHDYRHFCAVMEEFHRMRALADKLLELFLVALGLTGEQVAAVE	S-E	198	
<pre>19 p y cost 1 p y</pre>	5 SP (21409.1 G3023 KCMWS GTTFSPANLRSDL	S-E	198	
sp 996330.1 FLS1 AR KAWVDHIFHRIWPPSCVN	sh 293038.11 FL3H HOR VOUWERTVY FOR VVX START S	FAL.	189	
161 161 162 162 162 162 162 162 162 9 189 192 162 162 162 162 162 162 9 189 162 162 162 162 162 162 162 9 189 162 163 <td>SD 096330.1 FLS1 AR KKAWUHLFHRIWDPSCUN</td> <td>ALK</td> <td>189</td> <td></td>	SD 096330.1 FLS1 AR KKAWUHLFHRIWDPSCUN	ALK	189	
9 sp 09C614.1 G2OX7_ OYSVS AFHILSEVSRIS DRNNIRTIV TYVQ IARVAQMICEILGKQVNVSSEYF 185 sp 0900000000000000000000000000000000000	66 sp 221W9.1 ACC01 -LDWESTFFVRHLPESNIADIPULDDDWRRLMKRFAAELETLAERLLDLLCENIGLEKGVLT	KAF	152	
<pre>sp P10967.1 ACCH3 S0 KDYMCCFH-LFCSCNYYPPCQPELINGTICHTD IGFVTILLQDD-MGGLQVLHQNHWDVDPTPGGLVVNYGDLFQL 281 4 sp Q84MB3.1 ACCH1_ KD-MCTNSLLLGHYPPCQPDLTLGTCFHSDSFTILLQDH-IGGLQVLHQNHWDVDPTPGGLVVNYGDLLQL 283 6 sp Q39110.2 GAOX1_ RFFFERDD-SIMKLNYYPPCTPDLTLGTGPHCDFSSLTILHQDH-VNGLQVFVDNCWQSIRPNPKAFVVNYGDLTMA 292 7 sp Q39112.1 GAOX2_ RGFFERDD-SIMKLNYYPPCTPDLTLGTGPHCDFSSLTILHQDH-VNGLQVFVDNCWQSIRPNPKAFVVNYGDTFMA 291 8 sp Q39112.1 GAOX3_ KEFFEDD-SIMKLNYYPQCKQPLALGTGPHCDFSSLTILHQDH-VNGLQVFVDNCWQSIRPNPKAFVVNYGDTFMA 291 9 sp Q39112.1 GAOX3_ KEFFEDD-SIMKLNYYPQCKQPLALGTGPHCDFSSLTILHQDN-VGGLQVFVDNCWQSIRPNPKAFVVNYGDTFMA 291 9 sp Q39112.1 GAOX3_ RFFFGND-SIMKLNYYPQCKQPLALGTGPHCDFSSLTILHQDN-VGGLQVFTEGKWRSIRPRADAFVVNYGDTFMA 291 9 sp Q3910.2 GAOX1_ RFFFGND-SIMKLNYYPQCKQPLETLGTGPHCDFTSSLTILHQDN-VGGLQVHTEGKWRSIRPRADAFVVNIGDTFMA 269 9 sp Q3701.1 GAO1B_ RRFFGND-SIMKLNYYPQCRPLETLGTGPHCDFTSLTILHQDN-VGGLQVHTEGKWRSIRPRADAFVVNIGDTFMA 269 9 sp Q3703.2 GAOX1_ RFFFGND-SIMKLNYYPCQRPLETLGTGPHCDFTSLTILHQDN-VGGLQVHTEGKWRSIRPRADAFVVNIGDTFMA 269 9 sp Q39103.2 GAOX1_ SSDLNNAQ-AALQLNNYYCPCPPEPRTLGTGPHCDFTSLTILLQDD-VGGLEVLVDGEWRPVSPVPGAMVINIGDTFMA 294 9 sp Q39103.2 GAOX1_ SSDLNNAQ-AALQLNNYYCPCPPEPRTLGTGPHCDFTSLTILLQDD-VGGLEVLVDGEWRPVSPVPGAMVINIGDTFMA 294 9 sp Q3911.1 GAOX2_ REFFADSS-SIMRCNYYPCPEPRTAMGLAAHTDSTLLTILLQDD-VGGLEVLVDGEWRPVSPVPGAMVINIGDTFMA 294 9 sp Q39103.2 GAOX1_ SSDLNNAQ-AALQLNNYYCPEPPRAMGLAAHTDSTLLTILLQDD-VGGLEVLVDGEWRPVSPVPGAMVINIGDTFMA 294 9 sp Q3914.0.1 GAO22_ GXIA_MSGSKVGR-GAIRLNHYPVCPEPPRAMGLAAHTDSTLMTILVQNN-TAGLQVFREDG-WVTVPAPPGGLVVNVGDLFQI 276 9 sp Q3140.1 GAO22_ QKIAETMT-ATMHLNNYPKCPDPKRALGLIAHTDSGFFFFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGDLFQI 276 9 sp Q3140.1 GAO22_ QKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFFFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGDLFQI 276 9 sp Q3140.1 GAO22_ QKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFFFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGDLFQI 276 9 sp Q388.1 FL3H_A TNACVDMD-QKIVVNYPYPCPDPLLLGLIAKTHTPGTTITLLLQDC-VGGLQATNGGKNNITVQCPLGAFVNLGDHCHF 267 9 sp Q388.1 FL3H_A TNACVDMD-</pre>	9 sp 29614.1 22017 VSEAFHILLSEVSRIS	EYF	185	
* * * * * * * * * * * * * * * * * * *				
sp p10967.1 ACCH3 SO KDYMDCFH-LFCSCNTY PP CQPELTMGTIQHTDIGFYTILLOD-MGGLQVLHONHWVDVPPTPGSLVVNUGDLQL 281 4 sp Q84MB3.1 ACCH1 KD-MDCTNELLLGHYYPP CQPPLTLGLTKHSDNSFLTILLODH-IGGLQVLHDQYWVDVPPVPGALVVNUGDLQL 283 6 sp Q39110.2 GAOX1 REFFEND-SIMRLNY PP CIFPLIFGTGPHCDPTSLTILHODH-VNGLQVFVENQWSEIRPNPKAFVVNIGT FMA 292 9 sp Q39112.1 GAOX3 KEFFEDD-SIMRLNY PP CTPLILGTGPHCDPTSLTILHODH-VNGLQVFVDNWQSEIPPNFAFVVNIGT FMA 291 1 sp 004705.1 GAO1D REFFEDD-SIMRLNY PP CQRPLETLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGT FMA 291 1 sp 004707.1 GAO1A REFFEGND-SIMRLNY PP CQRPLETLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGT FMA 269 6 sp 004707.1 GAO1A REFFEGND-SIMRLNY PP CQRPLETLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGT FMA 269 9 sp 004706.1 GAOLB REFFEGND-SIMRLNY PP CQRPLDTLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGT FMA 269 9 sp 004706.1 GAOLB REFFEGND-SIMRLNY PP CQRPLDTLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGT FMA 269 9 sp 004706.1 GAOX1 REFFEGND-SIMRLNY PP CQRPLDTLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGT FMA 269 9 sp 005555.1 GAOX2 REFFADSS-SIMRCNYY PP CPEPERIGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGT FMA 279 87 jsp 190555.1 GAOX2 REFFADSS-SIMRCNYY PP CPEPERAMGLAAHTD STLLTILY ONN-TAGLQVFR DUG-WVTVPPFPGSLVVNVGT FFMA 279 9 sp 092784.2 G3OX2 NSDFQGTQ-AVIQLNHY PC VPEPERAMGLAAHTD STLLTILLQTN-TAGLQVFR DUG-WVTVPPFPGSLVVNVGT FFMA 271 0 sp Q92781.1 G3OX4 KSGSKVGR-GAIRLNHY PC PPEP RAMGLAAHTD STLLTILLQSN-TGGLQVFR DVG-WVTXPPVPGSLVVNVGT LHI 271 0 sp Q31410.1 G3O22 QKIAFTMT-ATMHLNWY PC PDPKRALGLIAHTD SGFFTFVLQSL-VPGLQL FRHEDPRWVTVPAVPGAMVVNVGT LFQI 276 7 sp Q31410.1 G3O22 QKIAFTMT-ATMHLNWY PC PDPKRALGLIAHTD SGFFTFVLQSL-VPGLQL FRHEDPRWVTVPAVPGAMVVNVGT LFQI 276 7 sp Q31410.1 G3O22 QKIAFTMT-ATMHLNWY PC PDPKRALGLIAHTD SGFFTFVLQSL-VPGLQL FRHEDPRWVTVPAVPGAMVVNVGT LFQI 276 7 sp Q31410.1 G3O22 QKIAFTMT-ATMHLNWY PC PDPKRALGLIAHTD SGFFTFVLQSL-VPGLQL FRHEDPRWVTVPAVPGAMVVNVGT LFQI 276 7 sp Q31410.1 G3O22 QKIAFTMT-ATMHLNWY PC PDPKRALGLIAHTD SGFFTFVLQSL-VPGLQL FRHEDPRWVTVPAVPGAMVVNVGT LFQI				
4sp084MB3.1ACCH1KD-MDCTNBLLLIGHYIPPCPQFTLTIGTTKHSINEFLTILLQTH-IGGLQVLH-DQYWDVPPVPGALVVNVGTLLQL2836sp039110.2GAOXIREFFEND-SIMRLNYYPPCQFPTLTIGTGPHCDPTSLTILHQDH-VNGLQVFV-ENQWRBIRPNPKAFVVNIGTTFMA2927sp039112.1GAOX2RGFFEND-SIMRLNYYPPCQTPTLTIGTGPHCDPTSLTILHQDH-VNGLQVFV-DNQWSIRPNPKAFVVNIGTTFMA2908sp039112.1GAOX3KEFFEIDD-SIMRLNYYPPCQTPLTLGTGPHCDPTSLTILHQD-VGGLQVFV-DNWQSIRPNPKAFVVNIGTTFMA2911sp004705.1GAOIDRRFFGND-SIMRLNYYPPCQRPLETLGTGPHCDPTSLTILHQDN-VGGLQVFV-DRWQSIRPNPKAFVVNIGTTFMA2696sp004706.1GAOIARRFFGND-SIMRLNYYPPCQRPLETLGTGPHCDPTSLTILHQDN-VGGLQVHT-ERWRSIRPRADAFVVNIGTTFMA2695sp004706.1GAOX1RRFFGND-SIMRLNYYPCQRPLETLGTGPHCDPTSLTILHQDN-VGGLQVHT-ERWRSIRPRADAFVVNIGTTFMA2693spP93771.2GAOX1RRFFGND-SIMRLNYYPCQRPLTLGTGPHCDPTSLTILHQD-VGGLEVWT-ERWRSIRPRADAFVVNIGTTFMA27987sp039103.2G3OX1SSIMRCNYPPCPPPFRAFUGTGPHCDPTSLTILHQD-VGGLEVWTDGEWRSIRPRADAFVVNIGTTFMA27987sp039103.2G3OX2SSIMRCNYPPCPPFRAMGLAAHTDSTLTILLQD-VGGLEVWTDGEWRSIRPPGALVNVGTTFMA27987sp039103.2G3OX2SSIMRCNYPPCPPFRAMGLAAHTDSTLTILLQD-VGGLEVWTDGEWRVSPVPGAMVINVGTTFMA27987sp031410.1G3O22NSDFGTQ-AVIQLNHYPCPPFRAMGLAAHTDSTLTILLQD-VGGLQVFREDG-WVTWPPFPGSLVVNVGTLH12710sp031410.1G3O22QKIAFTMT-ATMHLNWYPCPPFPFRAMGLAAHTDSTLTILHQN-TAGLQVFREDG-WVTWPAPAGAWVNVGTLFQ1276 <t< td=""><td>SD P10967.1 ACCH3 SO KDWMDCFH-LFCSCNWW PPEPOPELEMGTTCHTE TOWTTLASD - MCGLAULH - ONHWUNDEDTOGTLUNTONS</td><td>LOL</td><td>281</td><td></td></t<>	SD P10967.1 ACCH3 SO KDWMDCFH-LFCSCNWW PPEPOPELEMGTTCHTE TOWTTLASD - MCGLAULH - ONHWUNDEDTOGTLUNTONS	LOL	281	
6 sp Q39110.2 GAOX1 REFFEEND-SIMRLNYPPCIRPLITLGTGPHCDPTSLTILHQDH-VNGLQVFVENQWRBIRPNPKAFVVNIGTTFMA 292 7 sp Q39112.1 GAOX2 RGFFEEND-SIMRLNYPPCIRPLITLGTGPHCDPTSLTILHQDH-VNGLQVFVDNGWQBIRPNPKAFVVNIGTTFMA 290 8 sp Q39112.1 GAOX3 KEFFEDD-SIMRLNYPPCQTPLILLGTGPHCDPTSLTILHQD-VGGLQVFVDNGWQBIRPNPKAFVVNIGTTFMA 291 1 sp C04705.1 GAOLD RRFFEGND-SIMRLNYPPCQRPLETLGTGPHCDPTSLTILHQD-VGGLQVHTEGRWRSIRPRADAFVVNIGTTFMA 269 6 sp O04707.1 GAOLA RRFFEGND-SIMRLNYPPCQRPLETLGTGPHCDPTSLTILHQDN-VGGLQVHTEGRWRSIRPRADAFVVNIGTTFMA 269 5 sp O04706.1 GAOLB RRFFEGND-SIMRLNYPPCQRPMETLGTGPHCDPTSLTILHQDN-VGGLQVHTEGRWRSIRPRADAFVVNIGTTFMA 279 87 sp [93771.2 GAOX1 RRFFGRD-SIMRLNYPPCQRPMETLGTGPHCDPTSLTILLQD-VGGLEVLVDEGWRPSPVSPVGALVNINGTTFMA 274 5 sp Q39103.2 G3OX1 SSDINWAQ-AALQLNHPVCPEPPRAMGLAAHTDSTLLTILQDN-TAGLQVFRDDVG-WVTVPPPGSLVNVGDLFHI 278 5 sp Q3710.1 G3OX2 NSDFQG-AVIQLNHPVCPEPPRAMGLAAHTDSTLMTILYQNN-TAGLQVFRDDVG-WVTVPPPGSLVNVGDLFHI 271 0 sp Q9C971.1 G3O22 QKIAETMT-ATMHLNWPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVNVGDLFQI 276 8 sp Q31410.1 G3O22 QKIAETMT-ATMHLNWPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVNVGDLFQI 276 8 sp Q31409.1 G3O22 QKIAETMT-ATMHLNWPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVNVGDLFQI 276 8 sp Q31409.1 G3O22 QKIAETMT-ATMHLNWPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVNVGDLFQI 276 7 sp Q9S818.1 FL3H A TNACVDMD-QXIVVNYPKCPQPTLLLGKRHTDPGTITLLLQD-VGGLQATRBGKWITVQVPEGAFVNLGBHEF 267 7 sp Q9S818.1 FL3H A TNACVDMD-QXIVVNYPKCPQPTLLGLGKRHTDGFTITLLLQD-VGGLQATRBGKWITVQVPEGAFVNLGBHEF 267 7 sp Q9S818.1 FL3H A TNACVDMD-QXIVVNYPKCPQPTLLGLGVFRHTDFTILLLQD-VGGLQATRBGKWITVQVPEGAFVNLGBHEF 267 8 p026330.1 FL3H A REGLGGEMAEYMMKINYPPCPRPTLALGVPAPHTDLSGTLLVPNE-VFGLQVFKDHWPDAEYIPSAVINLGDGLEV 2	4 SD 084MB3.1 ACCH1 KD-MDCTNSLLLIGHYYPPCPOPDITLGLTKHSDNSFLTILLADH-IGGLOVLH-BOYWUDVPPVDALLUNVANT	LOL	283	
7 sp Q39111.1 GAOX2 RGFFEEND-SIMRLNHYPPQTPLILGTGPHCDPSSLTILHQDH-VNGLQVFVDNQWQSIRPNPKAFVVNIGDTFMA 290 8 sp Q39112.1 GAOX3 KEFFEDSD-SIMRLNHYPPQTPLALGTGPHCDPTSLTILHQD-VGGLQVFVDNKWQSIRPNPKAFVVNIGDTFMA 291 1 sp 004705.1 GAOLD RRFFEGND-SIMRLNYYPPQRPLETLGTGPHCDPTSLTILHQD-VGGLQVFTEGRWRSIRPRADAFVVNIGDTFMA 269 5 sp 004706.1 GAOLD RRFFEGND-SIMRLNYYPPQRPLETLGTGPHCDPTSLTILHQDN-VGGLQVHTEGRWRSIRPRADAFVVNIGDTFMA 269 5 sp 004706.1 GAOLB RRFFEGND-SIMRLNYYPPQRPLETLGTGPHCDPTSLTILHQDN-VGGLQVHTEGRWRSIRPRADAFVVNIGDTFMA 269 7 sp P03771.2 GAOX1 RRFFEGND-SIMRLNYYPPQRPLETLGTGPHCDPTSLTILHQDN-VGGLQVHTEGRWRSIRPRADAFVVNIGDTFMA 269 87 sp P03771.2 GAOX1 REFFEGND-SIMRLNYYPPQRPLETLGTGPHCDPTSLTILHQDN-VGGLQVHTEGRWRSIRPRADAFVVNIGDTFMA 279 87 sp P03771.2 GAOX1 SSDLWAQ-AALQLNHYVPCPPPRTLGTGPHCDPTSLTILHQDN-VGGLEVWAEGRWRSIRPRADAFVVNIGDTFMA 279 87 sp Q39103.2 G3OX1 SSDLWAQ-AALQLNHYVCPPPRAMGLAAHTDSTLLTILTQNN-TAGLQVFRIDUG-WVTVPPPPGSLVVNVGLLFHI 278 5 sp Q92T84.2 G3OX2 NSDF0GTQ-AVIQLNHYPCPPPRAMGLAAHTDSTLLTILTQNN-TAGLQVFRIDUG-WVTVPPPPGSLVVNVGLLFHI 271 0 sp Q9C971.1 G3OX4 KSGSKVGR-GAIRLNNYPCPPPRAMGLAAHTDSTLLTILHQSN-TGGLQVFRIDVG-WVTVPPPPGSLVVNVGLLFHI 271 6 sp Q31410.1 G3O22 QKIAETMT-ATMHLNWYPCPPPRAMGLAAHTDSTLTILILQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 7 sp Q31410.1 G3O22 QKIAETMT-ATMHLNWYPCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 8 sp Q31410.1 G3O22 QKIAETMT-ATMHLNWYPCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 7 sp Q31410.1 G3O23 QKIAETMT-ATMHLNWYPCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 8 sp Q31410.1 G3O23 QKIAETMT-ATMHLNWYPCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 7 sp Q31410.1 G3O23 QKIAETMT-ATMHLNWYPCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 8 sp Q31410.1 G3O23 QKIAETMT-ATMHLNWYPCPDPKDPLFDFTULLQDL-VGGLQATRBGKNWITVQPVEGAFVVNLGDHFF 267 7 sp Q9818.1 FL3H ĀN AKCVDMD-QKVVNNFYPCPDPFLALGVPAHTDSGFTFVLQSL-VPGLQATRBGKNWITVQPVEGAFVVNLGDHFF 267 8 sp Q3630.1 FL3H ĀN AKCVDMD-QKVVNYFYPCPPPFPLALGVPAHTDAGTILLLODL-VGGLQATRBGKNWITVQPVEGAFVV	6 sp Q39110.2 GAOXI REFFEND-SIMRLNYYPPCIKPELTLGTGPHCDPTSLTILHODH-VNGLOVFVENOWRSIRPNPKAFVVNICOT	FMA	292	
<pre>8 sp Q39112.1 GAOX3 KEFFEDSD-SIFRINYYPQCKQPELALGTGPHCDPTSLTILHQDQ-VGGLQVFVDNKWQSIPPNPHAFVVNIGTTFMA 291 1 sp C04705.1 GAOLD RRFFEGND-SIMRLNYYPPQRPLETLGTGPHCDPTSLTILHQDN-VGGLQVHTEGRWRSIRPRADAFVVNIGTTFMA 269 5 sp 004706.1 GAOLB RRFFEGND-SIMRLNYYPPQRPLETLGTGPHCDPTSLTILHQDN-VGGLQVHTEGRWRSIRPRADAFVVNIGTTFMA 279 87 sp 005555.1 GAOX2 REFFADSS-SIMRCNYYPPCPPPRRAUGLAHTDSTLLTILLQDN-VGGLEVVAEGRWRSIRPRADAFVVNIGTTFMA 279 87 sp 039103.2 G3OX1 SSDLWAQ-AALQLNHYPVCPPPRAMGLAAHTDSTLLTILLQDN-TAGLQVFRDDG-WVTVPPPFGSLVVNVGDLFHI 271 5 sp 039103.2 G3OX1 SSDLWAQ-AALQLNHYPVCPPPRAMGLAAHTDSTLLTILQNN-TAGLQVFRDDG-WVTVPPPFGSLVVNVGDLFHI 271 6 sp 0392071.1 G3OX4 KSGSKVGR-GAIRLNHYPVCPPPRAMGLAAHTDSTLLTILLQSN-TGGLQVFRDDVG-WVTAPPVPGSLVVNVGDLFHI 273 6 sp 031410.1 G3O22 GKIAFTMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFFFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGDLFQI 276 7 sp 031410.1 G3O22 GKIAFTMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFFFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGDLFQI 276 7 sp 031410.1 G3O22 GKIAFTMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFFFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGDLFQI 276 7 sp 031410.1 G3O23 GKIAFTMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFFFVLQSL-VPGLQFFRHOPRWVTVPAVPGAMVVNVGDLFQI 276 7 sp 031410.1 G3O23 GKIAFTMT-ATMHLNWYPKCPDPKENE 266 7 sp 02303.1 FL3H A TNACVDMD-QKIVVNYYPYPCPPPDLALGUVPKAHTDAGGILLFQDVGGLQUFKDGHWVTVPAVPGAMVVNUGDLFQI 276 7 sp 02</pre>	7 sp 239111.1 GAOX2 RGFFEND-SIMRLNHYPPOTPOLTLGTGPHCDPSLTILLCDH-VNGLOVFVDNOWOSIRPNPKAFVVNIGD	FMA	290	
1 sp 004705.1 GAOLD REFFGND-SIMELNYPPCQRPLETLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGTTFMA 269 6 sp 004706.1 GAOLB REFFGND-SIMELNYPPCQRPLETLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGTTFMA 269 3 sp P93771.2 GAOX1 REFFGND-SIMELNYPPCQRPLETLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGTTFMA 279 87 sp P05555.1 GAOX2 REFFQSND-SIMELNYPPCQRPLETLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIRPRADAFVVNIGTTFMA 279 87 sp 039103.2 GAOX1 REFFQSND-SIMELNYPPCQRPLETLGTGPHCDPTSLTILHODH-VGGLEVUXEGRWRSIRPRADAFVVNIGTTFMA 279 87 sp 039103.2 GAOX1 REFFQSND-SIMELNYPPCPPFRANGLAAHTOSILTILLQD-VGGLEVUXDGEWRPVSPVPGAMVINIGTTFMA 294 5 sp 039103.2 GAOX1 SSDLWMAQ-AALQLNHYPVCPEPTRAMGLAAHTOSILTILLQD-VGGLEVUXDGEWRPVSPVPGAMVINIGTTFMA 271 0 sp 03071.1 GAOX2 NSDF0GTQ-AVIQLNHYPVCPEPTRAMGLAAHTOSILTILIQON-TAGLQVFRDUG-WVTAPPVPGSLVVNVGTLHI 271 0 sp 031410.1 G3021 HKIATMT-ATMHLNWYPVCPEPTRAMGLAAHTOSILTILIGSN-TGGLQVFRDUG-WVTAPPVPGSLVVNVGTLHI 273 6 sp 031410.1 G3022 QKIATMT-ATMHLNWYPVCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGTLFQI 276 8 sp 031410.1 G3022 QKIATMT-ATMHLNWYPVCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGTLFQI 276 8 sp 031410.1 G3023 QKIATMT-ATMHLNWYPVCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGTLFQI 276 8 sp 031410.1 G3022 QKIATMT-ATMHLNWYPVCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGTLFQI 276 8 sp 031410.1 G3023 QKIATMT-ATMHLNWYPVCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGTLFQI 276 8 p205038.1 FL3H A TNACVDMD-QKIVVNYPVPCPPPTLILGLKRHTDPGTITLLLQDL-VGGLQATRDGKNWITVQPVGAMVVNVGTLFQI 276 9 p205038.1 FL3H A RACVDMD-QKIVVNYPPCPPPTLALGVPAVHTDGTILLQDL-VGGLQATRDGKNWITVQPVFGAFVVNLGDQLEV 279 9 NTOVICUS COVVVNTPPCPPPTLALGVPAVPPPPPLANGTLSOTTLLVPNE-VFGLQVFKDHW	8 sp Q39112.1 GAOX3_ KEFFEDSD-SIFRLNYYPQCKQPELALGTGPHCPTSLTILHODQ-VGGLQVFVDNKWQSIPPNPHAFVVNIGT	FMA	291	
b sp 004706.1 GAOLA RRFFGND-SIMRLNYY PPCQRPMETLGTGPHCDPTSLTILHODN-VGGLQVHT - EGRWRSIRPRADAFVVNIGD TFMA 269 5 sp 004706.1 GAOLA RRFFGND-SIMRLNYY PPCQRPMETLGTGPHCDPTSLTILHODN-VGGLQVHT - EGRWRSIRPRADAFVVNIGD TFMA 269 3 sp P3771.2 GAOXI RRFFGND-SIMRLNYY PPCQRPMETLGTGPHCDPTSLTILHODN-VGGLQVHT - EGRWRSIRPRADAFVVNIGD TFMA 279 87 sp P30701.2 GAOXI RRFFGND-SIMRLNYY PPCPEP RALGIAFTO STLLTILDDD-VGGLEVLVA - EGRWRSIRPRADAFVVNIGD TFMA 279 87 sp Q39103.2 G3OXI SSDINWAQ-AALQLNHY PPCPEP RAMGLAAFTO STLLTILLODD-VGGLEVLVA - EGRWRSIRPRADAFVVNVGD LFMI 278 5 sp Q32784.2 G3OXI SSDINWAQ-AALQLNHY PPCPEP RAMGLAAFTO STLMTILYQNN-TAGLQV FR DVG-WVTAPPVPGSLVVNVGD LFMI 271 0 sp Q92784.2 G3OX4 KSGRVGR-GAIRLNHY PPCPEP RAMGLAAFTO STLMTILYQNN-TAGLQV FR DVG-WVTAPPVPGSLVVNVGD LFMI 271 0 sp Q31410.1 G3O21 KIAFTMT-ATMHLNWY PCPEP RAMGLAAFTO STLFTLGTGFFFVLQSL-VPGLQL FRHGPDRWVTVPAVPGAMVVNGD LFQI 276 8 sp Q31410.1 G3O22 QKIAFTMT-ATMHLNWY PCPEP RAMGLAAFTO SGFFFFVLQSL-VPGLQL FRHGPDRWVTVPAVPGAMVNVGD LFQI 276 8 sp Q31410.1	1 sp 004705.1 GAO1D_ RRFFEGND-SIMRINYYPPCQRPLETLGTGPHCPTSLTILHQDN-VGGLQVHTEGRWRSIRPRADAFVVNIGT	FMA	269	
b pp 004/00.1 GAOLB RKFFFGND-SIMKLNYIPPCQRPMETLGTGPHCDPTSLTILHQDN-VGGLQVHTEGRWRAIRPRPGALVVNVGTTFMA 269 3 sp p93771.2 GAOX1 RRFFGND-SIMKLNYIPPCQRPLDTLGTGPHCDPTSLTILHQDH-VGGLQVHTEGRWRAIRPRPGALVVNVGTTFMA 279 87 sp p0555.1 GAOX1 REFFADSS-SIMRCNYIPPCQRPLETLGTGPHCDPTALTIILQDD-VGGLEVUX EGRWRAIRPRPGALVVNVGTTFMA 294 5 sp Q39103.2 GAOX1 SSDLNWAQ-AALQLNHYPVCPEPERAMGLAAHTDSTLLTILLQDD-VGGLEVUX EGRWRAIRPRPGALVVNVGTLFHI 278 5 sp Q92T84.2 GAOX1 SSDLNWAQ-AALQLNHYPVCPEPERAMGLAAHTDSTLMTILTQNN-TAGLQVFREDUG-WVTVPPPPGSLVVNVGTLFHI 271 0 sp Q92791.1 GAOX4 KSGSKVGR-GAIRLNHYPVCPEPERAMGLAAHTDSTLMTILTQNN-TAGLQVFREDVG-WVTVPAPGVVVVGTLFHI 273 6 sp Q31411.1 G3021 HKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGTLFQI 276 7 sp Q31410.1 G3022 QKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGTLFQI 276 8 sp Q31409.1 G3033 QKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGTLFQI 276 8 sp Q31409.1 G3033 QKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGTLFQI 276 9 sp Q95818.1 FL3H TNACVDMD-QKVVVNYPKCPQPDILTGLKRHTDFGTTILLQD-VGGLQATRBGKNWITVQVPGAFVVNLGDHGHF 264 sp P28038.1 FL3H TNACVDMD-QKVVVNYPKCPQPDILTGLKRHTDPGTTILLQDL-VGGLQATRBGKNWITVQPISGAFVVNLGHGHF 266 66	5 SD 004707.1 GAO1A REFFEGND-SIMELNYPPPCRPLETLGTGPHCDPTSLTILHODN-VGGLQVHTEGRWRSIEPRADAFVVNIGT	FMA	269	
87 sp P357112 GAOL_ KREFVKND-OLEKLINI PARQKELLIGIGFHCDFISLILLEUDH-VGGLEVNA- EGKMKALKPKPGALVVNGG LFMA 279 87 sp Q351312 GAOL_ KREFVKND-OLEKLINI PARQKELLIGIGFHCDFISLILLEUDH-VGGLEVNA- EGKMKALKPKPGALVVNGG LFMA 294 5 sp Q351312 GAOL_ SSDLWAQ-AALQLNHI PVCPEPERAMGLAAHTDSTILLILLOD-VGGLEVIV- DGEWRPVSPVPGALVVNVGLLFHI 278 5 sp Q32784.2 GAOL_ KSGSKVGR-GAIRLNHI PVCPEPERAMGLAAHTDSTILLILVON-TAGLQVFRIDUG-WVTVPPFPGSLVVNVGLLFHI 271 0 sp Q32791.1 GAOL_ KSGSKVGR-GAIRLNHI PVCPEPERAMGLAAHTDSTILTILLOD-VGGLVFRIDUG-WVTVPAPGVLVVNIGLLFHI 271 6 sp Q31411.1 GAOL_ HKIAETMT-ATMHLNWI PKCPEPERAMGLAAHTDSTILTILLOD-VGGLVFRIDUG-WVTVPAVPGAMVVNVGLLFHI 273 6 sp Q31410.1 GAOL_ KKIETMT-ATMHLNWI PKCPEPERAMGLAAHTDSTILTILLOD-VGGLVFRIDUG-WVTVPAVPGAMVVNVGLLFQI 276 7 sp Q31410.1 GAOL_ QKIAETMT-ATMHLNWI PKCPEPERALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPERWTVPAVPGAMVVNVGLLFQI 276 8 sp Q31410.1 GAOL_ QKIAETMT-ATMHLNWI PKCPEPERALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPERWTVPAVPGAMVVNVGLLFQI 276 7 sp Q31410.1 GAOL_ QKIAETMT-ATMHLNWI PKCPEPERALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPERWTVPAVPGAMVVNVGLLFQI 276 8 sp Q31410.1 GAOL_ QKIAETMT-ATMHLNWI PKCPEPERALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPERWTVPAVPGAMVVNVGLLFQI 276 8 sp Q31410.1 GAOL_ QKIAETMT-ATMHLNWI PKCPEPERALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPERWTVPAVPGAMVVNVGLLFQI 276 8 sp Q31410.1 GAOL_ QKIAETMT-ATMHLNWI PKCPEPERALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPERWTVPAVPGAMVVNVGLLFQI 276 9 sp Q96330.1 FLSH_ A TNACVEMD-QKIVVNYFYPCPPEILTILLGEPERTENCEVEGLQATREGENWITVQPISGAFVVNLGHHFF 267 9 sp Q96330.1 FLSH_ A TAACVEMD-QKIVVNYFYPCPPEILALGYPAHTENCEVEGLQATREGENWITVQPISGAFVVNLGHHFF 267 9 sp Q96330.1 FLSH_ A REGLGGEMAEYMMKINYIPPEPEPERALGYPAHTELSGITLLVPNE-VFGLQVFK- DHWFDAEYIPSAVIHIGEQILR 266 9 sp Q000014 LCOOT_ REPAGAPT-FGTKVSSYPPEPEPERE VERAUVEREDETETETETETETETETETETETETETETETETETET	5 SP 004/00.1 GAOLE REFEREND SIMKINT PPEQREMETIGTEFECTPTELTILHODN-VGGLOVHT - GRWRBIRPRADAFVVNIGT	FMA	269	
5 sp (39103.2 G30X1 SSDINNAQ-AALCHNYPYEPERAMGLAAHTOSTLTILLOUD-VGGLEVLVDGEWEVSPVPGAMVINUG IFMA 294 5 sp (39103.2 G30X1 SSDINNAQ-AALCHNYPYEPERAMGLAAHTOSTLTILIQUN-TAGLQVFRDLG-WVTVPPFGGLVVNVGDILLHI 271 0 sp (392971.1 G30X4 KSGEVGR-GAIRLNNYPYEPERAMGLAAHTOSTLTILLGSN-TGGLQVFRDLG-WVTVPPAPGSLVVNVGDILLHI 271 6 sp (31411.1 G3021 HKIATMT-ATMLNWYPEPPRAMGLAAHTOSTLTILLGSN-TGGLQVFRDLG-WVTVPAPGAVVNVGDILFU 276 7 sp (31410.1 G3022 QKIATMT-ATMLNWYPEPPRAMGLAAHTOSTLTILGSN-TGGLQVFRDDVG-WVTVPAPGAVVNVGDILFU 276 8 sp (31410.1 G3022 QKIATMT-ATMLNWYPEPPRAMGLAAHTOSGFFTFVLQSL-VPGLQIFRHGPDRWVTVPAVPGAMVNVGDIFQI 276 7 sp (31409.1 G3023 QKIATMT-ATMLNWYPEPPRKALGLIAHTOSGFFTFVLQSL-VPGLQIFRHGPDRWVTVPAVPGAMVNVGDIFQI 276 7 sp (39818.1 FL3H A TNACVDMD-QKIVVNYYPEPPPCTILGLEKHTDPGTITLLLQDQ-VGGLQATRINGKNWITVQVPGGAFVVNLGHGHF 264 7 sp (29638.1]FL3H A TNACVDMD-QKIVVNYYPEPPPCTILGLKHTDPGTITLLLQDL-VGGLQATRINGKNWITVQPIGGAFVVNLGHGHF 264 7 sp (29633.1]FL3H A KACVDMD-QKIVVNYFYPEPPCPTILLGLKHTDPGTITLLLQDL-VGGLQATRINGKNWITVQPIGAFVVNLGHGHF 264 9 p26038.1]FL3H A KACVDMD-QKIVVNYFYPEPPCPTILAGVPAHTDLSGITLLVPNE-VPGLQVFKDHWFDAEYIPSAVIVHIGDQILR 266 6 sp (22199.1)ACCO1 RGPAGAPT-FGTKVSYPPCPPPCPR LALGVPAHTDLSGITLLVPNE-VPGLQVFKDGWVDVPPMRHSIVVNLGDQLEV 229 9 NDTON 000000000000000000000000000000000000	2) an Inocyte 1 anova press, and an and the and an	PMA	219	
5 sp Q9ZT84.2 G3OX2 NSDFQTQ-AVIQUNHY PK CPEPT RAMGLAAHTO STLMTILY QNN-TAGLQVFR DVG-WYTAPPVPGSLVVNVGD LLHI 271 0 sp Q9C971.1 G3OX4 KSGRVGF-GAIRLNHY PK CPEPT RAMGLAAHTO STLMTILY QNN-TAGLQVFR DVG-WYTAPPVPGSLVVNVGD LFHI 273 6 sp Q3I410.1 G3O21 HKIATMT-ATMHLNW PK CPDPKRALGLIAHTO SGFPTFVLQSL-VPGLQLFRHGPDRWTVPAVPGAMVVNVGD LFQI 276 8 sp Q3I410.1 G3O22 QKIATMT-ATMHLNW PK CPDPKRALGLIAHTO SGFPTFVLQSL-VPGLQLFRHGPDRWTVPAVPGAMVVNVGD LFQI 276 8 sp Q3I410.1 G3O23 QKIATMT-ATMHLNW PK CPDPKRALGLIAHTO SGFPTFVLQSL-VPGLQLFRHGPDRWTVPAVPGAMVVNVGD LFQI 276 7 sp Q9S818.1 FL3H A TMACVDMD-QRIVVNY PK CPDPKRALGLIAHTO SGFPTFVLQSL-VPGLQLFRHGPDRWTVPAVPGAMVVNVGD LFQI 276 7 sp Q9S818.1 FL3H A TNACVDMD-QRIVVNY PK CPDPKRALGLIAHTO SGFPTFVLQSL-VPGLQARR NGKWITVQPVEGAFVNLGFHGFP 264 7 sp Q9S818.1 FL3H A TNACVDMD-QRIVVNY PK CPQPTLLGLKRHTD PGTITLLLQDQ-VGGLQATR NGKWITVQPVEGAFVNLGFHGFP 264 7 sp Q9S830.1 FL3H A TNACVDMD-QRIVVNY PK CPQPTLLGLKRHTD PGTITLLLQDL-VGGLQATR NGKWITVQPVEGAFVNLGFHGFP 264 8 sp Q33440.1 G3023 QKIATMT-FTARK STARTA TAMHLNW PK CPDPKTLGGTLLGKRHTD PGTITLLLQDL-VGGLQATR NGKWITVQPVEGAFVNLGFHGFP 264 9 P2038.1 FL3H A TNACVDMD-QRIVVNY PK CPQPTLLGLKRHTD PGTITLLLQDL-VGGLQATR NGKWITVQPVEGAFVNLGFHGFP 264 9 p2038.1 FL3H A TNACVDMD-QRIVVNY PK CPQPTLLGLKRHTD PGTITLLLQDL-VGGLQATR NGKWITVQPVEGAFVNLGFHGFP 264 9 p2038.1 FL3H A TNACVDMD-QRIVVNY PK CPQPTLLGLKRHTD PGTITLLLQDL-VGGLQATR NGKWITVQPVEGAFVNLGFHGFP 264 9 p2038.1 FL3H A TNACVDMD-QRIVVNY PK CPQPTLLGLGVFFFFUGSTLLVPNE-VFGLQVFKDHWFDAEY IPSAVIVHLGFQLFF 267 9 p2038.1 FL3H A TNACVDMD-QRIVVNY PF PPPPPPPPPPP 100000000000000000000000	o/jsp/rvcsh5.1/GAVAZ KEFKALSS-SLEKCHTIPPPPERELGTEPPELTILLUUP-VGGLEVLVBGEWRVSPVPGAWVINTOT	FMA	279	
0 sp Q9C971.1 G3OX4 KSGSKVGR-GAIRLNHY PVC PEPERAMGLAAHT STILTILLG SN-TGGLQVFREESG-WVTVEPAPGVLVVNIGTLFHI 273 6 sp Q3I411.1 G3O21 HKIAETMT-ATMHLNWYPKC PDPKRALGLIAHT SGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 7 sp Q3I410.1 G3O23 QKIAETMT-ATMHLNWYPKC PDPKRALGLIAHT SGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 8 sp Q3I409.1 G3O23 QKIAETMT-ATMHLNWYPKC PDPKRALGLIAHT SGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 7 sp Q3S18.1 FL3H A TNACVDMD-QXIVVNYPKC PQPPLTLGLKRHT PGTITLLLQDQ-VGGLQATRDGKTWITVQVPGAFVVNLGEHGHF 264 sp P28038.1 FL3H HOR AKACVDMD-QKVVVNFYPRC PQPLTLGLKRHT PGTITLLLQDL-VGGLQATRDGKTWITVQPISGAFVVNLGEHGHF 267 [sp Q95318.1 FL3H HOR AKACVDMD-QKVVVNFYPRC PQPLLTLGLKRHT PGTITLLLQDL-VGGLQATRDGKKNWITVQPISGAFVVNLGEHGHF 267 [sp Q95303.1]FL3H HOR AKACVDMD-QKVVVNFYPRC PQPLLTLGLKRHT DGTITLLLQDL-VGGLQATRDGKNWITVQPISGAFVVNLGEHGHF 267 [sp Q95303.1]FL3H HOR AKACVDMD-QKVVVNFYPRC PQPLLTLGLKRHT DGTITLLLQDL-VGGLQATRDGKNWITVQPISGAFVVNLGEHGHF 267 [sp Q95303.1]FL3H HOR AKACVDMD-QKVVNFYPRC PQPLLTLGLKRHT DGTITLLLQDL-VGGLQATRDGKNWITVQPISGAFVVNLGEHGHF 267 [sp Q95303.1]FL3H HOR AKACVDMD-QKVVNFYPRC PQPLLTLGLKRHT DGTITLLLQDL-VGGLQATRDGKNWITVQPISGAFVVNLGEHGHF 267 [sp Q96330.1]FL3H HOR AKACVDMD-QKVVNFYPRC PQPLLAGVPAHT DGTITLLLQDL-VGGLQATRDGKNWITVQPISGAFVVNLGEHGHF 267 [sp Q96330.1]FL3H HOR AKACVDMD-QKVVNFYPRC PQPLLAGVPAHT DGTITLLQDL-VGGLQATRDGCKNWITVQPISGAFVVNLGEHGHF 267 [sp Q96330.1]FL3H HOR AKACVDMD-QKVVNFYPRC PQPLLAGVPAHT DGTITLLQDL-VGGLQVFKDGWVDVPPMPHSIVNLGEHGF 266 [sp]A221W9.1]ACCO1 RGPAGAPT-FGTKVSSYPPC PPP PPPLAGVVNFTFFTLGAGULVFTFTTVFTFTTPTT DGTUTVFTFTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	5 sp 092784.2 (GOX2 NSDF0GO-AVIOLINHYKOPEDERAMCLAHTSSTIMTILGAM. TAGGYFRUDG-WTSPEDUGGLUWWAUT	LHT	271	
6 sp Q31411.1 G3021 HKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGDLFQI 276 7 sp Q31410.1 G3022 QKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGDLFQI 276 8 sp Q31409.1 G3023 QKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGDLFQI 276 7 sp Q95818.1 FL3H A TNACVDMD-QKVVVNYPKCPQPDLTLGLKRHTDPGTITLLLQDQ-VGGLQATRINGKTWITVQPVGAFVVNLGDHGHF 264 sp P28038.1 FL3H A TNACVDMD-QKVVVNFYPKCPQPDLTLGLKRHTDPGTITLLLQDL-VGGLQATRINGKNWITVQPVSGAFVVNLGDHGHF 267 [sp Q95318.1]FL3H HOR AKACVDMD-QKVVVNFYPKCPQPDLTLGLKRHTDPGTITLLLQDL-VGGLQATRIGGKNWITVQPVSGAFVVNLGDHGHF 267 [sp Q95330.1]FL3H HOR AKACVDMD-QKVVNFYPKCPQPDLLGUPAHTDLSGITLLVPNE-VPGLQVFKDHWPDAEXIPSAVIVNLGDHGHF 266 [sp Q26330.1]FL3H REGLGGEMAEXMMKINYPPCPPPDLALGVPAHTDLSGITLLVPNE-VPGLQVFKDHWPDAEXIPSAVIVNLGDQILK 266 [sp Q26330.1]FL3H NOR GRAAPT-FGTKVSSYPPCPPP PPPLVKGLRAHTDAGGILLFQDDSVGGLQLLKDGEWVDVPPPMRHSIVNLGDQLK 269 [sp Q06534.1]FL3H NOR GRAAPT-FGTKVSSYPPCPPP PPPLVKGLRAHTDAGGILLFQDDSVGGLQLLKDGEWVDVPPPMRHSIVNLGDQLK 269	0 sp 09C971.1 G30X4 KSGSKVGR-GAIRLNHYPVCPEPERAMGLAAHTDSTILTILHOSN-TGGLOVFREESG-WVTVEPAPGVLVVNIGDI	FHI	273	
7 sp Q31410.1 G3022 QKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 8 sp Q31409.1 G3023 QKIAETMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGLLFQI 276 7 sp Q5818.1 FL3H A TNACVDMD-QKVVNYYPKCPQPTLTLGLKRHTDPGTITLLLQD-VGGLQATREGKNWITVQPVEGAFVVNLGTHGHF 264 sp P28038.1 FL3H BG AKACVDMD-QKVVNYYPKCPQPTLTLGLKRHTDPGTITLLLQDL-VGGLQATREGKNWITVQPISGAFVVNLGTHGHF 267 sp Q56330.1 FL3H BG AKACVDMD-QKVVNYYPKCPQPTLTLGLKRHTDPGTITLLLQDL-VGGLQATREGKNWITVQPISGAFVVNLGTHGHF 267 sp Q56330.1 FL3H BG AKACVDMD-QKVVNYPPCPPTLALGVPAHTDLSGITLLVPNE-VPGLQVFKDHWFDAEYIPSAVIVNLGDQLK 266 66 sp A221W9.1 ACCO1 RGPAGAPT-FGTKVSSYPPCPPPLVKGLRAHTDAGGILLFQDSVGGLQLLKDGEWVDVPPPRHSIVNLGDQLK 269	6 sp q31411.1 g3021 HKIAETMT-ATMHLNWYRCPDPKRALGLIAHTDSGFFTFVLGSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGI	FQI	276	
8 sp 31409.1 G3023 _ QKIATMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFFFVLGSL-VPGLQLFRHGPDRWTVPAVPGAMVVNVGLLFQI 276 7 sp Q98818.1 FL3H_A TNACVDMD-QKIVVNYPKCPQPILTGLKRHTDPGTITLLLQDQ-VGGLQATRIGKNWITVQPVGGAFVVNLGHGHF 264 sp P28038.1 FL3H_HOR AKACVDMD-QKIVVNYPKCPQPILTGLKRHTDPGTITLLLQDL-VGGLQATRIGKNWITVQPVGGAFVVNLGHGHF 267 sp Q96330.1 FL3I_A EGLGGEMAEYMMKINYPPCPRTLALGVPAHTDLSGITLLVPNE-VPGLQVFKDHWFDAEYIPSAVIVHLGDGLF 266 66 sp A221W9.1 ACCO1_RGPAGAPT-FGTKVSSYPPCPRTLALGVPAHTDLSGITLLVPNE-VPGLQVFKDGWVDVPPMRHSIVVNLGDQLEV 229 91 sp Q06514_1 G2007_RGPAGAPT-FGTKVSSYPPCPRTLAUGUNTUPTTUFTUFUTUFTUFUTUFTUFUTUFTUFUTUFTUFUTUFTUF	7 sp Q31410.1 G3022 QKIAFTMT-ATMHLNWYPKCPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGT	FQI	276	
/ sp Q95818.1 FL3H A TNACVDMD-QXIVVNYYFKCPQPBLTLGLKRHTDPGTITLLLQDQ-VGGLQATRINGKTWITVQPVEGAFVVNLGDHGHF 264 sp P28038.1 FL3H HOR AKACVDMD-QXVVNFYPRCPQPDLTLGLKRHTDPGTITLLLQDL-VGGLQATRINGKTWITVQPISGAFVVNLGDHGHF 267 sp Q96330.1 FL5I AR EGLGGEMAEYMMKINYYPPCPRPILALGVPAHTDLSGITLLVPNE-VPGLQVFKDHWFDAEYIPSAVIVHIGDQILR 266 66 sp A221W9.1 ACCO1 RGPAGAPT-FGTKVSSYPPCPRPILVKGLRAHTDAGGIILLFQDDSVGGLQLLKDGEWVDVPPMRHSIVVNLGDQLEV 229	8 SP Q31409.1 G3023 _ QKIATTMT-ATMHLNWYFKOPDPKRALGLIAHTDSGFFTFVLQSL-VPGLQLFRHGPDRWVTVPAVPGAMVVNVGI	FQI	276	
SP 22038.1 FL3H HOK AKAVUMD-QXVVVNFPFCPQPELLGLKKHTDFGTITLLQUL-VGGLQATRDGGKNWIFVQFISGAFVVNLGDHGHF 267 sp Q96330.1 FL51_AR EGLGGEMAEYMMKINYYPPCPRPLLALGVPAHTDLSGITLLVPNE-VFGLQVFKDHWFDGEYIPSAVIVHIGDQILR 266 66 sp A221W9.1 ACCO1 RGPAGAPT-FGTKVSSYPPCPRPLVKGLRAHTDAGGIILLFQDSVGGLQLKDGEWVDVPPMRHSIVVNLGDQLEV 229 sp Q96514_1 22027_PNTET FN_GTKVSSYPPCPRPLVKGLRAHTDAGGIILLFQDSVGGLQLKDGEWVDVPPMRHSIVVNLGDQLEV 229	7 SD QVS818.1 FL3H A TNACVDMD-QKIVVNY9KCPQPDLTLGLKHHTDPGTITLLLQQ-VGGLQATRDNGKTWITVQDVEGAFVVNLGF	GHF	264	
66 sp A221W9.1 ACCO1 RGPAGAPT-FGTKVSSYPPCPRP LVKGLRAHTDAGGILLLFODSVGGLQLLKDGEWVDVPPMRHSIVVNLGDQLEV 229	BP 22038.1 FL3H HOK AKACVDMD-QKVVVNFFFFPQPULIGLKRHTPGTITLLOGL-VGGLQATRUGGKNWITVQPISGAPVVNLGE	GHF	267	
CIPELING CONTRACT DE CONTRACTOR DE LA CO	6 s0 1221W9 1 2CC09 RODACADT. FORVESTIGNAL THE PEPEPEPELAG VARIALS	LEV	220	
	CUP LARGET 1 10007 ANTONE TO OFF DEFENSION OF A DESCRIPTION OF A DESCRIPTI	FOR	261	

Fig. 1 contd...

Contd		Fig	1
Conta	•••	1 15.	

The second s	* . *	* : :	*	
p P10967.1 ACCH3 SC	LSNDKYLSVEHRAIS	INNVGSRMSITCFFGESPYQSS	KLYGPITELLSEDNPPKYRATTVKDHTSYLHNRGLDGTSALS	359
sp Q84MB3.1 ACCH1	ITNDEFISVEHRVLA	NVAGPRISVACFFSSYLMANP	RVYGPIKEILSEENPPNYRDTTITEYAKFYRSKGFDGTSGLL	361
D 039110.2 GAOX1	LSNDRYKSCLHRAVV	NSESERKSLAFFLC PKKD	RVVTPPRELLDSITS RRYPDFTWSMFLEFTOKHYRADMNTLO	367
5D 039111.1 GAOX2	LSNGIFKSCLHRAVV	NRESARKSMAFFLC PKKD	KVVKPPSDILEKMKT RKYPDFTWSMFLEFTOKHYRADVNTLD	365
5D 039112.1 GAOX3	LINGRYKSCLHRAVY	NSERERKTFAFFLCPKGE	KVVKPPEELVNGVKSGERKYPDFTWSMFLEFTOKHYRADMNTLD	368
SD 004705.1 GA01D	LSNGRYKSCLHRAVY	NSRVPRKSLAFFLC PEMD	KVVAPPGTLVDAANP RAYPDFTWRSLLDFTOKHYRADMKTLE	344
SD 004707.1 GA01A	LSNGRYKSCLHRAVY	NSRVPRKSLAFFLCPEMD	KVVAPPGTLVDASNP RAYPDFTWRSLLDFTOKHYRADMKTLE	344
SD 004706.1 GA01B	LSNGRYKSCLHRAVY	NSKVPRKSLAFFLCPEMD	KVVAPPGTLVDAANP RAYPDFTWRSLLDFTOKHYRADMKTLE	344
SD P93771.2 GAOX1	LSNARYRSCLHRAVY	NSTAPERSLAFFLC PEMD	TVVRPPEELVDDHHP RVVPDFTWRALLDFTORHYRADMRLFO	354
SD POC5H5.1 GAOX2	LSNGRYKSCLHRAVY	NORRERRSLAFFLC PRED	RVVRPPPSAATPOHYPDFTWADLMRFTORHYRADTRTLD	366
sp 039103.2 g30X1	LSNGLFKSVLHRARY	NOTRARLSVAFLWGPOSD	IKISPVPKLVSPVES PLYOSVTWKEYLRTKATHFNKALSMIR	353
SD 09ZT84.2 G30X2	LINGIFPSVLHRARY	NHVRSRFSMAYLWGPPSD	IMISPLPKLVDPLOS PLYPSLTWKOYLATKATHFNOSLSIIR	346
SD 09C971.1 G30X4	LSNGKIPSVVHRAKV	NHTRSRISIAYLWGGPAGD	VOIAPISKLTGPAEP SLYRSITWKEYLOIKYEVFDKAMDAIR	349
SD 031411.1 G3021	LINGREHSVYHRAVY	NRDSDRISLGYFLG PPAH	VKVAPLEBALAG-TPAAYRAVTWPEYMGVEKKAFTTGASALK	350
SD 031410.1 G3022	LINGREHSVYHRAVY	NRDSDRISLGYFLG PPAH	VKVAPLERALAG-TPAAYRAVTWPEYMGVEKKAFTTGASALK	350
SD 03T409.1 G3023	LINGREHSVYHRAVI	NEESDETSLOVELG PPAH	VKVADLERALAG-TDAAVRAVTWDEVMGVRKKAFTTGASALK	350
SD 095818.1 FL3H 2	LSNGREENADHOAVI	NSNSSRLSTATEON PAPE	ATVYDLKVREGEKAT LEEDITEAEMVKRKMGRDLELARLKK	338
P28038 1 FT.3H HOF	MSNGREENADHOAVI	NGESSBLSTATEON PAPD	ARUWPLAUREGERPT LEEPITETEMVRRKMERDLDLAKRKK	341
096330 1 81.91 45	L CNOPVENUL HOTTL	THEFT PAGE DUFLE	KTUGDI DELTODOND - DEEKDEAFKDVSVPKINKT.DLD	226
0 0 3 2 7 1 W9 1 3 C C O 1	TTNOPVEGUMUPUU	OTDONEMOTACEVN DOCE	AUTODADATUVEEPA WUAVDEUVEENVEVUEENVEVUEENVEDDE	205
ap 090614 1 02077	TENGUNOSUDUDUT	DANTERMOTAREVC	AVISPAPADVALLA-VVAIPAPVPLDIMADIVALAALEAALEPAP	220
sp Q84MB3.1 ACCH1	YLKI		365	
sp Q39110.2 GAOX1	AFSDWLTK PI		377	
	OTHER DESIGNATION OF THE OWNER OWNER OF THE OWNER OWNE OWNER OWNE		378	
sp Q39111.1 GAOX2_	SFSNWVITNNNPI			
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3	EFSIWLKNRRSF		380	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp 004705.1 GAO1D	EFSIWLKNRRSF VFSSWIVQQQQP	-QPART	380 361	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp 004705.1 GAO10 sp 004707.1 GAO1A	SFSNWVIINNNPI EFSIWLKNRRSF VFSSWIVQQQQP VFSSWIVQQQQGQLI	-QPART	380 361 365	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp 004705.1 GAO10 sp 004707.1 GAO1A sp 004706.1 GAO18	SFSNWVITNNNPI EFSIWLKNRRSF VFSSWIVQQQQP VFSSWIVQQQQQQLI VFSSWIVQQQQQQLI	-QPART ALQPAMT	380 361 365 365	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp 004705.1 GAO1D sp 004707.1 GAO1A sp 004706.1 GAO1B sp P93771.2 GAOX1	SFSNWVITNNNPI- EFSIWLKNRRSF VFSSWIVQQQQQ VFSSWIVQQQQQQLI AFSDWLNHHRHL	-QPART	380 361 365 365 372	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp 004705.1 GAO1D sp 004707.1 GAO1A sp 004706.1 GAO1A sp P93771.2 GAOX1 sp P93771.2 GAOX1 sp P93771.2 GAOX1	SFSNWVITNNNPI EFSIWLKNRRSF VFSSWIVQQQQGQLA VFSSWIVQQQQGQLA VFSSWIVQQQQQQQLA AFSDWLNHHRHL AFTRWLAPPAADAAA	-QPART	380 361 365 365 372 389	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp O04705.1 GAO1D sp O04705.1 GAO1A sp O04706.1 GAO1A sp P93771.2 GAOX1 sp P0C5H5.1 GAOX2 sp Q39103.2 G3OX1	SFSNWVITNNNPI EFSIWLKNRRSF- VFSSWIVQQQQP- VFSSWIVQQQQQLI VFSSWIVQQQQQLI AFSDWLNHHRHL- AFTRWLAPPAADAAI NHREE-	-QPART LLQPAMT 	380 361 365 365 372 389 358	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp O04705.1 GAO1D sp 004707.1 GAO1A sp 004706.1 GAO1B sp P93771.2 GAOX1 sp P905H5.1 GAOX2 sp Q32T84.2 G3OX1 sp Q32T84.2 G3OX2	SFSNWVITNNPI EFSIWIKNRRSF- VFSSWIVQQQQQL VFSSWIVQQQQGQL AFSDWINHHRHL AFTRWIAPPAADAAA NHREE- N-	-QPART	380 361 365 365 372 389 358 347	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp O04705.1 GAO1D sp O04705.1 GAO1D sp 004706.1 GAO1E sp P93771.2 GAOX1 sp P905H5.1 GAOX2 sp Q39103.2 G3OX1 sp Q92T84.2 G3OX2 sp Q92T84.2 G3OX2	SFSNNVITNNNPI EFSINUKNRRSF VFSSWIVQQQQQL VFSSWIVQQQQQQL VFSSWIVQQQQQQQL AFSDWINHHRHL AFTRWIAPPAADAAA NHREE N	QPART	380 361 365 372 389 358 347 355	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp O04705.1 GAO1D sp O04705.1 GAO1D sp 004706.1 GAO1A sp P93771.2 GAOX1 sp P93771.2 GAOX1 sp Q39103.2 G3OX1 sp Q37184.2 G3OX2 sp Q92784.2 G3OX2 sp Q92791.1 G3OX4 sp Q31411.1 G3O21	SFSNNVITNNNPI EFSINLKNRRSF- VFSSWIVQQQQQL VFSSWIVQQQQQQL VFSSWINQQQQQQL AFSDWLNHHRHL AFTRWLAPPAADAAA NHREE- N- VVNPTN- MVAISTDNDAANDTI	QPART	380 361 365 365 372 389 358 347 355 370	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp O04705.1 GAO1D sp O04705.1 GAO1D sp 004706.1 GAO1A sp P93771.2 GAOX1 [sp P0C5H5.1 GAOX2 sp Q39103.2 G3OX1 sp Q92T84.2 G3OX2 sp Q9271.1 G3O22 sp Q31410.1 G3O22	SFSNWVITNNNPI EFSIWLKNRRSF- VFSSWIVQQQQQ VFSSWIVQQQQQL/ VFSSWIVQQQQQL/ AFSDWINHHRHL AFTRWLAPPAADAA/ NHREE- N- VVNPTN- WVAISTONDAANDTI MVAISTONDAANHTI	-QPART	380 361 365 365 372 389 358 347 355 370 370	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp O04705.1 GAO1D sp O04705.1 GAO1D sp 004706.1 GAO1D sp 93771.2 GAOX1 sp 93771.2 GAOX1 sp Q3103.2 G3OX1 sp Q3141.1 G3OX2 sp Q31410.1 G3O22 sp Q31410.1 G3O23	SFSNWVITNNNPI EFSIWLKNRRSF VFSSWIVQQQQQL VFSSWIVQQQQQQL VFSSWIVQQQQQQL AFSDWLNHHRHL AFTRWLAPPAADAAA NHREE N- VVNPTN MVAISTDNDAANDTI MVAISTDNDAANDTI MVAISTD-DAANDTI	QPART LQPAMT 	380 361 365 365 372 389 358 347 355 370 370 370 369	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp O04705.1 GAO1D sp O04705.1 GAO1D sp O04706.1 GAO1E sp P93771.2 GAOX1 sp Q3103.2 G3OX1 sp Q3143.2 G3OX2 sp Q92784.2 G3OX2 sp Q31411.1 G3O21 sp Q31411.1 G3O21 sp Q31410.1 G3O23 sp Q35818.1 FL3H_Z	SFSNWVITNNNPI EFSIMUKNRRSF- VFSSWIVQQQQQLJ VFSSWIVQQQQQQLJ VFSSWIVQQQQQQLJ AFSDWINHHRHL AFTRWIAPPAADAA NHREE- N- VVNPTN- MVAISTDNDAANHTI MVAISTDNDAANHTI LAKEERDHKEVD-	QPART 	380 361 365 372 389 358 347 355 370 370 370 369 358	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp O04705.1 GAO1D sp O04705.1 GAO1D sp 004706.1 GAO1A sp P93771.2 GAOX1 sp P93771.2 GAOX1 sp Q3103.2 G3OX1 sp Q3141.1 G3OX2 sp Q31411.1 G3O21 sp Q31410.1 G3O22 sp Q31410.1 G3O23 sp Q31410.1 G3O23	SFSNNVITNNNPI EFSINKKNRSF- VFSSWIVQQQQQL VFSSWIVQQQQQQL AFSDWLNHHRHL- AFTRWLAPPAADAAA NHREE- VVNPTN- MVAISTDNDAANDTI MVAISTDNDAANTI MVAISTDNDAANTI MVAISTD-DAANDTI LAKEERDHKEVD- QAKDQLMQQQLQLQ	-QPART	380 361 365 372 389 358 347 355 370 370 369 358 377	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp O04705.1 GAO1D sp O04705.1 GAO1D sp 004706.1 GAO1D sp 93771.2 GAOX1 sp 93771.2 GAOX1 sp Q39103.2 G3OX1 sp Q3784.2 G3OX2 sp Q3784.2 G3OX2 sp Q3784.1 G3OX4 sp Q3140.1 G3O22 sp Q3140.1 G3O23 sp Q3818.1 FL3H HOF P28038.1 FL3H HOF p Q96330.1 FLST_AF	SFSNWVITNNPI EFSIWVQQQQQD- VFSSWIVQQQQQQL VFSSWIVQQQQQQL AFSDWINHHRHL- AFTRWLAPPAADAA NHREE- N- VVNPTN- WAISTDNDAANDTI MVAISTDNDAANHTI MVAISTD-DAANHTI LAKEERDHKEVD- QAKDQLMQQQLQLQQ	QPART LQPANT 	380 361 365 372 389 358 347 355 370 370 369 358 358 377 336	
sp Q39111.1 GAOX2 sp Q39112.1 GAOX3 sp O04705.1 GAO1D sp O04705.1 GAO1D sp 004707.1 GAO1A sp 93771.2 GAOX1 sp P0C5H5.1 GAOX2 sp Q3103.2 G3OX1 sp Q9ZT84.2 G3OX2 sp Q9C971.1 G3OX4 sp Q31411.1 G3O21 sp Q31410.1 G3O22 sp Q31409.1 G3O23 sp Q9S818.1 FL3H AD p Q9S818.1 FL3H AD p Q9630.1 FL3I AD p Q9630.1 FL3I AD p Q9630.1 FL3I AD	SFSNWVITNNNPI EFSIWLKNRRSF- VFSSWIVQQQQQQL VFSSWIVQQQQQQL AFSDWLNHHRHL- AFTRWLAPPAADAAA NHREE- N- VVNPTN- MVAISTDNDAANDTI MVAISTDNDAANDTI LAKEERDHKEVD- QAKDQLMQQLQLQQ EAFKSMETETSNRIA	QPART	380 361 365 372 389 358 347 355 370 370 370 369 358 377 336 322	

alignment employing combined PAM 120 scoring matrices of FASTA and WU-Blast2 yielded gapped, linear and organized alignments with homologous sequences. The study represented selection of parameters and overlap residues for different scoring matrices that resulted in better explanation and understanding of evolutionary relationships with semidwarf protein of rice. The quality of alignments produced by BLAST reveals clear evolutionary relationship when compared with FASTA and WU-Blast2. The conserved residue pattern observed for homologous sequences scanned by BLAST (Blosum80) resulted in 21 conserved residues by ClustalX multiple sequence analysis in comparison with WU-Blast2 (PAM120) and FASTA (PAM120) that showed 12 and 11 conserved residues, respectively. The results suggests that selection of correct matrix for proper sequence analysis plays a major role in identifying an organized alignment pattern for predicting functional and evolutionary biology.

REFERENCES

- Altschul, S.F. (1993). A protein alignment scoring system sensitive at all evolutionary distances. J. Mol. Evol., 36: 290-300.
- Altschul, S.F., Gish, W., Miller, W., Eugene, W.M. and David, J.L. (1990). Basic local alignment search tool. J. Mol. Biol., 215: 403-410.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25: 3389-3402.
- Barton, G.J. and Stemberg, M.J.E. (1987). A strategy for the rapid multiple alignment of protein sequences. *J. Mol. Biol.*, **198**: 327-337.

- Feng, D.F. and Doolittle, R.F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol., 25: 351-360.
- Gonnet, G.H., Cohen, M.A. and Benner, S.A. (1992). Exhaustive matching of the entire protein sequence database. *Sci.*, **256**: 1443-1445.
- Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G. and Gibson T.J. (1994). CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positionspecific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22: 4673-4680.
- International Rice Research Institute (1967) Annual Report for 1966, 59–82.
- Ishii, T., Nakano, T. and Maeda, H. (1996). Phylogenetic relationships in A-genome species of rice as revealed by RAPD analysis. *Genes & genetic Systems*, **71**: 195-201.
- Jia, Q., Zhang, J., Westcott, S., Zhang, X.Q., Bellgard, M., Lance, R. and Li, C.(2009). GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. *Functional Integartive Genomics*, 9(2): 255-262.
- Karlin, S and Altschul, S.F.(1990). Methods for accessing the statistical significance of molecular features by using general scoring schemes. *Proceedings of National Academy of Sciences*, 87: 2264-2268.
- Monna, L., Kitazawa, N., Yoshino, R., Suzuki, J., Masuda, H., Maehara, Y., Tanji, M., Sato, M., Nasu, S. and Minobe, Y. (2002). Positional cloning of rice semidwarfing gene, sd-1 rice "green revolution gene" encodes a mutant enzyme involved in gibberellins synthesis. *DNA Res.*, 9: 11-17.

- Nagano, H., Onishi, K., Ogasawara, M., Horiuchi, Y. and Sano, Y. (2005). Genealogy of the "Green Revolution" gene in rice. *Genes Genetics Systems*, **80**: 351-356.
- Ogi, Y., Kato, H., Marayuma, K. and Kilkuchi, F.(1993). The effects of culm length and other agronomic characters caused by semidwarfing gene at the *sd-1* locus in rice. *Japanese J. Breed.*, **43**: 267-275.
- Pearson, W.R. and Lipman, D.J.(1988). Improved Tools for Biological Sequence Comparison. *Proceedings of National Academy of Sci.*, **85**: 2444-2448.
- Pearson, W.R.(1991). Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics*, **11**: 635-650.
- Sasaki, A., Ashikari, M., Tanaka, M., Itoh, H., Nishimura, A., Swapan, D., Ishiyama, K., Saito, T., Kobayashi, M., Khush, G.S., Kitano, H. and Matsuoka, M. (2002). Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature*, **416**(6882):701-702.
- Smith, T.F. and Waterman, M.Sc.(1981). Identification of common molecular subsequences. J. Mol. Biol., 147: 195-197.
- Timo Lassmann and Erik L. L. Sonnhammer (2002). Quality assessment of multiple alignment programs. *FEBS Letters*, **529**: 126-130.
- Wang, Y. and Li, J.(2005). The Plant Architecture of Rice (*Oryza sativa*). *Plant Mol. Biol.*, **59**(1): 75-84.

****** *****