Phylogeny of related functions: the case of polyamine biosynthetic enzymes

Agnieszka Sekowska,^{1,2} Antoine Danchin^{1,2} and Jean-Loup Risler³

Author for correspondence: Antoine Danchin. Tel: +852 2816 8402. Fax: +852 2168 4427. e-mail: adanchin@hkucc.hku.hk

- Regulation of Gene Expression, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France
- ² Hong Kong University Pasteur Research Centre, Dexter HC Man Building, 8 Sassoon Road, Pokfulam, Hong Kong
- ³ Genome and Informatics, Université de Versailles-Saint-Quentin, 45 Avenue des Etats Unis, 78035 Versailles Cedex, France

Genome annotation requires explicit identification of gene function. This task frequently uses protein sequence alignments with examples having a known function. Genetic drift, co-evolution of subunits in protein complexes and a variety of other constraints interfere with the relevance of alignments. Using a specific class of proteins, it is shown that a simple data analysis approach can help solve some of the problems posed. The origin of ureohydrolases has been explored by comparing sequence similarity trees, maximizing amino acid alignment conservation. The trees separate agmatinases from arginases but suggest the presence of unknown biases responsible for unexpected positions of some enzymes. Using factorial correspondence analysis, a distance tree between sequences was established, comparing regions with gaps in the alignments. The gap tree gives a consistent picture of functional kinship, perhaps reflecting some aspects of phylogeny, with a clear domain of enzymes encoding two types of ureohydrolases (agmatinases and arginases) and activities related to, but different from ureohydrolases. Several annotated genes appeared to correspond to a wrong assignment if the trees were significant. They were cloned and their products expressed and identified biochemically. This substantiated the validity of the gap tree. Its organization suggests a very ancient origin of ureohydrolases. Some enzymes of eukaryotic origin are spread throughout the arginase part of the trees: they might have been derived from the genes found in the early symbiotic bacteria that became the organelles. They were transferred to the nucleus when symbiotic genes had to escape Muller's ratchet. This work also shows that arginases and agmatinases share the same two manganese-ion-binding sites and exhibit only subtle differences that can be accounted for knowing the three-dimensional structure of arginases. In the absence of explicit biochemical data, extreme caution is needed when annotating genes having similarities to ureohydrolases.

Keywords: secondary metabolism, *Helicobacter pylori*, *Bacillus subtilis*, *Synechocystis* PCC6803, discriminant amino acid residue

INTRODUCTION

While the number of genome sequences increases exponentially it remains difficult to identify gene functions explicitly. Automatic annotation procedures rely mostly on sequence comparisons. They are used to build up phylogeny trees, where reference activities are assumed to spread to neighbours by contiguity. The corresponding functions are thus described tentatively

Abbreviation: FCA, Factorial Correspondence Analysis.

as identical to that of the known reference. However, these methods do not address the central question of enzyme recruitment for new activities (Jensen, 1976; Danchin, 1989; Roy, 1999). Furthermore, genes and proteins are not simply sequences of letters, they are made from chemicals derived from cell metabolism and a single gene alteration may result in a general base or amino acid content bias (Cox & Yanofsky, 1967), changing the 'style' of an organism, possibly altering its place in calculated phylogenies, thus leading to wrong assignments of enzyme activities.

From the chemical standpoint, the origin of life is often

ARG1 SCHPO	VSIINMEFSGG-QPKD	AELAEEMIEAAG	LPEDLER-LGYSVNVVQNP-	KFKSRPLKEG
ARG1_XENLA	VAVICAPFSKG-QKRR	VEHGPAAIRSAG	LIERLSN-LGCNVC	DFGDLHFSQV
ARG2_AGRTU	IRLIGAELQIG-AGQL	CEMGESAYRIAG	LTRALED-LGHRVV	DTGNVTPAPL
ARG2_HUMAN	VAVIGABFSQG-QKRK	VEHGRAAINEAG	LMKRLSS-LGCHLK	OF OF CENTY
ARG2 MOUSE	VALVEARTSRG-QARL	UEV CEAATEREAG	LIKRISK-LGCHIK	DE ODI SETNV
ARG2 SCHPO	VSTINMUESGG-OPKD	AFL ADEMVEKAG	LVDDLEH-LGYDVKLTONP-	EFKSRPSKEG
ARGI LEIAM	MSIVLAPFSGG-OPHS	VELGPDYLLKOG	LOODMEK-LGWDTRLERVFD	SKVVEARKASENEERI
ARG3 XENLA	VAVIEAFFSKG-QKRR	VEHGPAAIESAG	LIDRLSN-LGCNVC	DFGDLHFSQV
ARGI_AGRTU	CQILCARVQSG-ASQP	CLMGPDAFRTAG	LTQVLTE-LGWAVT	DLGDATPTVE
ARGI_ARATH	TSLLCVPLGHNSSFLQ	PAFAPPRIREAI	-WCGSTNSATE-EGKELKDPRVLT	DVGDVPVQEI
ARGI_BACCD	ISIICVEMDLG-QTRR	VDMGRSAMRYAG	VIERLER-LHYDIE	DLGDIPIGKA
ARGI_BACSU	ISVIGMEMDLG-QARR	VDMGESAIRYAH	LIERLSD-MGYTVE	DLGDIPINRE
ARGI_STAAU	IDIIGAESTFG-QRKL	VDLGETAIRYAG	LISRLKQ-LDLDVY	DKEDIKVPAV
ARGI BACHA	CKITEL BUORC-BCRY	CNMCHDCVHAAC	LLAWLAPDNDQPLIHVPVQA	THE REAL PROPERTY AND ADDRESS
ARGI BROAD	TRATECANCLA-CROL	CENAVEVIEASE	VIAGVOTPLDLEWCKTT	FFVN#CPUAS
ARGI CLOAC	IDIL VEIYYG-SDRK	VDLGESKLEEKN	LASLIAK-YNHNVK	
ARGI COCIM	LGVVAVGFSDG-OPNO	VDPSGLIEAG	LLDQLRDDLEYDIRHDGQVH	TYAE FVPEHDPNH-
ARGI DEIRA	ISILGIEMDLG-AGRR	VDMGPSALENAH	LAHTLRD-LGHDVT	DLGDIEVALP
ARGI_EMENI	LGVVAVGFNGG-QCKL	VVAAPMALVEAG	LLDQLRDDLDYEIHYDNTVH	YEKEVPAEDPDH-
ARGI_GLYMA	STLLEVELGHNSSFLE	PAFAPPFIREGI	-WCGSANSTTE-EGKDLKDLRIMV	DVGDIPIQEM
ARGI_HELPY	MILVELEAELG-ASKR	TDKGVRRLISEALSA	THGDVIKGMCV-LQRTITQEYKEF	RYAKNFEDYY
ARGI HOMAN	LETTENPERG-OPEG	VEKCEASIEVAC	LLEKLKE-QECDVR	DUGDIAFVDV
ARGI NEIGO	WVITEVEYDMAVSCRS	ARFCHEATURAS	VNLAWEHRR-FPWTFDVRERLN	TDCGDLVFSFG
ARGI NEUCR	LGIVAVGFSGG-OCKP	VDAAPSALIESG	LLTOLREELGYRLHGDDEVH	LYTDLVPKEDPPH
ARGI RANCA	VGVLEARFSKG-QARG	VEEGRIYIRRAG	LIEKLEE-LEYEVR	DYCOLHFPEL
ARGI_RATNO	IEIIGAPFSKG-QPRG	VEKGRAALRKAG	LVEKLKE-TEYNVR	DHGDLAFVDV
ARGI_RHOCA	SIILGARIQSG-THQP	CVMGRASLRTAG	LIASLTA-LGWRIE	DQGDLSIGPQ
ARGI_SCHPO	IAIIGVEFDTAVSHRP	ARFGEKGINSAS	SRQMAIRGFNPSLNVNPYESWAKI	LDCGDIPVSSY
ARGI STRPI	TALIGEKSDAGVIINN	CPU CEPAULENC	RTQLAK-FPWHLGNQVMVI	DC3
ARGI YENLA	VOVI SANFSKG-OPPP	VERCENVLERS	LIEKLRE-FONDVPDCCDI-	PLANSNKHPA
ARGI YEAST	LSIVLARFSGG-OCKL	VEKGREYMLEHG	LOTSIED-LGWSTELEPSMD	CAOFVGKLKMEKDSTTGGSSVM
SPEB ARCFU	YLIYCINYDATOSFKP	SRFAPNAISEAS	WNLESYSNLFDVELSLVKVG-	DAMNINCDGG
SPEB BACSU	AILYGMEMDWTVSYRP	SRFGESRIBEVS	IGLEEYSP-YLDRDLADLNFF	DACDIPLPFG
SPEB_DEIRA	VAALGVEFDIALGFRE	ARFAPRALREAS	LRSVMPP-FTGLDGKTRLQG	TFADAGDVILPSL
SPEB_ECOLI	WVITCVEFDMATSGRA	GRHGRAAIRQVS	TNLAWEHNR-FPWNFDMRERLN	VVDCGDLVYAFG
SPEB_METFE	FGLLGVPFDSTSTYKP	SRFGELMIRQAS	YNFENYSLHYRKKLDVPII	DLGDIEVILG
SPEB METJA	GVIFSIEYDETTSFKP	ARE GGNAINTAS	WGLETYSPIL-DRDLAELKYC	DLKBLDLYGS
SPEB METTH	VIILGIDEDGTTSVEP	APPCPVI.TROAT	ISPLAINLEP-SENVEVESF-	DA GOWAL DVS
SPEB RHOCA	IGLIGARWDGGTTNRP	PRHSEOGRADUBDAS	TMIRAVNGATE-VAPFDLARC	ADLGDVAPNPG
SPEB SCHPO	IAFLCAFFDTGTSYRP	ARFGPSGIREGS	RRLNLYGGYNVPMETNPFNNWAKI'	VDCGDIPLTSY
PAHA STRCL	VVVIGAPYDGGTSYRP	ARFGPQAIRSES	GLIHGVGIDRG-PGTFDLINC	VDAGDINLTPF
SPEB_SYNEC	VAIVEVEHDSGTTYRP	TRFGPQGIRRIS	ALYTPYNFEMGVDLREQISLC	DVGDIFTIPA
HUTG BACSU	PALICVELSKSSISHS	ASFAPGTIRQAL	KHSSAYSAELGEHVVSELLY-	DLCDIDIHVT
C279_PSEAE	AAFVEVELDIGTSLRS	TRFGURE IRAES	VMIRPYNMATG-AAPFDSLNV	ADICDVAINTF
CZ97_POLAL	VGLIGVEWDGGIINKA	SARAGIARE VIANES	SERRAVER-VSRIAPIDEVRV	Jan Stranger Martine Land APVNP1
	Innananana		II	
	TO TE TO TE	****** *********	****	
				↓
ARG1_SCHPO	PNQALMKNPL	YVSNVTROVRNIVOOELE	KQRIAVNIGG	HSLAIGTVEGVQAVYD
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU	PNQALMKNPLY PNDELYNSIVKHPR REFSHPNPAVHHLAI	YVSNVTRQVRNIVQQELE IVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA	KQRIAVNICC AGHTCVTLCC AAVPIFLCC	HSLAIGTVEGVQAVYD HSLAFGSITGHAQQCP HAISAGTVAGMARRVAO-SGR
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN	PNQALMKNPLY PNDELYNSIVKHPR REFSHPNPAVHHLAI PKDDLYNNLIVNPRS	YVSNVTROVRNIVOOELE TVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANOELAEVVSRAVS	KQRIAVNIGG AGHTCVTLGG AAVPIFLGG DGYSCVTLGG	DESLAIGTVEGVQAVYD DESLAFGSITGHAQQCP DEAISAGTVAGMARRVAQ-SGR DESLAIGTISGHARHCP
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE	PNQALMKNPLY PNDELYNSIVKHPRY REFSHPNPAVHHLAF PKDDLYNNLIVNPRS PQDDPYNNLVVYPRS	YVSNVTROVRNIVOOELE TVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS	KQRIAVNIGG AGHTCWTLGG AA	DESLAIGTVEGVQAVYD DESLAFGSITGHAQQCP DEAISAGTVAGMARRVAQ-SGR DESLAIGTISGHARHCP DESLAIGTISGHARHCP
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO	PNQALMKNPLJ PNDELYNSIVKHPR REFSHPNPAVHHLAN PKDDLYNNLIVNPRS PKDDPYNNLVVYPRS	YVSNVTROVRNIVOQELE TVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS	KQRIAVNIGG AG	DESLAIGTVEGVQAVYD DESLAFGSITGHAQVYD DEAISAGTVAGMARVAQ-SGR DESLAIGTISGHARHCP DESLAIGTISGHARHRP DESLAIGTISGHARHRP
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO	PNQALMKNPLY PNDELIN-SIVKHPR: REFSHPN-PAVHHLAI PKDDLYN-NLIVNPR: PKDDPYN-NLVVYPR: PKDPYN-NLVVYPR: PNQALMKNPL	YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS YVSNVTRQVRDAVQRELE	KQRIAVNIGG AGRICVTLGG AA	DESLAIGTVEGVQAVYD DESLAFGSITGHAQVYD DEAISAGTVAGMARVAQ-SGR DESLAIGTISGHARHCP DESLAIGTISGHARHRP DESLAIGTISGHARHRP DESLAIGTVEGVQAVYD
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO ARG2_SCHPO ARG1_LEIAM ARG2_VENLA	PNQALMKNPL PNDELYNSIVKHPR REFSHPNPAVHHLAI PKDDLYNNLVVYPR PKDDPYNNLVVYPR PNQ	YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVATEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS YVSNVTRQVRAVQRELE LTAECTEKIYKCVRRVAE	KQ RIAVNIGE AG HTCVTLCG AA -VPIFLCG DG YSCVTLCG GG YSCVTLCG GG YSCVTLCG QQ RVVNIGC QQ RVVNICC QC RVVNICC	V DESLAIGTVEGVQAVYD DESLAFGSITGHAQQCP DEAISAGTVAGMARRVAQ-SGR DESLAIGTISGHARHCP DESLAIGTISGHARHP DESLAIGTISGHARHP DESLAIGTVEGVQAVYD DESLAIGTVEGVQAVYD DESLAIGTVEGVQAVHP DESLAIGTVEGVQAVHP
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO ARG2_SCHPO ARG1_LEIAM ARG3_XENLA ARG3_XENLA	PNQALMKNPL) PNDELYNSIVKHPR' REFSHNNPAVHHLAI PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDPYNNLVVYPR PNQALMKNPL) 	YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGTANQELAEVVSRAVS LTACTEKIYKCVRRVAE TVGLACKVLAKEVGRAVG LVGWTRSISOKALEMAB	KQRIAVNIGG AG	V SLAIGTVEGVQAVYD DESLAFGSITGHAQQCP DESLAIGTVAGMARVAQ-SGR DESLAIGTVSGHARHCP DESLAIGTISGHARHP DESLAIGTVEGVQAVHD DESLAIGTVEGVQAVHD DESLAIGTVAGVLSVHD DESLAFGSITGHAQQCP DESLAFGSITGHAQQCP
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_RATNO ARG2_SCHPO ARG7_LEIAM ARG3_XENLA ARG1_AGRTU ARG1_AGRTU	PNQALMKNPLN PNDELYNSIVKHPX REFSHENPAVHHLAI PKDDLYNNLVVVPR PKDDPYNNLVVVPR PNQALMKNPL 	YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS IVSNVTRQVRDAVQRELE TAECTEKIYKCVRVAE TVGLACKVLAKEVGKAVG ALVGWTRSLSQKALEMAR	KQRIAVNIGG AG	DESLAIGTVEGVQAVYD DESLAIGTVEGVQAQCP DEAISAGTVAGMARVAQ-SGR DESLAIGTISGHARHCP DESLAIGTISGHARHRP DESLAIGTISGHARHP DESLAIGTVEGVQAVYD DESLAIGTVEGVQAQCP DESLAFGSITGHAQQCP DESLAFGSITGHAQQCP DESLAFGSITGHAQQCP
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 AGRTU ARG1 ARATH ARGT BACCD	PNQALMKNPL PNDELYNSIVKHPR REFSHPNPAVHHLAI PKDDLYNNLVVYPR PKDDPYNNLVVYPR PNQ	YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVATEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLNQELAEVVSRAVS SVGIANQELAEVVSRAVS YVSNVTRQVRAVQRELE LTAECTEKIYKCVRRVAE ALVGWTRSLSQKALEMAR DDDRLMNVISESVLLVME ALVGWTRSLSQKALEMAR	KQ RIAVNIGE AG HTCVTLGC AG VPIFLGC DG YSCVTLGC GG YSCVTLGC GG YSCVTLGC QQ YSCVTLGC QQ RVVNNIGC QG RVVVNIGC QG RVVVNIGC QG DIPVFLTCG AG DIPVFLCG EEP LRPUVLCG RG RFPUVLCS	V DESLAIGTVEGVQAVYD DESLAFGSITGHAQVYD DESLAIGTISGHARHCP DESLAIGTISGHARHCP DESLAIGTISGHARHP DESLAIGTISGHARVYD DESLAIGTVEGVQAVYD DESLAGSITGHAQQCP DESLAGSITGHAQQCP DESLAGTVSGVAQRTAE-LGG DESISIPVVRAVSEKHGG
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG7 LEIAM ARG3 XENLA ARG1 ARATH ARG1 BACTU ARG1 BACSU	PNQALMKNPLI PNDELYN-SIVKHPR REFSHDN-PAVHLAI PKDDLYN-NLVVYPR PKDDPYN-NLVVYPR PNQALMKNPLJ PNQALMKNPLJ PNDEQYN-SIVKHPR PDLSHPN-SAVKNLDJ RDCGVI ERLHEQG-DSRLRNLKJ KIKNDEELKNLM	YVSNVTRQVRNIVQELE TVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS VVSNVTRQVRDAVQRELE LTAECTEKIYKCVRRVAE TVGLACKVLAKEVGKAVG DDDRLMNVISESVKLVME AVAEANEKLAAAVDQVVQ SVLAGNEKLAQKVMKVIE	KQ RIAVNIGG AG HTCVTLGG DG YSCVTLGG GG YSCVTLGG QQ YSCVTLGG QQ RVVNIGC QG RFPITICG AG HTCVTLGG QC RFPITICG AG HTCVTLGG EC DLPVFLGG RG RFPLVLGG RG RFPLVLGG	DESLAIGTVEGVQAVYD DESLAFGSITGHAQQCP DEAISAGTVAGMARRVAQ-SGR DESLAIGTISGHARHCP DESLAIGTIGGHARHP DESLAIGTIGGHARHP DESLAIGTVEGVQAVHD DESLAFGSITGHAQQCP DESLAFGSITGHAQQCP DESMSAGTVSGVAQRTAE-LGG DESIAIGTLAGVAKHYE DESLAIGTLAGVAKHYE
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_RATNO ARG2_RATNO ARG2_SCHPO ARG3_XENLA ARG3_XENLA ARG1_ARATH ARG1_BACCU ARG1_BACSU ARG1_STANU	PNQALMKNPL) PNDELYNSIVKHPR: REFSHNNPAVHHLAI PKDDLYNNLVVYPR: PKDDPYNNLVVYPR: PNQALMKNPL) 	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGIANQELAEVVSRAVS TVGLACKVLAKEVCRAVG ALVGWTRSLSQKALEMAR DDRLMNVISESVKLVME AVAEANEKLAAAVDQVVQ SVLAGNEKLAQKVNKVIE EIIDVNQKLNKEVSASIE	KQRIAVNIGG AG	V ESLAIGTVEGVQAVPD DESLAIGTVAGMARVAQ-SGR DESLAIGTISGHARHCP DESLAIGTISGHARHCP DESLAIGTISGHARHPD DESLAIGTVEGVQAVPD DESLAIGTVEGVQAVPD DESLAIGTVAGVLS
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 AGRTU ARGT BACCD ARG1 BACCD ARG1 BACAN	PNQALMKNPL PNDELYNSIVKHPR REFSHPNPAVHHLAI PKDDLYNNLVVYPR PKDDDYNNLVVYPR PKDDPYNNLVVYPR PNQ	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVATEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS YVSNVTRQVRAVQRELE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE ALVGWTRSLSQKALEMAR DDDRLMNVISESVRLVME VAEANEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVNEVIE SRKQLLEQLEAAQHIILA	KQ RIAVNIGE AG HTCVTLGC AG VPIFLGC DG YSCVTLGC GG YSCVTLGC GG YSCVTLGC QG YSCVTLGC QQ RFVVNMGC QQ RVVVNGC QG RVVVNGC QG DLPVFLTCG SC DLPVFLCG EEP LRPUVLGC EK KFPIVLGG NN RFPIVLGG NN DRIVMGC	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAIGTVAGMARRVAQ-SGR USLAIGTISGHARHCP USLAIGTISGHARHPD USLAIGTISGHAR
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARGT GARTU ARGT BACAU ARGT BACAU ARGT BACHA ARGT BACHA ARGT BACHA	PNQALMKNPLI PNDELYNSIVKHPR REFSHPNPAVHLAI PKDDLYNNLVVYPR PKQDDPYNNLVVYPR PNQALMKNPL PNQALMKNPL PNDEQYNSAVKNLD RDEGG	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGIANQELAEVVSRAVE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE ALVGWTRSLSQAALEMAR DDDRLMNVISESVKLVME DDDRLMNVISESVKLVME SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVNIVIE SIKQLLEQLEAAQHIIHA YAVANEASLEAAYRESA SVTOOCOLAETPOUTE	KQ	DESLAIGTVEGVQAVYD DESLAFGSITGHAQVYD DESLAFGSITGHAQVYD DESLAIGTISGHARHCP DESLAIGTISGHARHCP DESLAIGTISGHARHPD DESLAIGTISGHARVHD DESLAFGSITGHAQVYD DESLAFGSITGHAQ
ARG1 SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HOMUSE ARG2_SCHPO ARG2_SCHPO ARG1_EIAM ARG3_XENLA ARG1_ARATH ARG1_ARATH ARG1_BACSU ARG1_BACSU ARG1_BACHA ARG1_CAELL ARG1_CAEL ARG1_CAEL	PNQALMKNPL PNDELYNSIVKHPR REFSHNNPAVHHLAI PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDPYNNLVVYPR PNQALMKNPL PNEQYNSAVKNLD RDCG	YVSNVTRQVRNIVQELE YVSNVTRQVRNIVQELE ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGTANQELAEVVSRAVS SVJTANQELAEVVSRAVE UTACTEKIYKCVRRVAE ITACTEKIYKCVRRVAE ITACTEKIYKCVRRVAE SVSNASHEKLAAXDQVV SVLASHEKLAAXDQVV SVLASHEKLAAXDEV SIKQLLEQLEAAAYRESA SVTQTCRQLAHETRQVIE SIVAS	KQ RIAVNIGG AG HTCVTLGG DG YSCVTLGG GG YSCVTLGG QQ YSCVTLGG QQ RVVVNIGG QG RFPLTIGG AG HTCVTLGG QG RFPLTIGG AG HTCVTLGG RG RFPLTIGG RG RFPLVLGG RG RFPLVLGG RG RFPLVLGG NN RFPLVLGG NN RFPLVLGG NKP PFLVLGG NK FFLVFLGG SG NFPFVVGG	V ESLAIGTVEGVQAVPD DESLAIGTVAGMARVAQ-SGP DESLAIGTISGHARVPD DESLAIGTISGHARHCP DESLAIGTISGHARHPD DESLAIGTVEGVQAVPD DESLAIGTVSGVAQTAPH DESLAIGTVSGVAQTATE-LGK DESIAIGTLAGTAKHYD DESIAIGTLAGTAKHYD DESIAIGTLAGTAKHYD DESIAIGTLAGTAKYDG DESLAVGSVSAISKYDG DELAAGTVPGIARRAAE-KGR DESLAIGTVSGVAJTAMRPVG DESLGLGSITGASKAVG
ARG1 SCHPO ARGI XENLA ARG2 AGRTU ARG2 HOUSE ARG2 RATNO ARG2 SCHPO ARG1 AGRTU ARG1 AGRTU ARG1 AGRTU ARG1 BACHA ARG1 BACHA ARG1 BACHA ARG1 CAEEL ARG1 CLEEL ARG1 CCCIM	PNQALMKNPL PNDELYNSIVKHPR REFSHPNPAVHHLAI PKDDLYNNLVVYPR PKDDDYNNLVVYPR PKDDPYNNLVVYPR PNQ	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVAMTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS VSNVTRQVRDAVQRELE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE SVSAVELAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVZ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAQVINEVIE SVCUCQLAETRQVIE SVTQTCRQLAETRQVIE SVSAATQQLSRQVVEHAR	KQ RIAVNIGG AG HTCVTLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ YSCVTLGG QQ YSCVTLGG QQ NVVNNIGG QQ RVVVNIGG QG PLPVFLTGG EEP LRPLVLGG EK KFPLVLGG EK RFPLVLGG KK PLPVFLGG EG - RF PLVEGG SG NPFFGG NK ELLVFGG SG NFFFVVAGG	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARHCP USLAIGTISGHARHCP USLAIGTISGHARHPD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAFGSITGHAQTAE-LGG USISYDVVRAVSEKHYD USLAIGTLAGTAKHYD USLAIGTVSGVAQRTAE-LGG USIAIGTLAGTAKYDG USLAIGTVSGIARRAAE-KGR USLAGTVSGIARRAAE-KGR USLAGTVSGIARRAAE-LD USLAIGTSGTAKAIREFLGR
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARGT BACCU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 BACA ARG1 CAEL ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC	PNQALMKNPLJ PNDELYNSIVKHPR REFSHPNPAVHLAJ PKDDLYNNLVVYPR PKQDDPYNNLVVYPR PNQALMKNPLJ PNDEQYNSAVKNLDJ RDEGYNSAVKNLDJ RDCG	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVATEALTEAAVRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS IVSNVTRQVRAVQRELE LTAECTEKIYKCVRRVAE TVGLACKVLAKEVGRAV ALVGWTRSLSQKALEMAR ALVGWTRSLSQKALEMAR MARANEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDVEVASIE BIRQLLEQLEAAQHIIHA XAVAWIEAISEAAVRESA SVTQTCRQLAHETRQVIE PIVEANTELANKVYESLS TVSAATQQLSRQVYEMA	KQ RIAVNIGG AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RVVNIGC QQ RVVNIGC QG RVVNIGC SC DLPVFLGC EG RFPLVLGC KKP DRIMMFGG EG RVVIG SG NFPFVVGC EG RLVFGC ALPGG TFPELGC	V DESLAIGTVEGVQAVYD DESLAFGSITGHAQVYD DESLAFGSITGHAQVYD DESLAIGTISGHARHCP DESLAIGTISGHARHCP DESLAIGTISGHAR
ARG1 SCHPO ARG1 ZENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 CLEIAM ARG3 XENLA ARG1 GACAU ARG1 BACAU ARG1 BACAU ARG1 BACAU ARG1 BACAU ARG1 BACAU ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC	PNQALMKNPLI PNDELYNSIVKHPR' REFSHPNPAVHHLAI PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDPYNNLVVYPR PNQALMKNPLJ PNDEQYNSAVKNLDJ RDCG	YVSNVTRQVRNIVQQELE YVSNVTRQVRNIVQQELE ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGTANQELAEVVSRAVS VVSNVTRQVRDAVQRELE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE MVAEANEKLAAKVLKEV SVLAGNEKLAAKVNKVIE BITDVNQKLNKEVSASIE SIKQLLEQLEAAQHIIHA YAVAWIEAISEAAYRESA SVTQCRQLAHETRQVIE SVFFEPILDACRTAAREVM SVSAVTETLRSQVYEHSK	KQ RIAVNIGG AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG QQ SCVTLGG QQ RVVNIGC QG RFPITICG AG HTCVTLGG RG HTCVTLGG RG RFPUVLGG RG RFPUVLGG RG RFPUVLGG RG RFPUVLGG NN RFPUVLGG NK DRIVMFG EG SG SG NFPFVVGG EG RLPGG	V DESLAIGTVEGVQAVYD DESLAFGTVAGMARVAQ-SGR DESLAIGTVSGMAR-VQ-SGR DESLAIGTISGMARHCP DESLAIGTISGMARHCP DESLAIGTISGMARHPD DESLAIGTVSGVAQAVHD DESLAIGTVSGVAQATAE-LGK DESIAIGTVSGVAQRTAE-LGK DESIAIGTLAGTAKHYD DESIAIGTLAGTAKHYD DESIAIGTLAGTAKHYD DESIAIGTVSGVAJAKHYD DESIAIGTUSGTARAAE-KGR DESIAIGTUSGIARRAAE-KGR DESIGISTGASKAPVG DESIGISTGASKAPVG DESIGISTGASKAPVG DESIGISTGASKAPQ DESIGISTGASKATRELGR
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HOUSE ARG2 RATNO ARG2 RATNO ARG2 SCHPO ARG1 AGRTU ARG1 AGRTU ARG1 AGRTU ARG1 BACHA ARG1 BACHA ARG1 BACHA ARG1 CAEEL ARG1 CCCIM ARG1 CCCIM ARG1 CCCIM ARG1 CCCIM ARG1 CCCIM ARG1 CLYMA	PNQALMKNPL PNDELYNSIVKHPR REFSHPNPAVHHLAN PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDPYNNLVVYPR PNQ	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS USAVTRSLSQKALEMAR DDDRLMNVISESVKLVME SVLACKVLAKEVGKAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME SVLACKVLAKEVGKAVG SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAQXVNEVIE SVGLACQLSQVVYEAN VFFEPILDACRTAAERVM SVSAVTELRSQVVEHSK SDELMKVVSDSVKLVME	KQ RIAVNIGG AG HTCVTLGG AA VPIFLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RYCVNLGG QQ RVVVNIGG QQ RVVVNIGG QG PLPVFLGG EE LRPLVLGG EK KFPLVLGG EK RFPLVLGG EK RFPLVLGG KK PLPLVGG NN RFPLVLGG SG NFPFVVLGG GG RFPFVLGG SG NFPFVVLGG GG RFPFVLGG SG RFPFVLGG GG RFPFVLGG	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARHCP USLAIGTISGHARHCP USLAIGTISGHAR
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARGT BACCU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 CAEL ARG1 CLAEL ARG1 CLAEL ARG1 CLEIRA ARG1 CLEIRA ARG1 CLEIRA ARG1 CLEIRA ARG1 CLEIRA ARG1 CLEIRA ARG1 CLMA	PNQALMKNPLI PNDELYNSIVKHPR PKDLYNSIVKHPR PKDDLYNNLVVYPR PKDDYNNLVVYPR PNQALMKNPL PNQALMKNPL PNQALMKNPL PNQ	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVATEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS INGLACKVLAEVGRAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME DDDRLMNVISESVKLVME EIIDVNQKLNKEVSASIE SIKQLLEQLEAAQHIIHA YAVAWIEAISEAAYRESA SUTQTCRQLAHETRQVIE PIVEANTELANKVYESLS VSAATQQLSRQVYEHAR VFFEPILDACRTAAERVM SUSAKTETLRSQVYEHAR SDERLMKVVSDSVKLVME LFCKENLIPCMEVFE	KQ RIAVNIGE AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RYVNNIGG QQ RVVNIGG QG RFPLTIGG AG HTCVTLGG AG HTCVTLGG SC DLPVFLGG EBP LRPUVLGG NN RFPLVLGG NK PFPLVLGG KK FPFLVLGG SG NFPFVZ GG FPFLVLGG ALPGG FPFLVLGG SG NFPFVZ GG FPFTLGG SG NFPFVZ GG FPFTLGG SG KFTTLGG SG KFTTLGG SG KFTTLGG SG KFTTLGG SG KFTTLGG SG KFTTLGG	VESLAIGTVEGVQAVYD DESLAFGSITGHAQVYD DESLAFGSITGHAQVYD DESLAIGTISGHARHCP DESLAIGTISGHARHCP DESLAIGTISGHARHPD DESLAIGTISGHARVHD DESLAFGSITGHAQVYD DESMSAGTVSGVAGVTAT-LGK DESISYPVVRAVSEK
ARG1 SCHPO ARG1 SCHPO ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 SCHPO ARG1 EIAM ARG3 XENLA ARG1 BACCD ARG1 BACCD ARG1 BACAB ARG1 BACAB ARG1 COAE ARG1 COAE	PNQALMKNPLI PNDELYNSIVKHPR' REFSHPNPAVHLAI PKDDLYNNLVVYPR: PKDDPYNNLVVYPR: PKDDPYNNLVVYPR: PNQALMKNPLJ PNEQYNSAVKNED RDEQYNSAVKNED RDEQYNSAVKNED RDEQYNSAVKNED RDEQF	YVSNVTRQVRNIVQQELE YVSNVTRQVRNIVQQELE ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGIANQELAEVVSRAVS VSNVTRQVRAVQRELE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE AVGANTKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAQKVNKVIE EIIDVNQKLNKEVSASIE SIKQLLEQLEAAQHIIHA SVTQTCRQLAHETRQVIE SVFEPILAACVIESLS TVSAATQQLSRQVIEHAR SVSAVTETLRSQVYEHSK SDERLMKVVSDSVKLVME FCKENLIPCMKEVFE SVGKASEQLAGKVAQVKK	KQ RIAVNIGG AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG QQ RYVVMGC QQ RYVVMGC QG RYVVMGC QG RYVVMGC QG RFPLTICG AG HTCVTLGG EE DLPYFLGG RG RFPLVLGG EK VFPLVLGG EK PIFLVLGG SG PIFLVGG SG RLVVTLGG EG FPIFLGG SG RFPFUVLGG EG KTLTLGG ED RFPITLGG SG RFPITLGG KK EFPITLGG KK EFPITLGG	V DESLAIGTVEGVQAVYD DESLAFGSITGHAQQCP DESLAFGSITGHAQSCR DESLAIGTISGHARHCP DESLAIGTISGHARHCP DESLAIGTISGHARHPD DESLAIGTISGHARVHD DESLAFGSITGHAQVCD DESLAFGSITGHAQQCP DESLAFGSITGHAQVCD DESLAFGSITGHAQVCD DESLAFGSITGHAQVCD DESLAFGSITGHAQVCD DESLAFGSITGHAQVCD DESLAFGSITGHAQVCD DESLAFGSITGHAGTAKHYD DESLAFGSITGHAGTAKAHYD DESLAFGSITGHAGTAKAHYD DESLAFGSITGHASAADO DESLAFGSITGHASAALGG DESLAFGSISGIAKATRE-LGR DESLAFGSISGIAR
ARG1 SCHPO ARG1 XENLA ARG2 AGRTUU ARG2 HOUSE ARG2 RATNO ARG2 RATNO ARG2 SCHPO ARG1 AGRTU ARG1 AGRTU ARG1 AGRTU ARG1 BACHA ARG1 BACHA ARG1 CAEEL ARG1 CCCIM ARG1 CCCIM ARG1 CCCIM ARG1 CCCIM ARG1 GLYMA ARG1 GLYMA ARG1 GLYMA ARG1 GLYMA ARG1 HUMAN ARG1 MOUSE ARG1 MOUSE	PNQALMKNPL PNDELYNSIVKHPR PEDSLYNNLVVYPR PKDDLYNNLVVYPR PKDDYNNLVVYPR PKDDPYNNLVVYPR PNQ	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS USANTEX LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE SVGLASELAAATELAA SVTQTCRQLAHETRQVIE SVGANELLAACTIELA SVGANETLRSQVYEHSR SVGKASELAGVVAEVQE SVGKASELAGVAEVQE SVGKASELAGVAEVQE SVGKASELAGVAEVQE SVGKASELAGVAEVQE SVGKASELAGVAEVGKE SVGKASELAGVAEVGELAGVAEVQE SVGKASELAGVAEVQE SVGKASELAGVAEVGKELAS	KQ RIAVNIGG AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RYCVNIGG QQ RVVNNIGG QQ RVVNNIGG QQ RVVNNIGG QG PLPVFLGG EE LRPLVLGG EC PIPVFLGG EK KFPLVLGG NN RFPLVLGG NK FPLFLGG NK ELLVFGG SG NFPFVVLGG EG FPLTLGG ALPGG TPPFVGG EG KFTTLGG KK EFPLTLGG KK EFPLTLGG KK FPLTLGG KK FPLTLGG KK FPLTLGG KK FPLTLGG KK FPLTLGG KK FPLTLGG KG RVSTLGG GG RVSTLGG GG RVSTLGG FG KRCLELT	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARHCP USLAIGTISGHARHCP USLAIGTISGHAR
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARGI LEIAM ARG3 XENLA ARGT BACTU ARGT BACCU ARGT BACCU ARGT BACCU ARGT CAEL ARGT CAEL ARGT CLOAC ARGT CAEL ARGT CLOAC ARGT CAEL ARGT CLOAC ARGT CAEL ARGT CLOAC ARGT CLOAC ARGT CLOAC ARGT CLOAC ARGT CLURA ARGT HELPY ARGT HUMAN ARGT MEUCR	PNQALMKNPLJ PNDELYNSIVKHPR PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDPYNNLVVYPR PNQALMKNPLJ PNQALMKNPLJ PNDEQYNSIVKHPK RCG	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVATEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS TVSNVTRQVRAVQRELE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE VSGLACKVLAKEVGKAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME SVGLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVTEVS SVGLASEQLAGKVAQVKK SVGKANELAGVVAEVQK SVGKANEELAGVVAEVQK SVGKANELAGKLSSAV	KQ RIAVNIGE AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RYVNIGG QQ RYVNIGG QQ RYVNIGG QG RYVNIGG GG HTCVTLGG AG HTCVTLGG NN RFPLVLGG NN RFPLVLGG MKP DRIMMFGG EG RFPFVLGG KK EFPFVLGG SG NFPFVZGG SG RLVTLGG ALPGG FFPFVZGG EG KKTLTGG ALPGG FFPFVZGG SG RLVTLGG KK EFPFVZGG GG RVSVZGG FG KKCLSLGG EG RLVTLGG	VESLAIGTVEGVQAVYD DESLAFGSITGHAQVYD DESLAFGSITGHAQVYD DESLAIGTISGHARHCP DESLAIGTISGHARHCP DESLAIGTISGHARHPD DESLAIGTISGHARVHD DESLAFGSITGHAQVYD DESMSAGTVSGVAGVTAB-LGK DESISYPVVRAVSEK
ARG1 SCHPO ARG1 SCHPO ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 SCHPO ARG1 EIAM ARG3 XENLA ARG1 BACCD ARG1 BACCD ARG1 BACAB ARG1 BACAB ARG1 DACAB ARG1 CAEEL ARG1 COCIM ARG1 DEIRA ARG1 COCIM ARG1 DEIRA ARG1 COCIM ARG1 DEIRA ARG1 EMENI ARG1 CHENI ARG1 CHUMAN ARG1 MUSE ARG1 NEIGO ARG1 NEIGO ARG1 NEIGO	PNQALMKNPLI PNDELYNSIVKHPR' REFSHPNPAVHHLAI PKDDLYNNLVVYPR: PKDDPYNNLVVYPR: PKDDPYNNLVVYPR: PNQALMKNPLJ PNEQYNSAVKNLDJ RDCG	YVSNVTRQVRNIVQQELE YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGIANQELAEVVSRAVS VSNVTRQVRAVQRELE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE MSRQLLGUEAAQUKCK SVSNTKLAAAVDQVVQ SVLAGNEKLAAKVNKVIE BIDDNQKLNKEVSASIE SIKQLLGLEAAQUKI SVGAAELARVVESLS SVGRASEQLAGKVAQVKK SVGRASEQLAGKVAQVKK SVGRASELAGVVAEVQES SVGRASELAGVVAEVQES SVGRASELAGVAZVESS SVGRASELAGVAZVESS SVGRASELAGVAZVESS SVGRASELAGVAZVESS SVGRASELAGVAZVESS SVGRASELAGVAZVESS SVSRASELAGVAZVESS SVSRASELAGVAZVESS SVSNTKRIAEQVESSEV	KQ RIAVNIGG AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG QQ RYVVMGC QQ RYVVMGC QG RFPLTICG AG HTCVTLGG SC DLPYFLGG RG RFPLVLGG RG RFPLVLGG KK VFVVVGG EG RFPFVLGG SG NFPFVVGG EG RLVFTIGG EG FFFITICG EG RFTITICG EG FFFITICG SG RIVTTIGG EG RIVTTIGG SG RIVTTIGG SG RIVTTIGG SG RIVTTIGG SG <td>V DESLAIGTVEGVQAVYD DESLAFGSITGHAQQCP DESLAFGSITGHAQSQCP DESLAIGTISGHARVYD DESLAIGTISGHARHCP DESLAIGTIGHARVHD DESLAIGTISGHARVHD DESLAFGSITGHAQVCP DESLAFGSITGHAQ</td>	V DESLAIGTVEGVQAVYD DESLAFGSITGHAQQCP DESLAFGSITGHAQSQCP DESLAIGTISGHARVYD DESLAIGTISGHARHCP DESLAIGTIGHARVHD DESLAIGTISGHARVHD DESLAFGSITGHAQVCP DESLAFGSITGHAQ
ARG1 SCHPO ARG1 XENLA ARG2 AGRTUU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG1 AGRTU ARG1 AGRTU ARG1 BACCD ARG1 BACCD ARG1 BACCA ARG1 CAEL ARG1 CLAEL ARG1 CLAEL ARG1 CCIM ARG1 CCIM ARG1 CCIM ARG1 CCIM ARG1 CCIM ARG1 CCIM ARG1 CCIM ARG1 CCIM ARG1 CCIM ARG1 MOUSE ARG1 NEUCO ARG1 NEUCO ARG1 NEUCO	PNQALMKNPL PNDELYNSIVKHPR REFSHPNPAVHHLAI PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDPYNNLVVYPR PNDEQYNSIVKHPR PNDEQYNSIVKHPR PNDEQYNSIVKHPR RCG	YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGNTRSLSQKALEMAR DDDRLMNVISESVKLVME ITVGLACKVLAKEVGKAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME IIDVNQKLNKEVSASIE SRKQLLEQLEAAQHIIHA VAVAWIEAISEAAYRESA SVTQTCRQLAHETRQVIE SVSAVTELANKVYESLS SVSAVETLRSQVYEHSK SVGKANEELAGVVAEVGE SVGKANEELAGVALQVKK SVGKANEELAGVALQVKS SVSAVTEKMEAHAGKLLS AVSNVTKRIAEQVHSHAK TVGQAAEKVANAVSEVKE	KQ RIAVNIGG AG HTCYTLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QG YSCVTLGG QQ RYVNNIGG QQ RYVVNIGG QQ RYVVNIGG QG RFPLTICG EEP LRPLVLGG EK KFPLVLGG EK RFPLVLGG EG PIVFLGG EG RFPLVLGG EG RFPLVLGG EG RFPLVLGG GG RFPLVLGG SG NFPFVVGG EG RLVFTLGG SG RVTLGG SG RVTLGG SG RVTLGG SG RVSVLGG GG RLSULGG NG RVSVLGG	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARHCP USLAIGTISGHARHCP USLAIGTISGHARHPD USLAIGTVSGVAQTAE
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARGT BACCU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 CAEL ARG1 CLOAC ARG1 CAEL ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 MOUSE ARG1 MOUSE ARG1 MOUSE ARG1 RANCA ARG1 RANCA ARG1 RANCA ARG1 RANCA	PNQALMKNPLI PNDELYNSIVKHPR PKDLYNSIVKHPR PKDDLYNNLVVYPR PKQDDPYNNLVVYPR PKQDPYNNLVVYPR PKQALMKNPL PNQALMKNPL PNQALMKNPL PNQALMKNPL PNQSAKNDD RDCG	YVSNVTRQVRNIVQQELE YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVAMTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE SVGKALEAAAAA AVSWTRSLSQKALEMAR MYSAVLEAAAAAA SVSTCTCQLAETTRQVIE SVSAATQQLSRQVYEHAR SVSAATQLSRQVYEHAR SVGKANEQLAGVVAEVQK SVGAAELAGVVAEVQK SVGAAELAGVVAEVQK SVGAAELAGVVAEVQK SVGAAELAGVVAEVQK SVGAAELAGVVAEVQK SVGAAELAGVVAEVQK SVGAAELAGVVAEVQK SVGAAELAGVVAEVQK SVGAAELAGVVAEVQK SVGAAELAGVVAEVQKA	KQ RIAVNIGG AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RFVVNIGC QQ RVVNIGC QG RVVNIGC QQ RFPLTIGG AG HTCVTLGG AG HTCVTLGG CG RFPLVLGG EEP LRPUVLGG NN RFPLVLGG KKP DRIVMFGG EG RFPLVLGG KK FPFVLGG SG NFPFVZGG SG RFPFVZGG SG RFPFVZGG KK EFPITLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG FG KKCLSLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG GG </td <td>V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARHCP USLAIGTISGHARHCP USLAIGTISGHAR</td>	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARHCP USLAIGTISGHARHCP USLAIGTISGHAR
ARG1 SCHPO ARG1 SCHPO ARG2 AGRTU ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 SCHPO ARG1 SCHPO ARG1 BACHA ARG1 BACHA ARG1 BACHA ARG1 BACHA ARG1 CLOAC ARG1 RATNO ARG1 RATNO ARG1 RATNO ARG1 RATNO	PNQALMKNPLI PNDELYN-SIVKHPR' REFSHPN-PAVHLAI PKDDLYN-NLVVYPR: PKQDDPYN-NLVVYPR: PRQDPYN-NLVVYPR: PNQALMKNPL' PNQALMKNPL' PNQALMKNPL' PNQALMKNPL' PRLSHPN-SAVKNLD' RDLGG	YVSNVTRQVRNIVQQELE YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGIANQELAEVVSRAVE LTAECTEKIYKCVRRVAE TVGLACKVLAKEVGRAVE DDDRLMNVISESVKLVME DDDRLMNVISESVKLVME SVLAGNEKLAAVDOVVQ SVLAGNEKLAAVDOVVQ SVLAGNEKLAAVDVOVQ SVLAGNEKLAAVDVOVG SVLAGNEKLAAVNVKVIE EIIDVNQKLNKEVSASIE PIVEANTELENKVYESLS TVSAATQQLSRQVIEHAR SVGKANEQLAAVVEYES SVGKASEQLAGKVAQVKK SVGKANELAGVVAEVQE SVGKASELAGVVAEVQE SVSNVTKIAEQVHSHAA TVGQAAEKVANAVSEVKR SVGKANEQLAAVVAETQK STRAVILLAARESAAA	KQ RIAVNIGG AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG QQ RYVVMIGG QQ RYVVMIGG QG RYVVMIGG QG RYVVMIGG QG RYVVMIGG QG RYVVMIGG QG RYVVMIGG QG RYVVMIGG CG RYVVMIGG CG RYVVMIGG SC DLPYFLGG RG RFPLVLGG RG RFPLVLGG RG PFPLVLGG NN RFPLVLGG MK EFLLVFGG SG NFPFVV0G EG RLVFGG SG RLVFLGG NG RISUVLGG FG KK EFPILLS NG NG RLSVLGG GG RLVLTLGG SG RVCLLGG SG RVCLLGG SG DLPVFLGG SNNLXTAGLAKDGIFHPRLITLGG	V DESLAIGTVEGVQAVYD DESLAFGSITGHAQQCP DESLAFGSITGHAQQCP DESLAIGTISGHARHCP DESLAIGTISGHARHCP DESLAIGTISGHAR
ARG1 SCHPO ARG1 SCHPO ARG2 AGRTUU ARG2 AGRTUU ARG2 RATNO ARG2 SCHPO ARG2 SCHPO ARG3 SCHPO ARG3 SCHPO ARG3 SCHPO ARG3 SCHPO ARG1 AGRTU ARG1 BACSU ARG1 SCANU ARG1 SCANU ARG1 SCHPO ARG1 CCIM ARG1 CHPO ARG1 SCHPO ARG1 SCHPO ARG1 SCHPO	PNQALMKNPL PNDELYNSIVKHPR REFSHPNPAVHHLAI PKDDLYNNLVVYPR PKQDDYNNLVVYPR PKQDPYNNLVVYPR PNQ	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVAMTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS USAVENAELAAVDQVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVS SVGANELAAVDQVS SVGANELSAAVRESA SVGCANELAGVVAEVQE SVGCANELAGVVAEVGEN SVGCANELAGVVAEVS SVGCANELAGVVAEVS SVGCANELAGVVAEVS SVGCANELAAVVAETQK ETRAWIALLAARAEAAA MTEGYIDLSSKAIKRMCD	KQ RIAVNIGG AG HTCYTLGG AG YSCYTLGG DG YSCYTLGG GG YSCYTLGG GG YSCYTLGG GG YSCYTLGG QQ RYVNNIGG QQ RYVVNIGG QQ RYVVNIGG QQ RYVVNIGG QQ RYVVNIGG SG DLPVFLGG EEP LRPLVLGG EK KFPLVLGG EK RFPLVLGG EG PIFLGG EG PIFLGG EG RFPLVLGG GG RFPLVLGG SG NFPFVVGG GG RFPLVLGG SG NFPFVVGG GG RLVITLGG ALPGG TFPITLGG KK EFPITLSG RG RVSVLGG GG RVVLTLGG SG RVVLTLGG SG RVVLTLGG SG RVVLTLGG SG RVVLTLGG SG RVVLTLGG SG<	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARHCP USLAIGTISGHARHCP USLAIGTISGHARHPD USLAIGTVSGVAQTASVYD USLAIGTVSGVAQTASVYD USLAGSITGHAQ
ARG1 SCHPO ARG1 SCHPO ARG1 ZEPLA ARG2 AGRTU ARG2 AGRTU ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 ZENHO ARG1 EACSU ARG1 EACSU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 CLOAC ARG1 CLOAC	PNQALMKNPLJ PNDELYNSIVKHPR: REFSHPNPAVHLLJ PKDDLYNNLVVYPR: PKDDPYNNLVVYPR: PNQ	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVATEALTEAATRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME ITVGLACKVLAKEVGKAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME EIIDVNQKLNKEVSASIE SKQLLEQLEAAQMIIHA XVAWIEAISEAATRESA SVGCANEQLAEVTESLS SVGCANEQLASVTESS SVGKANEELAGVVEHAR VFFEPILDACRTAAERVM SVGKANEELAGVVENSU SVGKANEELAGVVAEVQE SVGKANEELAGVVAEVQE SVGKANEELAGVAETQKE SVGKANEELAGVAETQKE SVGKANEELAGVAETQKE SVGKANEELAGVAETQKE ETRAWIALLAARAEAAAA MTEGUIDLESKATASPA	KQ RIAVNIGE AG HTCVTLGG AG VPIFIGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RYVNIGG QQ RYVNIGG QG RYVNIGG QG RYVNIGG QG RYVNIGG QG RYVNIGG AG HTCVTLGG AG RFPLVIGG KG RFPLVLGG KK RFPLVLGG KK RFPLVLGG KK RFPLVLGG ALPGG RVSVLGG GG RVSVVLGG FG RLVITLGG SG RVSVVLGG FG RLVITLGG SG RVSVVLGG GG TISVVLGG GG TISVVLGG SG RVSVVLGG GG TISVVLGG GG TISVVLGG GG TISVVLGG GG TISVVLGG GG TISVVLGG GG TISVVLGG GG	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARVYD USLAIGTISGHARVPD USLAIGTISGHAR
ARG1 SCHPO ARG1 SCHPO ARG2 AGRTU ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 EIAM ARG3 XENLA ARGT BACCD ARG1 BACAD ARG1 BACAD ARG1 BACAD ARG1 BACAD ARG1 CLOAC ARG1 RENCA ARG1 RENCA ARG1 RATNO ARG1 SCHPO ARG1 SCHPO ARG1 SCHPO ARG1 SYNEC ARG1 SYNEC ARG1 SYNEC	PNQALMKNPLI PNDELYN-SIVKHPR' REFSHPN-PAVHLAI PKDDLYN-NLVVYPR: PKQDDPYN-NLVVYPR: PRQALMKNPL' PNQALMKNPL' PNQALMKNPL' PNQSUKHPR' PNQSUKHPR' RELSHPN-SAVKNLD' RELHEQG-DSRLRNLK' NIEKFHSEQKGLRNYD' NIEKFHSEQKGLRNYD' NIEKFHSEQKGLRNYD' NSGKKKPR' RDLGP	YVSNVTRQVRNIVQQELE YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVAWTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGIANQELAEVVSRAVS SVGIANGELAEVVSRAVE LTAECTEKIYKCVRRVAE TVGLACKVLAKEVGRAVE SVSTARSLSQAKLEMAR DDDRLMNVISESVKLVME SULAGNEKLAAVDVQVQ SVLAGNEKLAAVDVDVQ SVLAGNEKLAAVDVVV SVLAGNEKLAAVDVVV SVLAGNEKLAAVDVVV SVLAGNEKLAAVDVVV SVLAGNEKLAAVTVKV SVCANEQLAEVTROVIE PIVEANTELANKVYESLS TVSAATQQLSRQVYEHAR SVGCANEQLAGVAVVKNVE SVGCANELAGVVAEVQE SVGCASEQLAGKVAQVKK SVGCANELAGVVAEVQE SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVVAETQK SVGCANEQLAAVAETQK SVGCANEQLAAVAETQK SVGCANEQLAAVAETQK SVGCANEQLAAVAEYK	KQ RIAVNIGG AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG QQ RFVVNMIGG QQ RFVVNMIGG QG RFPLTIGG AG HTCVTLGG QG RFPLTIGG AG HTCVTLGG CG RFPLVLGG CG RFPLVLGG CR RFPLVLGG RG RFPLVLGG KK RFPLVLGG MK RFPLVLGG MK RFPLVLGG SG NFPFVVQG EG RLVFFVQG EG RLVFFVQG EG RLVFFVQG GG RFPTLGG KK EFPTLLGG KK EFPTLIGG KK EFPTLIGG GG RLVTTGG GG RLVTTGG GG RLVTTGG SS RCLSLTGG SG RVCLTLGG SG RVCLTLGG SS DLPVFLGC SNNLK	V DESLAIGTVEGVQAVYD DESLAFGSITGHAQQCP DESLAFGSITGHAQ
ARG1 SCHPO ARG1 SCHPO ARG2 AGRTUU ARG2 AGRTUU ARG2 AGRTU ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 AGRTU ARG1 AGRTU ARG1 BACSU ARG1 STAAU ARG1 BACSU ARG1 CLAEL ARG1 CLAEL ARG1 CLAEL ARG1 CLAEL ARG1 CLEEL ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 SCHPO ARG1 STRPY ARG1 STRPY ARG1 STRPY ARG1 STRPY ARG1 STRPY	PNQALMKNPL PNDELYNSIVKHPR REFSHPNPAVHHLAI PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDYNNLVVYPR PKD	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVG LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE SVGNAVELAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAVDQVVESLS STGTCCQLAHETRQVIE SVGCANELARSVAEVS SVGCANELAGVVAEVQK SVGCANELAGVVAEVQK SVGCANELAGVVAEVQK SVGCANELAGVVAEVQK SVGCANELAGVVAEVQK SVGCANELAGVVAEVQK SVGCANELAGVVAEVQK SVGCANELAGVVAETQK ETRAWIALLAARAEAAAA MTGYIDLSSKATKRMCD LSQAAKVANAVSEVKB SUGCANEULASVAEVAK SLGQLQNSLSKAIKRMCD FEQIVERTKEFLGEVEG	KQ RIAVNIGG AG HTCYTLGG AG YSCYTLGG DG YSCYTLGG GG YSCYTLGG GG YSCYTLGG GG YSCYTLGG QQ RYCYNMGG QQ RYVVNIGG QQ RYVVNIGG QQ RYVVNIGG QG PLPVFLGG EEP LRPLVLGG EK KFPLVLGG EK RFPLVLGG GG PIFLGG KK PIFLGG SG NFFFVVGG GG RFVILLGG KK PIFLGG SG NFFFVVGG GG RLVITLGG ALPGG TFPFVVGG GG RVSVLGG GG RVSVVLGG GG RVVLTLGG NG RVSVVLGG QS DLPVFLGG SMNLKAGLAKDGIFHPRLTLGG SNNLKTAGLAKDGIFHPRLTLGG SNNLKTAGLAKDGIFHPRLTLGG SNNLKTAGLAKDGIFHPRLTLGG QS DLPVFLGG QS DLPVFLGG	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARHCP USLAIGTISGHARHCP USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAFGSITGHAQ
ARG1 SCHPO ARG1 SCHPO ARG2 AGRTU ARG2 AGRTU ARG2 AGRTU ARG2 RATNO ARG2 SCHPO ARG1 EIAMA ARG3 XENLA ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACSU ARG1 BACSU ARG1 CLOAC ARG1 CLYAN ARG1 MUUSE ARG1 SCHPO ARG1 SCHPO	PNQALMKNPLI PNDELYNSIVKHPR PKDELYNSIVKHPR PKDDLYNNLVVYPR PKDDYNNLVVYPR PKDPYNNLVVYPR PKD	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVATEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE IVGLACKVLAKEVGRAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVTSSLS SVGCATEQLAETRAVISSLS SVGCATELGLAAVATEVS SVGCATELAGVVEHAR VFFEPILDACRTAAERVM SVGCATELAGVVAEVQK SVGCANEELAGVVAEVQK SVGCANEELAGVAETQK SVGCANEELAAVAETQK SVGCANEELAAVAETQK SVGCANEELAAVAETQK ETRAWIALLARATSAAA MTEQUIDLSRKATASPA SLEQLQNSLSKAIKRMCD AGDAMVTEVCDGIAPFVE VGCASELLANVASKVQ VFGETKEFLGEVEG	KQ RIAVNIGE AG HTCVTLGG AG YSCVTLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ YSCVTLGG QQ RYVVNIGG QQ RYVVNIGG QQ RYVVNIGG QG RFPLTIGG AG HTCVTLGG SC DLPYFLGG EEP LRPLVLGG NN RFPLVLGG NK RFPLVLGG KKP PFPVLGG ALPGG TFPLTGG SG NFPFVVGG GG RVSVVLGG FG KK EG RVSVVLGG FG RVSVVLGG FG RVSVVLGG FG RVSVVLGG GS RVSVVLGG FG RVSVVLGG SG RVSVVLGG SG RVSVVLGG GG RVSVVLGG GG RVSVVLGG SG RVSVVLGG SG RVSVVLGG SG <td>V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARVYD USLAIGTISGHARVYD USLAIGTISGHAR</td>	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHARVYD USLAIGTISGHARVYD USLAIGTISGHAR
ARG1 SCHPO ARG1 SCHPO ARG2 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARGT BACCU ARG1 BACCU ARG1 BACCU ARG1 BACCU ARG1 BACAU ARGT BACAU ARGT BACAU ARGT CAEL ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLUAC ARG1 HELPY ARG1 HELPY ARG1 HELPY ARG1 HELPY ARG1 RANCA ARG1 RANCA ARG1 RANCA ARG1 RANCA ARG1 RANCA ARG1 SCHPO ARG1 SCHPO A	PNQALMKNPLI PNDELYNSIVKHPR' REFSHPNPAVHHLAI PKDDLYNNLVVYPR: PKQDDPYNNLVVYPR: PKQDPYNNLVVYPR: PKQALMKNPLJ PNEQYNSAVKNLDJ RDCG	YVSNVTRQVRNIVQQELE YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVATEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANGELAEVVSRAVE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE MODORLMNVISESVKLVME SUVAGNEKLAAVNEVS SVLAGNEKLAAVNEVS SVLAGNEKLAAVNEVS SVLAGNEKLAAVNEVSE SVGCANELAEVNEVSES SVGCANEQLAATRESA SVGCANEQLAATRESA SVGCANEQLAAVVAEVGE SVGCASEQLAGTVAEVQE SVGCASEQLAGTVAEVQE SVGCASELAGVVAEVQE SVGCASELAGVVAEVQE SVGCASELAGVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCASELAGVVAEVVE SVGCANELAAVVAEVQE SVGCASELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVQE SVGCANELAAVVAEVVE SVGCANELAAVVAEVQE SVGCANELAAVVAEVCE SVGCANELAAVVAEVE SVGCANELAAVVAEVCE SVGCANELAAVVAEVE SVGCANEVE SVGCANEVE SVGC	KQ RIAVNIGG AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RVVNMIGC QQ RVVNMIGC QG RVVNMIGC QG RFPLTIGG AG HTCVTLGG CG RFPLVIGG CG RFPLVLGG CR RFPLVLGG RG RFPLVLGG KK RFPLVLGG MK RFPLVLGG MK RFPLVLGG SG NFPFVVGG ALPGG FFPTLGG KK EFPITLGG GG RLVFTLGG KG RVSVVLGG FG RVSVVLGG FG RVSVVLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG SNMLKTAGLAKDGITHPRLITLGG SNVLGG GG KFVVAIGG AN RFPTLCG AN RFPTLGG	V ESLAIGTVEGVQAVYD ESLAIGTVAGMARVAQ-SGR DSLAIGTISGHARVYD DSLAIGTISGHARVYD DSLAIGTISGHARVPD DSLAIGTISGHAR
ARG1 SCHPO ARG1 XENLA ARG2 AGRTUU ARG2 AGRTUU ARG2 RATNO ARG2 SCHPO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 AGRTU ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 CAEEL ARG1 CLEL ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 SCHPO ARG1 XENLA ARG1 XENLA	PNQALMKNPL PNDELYNSIVKHPR REFSHPNPAVHHLAI PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDYNNLVVYPR PKD	YVSNVTRQVRNIVQELE YVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVSRAVS SVGLANQELAEVSRAVS USANTELSQKALEMAR DDDRIMNVISESVKLVME LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE SVGANELAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAKVDVVE SVGANELAAVDQVVYESLS STQTCCQLAHETRQVIE SVGANEQLARVYEHSK SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELARVAEVSE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANELAGVVAEVQE SVGANEQLAVVAETQE TVGAAEKVANAVSEVER SVGANELAGVAEVAEVE SVGANEQLAAVVAETQE TVGAAELLANAVAEVKE LGEATKLVYNSVSKVQ DAEMSEKLQAHAEKLAA	KQ RIAVNIGG AG	V SILAIGTVEGVQAVYD SILAIGTVEGVQAVYD SILAIGSITGHAQVYD SILAIGTISGHARHCP SILAIGTISGHARHCP SILAIGTISGHARHCP SILAIGTVAGVLSVYD SILAIGTVAGVLSVYD SILAIGTVAGVLSVYD SILAIGTUAGVAK
ARG1 SCHPO ARG1 SCHPO ARG2 AGRTU ARG2 AGRTU ARG2 AGRTU ARG2 CATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 CAEL ARG1 CAEL ARG1 COCIM ARG1 CAEL ARG1 COCIM ARG1 DEIRA ARG1 CCEL ARG1 COCIM ARG1 DEIRA ARG1 CAEL ARG1 COCIM ARG1 DEIRA ARG1 CAEL ARG1 SCHPO ARG1 SC	PNQALMKNPLI PNDELYNSIVKHPR REFSHPN PAVHHLAJ PKDDLYNNLVVYPR PKDDPYNNLVVYPR PNQ	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVATEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS MODELINVISESVLIVRE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE SVGANEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVSASIE GRKQLLEQLEAAQHIIHA LAVAWIEAISEAAYRESA SVGCATEQLAHETRQVIE SVSAVIETLBSQVYEHAR VFFEPILDACRTAAERVW SVSAVTETLRSQVYEHAR VFFEPILDACRTAAERVM SVSAVTELARVVSDSVKLVME SVCAASEQLAGVVAEVQK SVGANEELAGVVAEVQK SVGCANEELAGVVAEVQK SVGANEELAAVVAETQK SVGANELAZVVAETQK SVGCANEQLAVVAETQK ETRAWIALLSRATASPA SLEQLQNSLSKAIKRMCD ACDAMVTEVCDGIAPFVE HOGASELLANVAETQK LVGEATKLVNSVSKVQ VFGIDLISKAATRANA SLEQLANVTEXLEVSKILA	KQ RIAVNIGE AG HTCYTLGG AG YSCYTLGG DG YSCYTLGG GG YSCYTLGG GG YSCYTLGG GG YSCYTLGG QQ RYVNNGG QQ RYVNNGG QQ RYVNNGG QG RFPLTIGG AG HTCYTLGG SC DLPYFLGG EEP LRPFLVGG KG FFPLVLGG NN RFPLVLGG KK FPLFLGG EG FPLFLGG EG FPLFLGG EG FPLFLGG GG RFPLVLGG SG NFPFVVGG SG RLVFTLGG EG FFLTLGG EG RLVFTLGG CG RLVFTLGG SG RLVFTLGG SG RLVFTLGG SG RLVFTLGG SG RLVFTLGG SG RVVLGG SG RVVLGG SG RVVLGG SG <	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAIGTISGHAR
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 CHPO ARG1 LEIAM ARG3 XENLA ARGT BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 CLOAC ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 SCHPO ARG1 SCHPO A	PNQALMKNPLI PNDELYNSIVKHPR' PKDLYNSIVKHPR' PKDDLYNNLVVYPR: PKDDPYNNLVVYPR: PKQDPYNNLVVYPR: PKDR	YVSNVTRQVRNIVQQELE YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVAMTEALTEAAYRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE UNGLACKVLAKEVGRAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME EIIDVNQKLNKEVSASIE SRCQLEQLEAAQHIIHA YAVAWIEAISEAAYRESA SVTQTCRQLAHETRQVIE PIVEANTELANKVYESLS SVGCANEQLARVNEYSL SVGCANELAGVVEHAR VSAATQUSRQVYEHAR VSGANELAGVVAEVQE SVGRASEQLAGKVAQVKE SVGCANELAGVVAEVQE SVGCANELAGVVAEVC SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVC SVGCANELAGVAEVC SVGCANECVC SVGCANECVC SVGCANECVC SVGCANECVC SVGCANECVC SVGCANECVC SVGCANECVC SVGCANECVC SVGCANECVC SVGCANECVC	KQ RIAVNIGG AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RVVNNIGC QQ RVVNNIGC QG RVVNNIGC QG RVVNNIGC QG RVVNNIGC QG RVVNNIGC QG RVVNNIGC QG RVVNIGC RG RFPLVLGG RG RFPLVLGG NK RFPLVLGG MK RELLYFGC SG RLVITLGG SG RLVITLGG SG RLVITLGG SG RVSVLGG GG RVSVLGG GG RVSVLGG GG RVSVLGG SG RVSVLGG SG RVSVLGG SG <t< td=""><td>V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD DSLAIGTISGHARVYD DSLAIGTISGHARVYD DSLAIGTISGHAR</td></t<>	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD DSLAIGTISGHARVYD DSLAIGTISGHARVYD DSLAIGTISGHAR
ARG1 SCHPO ARG1 XENLA ARG2 AGRTUU ARG2 AGRTUU ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 AGRTU ARG1 AGRTU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 CAEEL ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 SCHPO ARG1 XENLA ARG1 XENLA ARG1 XENLA ARG1 XENLA ARG1 SCHPO ARG1 XENLA ARG1 XENLA	PNQALMKNPL PNDELYNSIVKHPR REFSHPN- PAVHHLAI PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDYNNLVVYPR PKD	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVSRAVS SVGLANQELAEVSRAVS SVGLANGELAEVSRAVS ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME ITVGLACKVLAKEVGKAVG ALVGWTRSLSQKALEMAR SVGLEQLEAAQHTIHA GVTQTCRQLAHETRQVIE SRGQLEQLEAAQHTIHA SVTQTCRQLAHETRQVIE SVGKANELAAVDQVVQ SVLAGNEKLAQKVNKVIE SVGCASEQLAGVVEHAR VFFEPILDACRTAAERVM SVGCASEQLAGVVAEVGE SVGCASEQLAGVVAEVE SVGCASEQLAGVVAEVGE SVGCASELAGVVAEVGE SVGCASEQLAGVVES SVGCASELAGVVES SVGCASELAGVVES SVGCASELAGVVES SVGCASELAGVVES SVGCASELAGVVES SVGCASELAGVVES SVGCASELAGVVES SVGCASELAGVVES SVGCASELAGVVES SVGCASELAGVVES SVGCASE SVGCASELAGVVES SVGCASE SVGCASELAGVVES SVGCASE SVG	KQ RIAVNIGG AG HTCVTLGG AG YSCVTLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RYCVNLGG QQ RYVVNIGC QQ RYVVNIGC QQ RYVVNIGC QQ RYVVNIGC QQ RYVVNIGC QQ RYVVNIGC QG RYVVNIGC QG RYVVNIGC SC DLPVFLGG EE RFPUVLGG KK FPUVLGG KK FPUVLGG SG NFFUVLGG GG RFPUVLGG GG RFPUVLGG GG RFPUVLGG GG REVETLGG NK EELLVFGG SG RLVFTLGG ALPGG TFPITLGG KK EFPITLGG SG RVSUTLGG GG RVSUTLGG GG RVSUTLGG SG RVSUTLGG SG RVSUTLGG S	V SILAIGTVEGVQAVYD SILAIGSITGHAQVYD SILAIGSITGHAQVYD SILAIGTISGHARHCP SILAIGTISGHARHCP SILAIGTVAGVASVYD SILAIGTVAGVLSVYD SILAIGTVAGVLSVYD SILAIGTVAGVLSVPD SILAIGTVAGVASEKLGG SILAIGTLAGVAK
ARG1 SCHPO ARG1 SCHPO ARG2 AGRTU ARG2 AGRTU ARG2 AGRTU ARG2 CATNO ARG2 SCHPO ARG1 LEIAM ARG3 CANNA ARG1 ARGTU ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 CAEL ARG1 CAEL ARG1 COCIM ARG1 CAEL ARG1 COCIM ARG1 CAEL ARG1 COCIM ARG1 CAEL ARG1 COCIM ARG1 CAEL ARG1 COCIM ARG1 CAEL ARG1 CA	PNQALMKNPL PNDELYNSIVKHPR PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDPYNNLVVYPR PNQ	YVSNVTRQVRNIVQELE TVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS MIDDRLANVISESVLIVREVE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE SVGANEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVTEVIESLS GRKQLLEQLEAAQHIIHA LAVAWIEAISEAATRESA SUTQTCRQLAHETRQVIE SVSAVIETLESQVYEHAR SVGCANEQLAGVVAEVQK SVGANELAGVAEVVAEVQK SVGCANEQLAGVVAEVQK SVGCANEQLAAVVAETQK SVGCANEQLAAVAETQK SVGCANEQLAAVAETQK SVGCANEQLAAVAETQK SVGCANEQLAAVAETQK SVGCANEQLAAVAETQK SVGCANEQLAAVAETQK SUGCANEQLAVACETTAK SUGCANEQLAVAVAETQK SUGCANEQLAVAVAETQK S	KQ	V ESLAIGTVEGVQAVYD DSLAFGSITGHAQVYD DSLAFGSITGHAQVYD DSLAIGTISGHARHCP DSLAIGTISGHARHCP DSLAIGTVSGVAQLSVYD DSLAIGTVSGVAQRTAF-LGK DSSAFTVSGVAQRTAF-LGK DSSAFTVSGVAQRTAF-LGK DSSAFTVSGVAQRTAF-LGK DSSAFTVSGVAQRTARYG DSLAIGTVSGVAARAFYG DSLAIGTVSGVAARAFVHP DSLAIGTVSGVAARAFYG DSSAFTVSGVAGSISGHAR
ARG1 SCHPO ARG1 SCHPO ARG1 ZENLA ARG2 AGRTU ARG2 AGRTU ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 ZENLA ARGT BACCD ARGI BACSU ARGI BACSU ARGI BACSU ARGI BACSU ARGI CAEL ARGI CLOAC ARGI COLOC ARGI COCIM ARGI CLOAC ARGI CLOAC A	PNQALMKNPLI PNDELYNSIVKHPR' PKDELYNSIVKHPR' PKDDLYNNLVVYPR: PKDDPYNNLVVYPR: PKQDPYNNLVVYPR: PKQALMKNPLI PNDEQYNSIVKHPR' PLSHPN-SAVKNLD/ RDCG	YVSNVTRQVRNIVQQELE YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVATEALTEAATRESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS AUGMTRSLSQKALEMAR DDDRLMNVISESVKLVME LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE UNGLACKVLAKEVGKAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME EIIDVNQKLNKEVSASIE SKRQLLEQLEAAQHIIHA XAVANIELAALGKUNKVIS SVGKANELAAAVGUSS SVGLAGUSSQUYEHAR SVGCANEQLASVXEHAR SVGAADELAGVVAEVQE SVGKANEELAGVVAEVQE SVGKANEELAGVVAEVQE SVGKANEELAGVVAEVQE SVGKANEELAGVVAEVQE SVGKANEELAGVVAEVQE SVGKANEELAGVVAEVQE SVGKANEELAGVVAEVQE SVGKANELAAVVAETURS SVGKANELAAVVAETURSJE ETRAWIALLAARAEAAAA MTEGYIDLLSRKATASPA AGDAMVTEVCDGIAPFVE ETRAWIALLAARAEAAAAA MTEGYIDLLSRKATASPA AGDAMVTEVCDGIAPFVE ETRAWIALLAARAEAAAAA MTEGYIDLLSRKATASPA AGDAMVTEVCDGIAPFVE ETQUVETKEFLGEVEG DAEMSELQAAHAEKLAA DFKNTCRNISEKVQEVLK QUEIFGTIHSVSREILM DFKNTCRNISEKVQEVLK SVGCANEFICDVAEVLS	KQ RIAVNIGE AG HTCVTLGG AG VPIFLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RYVNIGC QQ RYVNIGC QG RYVNIGC QQ RYVNIGC QG RYVNIGC QQ RYVNIGC QG RFPLVLGG RG RFPLVLGG NK RFPLVLGG SG RFPFVVLGG SG RIVITLGG SG RVSVVLGG GG RVSVVLGG GG <t< td=""><td>V ESLAIGTVEGVQAVYD DSLAFGSITGHAQVYD DSLAFGSITGHAQVYD DSLAIGTISGHARHCP DSLAIGTISGHARHCP DSLAIGTISGHAR</td></t<>	V ESLAIGTVEGVQAVYD DSLAFGSITGHAQVYD DSLAFGSITGHAQVYD DSLAIGTISGHARHCP DSLAIGTISGHARHCP DSLAIGTISGHAR
ARG1 SCHPO ARG1 SCHPO ARG1 ZENLA ARG2 AGRTU ARG2 AGRTU ARG2 RATNO ARG2 SCHPO ARG1 AGRTU ARG1 AGRTU ARG1 AGRTU ARG1 AGRTU ARG1 BACHU ARG1 BACHU ARG1 BACHU ARG1 CAEEL ARG1 CAEEL ARG1 CAEEL ARG1 CAEEL ARG1 CCEM ARG1 CLEAL ARG1 CLEAL A	PNQALMKNPL PNDELYNSIVKHPR REFSHPN- PAVHHLAN PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDYNNLVVYPR PKDCPYNSIVKHPR PNDEQYNSIVKHPR PNDEQYN-SIVKHPR PNDEQYN-SIVKHPR PNDEQYN-SIVKHPR PNDEQYN-SIVKHPR RELHEQC-DSRLENLKI KIKN	YVSNVTRQVRNIVQELE YVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS UTOSACKVLAKEVGKAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME VALAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVQLAGUSANVEN SVGCANELASVANVVESLS SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVQE SVGCANELAGVVAEVVE SVGCANELAGVVAEVVE SVGCANELAGVVAEVVE SVGCANELAGVVAEVVE SVGCANELAGVVAEVVE SVGCANELAGVVAEVVE SVGCANELAGVVAEVE SVGCANELAGVVAEVEN SVGCANELAGVNAVEN SVGCANELAGVVAEVEN SVGCANELAGVNAVE	KQ RIAVNIGG AG HTCVTLGG AG YSCVTLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RFVUNIGG QQ RFVUNIGG QQ RFPLTIGG EE DLPVFLGG EE P RG RFPLVLGG EK KFPLVLGG EK RFPLVLGG EK RFPLVLGG EG PFLTGG GG RFPLVLGG GG RLVFTLGG GG RLVTLGG GG RFPLTIGG KK EFPLTIGG GG RVSVLGG GG RVVLGG GG RVVLGG GG RVVLGG SG RVVLGG SG RVVVLGG SG RVVV	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAFGTISGHAR
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 AGRTU ARG2 RATNO ARG2 SCHPO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 CAEL ARG1 COCIM ARG1 CAEL ARG1 CCIM ARG1 CCEL ARG1 CCEL ARG1 CCEL ARG1 CCEL ARG1 CCEL ARG1 CEL ARG1 CEL ARG1 CEL ARG1 CEL ARG1 RENCA ARG1 RENCA ARG1 RENCA ARG1 STRPY ARG1	PNQALMKNPL PNDELYNSIVKHPR PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDYNNLVVYPR PKDOPYNNLVVYPR PNDEQYNSIVKHPR PNDEQYNSIVKHPR PNDEQYNSIVKHPR PLSHPNSAVKNLD RDCG	YVSNVTRQVRNIVQELE TVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANGELAQVARVA LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE LTAECTEKIYKCVRRVAE SVGANELAAADDQVQ SVLAGNEKLAAADDQVQ SVLAGNEKLAAQVNKVIE SUCASCQLAHETRQVIE SVGCANEQLARVTESLS SVGCANEQLASVYEHAR VFFEDILDACRTAAERVM SVSAVTETLSSVKLVME SVGCANEQLASVYEHSK SVGCANEQLAGVVAEYQE SVGCANEQLAGVVAEYQE SVGCANEQLAGVVAEYQE SVGCANEQLAAVVAEYQE SVGCANEQLAAVVAETQE SVGCANEQLAAVVAETQE SUGCANEQLAAVVAETQE SUGCANEQLAAVVAETQE TVGCATEILANAVTAVKE LGQLINSLSKAIKEMCD AGDAMVTEVCDGIAPPVE SVGCANEQLAAVVAETQE ETRAWIALLAARAENAAM SLEQLQNSLSKAIKEMCD DAINVAVETIKEVESLIK NFMKTAGFIGDSVSEVLD DAINVAVETIKEVRSIP DLMDSLARIEAFYDRVMA IKQIENGHFELLTRNPTS DIVSSHHIFOTMALLS	KQ RIAVNIGG AG HTCYTLGG AG YSCYTLGG DG YSCYTLGG GG YSCYTLGG GG YSCYTLGG GG YSCYTLGG QQ RYVVNIGG QQ RYVVNIGG QQ RYVVNIGG QQ RYVVNIGG QG RPFDITIGG EEP LRPLVLGG EK KFPLVLGG EG RFPLVLGG KK FPLVLGG SG NFPFVVLGG GG RFPLVLGG SG NFPFVVLGG GG RFPLVLGG GG RFPLVLGG GG RFPLVLGG GG RFPLVLGG GG RLVFTLGG GG RLVFTLGG SG RVVLGG SG RLVTLGG SG RVVLTLGG SG RVVLTLGG SG RVVLTLGG SG RVVLTLGG SG RVVLTLGG AD KCELSCG CQ<	V SILAIGTVEGVQAVYD SILAIGTVEGVQAVYD SILAIGTIGHARVYD SILAIGTIGHARVYD SILAIGTISGHAR
ARG1 SCHPO ARG1 SCHPO ARG1 ZENLA ARG2 AGRTU ARG2 AGRTU ARG2 RATNO ARG2 SCHPO ARG3 SCHPO ARG1 LEIAM ARG1 GACTU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 CLOAC ARG1 RENCA ARG1 SCHPO ARG1 SCHPO	PNQALMKNPLI PNDELYNSIVKHPR' PKDELYNSIVKHPR' PKDDLYNNLVVYPR' PKDDYNNLVVYPR' PNQ	YVSNVTRQVRNIVQQELE YVGLACKVLAEEVSKAVG ETVATEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS AUSANTRSLSQKALEMAR DDDRLMNVISESVKLVME ITVGLACKVLAKEVGKAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME EIIDVNQKLNKEVSASIE SRKQLLEQLEAAQHIIHA IAVAEANEKLAAAVDQVVQ SVLAGNEKLAAVGUNQVY SVLAGNEKLAAVGUNGVI SVLAGNEKLAAVTSSLS SVGCLAETQLAETRQVIE SVGCATEQLAETRQVIE SVGCATGUSSVLVME SVGKANGLAGVVEHAR VFFEPILDACRTAAERVM SVGCATGUSSVUEHAR VFFEPILDACRTAAERVM SVGCATGUSSVUEHAR SVGKANEELAGVVAEVQK SVGKANEELAGVVAEVQK SVGKANEELAGVVAEVQK SVGKANEELAGVVAEVQK SVGKANEELAGVAETQK SVGKANEELAGVAETQK SVGKANEELAGVAETQK TVGQAEKVANAVSEVKR MTEGYIDLSRKATASPA MTEGYIDLSRKATASPA MTEGYIDLSRKATASPA DFNTCRNISEKUQAHAEKLA DFNTCRNISEKUQALSTVD AGDAMVTEVCDGIAPFVE EPQLAHDRITEAARQVRG SVGKANEELASVVAETQK QEIFGTIHSVSREVLD DAIKVAVETIKEVSILE DAIKVAVETIKEVSILE DAIKVAVETIKEVSILE DAIKVAVETIKEVSILE SVSCARELLARVASILL SVGCANESKUQAHAEKLAN SUGCASSVLONG SVGCANESKUQAHAEKLAN SUGSSVLONG SVGCANESKUQAHAEKLASSA SUGSSVLONG SVGCANESKUQAHAEKLAN SUGSSVLONG SVGCANESKUQAHAEKLAN SUGSSVLONG SVGCANESKUQAHAEKLAN SUGSSVLONG SVGCANESKUQAHAEKLAN SUGSSVLONG SVGCANESKUQAHAEKLAN SVGCANESKUQAHAEKLAN SUMTAIDASSHLSSCLONG NMIXAIDTASSHLSSCLONG SVSSTANESKUCAHAESSL SVGCANESKUCANASSEVLD SVSSTANESKUCANASSEVLD SUMSANTAGFIGDSVSEVLD SVSSTANESKUCANASSEVLD SVSSTANESKUCANASSEVLD SUMSANTAGFIGDSVSEVLD SUMSANTAGFIGDSVSEVLD SVSSTANESKUCANASSEVLD SVSSTANESKUCANASSEVLD SUMSANTAGFIGDSVSEVLD SVSSTANESKUCANASSEVLD SVSSTANESKUCANA	KQ RIAVNIGE AG HTCVTLGG AG YSCVTLGG DG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RYEVNIGG QQ RYEVNIGG QQ RYEVINIGG QQ RYEVINIGG QG RYEVINIGG AG HTCVTLGG AG RFPLVLGG EEP LRPEVVLGG MN RFPLVLGG MKP DRIMMFGS EG FPIFLUGG KK RFPEVVLGG MK EFDITLGG SG NFPFVVGG SG RVSVVLGG FG KKIVETLGG FG RVSVVLGG FG RVSVVLGG FG RVSVVLGG SG RVSVVLGG SG RVSVVLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG GG RVSVVLGG SG RVSVVLGG	V ESLAIGTVEGVQAVYD DSLAFGSITGHAQVYD DSLAFGSITGHAQVYD DSLAIGTISGHARVYD DSLAIGTISGHAR
ARG1 SCHPO ARG1 SCHPO ARG1 ZENLA ARG2 AGRTU ARG2 AGRTU ARG2 RATNO ARG2 SCHPO ARG1 AGRTU ARG1 AGRTU ARG1 AGRTU ARG1 AGRTU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 CAEEL ARG1 CAEEL ARG1 CAEEL ARG1 CAEEL ARG1 CLYMA ARG1 DEIRA ARG1 CLYMA ARG1 CLYMA	PNQALMKNPL PNDELYNSIVKHPR REFSHPN- PAVHHLAN PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDYNNLVVYPR PKDCPYNSIVKHPR PNQ	YVSNVTRQVRNIVQQELE TVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS UTGLACKVLAKEVGRAVG ALVGWTRSLSQKALEMAR DDDRLMNVISESVKLVME VALAGNEKLAAAVDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAAKDQVVQ SVLAGNEKLAGXVNEVIESLS SKGULEQLEAACHIIHA SVSAVTETLSQVVEHSK SVGKANEGLAGVVAEVGK SVGKANEGLAGVVAEVGK SVGKANEGLAGVVAEVGK SVGKANEGLAGVVAEVGK SVGKANEGLAGVVAEVGK SVGKANEGLAGVVAEVGK SVGKANEGLAGVVAEVGK SVGKANEGLAGVVAEVGK SVGKANEGLAGVVAEVG SVGKANEGLAGVVAEVG SVGKANEGLAGVVAEVG SVGKANEGLAGVVAEVGK SVGKANEGLAGVVAEVG SVGKANEGLAGVVESLSK SVGKANEGLAGVVAEVG SVGKANEGLAGVVAEVG SVGKANEGLAGVVAEVG SVGKANEGLAGVVAEVG SVGCANEGLAGVVAEVG SVGKANEGLAGVVAEVG SVGCANEGLAGVVAEVG SVGCANEGLAGVVAEVG SVGCANEGLAGVAGVK SVG	KQ RIAVNIGG AG HTCVTLGG AG YSCVTLGG GG YSCVTLGG GG YSCVTLGG GG YSCVTLGG QQ RYCVNIGG QQ RYVVNIGG QQ RYVVNIGG QQ RYVVNIGG QQ RYVVNIGG SG JSCVTLGG EE REPEVIGG EE REPEVIGG SG PPVFUGG EK KFPUVLGG SG PPVFUGG GG PPVTIGG GG PP	V USLAIGTVEGVQAVYD USLAFGSITGHAQVYD USLAFGSITGHAQVYD USLAFGSITGHAQ
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 AGRTU ARG2 AGRTU ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 BACCD ARG1 CAEL ARG1 CLOAC ARG1 CCIM ARG1 CLOAC ARG1 CCIM ARG1 GLYMA ARG1 GLYMA ARG1 GLYMA ARG1 GLYMA ARG1 GLYMA ARG1 GLYMA ARG1 RENENI ARG1 GLYMA ARG1 RENENI ARG1 RENENI ARG1 STRPY ARG1 XENNO ARG1 STRPY ARG1 STRPY	PNQALMKNPL PNDELYNSIVKHPR PKDDLYNNLVVYPR PKDDLYNNLVVYPR PKDDYNNLVVYPR PKDOPYNNLVVYPR PNDEQYNSIVKHPR PNDEQYNSIVKHPR PNDEQYNSIVKHPR PNDEQYNSIVKHPR PLSHPNSAVKNLD RDCG	YVSNVTRQVRNIVQELE YVGLACKVLAEEVSKAVG ETVAMTEALTEAATKESA SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANQELAEVVSRAVS SVGLANGELAEVSRAVS USANTELSQKALEMAR AVGANEKLAAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVQ SVLAGNEKLAAVDQVVS SVGANELAAVDQVVS SVGANELAAVDQVVS SVGANELAAVDQVVS SVGANELAAVDQVVS SVGANELSQUSSVKLVME SVGKANELASVLAV SVGANELAGVVAEVQS SVGANELAGVVAEVQS SVGANELAGVVAEVQS SVGANELAGVVAEVQS SVGANELAGVVAEVQS SVGANELAGVVAEVQS SVGANELAGVVAEVQS SVGANELAGVVAEVQS SVGANELAGVVAEVQS SVGANELASVAAVSEVKS SVGANELASVAAVSEVKS SVGANELASVAAVSEVKS SVGANELASVAAVSEVKS SVGANELASVAAVSEVKS SVGANEQLASVVAETQK TVGQAAEKVANAVSEVKS SUGANEQLAAVVAETQK SLGQLQNSLSKAIKSMCD DFNTCRNISEKVQEVLS DFNTCRNISEKVQEVLS DFNTCRNISEKVQEVLS DFNTCRNISEKVQEVLS DFNTCRNISEKVQEVLS DFNTCRNISEKVQEVLS DFNTCRNISEKLQAHAEKLLA SVSAVETIKEVZSIN DFNTCRNISEKVQEVLS DFNTCRNISEKVQEVLS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAHIFS DIVSSHHEIGTISSINGLAFFINGLISSINGLAFFINGLASSINGLAFFINGLASSINGLAFFINGLASSINGLAFFINGLASSINGLASINGLASSINGLASSINGLASSINGLASSINGLASSINGLASSINGLASSIN	KQ RIAVNIGG AG HTCYTLGG AG YSCYTLGG DG YSCYTLGG GG YSCYTLGG GG YSCYTLGG QG YSCYTLGG QQ RYVNNIGG QQ RYVVNIGG QQ RYVVNIGG QQ RYVVNIGG QQ RYVVNIGG SG DLPVFLGG EEP LRPLVLGG EK KFPLVLGG EG PIPTGG GG RFPLVLGG EG PIPTGG GG RFPLVLGG GG RLVFTLGG SG RLVFTLGG SG RLVFTLGG SG RVVLTLGG SG RVVLTLGG SG RVVLTLGG SG RVVLTLGG SG RVVLTLGG SG <td>V SIAIGTVEGVQAVYD SIAIGSISIGHAQVYD SIAIGSISIGHAQ</td>	V SIAIGTVEGVQAVYD SIAIGSISIGHAQVYD SIAIGSISIGHAQ

Fig. 1. For legend see page 1818.

	44.	
ARG1_SCHPO	DACVLWIDAHA	INTPDSSPSKNLHGCPLSFSLGYAEPLPEEFAWTRRV
ARG1_XENLA ARG2_AGRTU	PFFVLWLDAHT	INTPLTTPSGNLHCQPVSFLLRELQDKVPPIPGFSWAKP YHTLETTRSGNLHCTPVAYFSGRDGFSG-YFPPLSHAVP
ARG2 HUMAN	DLCVVWVDAHA	INTPLTTSSGNLHGQPVSFLLRELQDKVPQLPGFSWIKPC
ARG2_RATNO	DLCVIWVDAHA	INTPLTTVSGNIHCQPLSFLIRELQDKVPQLPGFSWIKPC
ARG2_SCHPO ARGI LEIAM	DACVLWIDAHA	INTPESSPSKNLHCCPLSFSLGYAEPLPEEFAWTKRV INTMSGTVSGNLHCCPLSILLGLDRENIPECFSWVPOVSGNLHCCPLSILLGLDRENIPECFSWVPOV
ARG3 XENLA	DLCVIWVDAHA	INTPLTTPSGNLHGQPVSFLLRELQDKIPPIPGFSWAKPC
ARGI_AGRIO	PVDILHLDAHP	DIYDCFEGNKYSHASSFARIMEGG
ARGI BACCD	RLGVIWYDAHG	VNTAETSPSGNIHCMPLAASLGFGHPALTQIGGYSPR UNTLETSPSGNIHCMPLAVSLGIGHESLVNLEGYAP
ARGI_STAAU	NLGVIWYDAHG	LNIPEESPSGNIHGMPLRILTGEGPKELLELNSNV
ARGI_BACHA	KQFVLWLDAHT	LUKYAGYDNGHTLPLGNLLGEEDEEFARHVKIPLKPENVFIAGLATPTEEETNVISKELQKLGVAPTE FHTLETTTSGNLHCTPVAYYTGQKGFEG-YFPKLAAPID
ARGI CAEEL	DIGLIWVDAHM	AHTPDTSDTGNIHGMPVAHLLGFGDKTLVKIGDRLPKIGDRLP
ARGI_COCIM	EMAVIWVDAHA	INRPEDSVSGNIHEMPLAFLTGLAKDD-NEDMFGWLQPDNL
ARGI_DEIRA ARGI EMENI	RTGVIWVDAHT EIGVIWVDAHA	YNTPESSPSGNIHCMPVAHLTGRGDERLTRLGGLVTGPSGNIHCMPVAHLTGRGDERLTRLGGLVTGPSGNIHCMPMAFLTRLATEEKKDIFGWLQEPSGNIHCMPMAFLTRLATEE-KKDIFGWLQEPSGNIHCMPMAFLTRLATEE-
ARGI GLYMA	PVDVLHFDAHP	LYDEFEGNYYSHASSFARIMEGG
ARGI_HUMAN	DLGVIWVDAHT	DINTPLTTTSGNLHGQPVSFLLKELKGKIPDVPGFSWVTPC
ARGI MOUSE ARGI NEIGO	DLCVIWVDAHT KLALIHFDAHT	INTPLTTSSGNLHCQPVSFLLKELKGKFPDVPGFSWVTPC TYDNGSEYDHCTMFYTAPKEGL
ARGI NEUCR	EIAVIWVDAHA	INTPETSGSGNIHCMPVSFLTGLASEDKEEFFGWLKPDHL
ARGI RANCA	DLCVIWVDAHT	INTPLTSSSGNLHGQPVSFLIKELQTKVPAIPGFSWVQPC DINTPLTTSSGNLHGQPVAFLLKELKGKFPDVPGFSWVTPC
ARGI RHOCA	PLFVLWLDAHP	LHSLDTTESGNLHETPVAYACGLGDFAAYPPLAHAID
ARGI_STRPY	DLAVINMON	LRPYDQTGPNSGTGFRQMFDDAVAD
ARGI_SYNEC ARGI_XENLA	PFTVVQIDAHG NLCVVWVDAHA	MRDKFEGSCHNHACVMRRVLELG
ARGI YEAST	DAGLLWIDAHA	INTIESTPSGNLHCCPVSFLMGLNKDVPHCPESLKWVPGN
SPEB_BACSU	DLAIIHFDAHT	LEVEFD
SPEB DEIRA	DLHVVQLDAHL	DFTDTRNDTKWSNSSPFRRACELPN
SPEB_METFE	DVTILHFDAHM	MANTYAGKKFSHATVMRRIYELHP
SPEB_METJA SPEB_METTH	SLTVVHLDAHM	LLRDEYLGRKLSHACVMRRVYELT
SPEB PYRHO	PKSYVVFDAHL	LRDSYQGSRFNHACVARRIHEMG
SPEB_SCHPO	PVSIIHFDSHL	SWKPKVFGGGKSSVGSINHCTYFYHASQEGL
PAHA_STRCL SPEB_SYNEC	PLAVVHLDAHS KVGIIHFDRHV	TNPAFYGGRYHHCTPFRHGIDEKL TQETDLDERMHTCPWFHATNMANAPL
HUTG BACSU	TTAVIQEDAH	VRNTEDGGPTNGTPFRRLLDEEI
C297_PSEAE	PLGMVHFDAHS	TNDRYFGDNPYT <mark>HG</mark> TPFRRAIEEGL

		↓ ↓
ARG1_SCHPO	IEERRLAFICL	CLDPMERAFLRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLEFIVEACEIVAE
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU	IEERRLAFIGL LSKSDIVYIGL EENIGMIGI	♥♥ RoldpmBraflrersitaytmhdvdkygiarvvemalehinpgrrrpihl§fovdacdeivae Roldpabgfilknydisyysmrhidcmgikkvmektfdqllgrrdrpihl§fdidafdbalar Svdpabraaledsgitvhdmrsidehgvavlraflarvqaanglhv§ldvgflesitae
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE	IEERRLAFICL LSKSDIVYICL EENIGMICI ISSASIVYICL LSPPNIVYICL	↓↓ CLDPMERAFLRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFDVDACDEIVAE CLDPABQFILKNVDISYYSMRHIDCMGIKKVMEKTTDQLLGRRDRPIHLSFDIDAFDFALAE SVDPABRAALEDSGITVHDMRSIDEHGVAVLLRAFLARVQAANGLHVSLDVDFLEPSIAE CVDPPEHFILKNVDIQYFSMRDIDRLGIQKVMERTTDLLIGKRQRPIHLSFDIDAFDFALAE DVDPDEHFILKNVDIQYFSMRDIDRLGIQKVMERTTDLLIGKRQRPIHLSFDIDAFDFALAE
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO	IEERRLAFICL LSKSDIVYICL EENIGMICI ISSASIVYICL LSPPNIVYICL LSPPNLVYICL	↓↓ DLDPMERAFLRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFDVDACDEIVAE DLDPABQFILKNYDISYYSMRHIDCMGIKKVMEKTTDQLLGRRDRPIHLSFDIDAFDFALA SVDPABRAALEDSGITVHDMRSIDEHGVAVLLRAFLARVQAANGLHVSLDVDFLEBSIA DVDPPHFHFI-LKNYDIQYFSMRDIDRLGIQKVMERTFDLLIGKRQRPIHLSFDIDAFDFTLAE DVEPPHFHFI-LKNYDIQYFSMREIDELGIQKVMEQTFDLLIGKRQRPIHLSFDIDAFDFTLAE DVEPAHFHFI-LKNYDIQYFSMREIDRLGIQKVMEQTFDRLIGKRKRPIHLSFDIDAFDFKLAE DVEPAHFHFI-LKNYDIQYFSMREIDRLGIQKVMEQTFDRLIGKRKRPIHLSFDIDAFDFKLAE
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO ARG2_SCHPO ARG1_LEIAM	IEERRLAFIGL LSKSDIVYIGL EENIGMIGI ISSASIVYIGL LSPPNIVYIGL IEERRLAFIGL LKPNKIAYIGL	↓ CLDPMERAFLRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFVPACOPIVAE CLDPABQFILKNYDISYYSMRHIDCMGIKKVMEKTTDQLLGRRDRPIHLSFDIDAFDFALA RSVDPABRAALEDSGITVHDMRSIDEHGVAVLLRAFLARVQAANGLHVSLDVDFLEBSIAF DVDPPHFHFI-LKNYDIQYFSMRDIDRLGIQKVMERTTDLLIGKRQRPIHLSFDIDAFDFTLAF DVEPPHFHFI-LKNYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQRPIHLSFDIDAFDFTLAF DVEPPHFHFI-LKNYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQRPIHLSFDIDAFDFTLAF DVEPABHFILKSYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQRPIHLSFDIDAFDFTLAF DVEPABHFILKSFDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQRPIHLSFDIDAFDFTLAF DVDDMERAFLRERNIAAYTMHHVDKYGIGRVVEMAMEHINPGKRRPVHLSFDIDAFDFTLAF DVDDEKKI-LHLNIAAFSMHHVDRYGIDKVVSMAIEAVSPKGT-EPVMVGYDVDTUDLVV
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO ARG2_SCHPO ARG1_LEIAM ARG3_XENLA ARG3_XENLA	IEERRLAFIGL ISKSDIVYIGL -EENIGMIGI ISSASIVYIGL ISPPNIVYIGL IEERRLAFIGL IEERRLAFIGL LKPNKIAYIGL SKSDIVYIGL -PENVSMMGI	↓ ↓ DLD PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRR RPIHLSFVØAC DEIVAE DLD PABQFI LKNYDISYYSMRHIDCMGIKKVMEKTTDQLLGRRD RPIHLSFDIDAFDFALA SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLHVGLDVDFLERSIAF DVD PFBHFI LKNYDIQYFSMRDIDRLGIQKVMERTTDLLIGKRQ - RPIHLSFDIAFDFTLAF DVE PFBHFI LKNYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQ - RPIHLSFDIAFDFTLAF DVE PABHFI LKSYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQ - RPIHLSFDIAFDFTLAF DVE PABHFI LKSYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQ - RPIHLSFDIAFDFTLAF DUE PABHFI LKSYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQ - RPIHLSFDIAFDFTLAF DLD PMBRAF LRERNIAAYTMHHVDKYGIGKVVSMAIEANVSKGT - EPVMVSYDVDTIDLIVW DLD PABQFI LHDLNIAAFSMHHVDRYGIDKVVSMAIEANSFKGT - EPVMVSYDVDTIDLIVY DLD PABQFI LKNYISYYSMRHIDCMGIRKVMEKTTDQLLGRRD - RPIHLSFDIAFDFILAFD SVD PFBRE VABIGYONMDVIDGEQVVE DIAFUSAVSS GBLHUGIDVOTIDALIV
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO ARG2_SCHPO ARG1_LEIAM ARG3_XENLA ARG1_AGRTU ARG1_ARATH	IEERRLAFIGL LSKSDIVYIGL SKSDIVYIGL LSPPNIVYIGL LSPPNLYYIGL LSPRLAFICL LKPNKIAYIGL LSKSDIVYIG -PRNVSMGI -YARRLLQVGI	↓ ↓ DLD PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRR RPIHLSFOVDACDEIVAE DLD PABQFI - LKNYDISYYSMRHIDCMGIKKVMEKTTDQLLGRRD RPIHLSFDIDAFDZALA SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLHVGLDVDFLERSIAF DVD PPBHFI - LKNYDIQYFSMRDIDRLGIQKVMEQTTDLLIGKRQ - RPIHLSFDIDAFDZTLA DVE PPHFI - LKNYDIQYFSMRDIDRLGIQKVMEQTTDLLIGKRQ - RPIHLSFDIDAFDZKLA DVE PABHFI - LKSYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQ - RPIHLSFDIDAFDZKLA DVE - PABHFI - LKSYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQ - RPIHLSFDIDAFDZKLA DUD - PMBRAF - LRERNIAAYTMHHVDKYGIGKVVEMAMEHINPGKR - RPIHLSFDIDAFDZKLA NVD DEEKKI - LHDLNIAAFSMHHVDRYGIDKVVSMAIEAVSRGT - EPVMVSYDVD TIDLZVU DLD - PABQFI - LKNYNISYYSMRHIDCMGIRKVMEKTTDQLLGRRD - RPIHLSFDIDAFDZALAF SVD - PEBRRR - VAEIGVQVAMRYLDEGGVVPLAFIDRSKVS GRHVSLUP VD VD VD SVD - PEBRRR - VAEIGVQVAMRYLDEGGVVPLAFILGKRV - RVHUSYDVD VD VD SVD - PEBRRR - VAEIGVQVAMRYLDEGGVVPLAFILGKRV - SVS GRHVSLUP VD VD SIN - QEGREQ - GKRFGVEQYEMRTFSKD RPMLENLKLGEGV KGVYISIDV CLDPAFA
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_MOUSE ARG2_CHPO ARG1_LEIAM ARG3_XENLA ARG1_ACRTU ARG1_RACHTU ARG1_BACSU	IEERRLAFIGL SKSDIVYIGL SKSDIVYIGL SPPNIVYIGL SPPNLYYIGL LSPPNLYYIGL LSPKIATICL SKSDIVYIGL SKSDIVYIGL YARRLLQVGI IKPENVVIGA	↓ ↓ DLD PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRR RPIHLSFOVACOPIVAE DLD PABQFI LKNYDISYYSMRHIDCMGIKKVMEKTTDQLLGRRD RPIHLSFDIDAFD2ALA SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN CLHVSLDVD/LESSIAF DVD PPBHFI LKNYDIQYFSMRDIDRLGIQKVMEQTTDLIGKRQ RPIHLSFDIDAFD2HLA DVE PPHFI LKNYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQ RPIHLSFDIDAFD2KLAF DVE PABHFI LKSFDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRQ RPIHLSFDIDAFD2KLAF DUD PMBRAF LRERNIAAYTMHHVDKYGIGKVVEMAMEHINPGKR RPVHLSFDIDAFD2KLAF DLD PMBRAF LHDINIAAFSMHHDDRYGIDKVVSMAIEAVSKGT - EPVHVSYDVACDPIVAF DLD PABQFI LKNYNISYYSMRHIDCMGIRKVMEKTFDQLLGRRD RPIHLSFDIDAFD2KLAF SVD PEBRKI LHDINIAFSMHHDDRYGIDKVVSNAIEAVSKST GRHLSFDIDAFD2ALAF SVD PEBRRR - VAEIGVQVADMRVLDEGGVVFLAFLAFVSKSVS GRHLSFDIDAFD2ALAF SSIN - QEGREQ - GKRFGVEQYEMRTFSKD RPMLENLKLGEGV KGVYISIDVD CLDBAFAF SLD - EGEKKF - IREKGIKIYTMHEUDRLGMTRVMEETIAYLKERT DGVHLSLDLDGLDBSDAF
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO ARG2_SCHPO ARG1_LIAM ARG1_XENLA ARG1_AGRTU ARG1_BACSU ARG1_BACSU ARG1_STAAU	IEERRLAFICL LSKSDIVYICL SKSDIVYICL ISSASIVYICL LSPPNIVYICL IEERRLAFICL LKPNKIAYICL LSKSDIVYICL SKSDIVYICL PRNVSMGI IKPENVVICA IKPENVVICA	↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO ARG2_SCHPO ARG1_LIAM ARG3_XENLA ARG1_AGRTU ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_BACSA ARG1_BRCAB	IEERRLAFICL SKSDIVYICL SKSDIVYICL ISSASIVYICL SSPPNIVYICL ISPPNIVYICL IEERRLAFICL LKPNKIAYICL SKSDIVYICL PRNVSMGI IKPEHVVLICA IKPENIVLICA IKPENIVLICA PHNVCALCI	ULD - PMERAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOVACOPIVA DLD PABQFI LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOTAAPDALA SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN CLLHVSLUVPLEBSIA DVD PPBHFI LKNYDIQYFSMRDIDRLGIQXVMERTTDLLIGKRQRPIHLSFOTAAPDALA DVD PPBHFI LKNYDIQYFSMRDIDRLGIQXVMEQTTDRLIGKRKRPIHLSFOTAAPDALA DVE PABHFI LKSYDIQYFSMRDIDRLGIQXVMEQTTDRLIGKRKRPIHLSFOTAAPDALA DVD PPBHFI LKSYDIQYFSMRDIDRLGIQXVMEQTTDRLIGKRKRPIHLSFOTAAPDALA DVD PABHFI LKSYDIQYFSMRDIDRLGIQXVMEQTTDRLIGKRKRPIHLSFOTAAPDALA DUD PMERAFLRERNIAAYTMHHVDKYGIDXVVSMAIEAV3PKGT EPVMV3VV0 TIDPLYVA DLD PAGQFI LKNYNISYSMRHIDCMGIRXVMEKTTDQLLGRRD RPIHLSFOTAAPDALA SVD PEBRRVAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKVS GRLHVSLUVDFLDAFAF SLD EGEKKF IREKGIKIYTMHEVDRLGMTKVMEETIAYLKERT DGVHLSLUVDFLDAFAF SLD EGEKKF IRESGIKIYTMHEVDRLGMTKVMEETIAYLKERT DGVHLSLUUDGLDSAA SUD FGERRY LKESGKKVYTMHEIDRLGMTKVMEETIAYLKERT DGVHLSLUDGDDSAA SLD EGEKKF IRESGIKIYTMHEVDRLGMTKVMEETIAYLKERT DGVHLSLUDGDDSAA SLD KGERQF IKDHNIKTFTMSDIDKLGIKEVIENTIEYLKSRNVDGVHLSLDUGDDSAA SUD KGERQF IKDHNIKTFTMSDIDKLGIKEVIENTIEYLKSRNVDGVHLSLDV
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO ARG2_SCHPO ARG1_LEIAM ARG3_AGRTU ARG1_AGRTU ARG1_BACSU ARG1_STAAU ARG1_BACSU ARG1_STAAU ARG1_CAEL ARG1_CAEL ARG1_CAEL	IEERRLAFICL LSKSDIVYIGL SKSDIVYIGL ISSASIVYIGL LSPPNIVYIGL LSPPNIVYIGL LSPPNIVYIGL LKPNKIAYIGL PRNVSMMGI IKPENVUIGA IKPENVUIGA IKPENIVLIGA PDTEVIGRLGI LLPHNLCMVGI LLPHNLCMVGI	ULD - PMERAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOVACOPIVA DLD PABQFI LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOTAAPDALA SVD PABRA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLLHVSLVV DVD PPHFI LKNYDIQYFSMRDIDRLGIQKVMERTTDLIGKRCRPIHLSFOTAAPDALA DVD PPHFI LKNYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRCRPIHLSFOTAAPDALA DVE PABHFI LKSYDIQYFSMRDIDRLGIQKVMEQTFDRLIGKRCRPIHLSFOTAAPDALA DVE PABHFI LKSYDIQYFSMRDIDRLGIQKVMEQTFDRLIGKRCRPIHLSFOTAAPDALA DVE PABHFI LKSYDIQYFSMRDIDRLGIQKVMEQTFDRLIGKRCRPIHLSFOTAAPDALA DUD PMERAF LRERNIAAYTMHHVDKYGIGRVVEMAMEHINPGKR RPVHLSFOTAAPDALA DLD PMERAF LRERNIAAYTMHHVDKYGIGRVVEMAMEHINPGKR RPVHLSFOTAAPDALA DLD PMERAF LRERNIAAYTMHHVDKYGIGRVVEMAMEHINPGKR RPVHLSFOTAAPDALA DLD PAGQFI LKNYNISYYSMRHIDCMGIRKVMEXTIDQLGRD RPIHLSFOTAAPDALA SVD PEBRR VAEIGVQVADMRVLDEQGVVPLEAFLDRVSKVS GRLHVSLVVD LD SATA SUD EGERKF IREKGIKIYTMHEVDRLGMTRVMETIAVLSKSV DGVHLSLD UD DLD SATA SLD EGERKY IKESGMKVYTMHE LHLGMTKVIETIAVLKERT DGVHLSLD LD GLD SBAR DLD RGBRQ IKDENKYTMHE UNRLGMTKVIETIAVLKERT DGVHLSLD LD GLD SBAR DLD KGBRQ IKDENKYTMHE TINSDIDKLGIKEVIENTIEYLKSRNV DGVHLSLD V ALDPLET NAG TK
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_MOUSE ARG2_RATNO ARG2_SCHPO ARG1_LEIAM ARG1_AGRTU ARG1_ARATH ARG1_BACSU ARG1_STAAU ARG1_BACSU ARG1_CAEL ARG1_CAEL ARG1_COCIM	IEERRLAFICL SKSDIVYIGL SKSDIVYIGL ISSASIVYIGL SSPPNIVYIGL SSPPNIVYIGL SSPPNIVYIGL LSPPNIVYIGL IEERRLAFIGL LKPNKINYIGL PRNVSMMGI IKPENVUIGA IKPENVUIGA IKPENVUIGA IKPENVUIGA SPRKLIVIGA ISPRKLIVYIGL	ULD - PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOVACOPIVA DLD PABQFI LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOTAAPDALA SVD PABQAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLLHVSLVVPILESSIA DVU PPBHFI LKNYDIQYFSMRDIDRLGIQXVMERTTDLLIGKRQRPIHLSFOTAAPDALA DVU PPBHFI LKNYDIQYFSMRDIDRLGIQXVMEQTTDRLIGKRKRPIHLSFOTAAPDALA DVU PPBHFI LKSYDIQYFSMRDIDRLGIQXVMEQTTDRLIGKRKRPIHLSFOTAAPDALA DVU PPBHFI LKSYDIQYFSMRDIDRLGIQXVMEQTTDRLIGKRKRPIHLSFOTAAPDALA DVU PPBHFI LKSYDIQYFSMRDIDRLGIQXVMEQTTDRLIGKRKRPIHLSFOTAAPDALA DU PMGRAF LRERNIAAYTMHHVDKYGIGRVVEMAMEHINPGKR RPVHLSFOTAAPDALA DU PADQFI LKNYNISYISMRHIDCMGIRXVMEXTDQLLGRND RPIHLSFOTAAPDALA SVD PEBQFI LKNYNISYISMRHIDCMGIRXVMEXTDQLLGRND RPIHLSFOTAAPDALA SVD PEBRR VAEIGVQVADMRVLDEQGVVPLEAFLDRVSKVS GRLHVSLUVPLDBAIA SUD EGGRKF IREKGIKIYTMHEVDRLGMTKVMEETIAYLKERT DGVHLSLUVDLDBAAA SLD EGGRKF IREKGIKIYTMHEUDRLGMTKVIEETIAYLKERT DGVHLSLD LD GLDENDAR SLD EGGRKF IKESGMKVYTMHEIDRLGMTKVIETIAYLKERT DGVHLSLD LD GLDENDAR SUD FABRA LKESGMKVYTMHEIDRLGMTKVIETIAYLKERT DGVHLSLD LD GLDENDAR SVD PABRA IKESGMKVYTMHEIDRLGMTKVIETIAYLKERT DGVHLSLD LD GLDENDAR SVD PABRA IKESGMKVYTMHEIDRLGMTKVIETIAYLKERT CHVHLSLD VALDPLATK SVD FABRA VKKTEVIVDMRLIDEHGVALLRERTERVKAED CLLWSLV VDIDKAFR SVD FABREA VKKTEVIVDMRLIDELGVYLENTIEYLKSRNV DGVHLSLD VALDPLATK SVD SABREA VKKTEVIVDMRLIDELGVYLENTIEYLKSRNV DGVHLSLD VALDPLATK SVD SABREA VKKTEVIVDMRLIDELGVYLENTIEYLKSRNV DGVHLSLD VALDPLATK SVD SABREA VKKTEVIVDMRLIDELGVYLENTIEYLKSRNV DGVHLSLD VALDPLATK SVD SABREA VKKTEVIVDMRLIDELGVYLENTIEYLKSRNN SYLAAHLDL VLD KAFR SVD SABREA VKKTEVIVDMRLIDELGVYLENTIEYLKSRNN SYLAAHLDL VLD KAFR SVD SABREA VKKTEVIVDMRLIDELGVYLGVYLKESTNI NSVHLSFFOTOLOCUTUR VLD VDOX LD SYNA
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 AGRTU ARG1 BACSU ARG1 STAAU ARG1 STAAU ARG1 STAAU ARG1 STAAU ARG1 CAEL ARG1 COAC ARG1 COAC ARG1 COCIM	IEERRLAFICL SKSDIVYIGL SKSDIVYIGL ISSASIVYIGL ISSASIVYIGL SPPNIVYIGL SPPNIVYIGL LSPDIVYIGL IKPERLAFICL KKDIVYIGL PRNVSMKGI IKPENVVIG IKPENVVIGL IKPENVVIGL LIPHNLCMVGI LIPHNLCMVGI IRPEDVVIGL IRPEDVVIG IRPENVG	↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 AGRTU ARG1 ARATH ARG1 BACSU ARG1 STAAU ARG1 STAAU ARG1 CAEL ARG1 COAC ARG1 COCIM ARG1 COCIM ARG1 COCIM ARG1 CHENA ARG1 EMENI ARG1 CLYMA	IEERRLAFICL LSKSDIVYIGL SKSDIVYIGL ISSASIVYIGL LSPPNIVYIGL LSPPNIVYIGL LSPPNIVYIGL LSPPNIVYIGL PRNVSMMGI IKPENVVIG YARLLQVGI IKPENVVIGL LIPHNLCMVGI LIPHNLCMVGI ISPRKLVYIGL IRPEDVVMIGI IRPEDVVMIGI PHNVCMLGI LIPHNLCVVGI IRPEDVVMIGA ISPRKLVYIGL YARRLQVGI YARRLQVGYFGW	ULD - PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOVACOPIVAE DLD PABQFI LKNYDISYYSMRHIDCMGIKKVMEKTTDQLLGRRD RPIHLSFOIAAPDBALA SVD PABRA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLLHVSLUVPIESTA DVU PPHFFI LKNYDIQYFSMRDIDRLGIQKVMERTTDLIGKRQRPIHLSFOIAAPDBALA DVU PPHFFI LKNYDIQYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOIAAPDBALA DUU PPHFFI LKNYDIQYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOIAAPDBALA DUU PABRAF LRENNIAAYTMHHVDKYGIGRVVENAMEHINPGKR RPIHLSFOIAAPDBALA DUU PMBRAF LRENNIAAYTMHHVDKYGIGRVVENAMEHINPGKR RPIHLSFOIAAPDBALA DLD DMBRAF LRENNIAAYTMHHVDKYGIGRVVENAMEHINPGKR RPIHLSFOIAAPDBALA SUU PEBKFI LKNYNISYYSMRHIDCMGIRKVMEKTTDQLLGRRD RPIHLSFOIAAPDBALA SUD DEGKFI LKNYNISYYSMRHIDCMGIRKVMEKTTDQLLGRRD RPIHLSFOIAAPDBALA SUD DEGREG GKRFGVEGYIMMTFSKD RPHLENKLKGEGV LGUVISIDV O'LDBAFA SUD EGERKF IRESGIKIYTMHEVDRLGMTRVMEETIAYLKRRT DGVHLSLD UD DLBAFA SUD EGERKF IRESGIKIYTMHEIDRLGMTRVMEETIAYLKRRT DGVHLSLD UD GLDESDAF SLD EGERKF IKESGIKIYTMHEIDRLGMTRVIEETLAYLKSRV DGVHLSLD UG GLDPNDA SUD RGRQF IKDENNIKTFTMSDIDKLGIKEVIENTIEYLKSRNV DGVHLSLD UG GLDPNDA SUD SABQEL LEKLGVRIFYAHEUDRLGMTRVIESTAYLKSRNV DGVHLSLD UG ALDBLAF SVD PABREA VKKTEVIVVDRLLDEHGVALLRFRIERVKAED CLLWSLDV FIDSINA SVD PABREA VKKTEVIVVDRLLDEGVVENGGIQVVMQLAQYLVTRNT GVHLSLD UG DDFV SVD RABREA VKKTEVIVDRKLIDEGVENGGIGVVEMALAHIGQD TPIHLSFOID OFDVSYA SVD RABREA VKKTEVIVDRKLIDEGVENGGIGVVEMALAHIGQD TPIHLSFOV ALDBLAF SVD RABREA VKKTEVIVDRKLIDEGVENGGIGVVEMALAHIGQD TPIHLSFOV ALDBQWA SVD RABREA LRENGIKAFSMHDIDKYGIGRVVEMALAHIGQD TPIHLSFOV ALDBQWA SVD RABREL LRENGIKAFSMHDIDKGIGRVVEMALAHIGQD TP
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 SCHPO ARG1 SCHPO ARG1 AGRTU ARGI ARATH ARGI BACSU ARGI STAAU ARGI STAAU ARGI CAEEL ARGI COORC ARGI COCIM ARGI COCIM ARGI CLYA ARGI EMENI ARGI GLYMA ARGI ELPY ARGI HUMAN	IEERRLAFICL LSKSDIVYIGL SKSDIVYIGL ISSASIVYIGL LSPPNIVYIGL LSPPNIVYIGL LSPPNIVYIGL LSPPNIVYIGL PRNVSMMGI PRNVSMMGI IKPENVVIG IKPENVVIG IKPENVVIGL ILPHNLCMVGI LLPHNLCMVGI ISPRKLVYIGL IRPEDVVMIGI IRPEDVVMIGI IRPEDVVMIGI IRPEDVVMIGI IRPEDVVMIGI IRPEDVVMIGI IRPEDVVMIGI IRPEDVVMIGI 	DLD - PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOVACOPIVA DLD PABQFI LKNYDISYYSMRHIDCMGIKKVMEKTTDQLGRRD RPIHLSFOTAAPD3ALA SVD PABRA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLLHVSLUVPIESSIA DVD PPHFFI LKNYDIQYFSMRDIDRLGJQKVMEQTFDRLIGKRQ RPIHLSFOTAAPD3KLA DVE PPHFFI LKNYDIQYFSMRDIDRLGJQKVMEQTFDRLIGKRQ RPIHLSFOTAAPD3KLA DVE PPHFFI LKNYDIQYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOTAAPD3KLA DVE PABRAF LRENIAAYTMHHVDKYGIGKVMEQTFDRLIGKRC RPIHLSFOTAAPD3KLA DVE PABRAF LRENIAAYTMHHVDKYGIGKVVEMAMEHINPGKR RPVHLSFOTAAPD3KLA DUD DEBKKI LHDLNIAAYTMHHVDKYGIGKVVEMAMEHINPGKR RPVHLSFOTAAPD3KLA SUVE PABQFI LKNYNISYYSMRHIDCMGIRKVMEXTFDQLUGRRD RPIHLSFOTAAPD3KLA SUD DEGRAF LRENIAAYTMHHVDKYGIGKVVEMAITAYDKGTD RPVHLSFOTAAPD3ALA SUD DEGRAF LRENIAAYTMHHUKYGIGKVVEMAITAYDKGTD RPVHLSFOTAADJALUV DUD RAGGFI LKNYNISYYSMRHIDCHGIRKVMEXTTDQLUGRRD RPIHLSFOTAADJALUV SUD PESRR VAEIGVQVADMKVLDEQGVYRPLEAFLDRVSKVS CRLHVSLUV PILDSAIA SUD EGRAF IREKGIKIYTMHEVRLGMTKVIEETIAYLKERT DGVHLSLUV DID GLDSAAP SLD EGRAKY IREKGIKIYTMHEVRLGMTKVIEETIAYLKERT DGVHLSLUV DID GLDSAAP SLD EGRAKY IREKGIKIYTMHEVRLGMTKVIEETIAYLKERT DGVHLSLU DGUDSAAP SLD EGRAKY IREKGIKIYTMHEVRLGMTKVIEETIAYLKERT DGVHLSLU DGUDSAAP SLD EGRAKY IREKGIKIYTMHEVGRLGMTKVIEETIAYLKERT DGVHLSLU DGUDSAAP SVD PABREA VKKTEVIVVDNKLLDEHGVANLGRIKVENTITIYLKERSNV DGVHLSLU DGUDSAAP SVD RABREA VKKTEVIVVDNKLLOEGKVVENTITIYLKERSNV DGVHLSLU DGUDSAAP SVD RABREA VKKTEVIVVDNKLLOEGKVENTIETITIYLESSNV DGVHLSLU DGUDSAAP OYE SABQEL LEKLGVRIFYAHEVERGIQUVMQEAQYLVTRNT IGVULSU DGUDSYN SUD RABREA VKKTEVIVNDRLIDEGKGVENALAHIGQD TPIHLSFOV ALD GUDYS SVD RABREA LREHGIKAYTMKNDQQLGTNKIEKSINNINSVHLSFOTI CLDETVV DVD RABREL LREHGIKAYTMKNDQQLGTNKIEKSKNNL SKUVISFOTI DUOGNA SVD RABREL LREHGIKAYTMKNDQQLGTNKIEKSKNNL SVHLSFONAALDGWY SIN NEGRAQ AKKFGVEQFESMHPHFSKD RFFLENLNLGGGA TPIHLSFONAALDGWY SIN NEGRAP
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 SCHPO ARG1 SCHPO ARG1 AGRTU ARGT AGRTU ARGT BACHA ARGT BACHA ARGT BACHA ARGT CAEEL ARGI CLOAC ARGI CLOAC ARGI CCOIM ARGI CLOAC ARGI CLOAC	IEERRLAFICL LSKSDIVYICL ISSASIVYICL LSPPNIVYICL LSPPNIVYICL IEERRLAFICL LKPNKINYICL IKPENVVICN IKPENVVICN IKPENVVICN IKPENVVICN IKPENVVICN ILPHNLCMVCIC ISPRLVYICL ISPRLVYICL ISPRLVYICL ISPRLVYICL ISAKDIVYICL ISAKDIVYICL ISAKDIVYICL ISAKOVCICI	DLD PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOVACOPIVAE DLD PABQFI LKNYDIGYYSMRHIDCMGIKKVMEKTTDQLLGRRD RPIHLSFOIAAPDALM SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLLHVSLUVPILESSIA DVU PPHFFI LKNYDIGYFSMRDIDRLGIQKVMERTTDLLIGKRQ RPIHLSFOIAAPDALM DVE PPHFFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOIAAPDALM DVE PABRAF LRENTAAYTMHHDKYGIGRVVEQTFDRLIGKRQ RPIHLSFOIAAPDALM DVE PABRAF LRENNIAAYTMHHDKYGIGRVVEQTFDRLIGKRQ RPIHLSFOIAAPDALM DVE PABRAF LRENNIAAYTMHHDKYGIGRVVEMAMEHINPGKR RPVHLSFOIAAPDALAY NVD DEBKKI LHDLNIAAFSMHHVDRYGIGRVVEMAMEHINPGKR RPVHLSFOIAAPDALAY NVD DEBKKI LHDLNIAAFSMHHVDRYGIGRVVEMAINENTYDQLLGRRD RPIHLSFOIAAPDALAY SUD PABQFI LKNYNISYYSMRHIDCMGIRKVMEKTTDQLLGRRD RPHHLSFOIAAPDALAY SUD PEBRR VAEIGVQVADMRVLDEQGVYRPLEAFLDRVSKVS GRLHVSLUVPIDBAIA SUD EGERKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLUVDIDDAIA SLD EGERKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLUVDIDDAIA SLD EGERKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLUVDIDAAADDAX SLD EGERKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLUVDIDAAADDAX SLD EGERKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLUVDIDAAADDAX SLD EGERKF IREKGIKIYTMHEVDRLGUTVINTEITIETIASSRV DGVHLSLUVDAAADDAX SVD PABRAA VKKTEVIVVDMRLIDEHGVAALLRFFLERVKAED CLLHVSLUVDIDAAADDAX SVD RABRAA LREHGIKAFSMHDIDKYGIGVVEMALAHISSINI NSVHLSFOID LD LGOFDVSVA SID KGB LAL IYEKNITFYARHVERGIGVVEMALAHISGO TFIHLSFOID CLDEFIV SVD RABRAA LREHGIKAFSMHDIDKYGIGVVEMALAHIGQD TFIHLSFONAALDOGVAA SVD RABRAA LREHGIKAFSMHDIDKHGIGVVEMALAHIGQD TFIHLSFONAALDOGVAA SVD RABRAA LREHGIKAFSMHDURHFGIGVVEMALAHIGQD TFIHLSFONAALDOGVAA SVD RABRAA LREHGIKAFSMHDURHGIGVVEMALAHIGQD TFIHLSFONAALDOGVAA SVD RABRAA LREHGIKAFSMHDURHGIGVVEMALAHIGQD TFIHLSFONAALDOGVAA SVD RABRAA LREHGIKAFSMHDURHFGIGVVEMALAHIGQD TFIHLSFONAALDOGVAA SVD RABRAA LREHGIKAFSMTEVDRLGIGVVEMALAHIGQD TFIHLSFO
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 LEIAM ARG1 AGRTU ARGT AGRTU ARGT AGRTU ARGT BACSU ARGT BACSU ARGT CAEEL ARGT CAEEL ARGT CLOAC ARGT CCIM ARGT GLYMA ARGT EMENI ARGT HUMAN ARGT HUMAN ARGT HUMAN ARGT HUMAN	IEERRLAFICL LSKSDIVYICL ISASJVYICL LSPPNIVYICL LSPPNIVYICL LSPPNIVYICL LSPPNIVYICL IKPENVYICL IKPENVVICN IKPENVVICN IKPENVVICN IKPENVICN IKPENVICN INFRNIVICN SARLUVICH ISARLUVICL ISARLUVICL ISARLUVICL ISARLUVICL ISARLUVICL ISARLUVICL ISARCIVYICL ISARCIVICL SARCIVYICL SARCIVICL	ULD - PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOVACOPIVAE DLD PABQFI LKNYDIGYYSMRHIDCMGIKKVMEKTTDQLLGRRD RPIHLSFOIAAPD3ALA SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLLHVSLUVPILESSIA DVD PPHFFI LKNYDIGYFSMRDIDRLGIQKVMERTTDLLIGKRQ RPIHLSFOIAAPD3ALA DVE PPHFFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOIAAPD3KLAR DVE PABRAF LRENTAAYTMHHDKYGIGRVVEQTFDRLIGKRQ RPIHLSFOIAAPD3KLAR DVE PABRAF LRENNIAAYTMHHDKYGIGRVVEQTFDRLIGKRQ RPIHLSFOIAAPD3KLAR DVE PABRAF LRENNIAAYTMHHDKYGIGRVVENAMEHNPGKR RPVHLSFOIAAPD3KLAR DLD MBRAF LRENNIAAYTMHHDKYGIGRVVENAMEHNPGKR RPVHLSFOIAAPD3KLAR SVD PESKR VLRIGVQVADMRVLDEQGVVRPLEAFLDRVSKVS CRLHVSLUVPID9LIVN SVD PESRR VAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKVS CRLHVSLUVPID9LAFA SVD FESRR VAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKVS CRLHVSLUVPID9LAFA SUD EGRKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLUVD -LDBATAR SVD FASRA VAEIGVQVIDMRLDELGMTRVMEETIAYLKERT DGVHLSLUVD 2LDBATAR SUD EGRQF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLUVD 2LDBATAR SUD EGRQF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLUVD 2LDBATAR SUD SARQEL LKLGVRIFYNSDINKGIKKVVENTNIT YLKERSNV DGVHLSLUVD 2LDBATAR SVD FASRA VKKTEVIVVDMRLLDEHGVANLARFILERVKAED CLLHVSLUVD 2LDBATAR SVD FASRA VKKTEVIVVDMRLLDEHGVANLARFILERVKAED CLLHVSLUVD 2LDSIAR SVD RASRA LREHGIKAFSMHDINKGIGVVENALAHIGQD TFHLSFVOALD 2LDSIAR SVD RASRA LREHGIKAFSMHDINKGIGVVENALAHIGQD TFHLSFVOALD 2000 SID KGGRL LKELGVRIFYSKTVKRLGVEYILKEITEKLSKNNI NSVHLSFU ALDD 2UDSIAR SVD RASRAL LREHGIKAFSMHDINKGIGVVENALAHIGQD TFHLSFVOALD 2000 SID KGGRQ AKKFCVEQFEMHFSKD RFFLENLNLGGGA TFHLSFVOALD 2000 SID KGGRQ AKKFCVEQFEMHFSKD RFFLENLNLGGGA TFHLSFVVALD 2000 SID KGGRQ AKKFCVEQFEMHFSKD RFFLENLNLGGGA TFHLSFVVALD 2000 SID RGGRKR LREHGIKAFSMHDINKHGIGVVENALAHIGND TFHLSFVVALD 2000 SIT CGGRR LREHGIKAFSMHDINKHGIGVVENKALGHIGND TFHLSF
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 SCHPO ARG1 SCHPO ARG1 SCHPO ARG1 SCHPO ARG1 AGRTU ARGT AGRTU ARGT AGRTU ARGT BACCD ARG1 BACCD ARG1 CAEL ARGT CAEL ARGT CLYMA ARGT CLYMA	IEERRLAFICL LSKSDIVYICL ISRASIVYICL LSPPNIVYICL LSPPNIVYICL ISPNIVYICL ISPRLAFICL LSPPNIVYICL IKPENVYICL IKPENVVICV IKPENVVICV IKPENVVICV IKPENVVICV ISPRLVYICL ISPRLVYICL ISPRLVYICL ISPRLVYICL ISARDIVYICL ISA	DLD PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOVACOPIVAE DLD PABQFI LKNDDISYYSMRHIDCMGIKKVMEKTTDQLGRRD RPIHLSFOIAAPD3ALA SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLHVS LUVPJLESSIA DVU PPHFFI LKNVDIQYFSMRDIDRLGIQKVMERTTDLIGKRQ RPIHLSFOIAAPD3ALA DVE PDHFFI LKNVDIQYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOIAAPD3KLA DVE PABHFI LKNVDIQYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOIAAPD3KLA DLD DEBKKI LHDLNIAAFSMHHVDRYGIORVVEMAMEHTNPGKR RPVHLSFOVAACDPIVA NVD DEBKKF LRENNIAAYTMHHVDRYGIORVVEMALETIDRVSKVS GRLHVSLUVD FLDBAIA SVD PEBRR VAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKVS GRLHVSLUVD FLDBAIA SVD PEBRR VAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKVS GRLHVSLUVD FLDBAIA SUD EGRQF IKESGMKVYTMHEIDRLGMTRVMEETIAYLKERT DGVHLSLUVD FLDBAIA SLD EGRQF IKESGMKVYTMHEIDRLGMTRVMEETIAYLKERT DGVHLSLUVD GLDBAAA SLD EGRQF IKESGMKVYTMHEIDRLGMTRVMEETIAYLKERT DGVHLSLUD GLDBAAA SLD EGRQF IKESGMKVYTMHEIDRLGMTRVMEETIAYLKERT DGVHLSLUD GLDBAAA SVD RABRA VKKTEVIVVDMRLIDEHGVAALLRFFLERVKAED CLHVSLUVD JLDSLAA SVD RABRA VKKTEVIVVDMRLIDEHGVAALLRFFLERVKAED CLHVSLUVD JLDSLAA SVD RABRA IKESGIKAYTMKDVQLGTTRIHEETQERLMDV GVHLSLUD GLDBAAA SVD RABREA VKKTEVIVTMRHEIDRKGIGVVEMALAHIGQD TPHLS FVVAALDBQWA SVD RABREA LREHGIKAFSMHDIDKHGIGVVEMALAHIGQD TPHLS FVVAALDBQWA SVD RABREA LREHGIKAFSMHDIKHGIGVVEMALAHIGQD TPHLS FVVAALDBQWA SVD RABRE
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 SCHPO ARG1 SCHPO ARG1 SCHPO ARG1 SCHPO ARG1 AGRTU ARGT AGRTU ARGT AGRTU ARGT BACSU ARG1 BACSU ARG1 BACSU ARG1 CLYMA ARG1 CLYMA	IEERRLAFICL LSKSDIVYICL ISKSDIVYICL ISSASIVYICL LSPPNIVYICL ISPPNIVYICL ISPPNIVYICL IKPENVYICL IKPENVYICL IKPENVVICV IKPENVVICV IKPENVVICV IKPENVVICV IKPENVVICL ISPRKLVYICL ISPRKLVYICL ISPRKLVYICL ISAKDIVYICL	DLD PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOVACOPIVAE DLD PABQFI LKNYDISYYSMRHIDCMGIKKVMEKTTDQLGRRD RPIHLSFOTAAPDALA SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLHVS LUVP JLESSIA DVU PPHFFI LKNYDIGYFSMRDIDRLGIQKVMERTTDLIGKRQ RPIHLSFOTAAPDALA DVE PDHFFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOTAAPDALA DVE PABHFI LKNYNISYSMRHIDCKGIGRVVEMAMEHINPGKR RPIHLSFOTAAPDALA ND DEBKKI LHDLNIAAFSMHHVDRYGIDRVVSMAIEAV3PKGT SPVHLSFOVAACDPIVA NVD DEBKKF LRENNIAAYTMHHUDRYGIDRVVSMAIEAV3PKGT SPVHLSFOVAACDPIVA NVD PEBRR VAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKVS GRLHVSLOV PLDBAIA SUD EGERKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLOV PLDBAIA SUD EGERKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLOU OLD BADA SLD EGERKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLOU DGLDBADA SLD EGERKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLOU DGLDBADA SUD RABRA VKKTEVIVVDMRLIDEHGVALLRFFLERVKAED CULVSLOU DGLDBADA SVD RABREA VKKTEVIVVDMRLIDEHGVALLRFFLERVKAED CULVSLOV DJDSIA SLD EGERKF IREKGIKAYTYMHEVGRIGVVWEMALAHIGO TFHLSFOVALDBETV DVD RABREA VKKTEVIVDMRLIDEHGVALLRFFLERVKAED CULVSLOV DJDSIA SVD RABREA VKKTEVIVDMRLIDEGT VVEMALAHIGO TFHLSFOVALDBETV SVD PABREA VKKTEVINTFYSTKVRLGVAYILKEITEKLSRNN DGVHLSLU DVLDKAFR SVD ARBREL LREGIKAYTYKNDUQLGITRIHEETGERLNDV SRLHVSFOVALDBETV SVD RABREL LREGIKAYTYKNDUQLGGTNKHEGTSYLLGRKK RPIHLSFOVALDBETV SVD RGBKKI LREGIKAFSMHDINKGIGVVERMALAHIGD TFHLSFOVALDBETV SVD RGBKKI LKELGIKAYTYKNDUQLGGTNEHFSYLLGRKK RPIHLSFOVALDBETV SVD PGBHYI LKTLGIKYFSMTEVDKLGIGVVMEETFSYLLGRKK RPIHLSFOVALDBETTV DVD FGBKKI LKTLGIKYFSMTEVDKLGI
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 SCHPO ARG1 SCHPO ARG1 AGRTU ARGT AGRTU ARGT AGRTU ARGT BACCD ARG1 BACCD ARG1 CAEEL ARG1 CAEEL ARG1 COCIM ARGT COCIM ARGT CLYMA ARGT EMENI ARGT CLYMA ARG1 EMENI ARG1 CLYMA ARG1 CLYMA	IEERRLAFICL LSKSDIVYICL ISKSDIVYICL ISSASIVYICL LSPPNIVYICL ISPPNIVYICL ISPPNIVYICL ICERRLAFICL LSPPNIVYICL ICERRLAFICL VKPENVVICL IKPENVVICL IKPENVVICL IKPENVVICL IKPENVVICL ILPHNLCMVCIC ILPHNLCMVCIC ISPRLIVYICL ISPRLIVYICL ISAKDIVY	DLD PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOVACOPIVAE DLD PABQFI LKNYDISYSMRHIDCMGIKKVMEKTTDQLGRRD RPIHLSFOTAAPDALA SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLHVSLOVP JEBSIA DVD PPHFI LKNYDIGYFSMRDIDRLGIQKVMERTTDLIGKRQ RPIHLSFOTAAPDALA DVE PDHFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOTAAPDALA DVE PABHFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOTAAPDALA DVE PABHFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOTAAPDALA DVE PABHFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRQ RPIHLSFOTAAPDALA DVD DEBKAF LRENNIAAYTMHHDKYGIGRVVEMAMEHINPGKR RPVHLSFOTAAPDALA ND DEBKAF LRENNIAAYTMHHDKYGIGRVVEMAMEHINPGKR RPVHLSFOTAACDDIVA ND DEBKKF LRENNIAAYTMHHDKYGIGRVVEMATTDQLGRRD RPIHLSFOTAAPDALA SVD PEBRR VAEIGVQVADMRVLDEQGVVPLEAFLDRVSKVS GRLHVSLOV PLDBAIA SUD EGBKKF IREKGIKIYTMHEDRLGMTRVMEETIAYLKERT DGVHLSLOV PLDBAIA SUD EGBKKF IREKGIKIYTMHEDRLGMTRVMEETIAYLKERT DGVHLSLOV PLDBAIA SLD EGBRKF IREKGIKIYTMHEDRLGMTRVMEETIAYLKERT DGVHLSLOL DGLDBAAA SLD EGBRKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLOL DGLDBAAA SLD EGBRKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLOL DGLDBAAA SLD EGBRKF IREKGIKIYTMHEVDRLGMTRVMEETIAYLKERT DGVHLSLOL DGLDBAAA SVD PABREA VKKTEVIVVDMRLIDEHGVAALLRFILERVKAED CLLVS DV DVLDSIAA SVD PABREA VKKTEVIVVMRLIDEGGV KWIKESII
ARG1 SCHPO ARG1 ZENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARG1 SCHPO ARG1 AGRTU ARGI AGRTU ARGI AGRTU ARGI BACCD ARG1 BACCD ARG1 CAEEL ARG1 CAEEL ARG1 CLYA ARGI CLYA A	IEERRLAFICL LSKSDIVYICL ISKSDIVYICL ISSASIVYICL LSPPNIVYICL ISPPNIVYICL ISPPNIVYICL ISPRILAFICL LSPPNIVYICL -PRNVSMGI -YARRLLQVCI IKPENVVICQU IKPENVVICQU IKPENVVICQU IKPENVVICQU ILPHNLCMVCI VKPENVFICA -PHNVCMCI ILPHNLCMVGI ISPRLIVYICL ISAKDIVYIC	DLD PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFUVACOPIVAE DLD PABQFI LKNYDISYYSMRHIDCMGIKKVMEKTTDQLLGRRD RPIHLSFUTAFDALAS SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLHVSLUVP JLESSIA DVU PPHFIFI LKNYDISYYSMRHIDLGIGVVMERTTDLLIGKRQ RPIHLSFUTAFDALAS DVU PPHFIFI LKNYDISYSMRDIDRLGIQVVMERTTDLLIGKRQ RPIHLSFUTAFDALAS DVE PABHFI LKNYDISYSMRDIDRLGIQVVMEQTFDRLIGKRQ RPIHLSFUTAFDALAS DVE PABHFI LKNYDISYSMRDIDRLGIQVVMEQTFDRLIGKRQ RPIHLSFUTAFDA DVE PABHFI LKNYDISYSMRDIDRLGIQVVENAMETTDLLGKRQ RPIHLSFUTAFDA DVD PEBRAF LRENNIAAYTMHHDSKYGIGSVVEMAMEHNPGKR RPVHLSFUTAFDA ACDDY VA ACDDYNA ND DEEKKI LHDLNIAFSMHHDSKYGIGSVVEMALENYSSKS GRLHVSLUV STORA SUD PEBRR VAEIGVQVADMRVLDEQGVVEPLEAFLDRVSKVS GRLHVSLUV SUDA CDDY LD SIN QEGREQ GKRFGVEQYEMETFSKD RPMLENIKLGEGV KGVIISIDV SUDAFA SUD EGBKKF IREKGIKIYTMHEDRLGMTKVIEETIAYLKSRT DGVHLSLU SUD SUDAFA SUD EGBRKF IREKGIKIYTMHEDRLGMTKVIEETIAYLKSRT DGVHLSLU SUD SUDAFA SUD EGBRKF IREKGIKYTTMHEDRLGMTKVIEETIAYLKSRT DGVHLSLU SUD SUDAFA SUD RD FRA VKKTEVIVVMRLIDERLGMTKVIEETIAYLKSRT DGVHLSLU SUD SUDAFA SUD ND REA VKKTEVIVVMRLIDEHGVALLRFILERVKAED CLHVSLUV SUDAFA SVD NA BREA VKKTEVIVVMRLIDERLGMTKVIEETIAYLKSRT DGVHLSLU SUD SUDAF SVD NA BREA VKKTEVIVVMRLIDERLGMTKVIEETIAYLKSRT DGVHLSLU SUD SUDAF SVD RA BREA VKKTEVIVVMRLIDEKGGIQVVMEAQYLVTRNT IGYGI SUD SUDAF SVD RA BREA VKKTEVIVVMRLIDERLGRVKAED CLLVSLUV SUD SUDAFA SVD RA BREA VKKTEVIVDRLIDE SUS SUD SUDAFA SVD RA BREA VKKTEVIVDRLIDE SUS SUS SUD SUDAFA SVD RA BREA VKKTEVIVDRLIDE SUS SUS SUD SUDAFA SVD RA BREA VKKTEVIVDRLIDE SUS SUS SUDAFA SVD RA SKRL LERGIKAYTMKDVDLGIGTKIHEGTGERLMDV ERLHVSFDAALDSGWC SIN KEGEQ KKKFGVEGFENHFSKD RFFLEMLNGGG TFIHLSFVV SUD SUD SUS SUS SUDAFA SVD RO BRA LERGIKAYTMKDVDLGIGTKVHEETISYLLGRKK RPIHLSFVV SUD SUS SUS SUDAFA SVD PGBRDV IRELGIPISSMEDDKLGIGTKVMEETISYLLGRKK RPIHLSFVV SUD S
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 SCHPO ARG7 AINO ARG7 LEIAM ARG3 XENLA ARG3 XENLA ARG5 BACAU ARG1 BACAU ARG1 BACAU ARG1 BACAU ARG1 BACAU ARG1 BACAU ARG1 BACAU ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLOAC ARG1 CLUAC ARG1 CLUAC A	IEERRLAFICL LSKSDIVICL ISKSJVYICL LSPPNIVYICL LSPPNIVYICL ISPNIVYICL ISPNIVYICL ISPNIVYICL ISPRIVYICL -PRNVSMGI -YARRLLQVCI IKPENVVICV IKPENVVICV IKPENVVICV IKPENVVICV IKPENVVICV ISPRIVVICV ISPRIVVICV ISPRIVVICU ISPRIVVICU ISPRIVVICL SAKDIVYICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAKDIVICL ISAK	DLD PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFUVACOPIVAE DLD PABQFI LKNYDISYYSMRHIDCMGIKKVMEKTTDQLLGRRD RPIHLSFUTAFDALAS SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN GLLHVSLUVP JLESSIA DVU PPHFI LKNYDIGYFSMRDIDRLGIQKVMERTTDLLIGKRQ RPIHLSFUTAFDALAS DVE PABHFI LKNYDIGYFSMRDIDRLGIQKVMEQTTDRLIGKRQ RPIHLSFUTAFDALAS DVE PABHFI LKNYDIGYFSMRDIDRLGIQKVMEQTTDRLIGKRQ RPIHLSFUTAFDALAS DVE PABHFI LKNYDIGYFSMRDIDRLGIQKVMEQTTDRLIGKRQ RPIHLSFUTAFDA DVE PABHFI LKNYDIGYFSMRDIDRLGIQKVMEQTTDRLIGKRQ RPIHLSFUTAFDA DVD DEBKKI LHDLNIAAFSMHHVDRYGIORVVEMALENVSPKGT EVMVSYVV TIDLY DLD PABGFI LKNYNISYYSMRHIDCMGIRKVMEKTTDQLLGRRD RPIHLSFUTAFDA SVD PEBRR VAEIGVQVADMRVLDEQGVVPLEAFLDRVSKVS GRLHVSLVV DVD TLDAFDA SVD PEBRR VAEIGVQVADMRVLDEQGVVPLEAFLDRVSKVS GRLHVSLVV DVD TLDAFA SSIN QEGREQ GKRFGVEQYEMTFSKD RPMLENIKLGEGV KGVII SIDVO - LDAFAR SSID EGBKKF IREKGIKIYTMHEDRLGMTKVIEETIAYLKSRT DGVHLSLDL GLDSDA SLD EGBRKF IREKGIKYTTMHEDRLGMTKVIEETIAYLKSRT DGVHLSLDL GLDSDAA SLD EGBRKF IREKGIKYTTMHEDRLGMTKVIEETIAYLKSRT DGVHLSLDL GLDSDAA SLD EGBRKF IREKGIKYTTMHEDRLGMTKVIEETIAYLKSRT DGVHLSLD LG GLDSDAA SVD PABREA VKKTEVIVVDMRLIDEHGVAALLRFILERVKAED CLLHVSLDV -LDGVHLSLD LG GLDSDAA SVD RABREA VKKTEVIVVDMRLIDEHGVAALLRFILERVKAED CLLHVSLDV -DGVHLSLD LG GLDSDAA SVD RABREA VKKTEVIVDMRLIDENGGIGVVEMALAHIGGD TPHLSFV LD KGBLAL IYEKNLTFYSTKTVKLGVIYILKEITEKLSKNNIN NSVHLSFV LD KGBLAL IYEKNLTFYSTKVKLGVIYILKGIGVVEMALAHIGGD TPHLSFV LD ROBLKK LREGIKAYTYKDVDLGIGVKMEATTSYLLGRKK RPHLSFV LD CGBKKI LKELGIKAYTYKDVDLGIGVVEMALAHIGGD TPHLSFV LD CGBKKI LKELGIKAYTYKDVDLGIGVVEETTSYLLGRKK RPHLSFV VLDGVAA STE QSBRDV IRELGIPLSANDDVRLGIGVVEETTSYLLGRKK RPHLSFV VLDGUDAFTT DVD FGBHYI IKTLGIKYFSMTEUDKLGIGVVEETTSYLLGRKK RPHLSFV VLDGVAA STE QSBRDV IRELGIPVEARAMDDK WFWEALAHIGHIGND TPHLSFV VLDGUDAFTT DVD FGBHYI IKTLGIKYSMSDDVRLT
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 SCHPO ARG7 AINO ARG7 XENLA ARG3 XENLA ARG7 BACSU ARG7 BACSU ARG7 BACSU ARG7 BACSU ARG7 BACSU ARG7 BACSU ARG7 CASEL ARG7 COLIM ARG7 STRPY ARG7 SYNEC ARG7 SYNEC ARG7 SYNEC ARG7 SYNEC ARG7 SYNEC ARG7 SYNEC	IEERRLAFICL LSKSDIVYICL LSKSDIVYICL LSPPNIVYICL LSPPNIVYICL LSPPNIVYICL LSPPNIVYICL LSPPNIVYICL LSPNIVYICL CONTRACTOR CONTRACTOR LKPNKIAYICL LKPNKIAYICL LKPNKIAYICL LSKSDIVYICL LLPHNLCWVICL LLPHNLCWVICL LLPHNLCWVICL LSFRLUVICL ISPRCLYFCL ISAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKLIV	ULD - PM®RAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHL DLD PABQFI LKNYDISYYSMRHIDCMGIKKVMEKTPDQLGRRD RPIHL SVD PFABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN CLLHVSLDVPIEBSIA SVD PFABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN CLLHVSLDVPIEBSIA DVU PFBHFI LKNYDIQYFSMREIDRLGIQXVMERTTDLIGKRK RPIHLSFDIAAFDBALA DVU PFBHFI LKNYDIQYFSMREIDRLGIQXVMEQTTDRLIGKRK RPIHLSFDIAAFDBALA DVU PFBHFI LKNYDIQYFSMREIDRLGIQXVMEQTTDRLIGKRK RPIHLSFDIAAFDBALA DUD PM®RAF LRERNIAAYTMHHVDKYGIGRVVEMAMEHINPGKR RPVHLSFDV ACDDIVAR AVD DEGKKI LHDLNIAAFSMHHVDRYGIDKVVSMAIEAVSPKGT EPVMVSYV VTIDPLYV DLD PABQFI LKNYNISYYSMRHIDCMGIKKVMEKTYDQLLGRRD RPVHLSFDV ACDDIVAR SVD PEBRRR VAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKVS GRLHVSLUV FLDPALAF SLD EGKKF IKESGMKVYTMHEIDRLGMTKVMESTIAYLKERT DGVHLSLUV FLDPALAF SLD EGFKKF IKESGMKVYTMHEIDRLGMTKVMESTIAYLKERT DGVHLSLUV FLDPALAF SLD EGFKKF IKESGMKVYTMHEIDRLGMTKVMESTIAYLKERT DGVHLSLUV FLDPALAF SLD EGFKKF IKESGMKVYTMHEIDRLGMTKVMESTIAYLKERT DGVHLSLUV FLDPALAF SLD KOBRQF IKDHNIKTFTMSDIKLGIKEVIENTIEYLKSRNV DGVHLSLUV FLDPALAF SLD SGFKY IKESGMKVYTMHEIDRUGMTKVMESTIAYLKERT WSULSID VUD VLDV SLD KOBRQF IKDHNIKTFTYSDIKLGIKKVIENTIEYLKSRNV
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 SCHPO ARGT XENLA ARG1 XENLA ARG1 AGRTU ARG1 BACAU ARGT BACAU ARGT BACAU ARGT BACAU ARGT BACAU ARGT BACAU ARGT BACAU ARGT COLIM ARGT COLIM ARGT COLIM ARGT COLIM ARGT COLIM ARGT COLIM ARGT COLIM ARGT COLIM ARGT COLIM ARGT MOUSE ARGT RANCA ARGT SCHPO ARGT SCHPO	IEERRLAFICL LSKSDIVYICL LSKSDIVYICL LSPPNIVYICL LSPPNIVYICL LSPPNIVYICL LSPPNIVYICL LSPNIVYICL LSPNIVYICL PRNVSMGI IKPENVULCV IKPENVULCV IKPENVULCV IKPENVULCA IKPENVULCA IKPENVULCA IKPENVULCA IKPENVULCA ISPRLVVICL LPHNLCAVOICA ISPRLVVICL ISAKDIVYICL ISAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL LSAKDIVYICL IDPRSVCICA 	DLD - PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFUVACD PIVA DLD PABQFI LKNYDISYYSMRHIDCMGIKKVMEKTPDQLGRRD RPIHLSFUTAFDALA SVD PFABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN CLLHVSLUVPILESSIA DVU PFBHFI LKNYDIQYFSMRDIDRLGIQKVMERTTDLIGKRC RPIHLSFUTAFDALA DVU PFBHFI LKNYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRC RPIHLSFUTAFDALA DVU PFBHFI LKNYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRC RPIHLSFUTAFDALA DVU PFBHFI LKNYDIQYFSMRDIDRLGIQKVMEQTTDRLIGKRC RPIHLSFUTAFDALA DLD PMGRAF LRERNIAAYTMHHVDKYGIGRVVEMAMEHINPGKR RPVHLSFUTAFDALA DLD PABQFI LKNYNISYYSMRHIDCMGINKVVSMAIEAVSPKGT RPVHLSFU VACDD IVA AVD DESKKI LHDLNIAAFSMHHVDRYGIDKVVSMAIEAVSPKGT RPVHLSFUTAFDALA SUD PEBRR VAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKVS GRLHVSLUV FLDFALA SUD EGSKKF IKESGMKVYTMHEIDRGIKKVMEKTIAQLGGV KGVVISIDV CLDFAFA SLD EGSKKF IKESGMKVYTMHEIDRLGMTKVMESTIAYLKERT DGVHLSLU VG LDFAFA SLD EGSKKF IKESGMKVYTMHEIDRLGMTKVMESTIAYLKERT DGVHLSLU VG LDFAFA SLD EGSKKF IKESGMKVYTMHEIDRLGMTKVMESTIAYLKERT DGVHLSLU VG LDFAFA SLD KOBRGF IKDHNIKTFTMSDIDKLGIKEVIENTIEYLKSRNV - DGVHLSLU VG LDFAFA SLD KOBRGF IKDHNIKTFTMSDIDKLGIKEVIENTIEYLKSRNV DGVHLSLU VG LDFAFA SLD KOBRGF IKDHNIKTFTYSBIDKLGIKKVIENTIEYLKSRNV NSVLESFU DVD -ASBCKL LEGGIKAFSMHDUKGIGIGVVEMALARFLERVKAED CLLHVSLU VG PLDFSIA DYE SABQEL LEKGIKAFSMHDIKKGIGUVMQEAQYLVTRNT NSVLESFU DVD RGBKL LREAGIKAFSMHDIKKGIGUVMERTALHESTGENDV ERLHVSFDALDGGVA STD KGBRCH IKEGIKAFSMHDIKKGIGUVMERTALHIGGD TPIHLSFDV ALDGGVA STD RGBKL LREAGIKAFSMHDIKKGIGUVMERTALHIGGD FIHLSFDV ALDGGVA STD RGBKL LREAGIKAFSMHDIKKGIGUVMERTSYLLGRKKRPIHLSFDV ALDGGVA STD RGBKL LREAGIKAFSMHDIKKGIGUVMERTSYLLGRKKRPIHLSFDV ALDGGVA STD RGBKK KKFGVEGUERTFSKLGUKKGIGUVMERTSYLLGRKKRPIHLSFDV ALDGGVA STD RGBKK KKFGVEGUERTFSKLGUKKGIGUVMERTSYLLGRKKRPIHLSFDV ALDGGVA STD RGBKKF IKLSFFTVLSAFKVMERTEDFLKARALGHTSYLLGRKKRPIHLSFDV ALDGGVA STD RGBKKF KKFGV
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE ARG2 RATNO ARG2 SCHPO ARGI LIAM ARG1 AGRTU ARGI AGRTU ARGI BACSU ARGI BACSU ARGI BACSU ARGI BACAN ARGI BACAN ARGI DACHA ARGI COCIM ARGI COCIM ARGI CLOAC ARGI CLOAC ARGI CLOAC ARGI CLOAC ARGI CLOAC ARGI CLYMA ARGI HELPY ARGI HELPY ARGI HELPY ARGI HELPY ARGI HELPY ARGI RHOUSE ARGI RHOCA ARGI RHOCA ARGI RHOCA ARGI RHOCA ARGI SCHPO ARGI SCHPO	IEERRLAFICL LSKSDIVYICL LSKSDIVYICL LSPPNIVYICL LSPPNIVYICL LSPPNIVYICL LSPPNIVYICL LKPNKIAYICL LKPNKIAYICL LKPNKIAYICL SKSDIVYICL -PRNVSMGI IKPENVVICA IKPENVVICA IKPENVVICA IKPENVVICA ISPRLVYICL -PHNVCMLCI VKENVYICL SKSDIVYICL ISAKDIVYICL ISAKDIVYICL ISAKDIVYICL ISAKDIVYICL ISAKDIVYICL ISAKDIVYICL LSKKIVYICL ISAKDIVYICL ISAKDIVYICL SKLFKYVCL PRRLCMCI INNGTSIAACD PRNVSVCICI MRVANFCV MRVANFCV MRVANFCV MRVANFCV KNIVCICI KNIVCICI	DLD - PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHLSFOUA CD PIVA DLD PABQFI LKNYDIGYYSMRHIDCMGIKKVMEKTFDQLLGRRD RPIHLSFOUA FDAFD SVD PFBHFI LKNYDIGYFSMRDIDRLGIQKVMEKTFDQLLGRRD RPIHLSFOUA FDAFD SVD PFBHFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRK RPIHLSFOUA FDAFD SVD PFBHFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRK RPIHLSFOUA FD SVD PFBHFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRK RPIHLSFOUA CD SVD PFBHFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRK RPIHLSFOUA CD SVD PFBHFI LKNYDIGYFSMRDIDRLGIQKVMEQTFDRLIGKRK RPIHLSFOUA CD SVD PFBHFI LKNYNISYYSMRHIDCMGIKKVMEXTFDQLLGRRD RPIHLSFOUA CD SVD PFBHFI LKNYNISYYSMRHIDCMGIKKVMEXTFDQLLGRRD RPIHLSFOUA CD SVD PEBQFI LKNYNISYYSMRHIDCMGIKKVMEXTFDQLLGRRD RPIHLSFOUA CD SVD PEBRRR VAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKYS CRLHVSLUV SVD SD SUD SGRRG IKBSKNYYTHHEIDRLGMTKVMESTIAYLKSRT DGVHLSLD UG DS SAF SLD SGGRKF IKBSKNYYTHHEIDRLGMTKVMESTIAYLKSRT DGVHLSLD UG DS SAF SLD SGGRKF IKBSKNYYTHHEIDRLGMTKVMESTIAYLKSRT DGVHLSLD UG DS SAF SLD SGGRKF IKBSKNYTHHEIDRLGMTKVMESTIAYLKSRT DGVHLSLD UG DS SAF DLD KGGRGF IKBSKNYTHHEIDRLGMTKVMESTIAYLKSRT DGVHLSLD UG DS SAF SLD SGGRKF IKBSKNYTHHEIDRLGMTKVMESTIAYLKSRT DGVHLSLD UG DS SAF DS SGFRKF IKBSKNYTHHEIDRLGMTKVMESTIAYLKSRT DGVHLSLD UG DS SAF DS SAFQEL LEKGVKIYTHHEIDRLGMTKVESTI STLYLKSRT SYLAIHLD UV DF KAFF SVD PABREA VKKTEVIYDMRLDEGTA KWIKSSII
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 MOUSE AGG2 SCHPO ARGI LIAM ARG1 SCHPO ARGI LIAM ARG1 BACSU ARGI BACSU ARGI BACSU ARGI BACSU ARGI BACSU ARGI CLOAC ARGI CLOAC AR	IEERRLAFICL SKSDIVYICL SKSDIVYICL SSPSIVYICL SSPSIVYICL SSPPNIVYICL IEERRLAFICL LSPPNIVYICL IEERRLAFICL LKPNKIAYICL SSSDIVYICL SKSDIVYICL SKSDIVYICL SKSDIVYICL IKPENVVICA IKPENVVICA IKPENVVICA ISPRLVVICA SSFRLVYICL SKSDIVYICL SAKDIVYICL SAKDIVYICL SAKDIVYICL SAKDIVYICL SSKKIVYICL SSKKIVYICL SKLFKYVCC PRNCMCT ISAKDIVYICL SSKKIVYICL SSKLVYICL	DLD - PMBRAF LRERSITAYTMHDVDKYGIARVVEMALEHINPGRRRPIHL DLD PABQFI LKNYDIGYYSMRHIDCMGIKKVMEKTFDQLLGRRD RPIHL SVD PABRAA LEDSGITVHDMRSIDEHGVAVLLRAFLARVQAAN CLLHVSLDVP LEDSIA DVD PFBHFI LKNYDIGYFSMREIDRLGIGXVMERTFDLLIGKRC RPIHLSFDIAAFD3LAS DVE PFBHFI LKNYDIGYFSMREIDRLGIGXVMEQTFDRLIGKRC RPIHLSFDIAAFD3KLA DVD PFBHFI LKNYDIGYFSMREIDRLGIGXVMEQTFDRLIGKRC RPIHLSFDIAAFD3KLA DVD PFBHFI LKNYDIGYFSMREIDRLGIGXVMEQTFDRLIGKRC RPIHLSFDIAAFD3KLA DVD PFBHFI LKNYDIGYFSMREIDRLGIGXVMEQTFDRLIGKRC RPVHLSFDIAAFD3KLA DVD PABRF LRERNIAAYTMHHVDKYGIGVVVSMAIEAVSPKGT RPVHLSFDIAAFD3ALA DLD PAGQFI LKNYNISYISMRHIDCMGIKKVMEATFDQLLGRRD RPVHLSFDIAAFD3ALA SVD PEBRR VAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKYS GRLHVSLUV PLD3AIA SVD PEBRR VAEIGVQVADMRVLDEQGVVRPLEAFLDRVSKYS GRLHVSLUV PLD3AIA SUD EGRKF IRESGKIYTMHEDVRLGMTKVMEETIAYLKSRT DGVHLSLU DGLD SDA SLD EGRKF IRESGKIYTMHEDVRLGMTKVMEETIAYLKSRT DGVHLSLU DGLD SDA SLD EGRKF IRESGKIYTMHETTSISI KKWIKESII KYLAIHLDLV DV ALDD LST TAG TKG
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_SCHPO ARG2_SCHPO ARG1_LIAM ARG3_XENLA ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_RHOUSE ARG1_RHOUSE ARG1_RHOUSE ARG1_RHOCA ARG1_RHOC	IEERRLAFICL SKSDIYYICL SKSDIYYICL SSPSIYYICL SSPSIYYICL SSPPNIYYICL IEERRLAFICL LSPPNIYYICL IEERRLAFICL LSPDIYYICL SSPSIYYICL SSPSIYYICL SKSDIYYICL SKSDIYYICL IKPENVYICL IKPENVYICL IFPENVYICL IFPENVYICL ISPRKLYYICL ISPKLYYICL ISAKDIYYICL ISAKONYICL ISAKONYICL ISAKDIYYICL ISAKONY	DLD - PMERAF - LRERSTRAYTMHDVDKYGIARVVEMALEHINPGRR- RPIHLSTVAC DPIVAC DLVAE DLD - PAEQPI - LRENSTRAYTMHDVDKYGIARVVEMALEHINPGRR- RPIHLSTIAAD AC DPIVAE DVD - PAERA - LEDSGTVHDMRSIDEHGVAVLLRAFLARVQAAN - GLHVG V 97 LEBSIA DVD - PEBHFI - LKNYDIQYFSMREIDELGIQXVMERTFDLIGRRQ - RPIHLSTIAAD PKLAR DVD - PEBHFI - LKNYDIQYFSMREIDELGIQXVMEGTFDRLIGRRQ - RPIHLSTIAAD PKLAR DVE - PABHFI - LKNYDIQYFSMREIDELGIQXVMEQTFDRLIGRRA - RPIHLSTIAAD PKLAR DVE - PABHFI - LKNYDIQYFSMREIDELGIQXVMEQTFDRLIGRRA - RPIHLSTIAAD PKLAR DLD - PMERAF - LRERNIAAYTMHHVDKYGIDRVVSMAIEAVSPKCT - EPWWS UV 00 AC DPIVAE DVD - PEBKR - VAEIGVQVADMRVDEQGVVPLEAFLDRVSKVS - GKLHVGU VVDIDDEVLA AVD - DEBKKI - LHDLNIAAFSMHHVDRYGIDRVVSMAIEAVSPKCT - EPWWS UV 07 UD 10 LVA AVD - DEBKKI - LHENNIAYSMHHUCMGIRXVMETTADULKLGGRD - RPIHLSTIDAFDEXLAR SUD - EGERF - IREKGINYTMHEVDRLIGMTXVMETTADULKLGGRD - ROWHSDV 00 UP 10 DLVA SUD - EGERF - IREKGINYTMHEVDRLIGMTXVMETTADULKLGGRD - ROWHSDV 00 DEDAFA SLD - EGERF - IREKGINYTMHEVDRLIGMTXVMETALVLKERT - DGYHLSDIDGDENAR SUD - EGERF - IREKGINYTMHEUDRLIGMTXVMETALVLKERT - DGYHLSDIDGDENAR DD - KGERQF - IKDHNIKTFTMSDIDKLGIKEVIENTIEYLKSRNV - DGYHLSDIDGDENAR DD - KGERQF - VKKTEVIVVDRLIDEHGVAALLRFPLENVKAED - GLHVGIV 07 LD 20 DENAR DD - SAQQL - LKKGVNIFYTHEUDRLGUTXVIEKSII - NSVL SU - NGUHLSDIDGDENAR DD - SAQQL - LKKGVNIFYTHEURGIDGVAALLRFPLENVKAED - GLHVGIV 07 LD 20 DENAR DD - SAQQL - LKKGVNIFYTHEURGIGUVAALLRFPLENVKAED - GLHVGIV 07 LD 20 DV SIA DD - SAQQL - LKKGVNIFYTHEURGIGUVAALLRFPLENVKAED - GLHVGIV 07 LD 20 DV SIA DD - SAQQL - LKKGVNIFYTHEURGIGUVANALAHIGOD - TPIHLSDIDCD 20 SIA DD - SAQQL - LKKGVNIFYTHEURGIGUVANALAHIGUP - TPIHLSDIV AD AD DG QCC DVD - RAGKKL - LREHGIKAFSMHDVRHGIGRVEMALAHIGUP - TPIHLSDIV AD AD DG QCC DVD - RGEKKL - LREHGIKAFSMHDVNGIGITHHEGQENVKKSLAV - DIIYLG DL DTAA SID - KGERQ - KKNFGVEQFEMRHFSND - SFLUNKKSLKAV - DIIYLG DI CD 20 SIA DVD - PGEHYI - LKTLGIKYSMTEVDRLGICKVMESTFSYLLGRKK - RPIHLSDIVGID DQ SAA SID - KGERKF - ALENGIKAFSMHDVRHGIGRVMEATFSYLLGRKK - RPIHLSDIVGID DQ SAA SID - KEDFKR - AKNGYMSYNDYNDYNTY DVENKLGICKVMESTFSYLLGRKK - RPIH
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 SCHPO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 CAEEL ARG1 COCIM ARG1 CAEEL ARG1 COCIM ARG1 CAEEL ARG1 CAEEL	IEERRLAFICL SKSDIVYIGL SKSDIVYIGL SKSDIVYIGL SSPSNIVYIGL SSPSNIVYIGL SSPSNIVYIGL SSPSNIVYIGL SSPSNIVYIGL SSPSNIVYIGL SSPSNIVYIGL SKSDIVYIGL SKSDIVYIGL IKPENVUIGA IKPENVUIGA ISPRKLYYIGL IFREDVVIGA ISPRKLYYIGL IFREDVVIGA ISPKLYYIGL SAKDIVYIGL ISAKDIVYIGL SAKDIVYIGL SAKDIVYIGL SKSDIVYIGL SKSKIVYIGL SKKLYYIGL SKLFKYFVGI PRRCAMPION SKLFKYFVGI PRNCAMPION SKLFKYFVGI FNRCAMPION SKLFKYFVGI FNRCAMPION SKLFYFVGI FNRCAMPION SKSDIVYIGL ISPKKIAYIGL SKLFYFVGI FNRCAMPION SKSDIVYIGL SKSSSS SKSSSS SKSSS SKSSSSS SKSSSSS SKSSSSS SKSSSS SKSSSSSS SKSSSSS SKSSSS SKSSSSS S	DLD - PMERAF - LRERSTAYTMHDVDKYGIARVVEMALEHINPGRR - RPHLSTIA GEIVAE DLD - PARQFI - LRENSTAYTMHDVDKYGIARVVEMALEHINPGRR - RPHLSTIA GA DEIVAE DLD - PARGA - LEDSGTVHDARSIDEHGVAVLLRAFLARVQAAN - GLLHVGL VOFILESIA SVD - PPEHFI - LKNYDIQYFSMRE IDRLGIQXVMEGTFDLIGKQ - RPHLSTIA FDTAA DVE - PAGHFI - LKNYDIQYFSMRE IDRLGIQXVMEQTFDRLIGKRQ - RPHLSTIA FDTAA DVE - PAGHFI - LKNYDIQYFSMRE IDRLGIQXVMEQTFDRLIGKRQ - RPHLSTIA FDTAA DVE - PAGHFI - LKNYDIQYFSMRE IDRLGIQXVMEQTFDRLIGKRQ - RPHLSTIA FDTAA DVE - PAGHFI - LKRYDIQYFSMRE IDRLGIQXVMEQTFDRLIGKRX - RPHLSTIA FDTAA DVE - PAGKIT - LKRYNIAYYMH HVDKYGICRVVEMAMEHNPGKR - RPHLSTIA FDTAA DLD - PAGQFI - LKNYNISYSMRH IDRGYGIDXVYMAIEA/SPGT - EFYMVSI VV TID LYW DLD - PAGQFI - LKNYNISYSMRH HVDKYGICRVVEMAMEHNPGKR - RPHLSTIA VV GTD LYW DLD - PAGQFI - LKNYNISYSMRH HUCKGIRVMEKTFDQLLGRRD - RPHLSTIA SIN - QEGREQ - GKRFGVEQYEMRTFSKD - RPHLSNIKKSTO - GKLHVSI VV GTD ZAFA SLD - EGEKKF - IREKGIKIYTHHE UDRLGMTKVMEETIALLKERT - DGVHLSI VV GTD ZAFA SLD - EGEKKF - IREKGIKIYTHHE UDRLGMTKVMEETIALLKERT - DGVHLSI UV GLD ELSTA SLD - KGERGY - IKDHNIKTFTMSD IDRLGIKKVIENTIETIAVLSKISI - KVILAIH DUD - KGERGY - IKDHNIKTFTMSD IDRLGIKKVIENTIETIAVLSKISI - KVILAIH DUD - KGERGY - IKDHNIKTFTMSD IDRLGIKKVIENTIETIAVLSKISI - KVILAIH DUD - KGERGY - IKDHNIKTFTMSD IDRLGIKKVIENTIETIAVLSAC DGVHLSI DLG DESAFA SUD - EGEKKF - IREKGIKAYTMKUVGYGIGQVVEMALAHIGOD TFHLSI VOALDELSTS SUD - PAGRA - VKKTEVIVYDARLIDEHGVAALLRFSLERVKAED GKUHLSI DLG DEVSYA SUD - RAGKKL - LREHGIKAFSMHD UXGIGQVVEMALAHIGOD TFHLSI VOALDEWAFA SUD - PAGKKL - INEHGIKAFSMHD VRGIGRVVEMALAHIGOD TFHLSI VOALDEWAFA SUD - RG KKL - LREHGIKAFSMHD VRGIGRVVEMALAHIGOD TFHLSI VOALDEWAFA SUD - RG KKL - LREHGIKAFSMHD VRGIGRVVEMALAHIGOD TFHLSI VOALDEWAFA SUD - RG KKL - LREHGIKAFSMHD VRGIGRVVEMALAHIGON TFHLSI VOALDEWAFA SUD - RG KKL - LREHGIKAFSMHD VRGIGRVVEMALAHIGND TFHLSI VOALDEWAFA SUD - RG KKL - LREHGIKAFSMHD VRGIGRVVEMALAHIGND TFHLSI VOALDEWAFA SUD - RG KKL - LREHGIKAFSMHD VRGIGRVVEMALAHIGND TFHLSI VOALDEWAFA SUD - PG HYI - LKLGIYSMEVDRLGIGVVEN
ARG1_SCHPO ARG1_XENLA ARG2_AGRTU ARG2_HUMAN ARG2_SCHPO ARG2_SCHPO ARG1_EIAM ARG3_XENLA ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_BACSU ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAEEL ARG1_CAESU ARG1_RANCA ARG1_YENLA A	IEERRLAFICL LSKSDIYYIGL SKSDIYYIGL SKSDIYYIGL SSPNIYYIGL SPPNIYYIGL LSPPNIYYIGL LSPPNIYYIGL LSPNILYYIGL SPRILYYIGL SRSDIYYIGL SKSNIYYIGL SKSNIYYIGL S	DLD PMSRAFLRERSITAYTMHUVDKYGIARVVEMALEHINPGRRRPIHLSFUTACOFIVA DLD PASCFILKNYDISYYMRKIDCMGIKKVMEKTFDQLLGRRD RPIHLSFUTAFDQALA SVD PASCFILKNYDIGYYSMRKIDLGGICVKMETFDLLIGKQRPIHLSFUTAFDQALA DVC PASTFILKNYDIGYYSMRKIDLGGICVKMETFDLLIGKQRPIHLSFUTAFDQKLA DVE PASTFILKNYDIGYYSMRKIDLGGICVKMEGTFDRLIGKKQRPIHLSFUTAFDQKLA DVE PASTFILKNYDIGYYSMRKIDLGGICVKMEGTFDRLIGKKQRPIHLSFUTAFDQKLA DVE PASTFILKNYDIGYYSMRKIDLGGICVKMEGTFDRLIGKKKRPIHLSFUTAFDQKLA DVE PASTFILKNYNISYSMRKIDCKGIRVKMEGTFDRLIGKKKRPIHLSFUTAFDQKLA DVE PASTFILKNYNISYSMRKIDCKGIRVKMENTFDQLLGKRD PUHLSFUTAFDQKLA DVE PASGFILKNYNISYSMRKIDCKGIRVKMENTFDQLLGKRD PUHLSFUTAFDQLA DVE PASGFILKNYNISYSMRKIDCKGIRVKMENTFDQLLGKRD PUHLSFUTAFDQLA DVE PASGFI
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 SCHPO ARG2 SCHPO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARGT BACSU ARGT BACSU ARGT BACSU ARGT BACSU ARGT CAEEL ARGT CAEEL ARGT COCIM ARGT CAEEL ARGT COCIM ARGT CAEEL ARGT COCIM ARGT CAEEL ARGT CAEEL ARGT CAEEL ARGT CAEEL ARGT CAEEL ARGT CAEEL ARGT CAEEL ARGT CAEEL ARGT CAEEL ARGT RANCA ARGT RANCA ARGT RANCA ARGT SCHPO ARGT SYNEC ARGT YENLA ARGT SCHPO ARGT SYNEC ARGT YENLA ARGT SYNEC ARGT YENLA ARGT SYNEC ARGT YENLA ARGT SYNEC ARGT YENLA ARGT SYNEC ARGT YENLA ARGT SYNEC ARGT YENLA ARGT YENLA	IEERRLAFICL LSKSDIYJIGL EENIGMIGI ISSASIYYIGL LSPPNIYYIGL LSPPNIYYIGL LSPPNIYYIGL LSPPNIYYIGL PRNVSMMGI IKPENVUJIGY IKPENVULIGY IKPENVULIGY IKPENVULIGY IKPENVULIGY IKPENVULIGY IKPENVULIGY IKPENVULIGY IKPENVULIGY IKPENVULIGY ISAKDIYYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKADIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL ISAKDIYIGL	DLD PMERAFLRERSITAYTMHUVDKYGIARVVEMALEHINPORRRPIHLSFUTACOEIVAL DLD PADRAA LEDSGITVHDMRSIDEHGVAVLEARFLARVQAAN CHUSSFUTAFDEALA SVD PADRAA LEDSGITVHDMRSIDEHGVAVLEARFLARVQAAN CHUSSFUTAFDEALA SVD PADRAA LEDSGITVHDMRSIDEHGVAVLEARFLARVQAAN CHUSSFUTAFDEALA DV PADRAF LRENNIDGY SMREIDLGIQVNKGTFPOLIGKRQ RPIHLSFIDAFDEALA DV PADRAF LRENNIDGY SMREIDLGIQVNKGTFPOLIGKRQ RPIHLSFIDAFDEXLA DV PADRAF LRENNIAYTMHUVDKYGIQVVENAMEGITPOLIGKRQ RPIHLSFIDAFDEXLA DV PADRAF LRENNIAYTMHUVDKYGIQRVVENAMEHINPOKR RPVHSY DV TIDZIVY DLD PAD QFI LKNYNISYYSMRHUDRYGIQRVVENAMEHINPOKR RPVHSY DV TIDZIVY DLD PAD QFI LKNYNISYYSMRHUDRYGIQRVVENAMEHINPOKR RPVHSY DV TIDZIVY DLD PAD QFI LKNYNISYYSMRHUDRYGIGRVVENAMEHINPOKR RPVHSY DV TIDZIVY DLD PAD QFI LKNYNISYYSMRHUDRYGIGRVVENAMEHINPOKR RPVHSY DV TIDZIVY DLD GREQ GKRFGVEQYEMTFSKD RPHLENLKLGEGV GUHISL DV GLDPAFAA SID EGRKF IREKGIKIYTMHEDDLALGMTKVHEETIDZISAC DGVHISL DV GLDPAFAA SID EGRKF IREKGIKIYTMHEDDLALGMTKVHEETIDZISAC DGVHISL DV GLDPAFAA SUD EGRKF IREKGIKIYTMHEDDLALGMTKVHEETIDZISAC DGVHISL DV GLDPAFAA SUD KGRAF LKENGIKIYTMHEDDLALGMTKVHEETIDZISAC CLHVSL DV GLDPAFAA SUD KGRAF LKENGIKIYTMHEDDLALGMTKVHEETIDZISAC CUHISL DV GLDPAFAA SUD RGRAF LKENGIKATSMBDIDKIGUT SUNGKAN DISTISL DV GLDPAFAA SUD RGRAF LKENGIKATSMBDIDKIGUT SUNGKAN DISTISL DV GLDPAFAA SVD PAB REA VKKTEVIVYDMRIDIDKGIGVYENALALARIGGO TYHLSF DV GLDPYSIA SUD RGRAF LKENGIKATSMBDIDKGIGUT SUNGKAN DISTISL DV GLDPYSIA SUD RGRAF LKENGIKATSMBDIDKGIGUT SUNGKAN DISTISL DV GLDPYSIA SUD RGRAF LKENGIKATSMBDIDKGIGUT SVENTVRIKKESLKNN DISTISL DV GLDPYSIA SUD PG SHI LKILGIVENTSKIVKLGVYGIT THISF DV GLDPYSIA SUD RGRAF LEKNGIKATSMBDIDKGIGUT SVENTVRIKKESLKNN DISTISL DV GLDPYSIA SUD RGRAF LKENGIKATSMBDIDKGIGUT SVENTVRIKKESLKNN DISTISL DV GLDPYSIA SUD RGRAF LEKNGIKATSMBDIDKGIGUT SVENTVRIKKELHTYVGNG TIPHLSF DV GLDP
ARG1 SCHPO ARG1 XENLA ARG2 AGRTU ARG2 HUMAN ARG2 SCHPO ARG2 SCHPO ARG2 SCHPO ARG1 LEIAM ARG3 XENLA ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 BACSU ARG1 CAEEL ARG1 COCIM ARG1 CAEEL ARG1 COCIM ARG1 COCIM	IEERRLAFICL LSKSDIVIGL SKSDIVIGL SKSDIVIGL LSPPNIVIGL LSPPNIVYIGL LSPPNIVYIGL LSPPNIVIGL LSPNKINYIGL SKSDIVIGL -PRNVSMMGI IKPENVUIGV IKPENVUIGV IKPENVVIGL LPHNLCMVGIGL IRPEDVMIGN IRPEDVMIGN IRPEDVMIGN IRPEDVMIGN IRPEDVMIGN IRPEDVMIGN IRPEDVMIGN ISAKDIVIGL JPSRSVGIGI LSAKDIVIGL SAKDIVIGL SAKDIVIGL SKAKDIVIGL SKAKDIVIGL SKAKDIVIGL SKAKDIVIGL ISAKDIVIGL ISAKDIVIGL ISAKDIVIGL ISAKDIVIGL ISAKDIVIGL ISAKDIVIGL ISAKDIVIGL ISAKDIVIGL ISAKDIVIGL ISAKDIVIGL ISAKDIVIGL ISAKDIVIGL IDPRISVIGU 	DLD PMBRAF LRERSITAYTMHDUDKYGIARVVEMALEHINPGRR RJTHLFFVWACOFIVA DLD PAGAFI LKENYDISYYSMRHIDCMGIKKYMEKTFDOLLGRRD RJTHLFFVWACOFIVA SVD PAGAA LEDSGITYUDMESIDEHGYAVLLRAFLRVQANN GLHWYLFFFILS FLA DVD PFBHFI LKNYDIGYFSMRBIDRLGGIQKVMEGTFDRLGKRQ RJTHLFFV DAFDELAA DVU PFBRAF LKERNIAATYMHEVDKYGGOVVSKALKENTPOLLGRRD RJTHLFFV DAFDELAA DVU PFBRAF LKERNIAAYTMHEVDKYGGOVVSKALKANSVSKGT RJVHWYSYGT DZLYV DLD PAGOFI LKNYNSYYSMRHIDCMGIRKVMESTFDLGLGRRD RJTHLFFV DAFDELAA SVU PEGRRY LKENKIKYTMHEVDRJGMTKVLEETLJYLSKO SUD SGUKFG IKNYNSYYSMRHIDCMGIRKVMESTFDLGLGRRD RJHHVGIGO GLOSDA SLD GGRKG GKRFGVEGVISMTSTSDIDLGGIKKVIENTILYLKSRT DGVHLSLDVGLGDSDA SLD GGRKY IKSGKKVYTMHEJDRLGMTKVIESTLJYLSKO SLD GGRKY IKSGKKYYTMESIDIDLGGIKKVIENTILYLKSRT DGVHLSLDVGLDG GLOSDA SLD SGUG LKKNETYSTYKTKIKGVISVIENTILYLKSRT SKULLU GUG DOSDA SLD SGURY IKNISTYSTYKTVKIKGVISVIENTILYLKSRT SKULLU GUG GLOSDA SLD SGURY IKNISTYSTYKTVKIKGVISVIENTILYLKSRT SKULLU GUG GLOSDA SLD SGURY IKNISTYSTYKTVKIKGVISVIENTILYLKSRT SKULLU GUG GLOSDA SVD - RA FREA VKKTEVIVYDUGUGIGVVKENKISIL SVD - RA FREA VKKTEVITYKTVKIKGVISVI IKKSITKVIKSIL SVD - RA FREA VKKTEVIVTVKIKU SVI IKKSITKVIKSI SVD - RA FREA VKKTEVITYKIKVIKGVISVI IKKSIT SVD - RA FREA VKKTEVITYKIKVIKGVISVI IKKSIT SVD - RA FREA VKKTEVITYKIKVIKU SVI IKKSIT SVD - RA GKAL LKENGIKAFSMHDIDKVGGIGVVKENKSLKAN DIVHLSFVY SVD - PAGOKIL - LKENGIKAFSMHDIDKVGGIGVVKENKISIL SVD - RA SKL LKENGIKAFSMHDIDKVGGIGVVVKENKSLKKAN DIVHLSFVY SVD - RA SKL LKENGIKAFSMHDIDKVGGIGVVXTAKSLKKAN DIVHLSFVY SVD - RA SKL LKENGIKAFSMHDIDKVGGIGVVXTAKSLKKSNL TIVHLSFVY SVD - PAGOKIK LKENGIKAFSMHDIDKUGGIVKYKENSLKKAN

Fig. 1 (cont.) For legend see page 1818.

ARG1 SCHPO	ATGA	RVPGGLTI	FREAMYICE	SVAETGS	LVAVEVM	VNELLC	GNKEEAKTT	VD-LARSIVRTC	LGQTLL
ARG1 XENLA	ATCT	PVICCLTY	REGVYITE	EIHNTGM	LSAVDLVB	VNEVLAAT	PS-EEVKAT	AN-LAVDVIASC	FGQTREGAHTRAD
ARG2 AGRTU	AVC	TVPGGATE	FREAHLVMEN	ILHDSGL	VCSLOLVE	LNEFL	DERGRTA	TLMVDLAASL	MGKRVMDRPTRAG
ARG2 HUMAN	ATCH	PVVGGLTY	REGMYIAE	EIHNTGI	LSALDLWE	VNBOLA-7	TSEEEAKTT	AN-LAVDVIASS	FGOTREGGHI
ARG2 MOUSE	ATCH	PVVGGLTY	REGVYITE	THNTGI	LSALDINE	VNEHLA-7	SEEEAKAT	AR-LAVDVIASS	FGOTREGGHI
ARG2 RATNO	ATCHT	PVVGCLTY	REGLYITE	THSTGI	LSALDLVE	VNEHLA-	SEEEAKAT	AS-LAVDVIASS	FGOTREGGHI
ARG2 SCHPO	A THEMP-	RVPGCLTT	FREAMYTCE	VAESGT	LVAVIDVM	UNDIT.TC	NEFEARTT	VD-LARSTVRTS	LGOTLL
ARGT LETAM	A TRAN	PVRGELST	FREALFLOE	TAECGE	LUALINA	CNULLAAT	-ESHVNDT	TS-DGRATARCM	MGETLL
ARG3 XENLA	A TOW	PVTREITTY	REGUYITE	THNTGM	T.SALDIAS	UNEVIAT	-SEEVKAT	AN-LAVDVIASC	FGOTREGAUTRAD
APOT ACPTI	A Webs	TUDERATI	PPPAUL TMPI	AT UDG OT	VTCT DT AP	TANKT	-DEPORTA	PI.TTDI AGGI	FORDUFORVER
ADOT ADATU	OVCU	TEDGOTOT	PDDVINITU	ILOD	VUCADAT	ENROR	-DEVDONE	A-MUNAVIVERI	AVTOV
ABOT BACCD	CUTALI-	DUTOOLEY	DPCUT MMP	ILARA OT	TTCAPTO	UNDITT	-DEDNETA	CVAVATMONT	FORVIN
ARGT BACSI	CVC	DVVARTSY	PESHLAME	ILYDA GT	TTSAFFUS	UNDITT	-DHENETG	KTAVELVESL	CKKLL
ARGT STAAL	CT PU	RVADVHES	SPACTMOLR	T.I.HI.WKET	SEVTING!	LOTTERM	OWDATHNNH	SAFOAVSLOCTE	FOFPLISS
ARGI DIARO	OTTEAN	DEDEDATE	PPEAULTMPI	TUDOCI	VTOT NT	TNET	-DEL VULL	PETDIINK	EGEF EL
ARGI BACHA	STILL AN	TUDOGATI	PREASE THE	dunbe	VTSLOUGH	TATE D	-DEBCREA	A	CREUMBROWTEY
ARGI BROAD	11000	DONDOTH	TREATLINE	TEADSGL	TTAMBT	DIVISION D	DDRODTO	RVHVDLMASL	LGRSVHDRPTIST
ARGI CALL	CTOR -	DUCDALNI	IDDEPTINUE	TUVPCT	UVOLAT	There	-DEDIGKIS	D-INVERTOCIE	KINGEQINSVNET
ARGI CLOAC	G T G T	PVSDGLN	DDIKLAVE	SHVRSGL	VISPILIVI	ENGAL	DKDHQIE	D-BYMEPIDCIP	KNLK
ARGI COCIM	S161	PVREELTI	DREGDETAE.	STHETGS	LVAMULWI	VNETL	ETLGASE	TIRAGUSLVRSA	LGDTLL
ARGI DEIRA	GVGH	PVPGGLS	REGELLME.	LISESGP	VISMULWI	ANETT	DTRNQTA	EVMVGMAASL	LGORIL
ARGI_EMENI	S109	PVREELTI	LCEGDFICE	WHETGR	LISMOLVE	ANEST	VAVGASD	TIRTGCSLVRSA	LGDTLL
ARGI_GLYMA	GVSH	ILSCELSE	RDAMWWPG	ALKGL	IVGGUVWE	YNEQR	EPPDRMT	A-MVAARFVREL	AARMSK
ARGI_HELPY	STRV	RENNELSI	DELKQLLG.	LLESFRDE	LKAVEVTE	INSTVS-	IKENNEE	EKQVLEIL	DLIINSCRIKDKH
ARGI_HUMAN	ATCH	PVVEELT	REGLYITE.	GIYRTGI	-SGLIDIME	VNRSL-GI	TPEEVTRT	VN-TAVAITLAC	FGLAREGNERPI-
ARGI MOUSE	ATCH	PATHETS	REGLYITE.	GIYRTGI	LSGLOIME	VNETL-G	TALEVEST	VN-TAVALTLAC	FGTQREGNERPGT
ARGI_NEIGO	GIGH	PACHETSS	SDRALKILR	SLTDLL	IVGMOVWE	VARSY	DQSDITA	LAGATIALEMLY	LQGARRD
ARGI_NEUCR	STON	PVREELTI	LREGDFICE	CVHETGS	LVAVOLWE	VNETL	AAPNDVG	AHETVRAGCSLV	RSRSRRNVL
ARGI_RANCA	ATC:	PVPGGLTY	REGMYITE	2LYNTGI	LSAVDMM	VNES-RGE	ETERESKLT	VN-TSLNMILSC	FGKAREGFHASS-
ARGI_RATNO	A Treas	PVVGGLS	REGLYITE	SIYKTGI	LSGLUIM	VNETL-GI	TPEEVTRT	VN-TAVALTLSC	FGTKREGNHKPET
ARGI_RHOCA	AVIEW	TVPECATI	FREAHLIME	LCDAGV	VISLOLW	TNELT	DERGRTA	SLICDLAASL	FGRRVLDRQTRSF
ARGI_SCHPO	GTCH	PETGGWT	FREMKSILR	KLDGHLN	LVGAEVW	VSEP	YDDRAES	TSLAASD	FIFEILSSMVKEP
ARGI_STRPY	GVSA	IQSLEVDI	PNLAVLVLQ	HIAASGE	LVGFDVWB	VSEP	HDIDNHT	ANLAATF	IFYLVQIMAQHS-
ARGI_SYNEC	GVIER	PEPGELG	YEGLNFFR.	RLFQTKQ	VIGCULM	LARV-RGS	SVVSEFSTA	KLAYKLMGY	WGESQLRKK
ARGI_XENLA	ATCH	PCPEERTY	REGRILHE	2LHKTGI	LSGVDTIN	MESTSRGE	T-RDVEVT	VK-TALDMTLSC	FGKAREGFHAST-
ARGI_YEAST	ATCH	PVREELTI	LREGLFLVE.	RLAESGN	TIVIDAME	CNR	DLAIHDI	HVSNT-ISAGCA	IARCALGETLL
SPEB_ARCFU	GVS	PEPFELKI	PIDFIRFFA	GIADF	VVGFDVVF	VVR	DSNKVTQ	TLAAK-IILEAI	AAKVRCDIPK
SPEB_BACSU	GTC	VDAGCITS	SKELLASVH	EIARS-EVN	VKGADLWE	VARVY	DHSEQTA	NTASK-IIREML	LGFVK
SPEB_DEIRA	GTSS	PEPDELTY	AQGMKILA	AAAANNI	VVGLULVI	LAENL	DPTGRSE	LLMARLVMET	LCEVFDHV
SPEB_ECOLI	GTCH	PVICCLTS	SDRAIKLVR	GTKDTN	IVGMOVVE	VARAY	DQSEITA	LAAATLALEM	LYIQAAKKGE
SPEB_METFE	GVGN	PTPVGITI	PYHMEKFIE	KIARKF	IIGIDIVE	VATDR	IGDPAAM	NAAKI	LYDFLFAIKI
SPEB_METJA	GTGT	PEPCGEST	FRELFNSLY	LLEEV-KDR	IIGFDIVE	VSEIY	DIANITA	ITAAKIAREL	MLMIL
SPEB_METTH	SVGN	PAPAGLTI	PHIMEELVL.	ALSGKD	VVGLDVVI	VASGG	MADPTSV	NAAKI	-YDILTLLI
SPEB_PYRHO	ETGT	PEPGGLGI	FWEVIEALE	WLTKRKB	VAGEDIM	VSGDR	LGNSTSI	TAAKLLFY	VIGMSAR
SPEB_RHOCA	GTGT	PEVGGPTS	SWTALEVAR	GLRGLD	IIGADLVI	VSEPF	DPAGNTA	SEQWLGVNLMFE:	MLCVLAERIASA-
SPEB_SCHPO	ATGT	PESAGWT	FRELETILR	GLDGIB	LVGADIVE	VARAY	DFAEVTT	LAAADILFEVMS	IMVKTPVYKEQKQ
SPEB_STRCL	GTGT	PAPGGLLS	SREVLALLR	CVGDLB	PVGFOVM	VSPLY	DHGGITS	ILATEIGAELLY	QYARAHRTQL
SPEB_SYNEC	GTOW	PEPGGLLI	PREALYLLK	RIIRETN	VCGMEVVE	VSEPY	DISDMTS	LMATRVICDTMA	HLVVSGQLPRTEK
HUTG_BACSU	GCPA	IGPECLY	PDELLEAVK.	IAQQPN	VAGIEI	VDPTL	DFRDMTS	RAAAHVLLHALK	GMKLSPFK
C279_PSEAE	GTGT	PEIGGLT	FIQAMEIIR	SCQGLD	LIGCOLVE	VSEPY	DTTGNTS	LLGANLLYEMLC	V
C297_PSEAE	GTGT	PEIGGMTS	SLQAQQLVR	GLRGLD	LVGADVVE	VSEPF	DVGGATA	LVGATMMFELL-	
				TX					
	####	#########	**********	##### **	*********	*****	000000000	4.0	

Fig. 1. Multiple alignment of 50 sequences of ureohydrolases and related enzymes, as obtained from CLUSTAL W and DIALIGN, manually edited in the regions containing many non-contiguous gaps. The highly divergent N and C termini have been removed. Identical and similar residues in more than 75% of the sequences are drawn on black and shaded backgrounds, respectively. Those residues that were kept for the phylogenetic reconstructions are labelled '#', the excluded parts corresponding to regions where CLUSTAL W and DIALIGN gave highly different alignments. The roman numerals point to the regions that were used for building the 'gap tree' (see Results and Fig. 3). The arrows point to the residues involved in the binding of manganese ions. In this figure and the following ones, the names of the proteins follow the SWISS-PROT convention: ARGI and SPEB indicate proteins that are noted as arginases and agmatinases, respectively. The names of the species follow the five-letter code as indicated in Table 1.

reduced to the question of incorporation of carbon into the simple building blocks needed for biopolymer synthesis. However, the universal requirement for nucleotides and amino acids demonstrates that nitrogen metabolism was extremely important in the first steps of life (Granick, 1957). It is likely that a nitrogen-fixing cycle predated life, placing nitrogen-rich molecules in the limelight (Danchin, 1989), polyamines being crucial molecules (Cohen, 1998). In this context, Ouzounis & Kyprides (1994) constructed an interesting evolutionary tree of agmatinases, the polyamine-synthesizing enzymes, with emphasis on their universal presence. Since this seminal work, many new sequences have been obtained and annotated by their similarity with the known sequences. We therefore undertook a comparative analysis of the corresponding set of sequences. Genes that were deemed important were cloned and attempts were made to identify their functions. We reconsidered the phylogeny trees of arginases/agmatinases and constructed new ones for the enzymes involved in this first step in polyamine metabolism. To incorporate evolutionary constraints of different types, we first considered the usual types of phylogeny trees constructed based on the variation of the amino acid sequence in these proteins, without taking into account the presence of gaps in the sequences. Several discrepancies with respect to the expected position of some organisms in the trees were found. In a second approach, we reconstructed trees based only on the presence and evolution of gap-containing regions in the sequences (Baldauf & Palmer, 1993; Gupta, 1998a; see Fitch & Yasunobu, 1975, for appropriate caveats), because gaps would be much less sensitive to genetic drift or amino acid metabolism. The crucial enzyme activities that presumably evolved from ancestral ureohydrolases were validated by cloning, expressing and measuring activity of the corresponding enzymes. The emerging picture is consistent with a bacterial origin of hydrolases (ureohydrolases and related activities), which later evolved to those of the Archaea and the Eukarya.

METHODS

Sequence data collection and bacterial strains. Sequence data were collected from the GenBank/EBI/DDBJ DNA sequence database. Some sequences were also obtained from ongoing genome sequencing programs.

The reference *Escherichia coli* strain for the study of polyamine biosynthesis was obtained from the *E. coli* Genetic Stock Center, through the kind help of Dr Mary Berlyn. Strain MA255 was used: K12 *thr-1 leuB6 fhuA2? lacY1 glnV44(AS)? gal-6* λ^- *relA1? can-1 speB2 speC3 rpsL133(strR) xylA7 mtlA2 thi-1*. For cloning experiments strain XL-1 Blue was used (K12 supE44 hsdR17 recA1 endA1 gyrA46 thi relA1 Δlac F'[proAB⁺ lacI^q lacZ Δ M15 Tn10(tet^R)]; laboratory collection).

Synechocystis PCC6803, Helicobacter pylori 26695 and Neisseria gonorrheae MS11-E DNAs were kind gifts from Dr N. Tandeau de Marsac (Physiologie Microbienne, Institut Pasteur, Paris), Dr H. De Reuse (Pathogénie Bactérienne des Muqueuses, Institut Pasteur, Paris) and Dr M.-K. Taha (Unité des Neisseria, Institut Pasteur, Paris), respectively.

Sequence alignment and construction of trees. The 50 ArgI and SpeB amino acid sequences and the 22 spermidine (spermine) synthase sequences currently available were aligned with the programs DIALIGN version 2 (Morgenstern et al., 1998), as well as CLUSTAL W (Higgins et al., 1996), using default parameters (similarity matrix BLOSUM, 30; gap open penalty, 10; gap extension penalty, 0.1). A second putative ArgI sequence (encoded by the rocF gene) from H. pylori was extracted from the Astra server for comparison to the one present in our alignments. Because the differences were very small and did not change the positions of gaps, this sequence was not incorporated further in our analysis. The resulting multiple alignment was then checked for the conservation of important residues (Perozich et al., 1998) and manually edited. Regions where the two alignment programs gave widely different solutions were removed (see Fig. 1). Phylogenetic analyses were performed using the PHYLIP 3.57c suite of programs (Felsenstein, 1993). The pairwise distance matrices were calculated by **PROTDIST** with the Dayhoff option to estimate the expected amino acid replacements per position. The neighbour-joining (NJ) trees were obtained with NEIGHBOR. The most parsimonious trees were determined with PROTPARS. In each case, we performed 1000 bootstrap resamplings with SEQBOOT. The consensus trees were calculated by CONSENSE and drawn with the program TREEVIEW (Page, 1996).

To obtain a tree based essentially on structural features, we constructed a matrix where the presence and the length of insertions/deletions (nine values for each sequence; see Fig. 1) was given as the input. This was done without considering the amino acid variation in the sequences, only by counting the number of gaps in those regions delimited by highly conserved residues. Because evolution trends are to conserve the length of gene products, a last column with the length of the sequence as input was added in the matrix. This matrix was used to perform a multivariate analysis (Factorial Correspondence Analysis, FCA) to compute distances (χ^2) between the sequences (Hill, 1974; Lebart et al., 1984). The distances between each pair of sequences were calculated from the first three co-ordinates of the FCA (73% of the total inertia). Finally, a UPGMA (Unweighted Pair Group Method using Arithmetic Averages) distance tree was calculated from the resulting distance matrix (Sneath & Sokal, 1973). Here we chose UPGMA rather than NJ because we did not intend to build a phylogeny tree, but rather to perform a simple cluster analysis which would group those species that are least different, regardless of any further refinement of the tree.

Cloning procedures and biochemical assays. Cloning was performed by PCR amplification of the DNA regions of the putative *argl* or *speB* genes under study and followed by subsequent ligation of PCR products into plasmid p*Trc99A* (Pharmacia Biotech).

To clone the putative argl gene from Synechocystis PCC6803, a DNA fragment beginning at the translational start point and ending 5 bp after the stop codon of argI was amplified by PCR, using primers introducing a BspHI cloning site at the 5' end and a BglII cloning site at the 3' end of the fragment. The PCR product was ligated and inserted into the NcoI and BamHI sites of pTrc99A, creating pDIA5600. To clone the putative speB gene from Synechocystis PCC6803, a DNA fragment beginning at the translational start point and ending 52 bp after the stop codon of *speB* was amplified by PCR, using primers introducing an AfIII cloning site at the 5' end and a BamHI cloning site at the 3' end of the fragment. The PCR product was ligated as described above, creating pDIA5601. To clone the *rocF* gene (encoding ArgI) from *H. pylori* 26695, a DNA fragment beginning at the translational start point and ending at the stop codon of *rocF* was amplified by PCR using primers introducing an Ncol cloning site at the 5' end and a BamHI cloning site at the 3' end of the fragment. The PCR product was ligated as described above, creating pDIA5602. To clone the putative *argI* gene from N. gonorrheae MS11-E, a DNA fragment beginning at the translational start point and ending 4 bp after the stop codon of argI was amplified by PCR using primers introducing a BspHI cloning site at the 5' end and a BamHI cloning site at the 3' end of the fragment. The PCR product was ligated as described above, creating pDIA5603.

The plasmids containing these putative *argI* or *speB* genes were expressed in an *E. coli* mutant unable to synthesize polyamines. To assay agmatinase/arginase activities, strain MA255 was used.

For determination of enzymic activities, the bacteria from 200 ml LB overnight cultures were centrifuged, washed with PBS and centrifuged again; the pellets were then weighed. Extracts were prepared by grinding the bacterial paste for several minutes in a mortar with alumina (equal to twice the pellet weight). The mixtures were resuspended in 50 mM Tris/HCl, pH 7·6, containing 1 mM EDTA and 1 mM DTT and centrifuged at 10000 r.p.m. for 30 min in the cold. Supernatants were used for determination of enzyme activities. Urea was produced as described by Hirshfield *et al.* (1970) using either agmatine or arginine as substrates (10 mM each). The assay was slightly modified by raising the pH of the Tris/HCl buffer from 7·5 to 9·0 as described by Yamamoto *et al.* (1988). Urea measurement was performed using the Blood urea nitrogen assay kit (Sigma).

RESULTS

Multialignment of arginases, agmatinases and related sequences

A collection of 50 arginase- or agmatinase-like proteins were multi-aligned using DIALIGN (Morgenstern *et al.*,

1998) and CLUSTAL W (Higgins et al., 1996), and the alignments were further refined by realigning regions located between highly conserved segments (Fig. 1). Remarkably, the *H. pylori* sequence does not retain several features highly conserved in the other sequences. The sequence was checked by using the sequence published by Astra-Zeneca in addition to the sequence published at TIGR. As seen in Fig. 1, the multiple alignment is characterized by short highly conserved regions, separated by segments of much lower overall similarity. On the whole, the similarity between sequences is still large enough to identify related sequences using the BLAST programs, in particular PSI-BLAST. To be as conservative as possible, the number of gaps permitting alignment was kept to a minimum. Nine gap-containing regions were kept for further study, leaving aside gap regions that were uninformative, i.e. that either displayed too little difference between sequences or gave widely different results with the two alignment programs (see Fig. 1). Among the conserved regions are the two manganese-ion-binding sites present in these enzymes (Prosite; http://www.expasy.ch/cgibin/nicedoc.pl?PDOC00135). The critical residues are conserved in all but a few of the sequences (particularly in the case of *H. pylori*; see Discussion). As the mean error rate in DNA sequences is of the order of 10^{-3} or higher (especially when one considers the GC swaps), the sequences in these regions should certainly be verified if one had to comment on the role of the corresponding residues (see discussion of the SpeB sequence from Synechocystis PCC6803 below).

Phylogeny trees

This alignment allowed us to construct phylogenies with two commonly employed approaches for phylogenetic reconstructions, i.e. distance and parsimony (maximum likelihood was also used, but did not give significantly different results; data not shown). Fig. 2 displays the trees obtained by using these methods. The trees are, on the whole, rather similar. They are firmly split into two main parts, one corresponding to a majority of identified or putative arginases (ARGI) and the other to agmatinases (SPEB). However, the bootstrap scores within the SPEB cluster are generally poor, which means that a detailed view of the evolution of agmatinases and, in some parts, of arginases – cannot be obtained from this study. An interesting observation, however, is that the sequence of SpeB from Bacillus subtilis (Sekowska et al., 1998) groups in both trees with the corresponding enzymes from the Archaea. Most of the sequences from the Eukarya, on the other hand, are supposed to be of the ArgI type (no agmatinases have been identified among the vertebrates) and group far from the archaeal ones. The sequences from Schizosaccharomyces pombe are also grouped in both trees with the sequences from Gram-negative bacteria (presumably agmatinases; see below). Enzyme activities corresponding to other functions that are related to, but distinct from urea hydrolysis (e.g. HutG encodes formiminoglutamase and PahA encodes proclavaminate amidino hydrolase, which synthesizes a precursor of clavulanic acid, etc.) are generally scattered in the tree.



Fig. 2. Reconstructed phylogenies based on the multiple alignment shown in Fig. 1. (a) Distance tree (PROTDIST) after 1000 bootstrap resamplings; (b) most parsimonious consensus tree (PROTPARS) after 1000 bootstrap resamplings. The bootstrap scores are given for some key branches.

Functional identification of pivotal activities

An examination of the phylogeny trees revealed that some enzymes annotated as agmatinases are grouped with arginases, and *vice versa*, not in line with the internal consistency of the tree. To substantiate the validity of the trees, we cloned the presumed arginase genes from *H. pylori*, *N. gonorrheae* and *Synechocystis* PCC6803 (found in a group of archaeal enzymes containing the experimentally identified B. subtilis agmatinase) and the presumed agmatinase gene from Synechocystis and expressed them in an E. coli strain unable to synthesize putrescine (MA255). Urea production was subsequently measured in cell-free extracts with either agmatine or arginine as substrate. If the enzyme is an agmatinase, urea production will occur when agmatine is added; if it is an arginase, urea will be produced in the presence of arginine.

As shown in Table 1, the *H. pylori* enzyme is an arginase. In contrast, the gene product labelled ArgI in *Synechocystis* PCC6803 (Cyanobase; http://www. kazusa.or.jp/cyano/cyano.orig.html) is an agmatinase. The gene product labelled SpeB in this organism might have been a second agmatinase. However, we failed to detect this activity or to identify this gene product as an arginase under conditions where the other gene products produced urea from either arginine or agmatine. It is possible that this is an enzyme used in the degradation of histidine similar to the *hutG* gene product of *B*. *subtilis*. More likely, it is involved in secondary metabolism comparable to the *pahA* gene product, which is implicated in synthesis of clavulanic acid in Streptococcus clavuligerus (see below). The N. gonorrheae enzyme labelled ArgI is clearly identified as an agmatinase, not an arginase.

Discrimination between arginases and agmatinases

Arginases and agmatinases release urea and ornithine or putrescine, respectively, from substrates that differ from each other only by the presence of a carboxylate group (in arginine). Because arginases and agmatinases can be divided into two well-defined categories, we investigated whether they could be separated according to a consensus sequence. The manganese-binding sites, as shown above, make an identical consensus in both enzymes. This could be expected since these enzymes display the same catalytic activity. Ten invariant conserved residues are found in the Prosite motif (http://www.expasy. ch/cgi-bin/nicedoc.pl?PDOC00135), modified below, using the data from our study. Among these, six bind to the manganese-ion cofactors (indicated by #):

Pattern 1 [ILV]-X-[FILMV]-G-G-[ED]-H#-X-[ILMV]-[ASTV]-X-[AGP]-X(3)-[AGST];

Pattern 2 [ILMV](2)-X-[FILMVY]-D#-[AS]-H#-X-D#;

Pattern 3 [FHY]-[ILV]-[ST]-[FILMVY]-D#-[ILMV]-D#-X(3)-[APQ]-X(3)-P-[AGS]-X(7)-G.

A fourth conserved pattern also seems to be important:

Pattern 4 [AGSV]-[ACFILMV]-[DE]-[FILMTV]-[AIM-TV]-E-[FILMV]-[AGHNS]-[GPS].

If only the arginase side of the tree is retained, one uncovers several further constraints that specify the arginase family. These include restriction to [ILMV] at the third residue in pattern 1 (no F), a conserved W (Y in the widely deviant *H. pylori* sequence; see Discussion) at position 3 in pattern 2 and restriction to H at position 1 in pattern 3 (neither F nor Y). In contrast, several regions generally differ between the ArgI and SpeB families precisely where the patterns are restricted in arginases. In addition, the polypeptide sequence intervening between patterns 2 and 3 is significantly shorter (20 or so residues less) in agmatinases compared to arginases. In agmatinases, position 3 of pattern 2 is more variable than in arginases, H, Q or N being present instead of W (Y). Likewise, the H at position 1 of pattern 3 of arginases is replaced by an aromatic residue in agmatinases.

Once these general features are identified, one may also remark that several sequences appear to differ significantly from the others. This is particularly true in the case of the *H. pylori* arginase. Some sequences also appear to be much longer than the mean. These include certain sequences of *Schizosaccharomyces pombe*, the sequence of *Caenorhabditis elegans* and the sequence of *Synechocystis* PCC6803 (labelled SPEB).

Alignment of gaps in the sequences

Until now, the phylogeny trees constructed were based only on the differences between amino acids at equivalent positions. However, changes yielding gaps in the alignment also correspond to one or more mutational events. In general, gaps specify the insertion or deletion of loops in the three-dimensional structure of the protein (Briozzo et al., 1998). For this reason, their length is usually variable, whereas the place where they occur is fixed. This is due to structural constraints in the architecture of the protein active site(s). It is therefore interesting to consider gaps as the hallmark of some mutational events, noting that conservation of both the presence and the length of the insertions/deletions may indicate some kinship between the corresponding proteins. We therefore constructed a matrix based on the lengths of insertions occurring between well-conserved regions (see Fig. 1). As expected, the insertions/deletions often occur at places where the multiple alignment is less reliable. However, since they are anchored by highly conserved residues, their length is unequivocal. In addition, protein length was also included in the matrix because the length of the protein introduces a constraint in evolution of the same structural nature as the introduction of gaps. This matrix was used to calculate the co-ordinates of the sequences using FCA. The latter enabled us to obtain pairwise distances that were subsequently used as input in the program NEIGHBOR.

Sequence name	Organism	Function in database	Predicted function	Identified function
ARG2_AGRTU	Agrobacterium tumefaciens	Arginase	Arginase	Arginase*
ARGLARATH	Arabidopsis thaliana	Arginase	Agmatinase or secondary metabolism†	Arginase*
SPEB_ARCFU	Archaeoglobus fulgidus	Agmatinase	Agmatinase	
ARGI_BACCD	'Bacillus caldovelox'	Arginase	Arginase	Arginase
ARGI_BACHA	Bacillus halodurans	Arginase	Arginase	
ARGI_BACSU	Bacillus subtilis	Arginase (RocF)	Arginase	Arginase
SPEB_BACSU	Bacillus subtilis	Agmatinase (YwhF)	Agmatinase	Agmatinase
HUTG_BACSU	Bacillus subtilis	Formiminoglutamate	Secondary metabolism and	
ADCI DDUAD	During the strength of	nydrolase	A min and	A
ARGI_BRUAB	Brucella abortus	Arginase	Arginase	Arginase
ARGI_CAEEL	Claenorhaballis elegans	Arginase	Arginase	
ARGI_CLUAC	Clostriaium acetobutylicum	Arginase	Arginase	A
ARGI_COCIM	Cocciaioiaes immilis	Arginase	Arginase	Arginase
CDED DEIDA	Deinococcus radiodurans	Arginase	Arginase	
SPED_DEIKA	Easharishig soli	Agmatinase	Agmatinase	Agmatinass
APCI EMENI	Escherichia coli	Aginatinase	Aginatinase	Aginatinase
ARGI_GLYMA	Glycine max	Arginase	Agmatinase or secondary metabolism ⁺	Aiginase
ARGI HELPY	Helicobacter tylori	Arginase (RocF)	Arginase	Arginaset
ARGI HUMAN	Homo sabiens	Arginase (Roer /	Arginase	Arginase
ARG2 HUMAN	Homo sapiens	Arginase	Arginase	Arginase
ARGLIFIAM	Leishmania americana	Arginase	Arginase	Arginase
SPEB METEE	'Methanococcus fervidus'	Agmatinase	Agmatinase	mgmuse
SPEB METIA	Methanococcus jannaschii	Agmatinase	Agmatinase	
SPEB_METTH	Methanobacterium thermoautotrophicum	Agmatinase	Agmatinase	
ARGI_MOUSE	Mus domesticus	Arginase	Arginase	Arginase
ARG2_MOUSE	Mus domesticus	Arginase	Arginase	Arginase
ARGI_NEIGO	Neisseria gonorrheae	Arginase	Agmatinase ⁺	Agmatinase‡
ARGI_NEUCR	Neurospora crassa	Arginase	Arginase	Arginase
C279_PSEAE	Pseudomonas aeruginosa	Unknown	Secondary metabolism and other functions	
C297_PSEAE	Pseudomonas aeruginosa	Unknown	Secondary metabolism and other functions	
SPEB_PYRHO	Pyrococcus horikoshii	Agmatinase	Agmatinase	
ARGI_RANCA	Rana catesbeiana	Arginase	Arginase	Arginase*
ARGI_RATNO	Rattus norvegicus	Arginase	Arginase	Arginase
ARG2_RATNO	Rattus norvegicus	Arginase	Arginase	Arginase
ARGL_RHOCA	Rhodobacter capsulatus	Arginase	Arginase	Arginase
SPEB_RHOCA	Rhodobacter capsulatus	Agmatinase	Secondary metabolism and other functions‡	-
ARGI_SCHPO	Schizosaccharomyces pombe	Arginase	Arginase	Arginase*§
ARG1_SCHPO	Schizosaccharomyces pombe	Arginase	Arginase	Arginase*§
ARG2_SCHPO	Schizosaccharomyces pombe	Arginase	Arginase	Arginase*§
SPEB_SCHPO	Schizosaccharomyces pombe	Agmatinase	Arginase	Arginase*§
ARGI_STAAU	Staphylococcus aureus	Arginase	Arginase	
ARGI_STRPY	Streptococcus pyogenes	Arginase	Secondary metabolism and other functions [†]	
PAHA_STRCL	Streptomyces clavuligerus	Proclavaminate amidino hydrolase	Secondary metabolism and other functions	Proclavaminate amidino hydrolase
ARGI_SYNEC	Synechocystis PCC6803	Arginase	Agmatinase†	Agmatinase†

Table 1. Arginases, agmatinases and related activities extracted from public data libraries

Table 1 (cont.)

Sequence name	Organism	Function in database	Predicted function	Identified function
SPEB-SYNEC	Synechocystis PCC6803	Agmatinase	Secondary metabolism or other functions†	Secondary metabolism or other functions‡
ARGI-XENLA	Xenopus laevis	Arginase	Arginase	Arginase*§
ARG1-XENLA	Xenopus laevis	Arginase	Arginase	Arginase*§
ARG3-XENLA	Xenopus laevis	Arginase	Arginase	Arginase*§
ARGI-YEAST	Saccharomyces cerevisiae	Arginase	Arginase	Arginase*

* Activity identified by complementation studies.

†Conflict.

‡Activity identified experimentally in this study.

§ Activity identified collectively as a bulk cell arginase activity.



Fig. 3. Gap-tree of ureohydrolases and related enzymes, calculated by factorial analysis of the lengths of the insertions in the multiple alignments (roman numeral in Fig. 1). This is *not* a phylogenetic reconstruction, but a distance tree based on the conservation of the position and length of probable loops in the three-dimensional structure of the proteins.

The resulting distance tree is displayed in Fig. 3. One of its prominent features is that arginases and agmatinases are clearly separated. It should be noted that this separation does not merely come from the fact that the lengths of the proteins were used in the pairwise distance calculation, even if the agmatinases tend to be shorter than the arginases. For example, the protein labelled SPEB_SYNEC is one of the longest in the set (390 aa) and is nevertheless clearly not a member of the arginase family. In addition, it is probable that gap regions III and VI (Fig. 1) are the most discriminant. Interestingly, some of the sequences that were located in a continuum near the junction of the arginase and agmatinase trees when using evolution of the sequences without consideration of gaps, moved to less ambiguous positions. The ambiguous situation of the *H. pylori* enzyme has been solved by demonstrating experimentally that it is an arginase. Remarkably, the enzymes that differ both from arginases and from agmatinases also group together in this tree. The protein labelled SPEB from *Synechocystis* is located next to the agmatinase cluster in a group comprising poorly defined activities (a protein labelled ARGI from a *Streptomyces* sp. and the SPEB protein from *Rhodobacter capsulatus*). Our experiments suggest



that it might not be an agmatinase but an enzyme related to secondary metabolism activities linked to polyamines (Cohen, 1998). Furthermore, the plant enzymes may be a sister group of their cyanobacterial counterpart, labelled ARGI, but that we identified as an agmatinase. Finally, we observed that the *B. subtilis* agmatinase clustered with the archaeal enzymes, as in the previous phylogenetic reconstructions, while the enzymes from *E. coli* and *N. gonorrheae* are now correctly grouped together (the *N. gonorrheae* ARGI enzyme being in fact an agmatinase).

DISCUSSION

Living organisms have been categorized into two architectural domains, according to the way they manage envelopes, membranes and skins. The protozoologist Edouard Chatton separated microbes made of a more or less complex single envelope from those that were nucleated, the prokaryotes and the eukaryotes (Chatton, 1938). Based on a phylogenetic tree of rDNA, Woese proposed in 1977 a bold hypothesis that divided the prokaryotes into two domains, the eubacteria and another one for which he coined the suggestive name 'archaea' (Woese & Fox, 1977). Since then, this paradigm has been substantiated by rDNA phylogeny trees. However, the problem was complicated by phylogenetic analysis of genes encoding the translation and the transcription machinery, when compared to those for ATP synthesis or intermediary metabolism (Ouzounis & Kyrpides, 1994, 1996; Ibba et al., 1997; Koonin et al., 1997; Diaz-Lazcoz et al., 1998; Gupta, 1998a; Koonin & Aravind, 1998; Kyrpides & Woese, 1998; Mayr, 1998). Lack of consistency in data coming from protein sequence analysis, depending on the chosen family, suggested an important involvement of horizontal gene transfer in early evolution (Ribeiro & Golding, 1998; Woese, 1998). Therefore, identification of enzyme activities solely based on such phylogeny trees became questionable. Because nitrogen metabolism must have existed from the very beginning of life, we reinvestigated the tree of ureohydrolases, using a different approach, to evaluate the validity of phylogenies based on implicit functional homology. In this study, we found that these enzymes are split into two major families, agmatinases and arginases. This separation is observed both in standard phylogeny reconstructions and in those using gaps. It corresponds to fine alterations of the enzyme structure to accommodate the subtle differences in their substrates.

The *H. pylori* sequence is of particular interest because, although from a Gram-negative organism, it does not group with E. coli and N. gonorrheae. Furthermore, the SSEH motif differs significantly from the consensus GGD(or E)H motif present in all other arginases. Using a second type of *H. pylori* ARGI, it is evident that SSEH is specific to these bacteria and does not result from a sequencing error (data not shown). Indeed, the lifestyle of *H. pylori* is very unusual. Possibly, this is reflected in the 'style' of its genome, constraints on the availability of metabolites affecting the composition of its proteins (Danchin et al., 2000). Alternatively, catalytic mechanisms might be affected in a more extreme environment. Variation in the nature of the preferred amino acids (or codons) would account for the rather odd place of the sequence in the non-gapped trees. The sequence from Bacillus halodurans, which is also present in bacteria living in unusual environments, is also similarly biased, a GGDC sequence replacing the (almost) universal GGD(or E)H sequence (and it groups with H. *pylori* in the gap tree). Several methods have been devised to deal with composition biases in alignments, for example by taking into account biases in the G + Ccontent of genomes (Tourasse & Gouy, 1997; Foster & Hickey, 1999; Wilquet & Van de Casteele, 1999) or by



Fig. 5. Most parsimonious consensus tree (PROTPARS) after 1000 bootstrap resamplings for the family of spermidine synthase and related enzymes. Proteins labelled SPSY, SPEE and PUTR are spermine synthases, spermidine synthases and methyltransferases, respectively. In addition to the organisms listed in Table 1, several sequences are from other organisms: MYCTU, Mycobacterium tuberculosis; AQUAE, Aquifex aeolicus; COFAR, Coffea arabica; DANRE, Danio rerio; DATST, Datura stramonium; FUGRU, Fugu rubripes; HYONI, Hyoscyamus niger; LYCES, Lycopersicon esculentum; NICSY, Nicotiana sylvestris; NICTA, Nicotiana tabacum; NICTY, Nicotiana tomentosiformis; PISSA, Pisum sativum; TETFL, Tetraodon fluviatilis.

taking care of long-branch attraction artefacts in the statistical methods used (Lyons-Weiler & Hoelzer, 1997). But, as seen here, this is probably of limited interest when divergence is very ancient. As a consequence, we emphasize the utility of considering gaps when constructing trees, especially for proteins that have been evolving for a very long time (Gupta, 1998b). Gaps are less sensitive to the nature of the genetic code or amino acids, but only to the relationship between the three-dimensional structure of a protein and its function. Besides allowing better prediction of protein function, our alignment identified residues of major importance for the two binding sites of the manganese co-factor as well as residues and regions separating arginases from agmatinases.

Although not substantiated experimentally, some of the sequences which differ from arginases may have other important activities (Cohen, 1998). Secondary metabolism activities or degradation of histidine are placed on the agmatinase side in the trees. Interestingly, in the highly conserved region, [FYMLIV]-[MLIV]-[WHYN-QV]-[FMLIV]-D-[AS]-H, the alanine or serine residue before the final histidine is replaced by an arginine in the Synechocystis SPEB sequence (clustered in the 'secondary metabolism' sequences). This alteration replaces a small non-polar or polar amino acid with a very large basic one. We did not identify the corresponding enzyme as an arginase or an agmatinase. This prompted us to check whether this was due to a sequencing error: arginine and alanine are coded by CGN or GCN codons and it is well known that C and G residues often migrate at the same rate on gels, resulting in frequent GC swaps in the final sequence. After PCR amplification of the corresponding region from the chromosome, we found that the sequence is exact (data not shown). We therefore propose that this residue is involved in defining a new specificity for the enzyme, probably for some reaction of secondary metabolism (cyanobacteria are known to produce secondary metabolites that may derive from molecules related to polyamines; Cohen, 1998). This observation illustrates the identification of an amino acid residue that may play a discriminant role in catalysis and could be a target for oriented *in vitro* evolution of enzyme activities.

The question of the origin and fate of arginine early in evolution is important because it is related both to the fixation of nitrogen and to the universality of the genetic code. Remarkably, the Archaea are only present in the agmatinase tree. Also, the pathways leading to putrescine (the primary polyamine, Fig. 4) involve ornithine, an amino acid not incorporated into proteins. The cell concentration of ornithine (an analogue of lysine) must therefore be finely controlled to avoid misincorporation into proteins. As a consequence, the arginase pathway is likely to have evolved later than that involving arginine decarboxylation and agmatinase. In this respect, one should note that the consensus sequence corresponding to pattern 2 is highly variable in the nonarginase sequences while it is extremely conserved in arginases, suggesting indeed that these sequences are of more recent descent. In particular, the W residue in WXDAHXD could be chosen as a discriminant residue to identify arginases (Fig. 1). One can therefore confidently assume that agmatinases predated arginases. The latter would have appeared in the Bacteria by recruitment of a wide specificity agmatinase and then transferred to eukaryotes (perhaps through transfer from mitochondria).

Plants have sequences labelled as arginases, but they are grouped with agmatinases in all three tree types. An arginase activity has been identified indirectly by complementation studies of a yeast mutant (Krumpelman et al., 1995). Plants possess an arginine decarboxylase (Klein *et al.*, 1999), therefore they make agmatine, but the presence of agmatinase must be substantiated experimentally, because they would be the only eukaryotes with this feature. This may reflect a gene transfer from chloroplasts to the nucleus. Under the selective pressure of sexual reproduction, genes in the nucleus avert the deleterious effect of Muller's ratchet in organelles (Haigh, 1978; Bergstrom & Pritchard, 1998). This hypothesis is supported by the experiment where we showed that Synechocystis ARGI is in fact an agmatinase. However, the plant sequences may correspond to enzymes involved in secondary metabolism, known to be very active in plants. Noticeably, they are grouped in the gap tree near the enzymes involved in secondary metabolism. The absence of agmatinases in known sequences from the Eukarya is puzzling, particularly because agmatine does exist in the Eukarya (Sastre et al., 1998; Reis & Regunathan, 1999). Recently, an agmatinase activity from the mitochondrial matrix has been identified by Sastre and co-workers (Sastre et al., 1996), but we do not know whether it belongs to the same family of manganese enzymes. This is unlikely because the whole genome sequence of several eukaryotes is known. In fact, the Eukarya recruited another activity, agmatine iminohydrolase, to fulfil this requirement (Park & Cho, 1991; Klein et al., 1999; see Fig. 4). This is not an uncommon feature among the Eukarya (see for example the case of aminolevulinate synthase; Duncan et al., 1999). The results presented here indicate that the whole question should be reinvestigated in depth.

Remarkably, the only Gram-negative arginase is that from H. pylori (substantiated experimentally in this study). The Bacteria are split into several groups in the gap tree. In the case of agmatinases, those of *B. subtilis* and the Archaea go together. A group comprising the Synechocystis PCC6803 agmatinase (wrongly annotated as an arginase) together with enzymes involved in secondary metabolism, and in broader specificity hydrolases, is also prominent in the gap tree. If one reasons in terms of acquisitive evolution (Thompson & Krawiec, 1983), assuming that an ancient activity with broad specificity has been progressively specified during evolution (Jensen, 1976; Danchin, 1989; Roy, 1999), then the origin of the ureohydrolase tree would be near the PahA HutG group, leading to agmatinases and enzymes involved in secondary metabolism and, subsequently, to arginases. To substantiate the grouping of B. subtilis and archaeal polyamine enzymes, we performed a phylogenetic reconstruction of spermidine synthases. These enzymes are less universally distributed than ureohydrolases. They exert their function downstream from the synthesis of putrescine, the precursor of polyamines. Fig. 5 displays the most parsimonious tree obtained in this case. Again, *B. subtilis* groups with its archaeal counterparts, while the *Eukarya* form another well separated group. In addition, we find not only a group made of spermine synthases (SPSY and SPEE), which use spermine as the substrate instead of spermidine, but also a group of related methyltransferases, which use *S*-adenosylmethionine rather than decarboxylated *S*-adenosylmethionine as a substrate (PUTR).

Because it seems difficult to transfer systematically the corresponding genes horizontally unless at a very early stage of evolution, an examination of polyamine metabolism suggests that the empire of prokaryotes may be more a continuum than a clear-cut division between the two domains of life, perhaps reflecting an early pool of metabolic genes that would have subsequently been frozen after differentiation into specific families. The data presented here suggest that some machinery, presumably involving RNA molecules, contained polyamines at a very early time in the evolution of life. In this machinery, the ancestors of some significant part of Gram-positive bacterial and archaeal genomes appeared to cluster together. This should be taken into consideration when trying to reconstruct the tree of origin of life.

ACKNOWLEDGEMENTS

We wish to thank M. Berlyn and the *E. coli* Genetic Stock Center, N. Tandeau de Marsac, H. De Reuse and M.-K. Taha for the gift of strains or chromosomal DNA samples. We wish to express particularly our thanks to Professor Chun-Kwun Wun for critical and attentive reading of this manuscript and to Ivan Moszer for helping us in its preparation. This work was supported in part by European Union contract BIOTECH B104-CT96-0655.

REFERENCES

Baldauf, S. L. & Palmer, J. D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci USA* 90, 11558–11562.

Bergstrom, C. T. & Pritchard, J. (1998). Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. *Genetics* **149**, 2135–2146.

Briozzo, P., Golinelli-Pimpaneau, B., Gilles, A. M., Gaucher, J. F., Burlacu-Miron, S., Sakamoto, H., Janin, J. & Barzu, O. (1998). Structures of *Escherichia coli* CMP kinase alone and in complex with CDP: a new fold of the nucleoside monophosphate binding domain and insights into cytosine nucleotide specificity. *Structure* **6**, 1517–1527.

Chatton, E. (1938). *Titres et Travaux Scientifiques* (1906–1937). Sottano, Italy: Sete.

Cohen, S. (1998). A Guide to the Polyamines. Oxford: Oxford University Press.

Cox, E. C. & Yanofsky, C. (1967). Altered base ratios in the DNA of an *Escherichia coli* mutator strain. *Proc Natl Acad Sci USA* **58**, 1895–1902.

Danchin, A. (1989). Homeotopic transformation and the origin of translation. *Prog Biophys Mol Biol* 54, 81–86.

Danchin, A., Guerdoux-Jamet, P., Moszer, I. & Nitschké, P. (2000). Mapping the bacterial cell architecture into the chromosome. *Philos Trans R Soc B Biol Sci* **355**, 179–190.

Diaz-Lazcoz, Y., Aude, J., Nitschké, P., Chiapello, H., Landès-Devauchelle, C. & Risler, J. (1998). Evolution of genes, evolution of species: the case of aminoacyl-tRNA synthetases. *Mol Biol Evol* 15, 1548–1561.

Duncan, R., Faggart, M. A., Roger, A. J. & Cornell, N. W. (1999). Phylogenetic analysis of the 5-aminolevulinate synthase gene. *Mol Biol Evol* 16, 383–396, erratum following 883.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.57c. Seattle: Department of Genetics, University of Washington.

Fitch, W. M. & Yasunobu, K. T. (1975). Phylogenies from amino acid sequences aligned with gaps: the problem of gap weighting. *J Mol Evol* 5, 1–24.

Foster, P. G. & Hickey, D. A. (1999). Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J Mol Evol* 48, 284–290.

Granick, S. (1957). Speculations on the origin and evolution of photosynthesis. *Ann NY Acad Sci* 69, 292–308.

Gupta, R. S. (1998a). Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* **62**, 1435–1491.

Gupta, R. S. (1998b). What are archaebacteria: life's third domain or monoderm prokaryotes related to gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. *Mol Microbiol* **29**, 695–707.

Haigh, J. (1978). The accumulation of deleterious genes in a population – Muller's Ratchet. *Theor Popul Biol* 14, 251–267.

Higgins, D. G., Thompson, J. D. & Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* 266, 383–402.

Hill, M. O. (1974). Correspondence analysis: a neglected multi-variate method. *Appl Statistics* 23, 340–353.

Hirshfield, I. N., Rosenfeld, H. J., Leifer, Z. & Maas, W. K. (1970). Isolation and characterization of a mutant of *Escherichia coli* blocked in the synthesis of putrescine. *J Bacteriol* 101, 725–730.

Ibba, M., Morgan, S., Curnow, A. W., Pridmore, D. R., Vothknecht, U. C., Gardner, W., Lin, W., Woese, C. R. & Soll, D. (1997). A euryarchaeal lysyl-tRNA synthetase: resemblance to class I synthetases. *Science* **278**, 1119–1122.

Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* 30, 409–425.

Klein, R. D., Geary, T. G., Gibson, A. S. & 8 other authors (1999). Reconstitution of a bacterial/plant polyamine biosynthesis pathway in *Saccharomyces cerevisiae*. *Microbiology* **145**, 301–307.

Koonin, E. V. & Aravind, L. (1998). Genomics: re-evaluation of translation machinery evolution. *Curr Biol* 8, R266–R269.

Koonin, E. V., Mushegian, A. R., Galperin, M. Y. & Walker, D. R. (1997). Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol Microbiol* 25, 619–637.

Krumpelman, P. M., Freyermuth, S. K., Cannon, J. F., Fink, G. R. & Polacco, J. C. (1995). Nucleotide sequence of *Arabidopsis thaliana* arginase expressed in yeast. *Plant Physiol* 107, 1479–1480.

Kyrpides, N. C. & Woese, C. R. (1998). Archaeal translation initiation revisited: the initiation factor 2 and eukaryotic initiation factor 2B alpha-beta-delta subunit families. *Proc Natl Acad Sci USA* 95, 3726–3730.

Lebart, L., Morineau, A. & Warwick, K. A. (1984). Multivariate Descriptive Statistical Analysis. New York: Wiley.

Lyons-Weiler, J. & Hoelzer, G. A. (1997). Escaping from the Felsenstein zone by detecting long branches in phylogenetic data. *Mol Phylogenet Evol* **8**, 375–384.

Mayr, E. (1998). Two empires or three? *Proc Natl Acad Sci USA* 95, 9720–9723.

Morgenstern, B., Frech, K., Dress, A. & Werner, T. (1998). DIALIGN: finding local similarities by multiple sequence alignment. *Bio-informatics* 14, 290–294.

Ouzounis, C. A. & Kyrpides, N. C. (1994). On the evolution of arginases and related enzymes. *J Mol Evol* **39**, 101–104.

Ouzounis, C. & Kyrpides, N. (1996). The emergence of major cellular processes in evolution. *FEBS Lett* **390**, 119–123.

Page, D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.

Park, K. H. & Cho, Y. D. (1991). Purification of monomeric agmatine iminohydrolase from soybean. *Biochem Biophys Res Commun* 174, 32–36.

Perozich, J., Hempel, J. & Morris, S. M., Jr (1998). Roles of conserved residues in the arginase family. *Biochim Biophys Acta* **1382**, 23–37.

Reis, D. J. & Regunathan, S. (1999). Agmatine: an endogenous ligand at imidazoline receptors is a novel neurotransmitter. *Ann* NY Acad Sci 881, 65–80.

Ribeiro, S. & Golding, G. B. (1998). The mosaic nature of the eukaryotic nucleus. *Mol Biol Evol* 15, 779–788.

Roy, S. (1999). Multifunctional enzymes and evolution of biosynthetic pathways: retro-evolution by jumps. *Proteins* **37**, 303–309.

Sastre, M., Regunathan, S., Galea, E. & Reis, D. J. (1996). Agmatinase activity in rat brain: a metabolic pathway for the degradation of agmatine. *J Neurochem* 67, 1761–1765.

Sastre, M., Galea, E., Feinstein, D., Reis, D. J. & Regunathan, S. (1998). Metabolism of agmatine in macrophages: modulation by lipopolysaccharide and inhibitory cytokines. *Biochem J* 330, 1405–1409.

Sekowska, A., Bertin, P. & Danchin, A. (1998). Characterization of polyamine synthesis pathway in *Bacillus subtilis* 168. *Mol Microbiol* 29, 851–858.

Sneath, P. H. A. & Sokal, R. R. (1973). Numerical Taxonomy: the Principles and Practice of Numerical Classification. San Francisco: Freeman.

Thompson, L. W. & Krawiec, S. (1983). Acquisitive evolution of ribitol dehydrogenase in *Klebsiella pneumoniae*. *J Bacteriol* **154**, 1027–1031.

Tourasse, N. J. & Gouy, M. (1997). Evolutionary distances between nucleotide sequences based on the distribution of substitution rates among sites as estimated by parsimony. *Mol Biol Evol* **14**, 287–298.

Wilquet, V. & Van de Casteele, M. (1999). The role of the codon first letter in the relationship between genomic GC content and protein amino acid composition. *Res Microbiol* 150, 21–32.

Woese, C. R. (1998). The universal ancestor. *Proc Natl Acad Sci USA* 95, 6854–6859.

Woese, C. R. & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 74, 5088–5090.

Yamamoto, S., Nakao, H., Yamasaki, K., Takashina, K., Suemoto, Y. & Shinoda, S. (1988). Activities and properties of putrescine-

biosynthetic enzymes in Vibrio parahaemolyticus. Microbiol Immunol 32, 675–687.

Received 25 January 2000; revised 1 May 2000; accepted 5 May 2000.