



**bio** – informatics: bioinformatics is conceptualizing biology in terms of molecules (in the sense of physical chemistry) and applying informatics techniques (derived from disciplines such as applied maths, computer science and statistics) to understand and organize the information associated with these molecules, on a large scale. In short, bioinformatics is a management information system for molecular biology and has many practical applications



#### Structure of a Human Gene (PSA)

CTGRCTCTTRGCCCCRGRCTCTTCRTTCRTTCRTTCCCTTTCCTTRGGRRRRRCRTGRCCRCCRGCCCCCRGCTCCCTGRGTCCCCCCRTCCTTTTCCRRRRCCTRRRRCRRR CCACTGCATCAGGARGTGAGTAGGGGCCTGGGGTCTGGGGAGCAGGTGTCTGTGTCCCAGAGGAATAACAGCTGGGCATTTTCCCCCAGGATAACCTCTAAGGCCAGCCTTGGGACTGG GGGABAGAGGGAAASTTCTGGTTCAGGTCACATGGGGAGGCAGGGTTGGGGCTGGACCACCCCCCCATGGCTGCCTGGGTCTCCATCTGTGTCCCTCTTGTCTCGCTTCATTATGTCTC ARGTGGAGGATACAACCTTGGGCCTGCAGGCTACCTACCCACTTGGAAACCCACGCCAAAGCCGCATCTACAGCTGAGCCACTCTGAGGCCTCCCCGGCGGTCCCCACTCAGGCCACTCCAAAGT TCCCCCTGTCASACCCTGCCGASCTCACGGATGCTGTGAAGGTCATGGACCTGCCCACGCACCCAGGACCACCTGCGGACCACGCCTGCCGGGCCAGGACTGAACCAGAGGACTGTACGCC CGTRGTCTTGRCCCCRARGRARCTTCRGTGTGTGGGCCTCCATGTTATTTCCARTGRCGTGTGCGCRRGTTCRCCCTCRERRGGTGRCCRAGTTCRTGCTGTGTGCGCGRC ACCTCCCCCGTGTCTCATCTCATTCCCTCCTTCTTTGACCCCCTCAAGGCAAT AGGTTATTCTT ACAGCACACTCATCTGCTTCCGCGTCAGCACACGGTTACTAGGCACCTGCTATGCACC GTCCCTTCCTACCGGCCAGGACTGGAGCCCCTACCCCTCTGTTGGAATCCCTGCCCACCTTCTTCTGGAAGTCGGCTCTGGAGACATTTCTCTTTCTACAAAGCTGGGAACTGGCTATCTGTTATCTGC CTGTCCAGGTCTGAAAGATTAGGATTGCCCAGGCAGAAACTGGGACTGACCTATCTCACTCTCCCCTGCTTTACCCTTAGGGTGATTCTGGGGGGCCCACTTGTCTGTAATGGTGTGCC GCAGTGAACCATGTGCCCTGCCCGAAAGGCCTTCCCTGTACACCCAAGGTGCATTACCGGAAGTGGATCAAGGACACCATGGTGGCCAACCCCTGA<mark>GCACCCTATCAAGGTCCTATTG</mark> TGTTTCTTARATGGTGTARTTTTGTCCTCTGTGTCCTGGGGARTACTGGCCATGCCTGGAGACATATCACTCACTCTGAGGACACAGTTAGGATGGGGTGTCTGTGTTATTTGTGGGATACGGA GRTGRARGAGGGGTGGGRTCCRCRCTGRGRGRGTGRGRGTGRCRTGTGCTGGRCRCTGTCCRTGRAGCRCTGRGCRGRGGCRCCRCCCCRGRCRCTCCRCRCRCTGRRGRCTGRAGRCCCCCRGCRGGRTGGRGTGRRGCTGRRAGC AT A ACCC ACT OF GOOD GOOD CONTROL OF A CONTROL OF CONTROL OF CONTROL OF A CONTROL OF TATGARGTATTCCCRACTARGGARGCTCACCTGAGCCTTAGTGTCCAGAGTTCTTATTGGGGGTCTGTAGGGATAGGGGTACTGGGATAGCGGACCTGACCTTACTCCAGACCTGAGGTTCCCCAAGAG TTCRAGCAGATACAGCATGGCCTAGAGCCTCRGATGTACAAAAACRGGCATTCATGATGAATCGCRCTGTTAGCATGAATCATCTGGCACGGCCCCAGGCCCCAGGTATACCAGGCACTTGGGCCCGAAT 

![](_page_5_Figure_0.jpeg)

![](_page_6_Figure_0.jpeg)

![](_page_7_Figure_0.jpeg)

![](_page_7_Figure_2.jpeg)

Tertiary: the 3D shape of the<br/>fully folded polypeptide chainQuaternary: arrangement of<br/>several polypeptide chains.

Early History First protein amino acid sequence database in mid 1960's. Researchers developing algorithms to analyze these data for individual research projects in 1960's and 1970's. GenBank and other public databases with freely available analysis tools in the 1980's Enormous growth of GenBank and PDB in 1990's

Data Explosio These research trends, especially the sequencing projects, have resulted in huge amounts of raw data that must be stored, manipulated, and analyzed. Computer information technology has also advanced very rapidly and has provided the means for handling this data explosion.

![](_page_10_Figure_0.jpeg)

![](_page_11_Figure_0.jpeg)

に同じ 北京には 1012LE 1000

## Genome Sizes

教育的自由的人名法德尔 化氯化物 化化物化物 化物物化学 化化物化物化物 化合理合同化合理合理合理 化合理合理 化合理合理

Human Genome Mouse Genome

Fruit Fly Genome

Nematode Genome

Yeast Genome

E. coli Genome

![](_page_12_Figure_7.jpeg)

~3,000,000,000 bp ~160,000,000 bp ~100,000,000 bp ~15,000,000 bp ~5,000,000 bp

>100,000 species are represented in GenBank all species 128,941 6,137 viruses 31,262 bacteria 2,100 archaea eukaryota 87,147 COLUMN REPAIRED

![](_page_14_Picture_0.jpeg)

![](_page_15_Figure_0.jpeg)

の思想など CIACIANIII 如日在田村市 126 (包含香用)(3 SSEE CIACH 行法知道 20日前用目 50 E 5 CACAD 的现在分词 sigis 但已有自己 SSEE IN SUMPRICE REPORT OF THE

![](_page_16_Picture_1.jpeg)

Human Genome Sequencing Centers

оны пострекс серенина устанула соло селоности селоности состанов состановано переда и опередани и раз 1999 година и распарука состани и и постаности состанова состанование и селоности и и и и спортани состани и и и

#### ~3,000 bp (0.0001%) of Human Genome Sequence

GAACCCGACTAGGATCATCGGGAAAAAGGAGGAGGAGGAGGAGGAGGCAGCTCCGGGGAAGCTGGTGGCCAGCGGTCCTGGGTCTGGCGGACCCTGA GAAAGCCGCPAGAGCAAATTTGGGGCCCGGACCAGGCAGCACTCGGCT AGAGCAAAAGGAAGGGGGTGG GELTERGECATTAGEAGELTEGAGECCAGACGGECCTAGCAGGGACCCCGAGAGACCATGCAGAGECCCGGAGAGACCATGCAGAGECCCGGAGAGAAAAGGCCAGCGT AGAAGAGAGAGAAGAATGAACTGAACTGAATTGAATAGAGAGCCACATCTACTTGCAACTGAAAGTTAGAATCTCAAGACTCAAGTACGCTACT ATGCACTTGTTTTATTTCATTTTTCTAGAAACTAAAAATACTTGTTAATAAGTACCTAAGTATGGTTTATTGGTTTTCCCCCCTTCATGCCTTGG TTGGGGTAGGTAGAAAATATATGCTTAAAGTATTTATTGTTATGAGACTGGATATATCTAGTATTTGTCACAGGTAAATGATTCTTCAAAAATTG CTGACATTTAATAAATTGTATTCATAGCCTAATGTGATGAGCCACAGAAGCTTGCAAACTTTAATGAGATTTTTTAAAAATAGCATCTAAGTTCGG AATCTTAGGCAAAGTGTFGTTAGATGTAGCACTTCATATTTGAAGTGTTCTTTGGATATTGCATCTACTTGTTCCTGTTATTATACTGGTGTGA ATGAATGAATAGGTACTGCTCTCTCGGGACATTACTTGACACATAATTACCCAATGAATAAGCATACTGAGGT CAAAAAACTCAAATATGT TCCGGTGCTAAGGAGAGACTGTTGGCCC талслитителелет TGATGTTATCCACCTTTTPSTTACTCCACCTATAAAATCGGCTTATCTATTGATCTGTTTTCCTAGTCCTTATAAAGTCAAAATGTTAATTGGCAT AAATTATAGACTTTTTTTTAGCAGAGAACTTTGAGGAACCTAAATGCCAAC THE A PARTY TO A THE A THE GTTCTAAATACTAATGAACTTTAAAATAGCTTACTATTGATCTGTCAAAGTGGGTTTTTATATAATTTTCTTTTTACAAATCACCTGACACATTT AATATAGGTTAAAAAATGCTATCAGGCTGGTTTGCAAAGAAAATGTATTACAAAGGCTGCTAAGTG **HEADCHAAAATACCAATCA** TTACAAACTTCCTTCTGGCCTTCTGGACTGCAATTCTAAAAAGTGTAAAAAACATATTTTCTGCATTAAGTTAGGCAGTATTGCTTAGTTTTCAAA GTGCTAGCCTTTGCAGTCAGATTATTTTGATTCAGATCCTACATCTACTGTTTAGTAGCTCTGTTGCCTGAGGCAGGTCCCTTAACATCTCTGTG TGTGACTTGACCTTTAAAAATTTGGA ATGTGAAAAGTTAGCCTAATGTTAACTGCTATTATT CAPTCTCATACCCCTTAATCCCTT ATGGATTACCATATTTTCACATTCACCAGTACATGCACCTTGTTAATAAGATGCTCAATTCATCTTTGAGTATAATTTTTGTGACTCTCAAT CTGGATATGCAATGAGTGGGCCTGTATGAGAATTTAATTTATGAAAAATTGTGTTTCACATGGCCTTACCAGATATACAGGAAACACGTCACATG ͲͲͲϹͲϪͲϒ;ϨϷϒ;ϨͲͲϪϪϪϒ;;;ϹϹͲͲϪ;;ϪͲͲͳϪϽ;ϹͲͲͲ;;;;;;ϪϤϲ;;;ϫͲ;;;;;;;;;;;;;;;;;;;;;;;;;;

**Aims of bioinformatics** First Data organization researchers access to existing information submit new entries Second develop tools and resources that aid in the analysis of data Third interpret the results in a biologically meaningful manner.

опоненно спреко се стала да служити на стало слада спос стал свој са денот свој се са лиција Болуници ја риминици ја служи и случи ја си на случе се се се се са случе случе случе случе средски ници.

### "... ORGANISE the information on a LARGE SCALE ..."

CACAMERAGE/REA ACTACA AND CELECOMO A 2010 MULTINACIE COAMERICA ACTENCIACIÓN CURANA

10

TR. THURSDA	Database	URL	人の市上に日1日 2. 11日 火田市の
A REPART	Protein sequence		Tentere Discontinue
IRBIGHIS!	(primary)		adia a france a di
e marente	SWISS-PROT	www.expasy.ch/sprot/sprot-top.html	NUMBER OF STREET
a the second and	PIR-International	www.mips.biochem.mpg.de/proj/protseqdb	and spiral of the
STATICTAVAS	Protein sequence (composite)		SAN SCHOLLLC
10 633.01010	OWL	www.bioinf.man.ac.uk/dbbrowser/OWL	PARTA CAUTIONIC
COLUMN THE	NRDB	www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein	
PANANES	Protein sequence (secondary)		Sherrich and bee
P T (TI ST )	PROSITE	www.expasy.ch/prosite	BITTLE Y IL 25 4 19 2 W IT IT IT
Contraction of the	PRINTS	www.bioinf.man.ac.uk/dbbrowser/PRINTS/PRINTS.html	
1010102020	Pfam	www.sanger.ac.uk/Pfam/	to which we want
A A MARK	Maaramalaaular		UCICICA & COMM
	structures		
ETER	Protein Data Bank (PDB)	www.rcsb.org/pdb	an intersection of the
78 AT 11 15 28 200	Nucleic Acids Database (NDB)	ndbserver.rutgers.edu/	NAMES AND ALL
interest of the	HIV Protease Database	www.ncifcrf.gov/CRYS/HIVdb/NEW_DATABASE	
CERCIC	ReL1Base PDBsum	www2.ebi.ac.uk:8081/home.html	Personal and and
27.0 A E 6	CATH	www.biochem.ucl.ac.uk/bsm/cath	PHILES OF THE X C
married and some of	SCOP	scop.mrc-lmb.cam.ac.uk/scop	
E PAR EPP	FSSP	www2.embl-ebi.ac.uk/dali/fssp	A address to the
CANTARS	Nucleotido sogueneos		SCIENTIS TESTIC
Wat only out	GenBank	www.ncbi.nlm.nih.gov/Genbank	TIME PROPERTY AND ADDRESS OF THE PARTY OF
RABAR	EMBL	www.ebi.ac.uk/embl	addination of the
e maiene	DDBJ	www.ddbj.nig.ac.jp/	的复数的复数形式 机制度的
relieve on one of	C		
Failure 194 bes	Genome sequences	www.nchi.nlm.nih.cow/antraz/auany.fcgi?db=Ganoma	SAMPLE MOUNT
e stadele	GeneCensus	bioinfo.mbb.vale.edu/genome	经公司支付和契约期间
	COGs	www.ncbi.nlm.nih.gov/COG	
POTENCE			A cuto we
E FUR STA	Integrated databases		Pittin VTI 24/22/Thinking
- Internet	InterPro Sequence retrieval system (SRS)	www.ebi.ac.uk/interpro www.expasy.cb/srs5	Commission of the
101072V2V2	Entrez	www.ncbi.nlm.nih.gov/Entrez	to when we we we
N RIM NON		NEW CONTRACTOR AND	THE COMPANY COMMIN

ассернострасовська же тактиксти селохожност солденски и сласов до даниих аменицика интиграсти селоходи и признати селособиета са селоти и са селоти и с

#### "...the INFORMATION associated with these molecules..."

CTAVILLET,	Data source	Data size	Bioinformatics topics	11225/11225/119
e marries	Raw DNA sequence	million	Separating coding and non-coding regions	CONTRACTOR AND
SCH FWH		sequences	Identification of introns and exons	2 4 2 2 2 2 2 2
APICIAS		( binton bases)	Forensic analysis	SAAGE
eniger	Protein sequence	sequences	Sequence comparison algorithms	日本の日本
CMILLON 2ND	-	(~300 amino acids	Multiple sequence alignments algorithms	SCACEN
edade		each)	Identification of conserved sequence motifs	a A GROWING
A STATISTICS	Macromolecular	structures	Secondary, tertiary structure prediction	Same and second process
a states	structure	(~1,000 atomic	3D structural alignment algorithms	and a set of the set
E FUTUES		coordinates each)	Protein geometry measurements	(为11234位2011)
MURLEY,			Intermolecular interactions	sammers and so
NEW LEA			Molecular simulations	ion and the
sinte (est			(force-field calculations,	the state of the second
			docking predictions)	
T AUTO TA .				<b>7636376</b> 首直
REPART	Genomes	complete	Characterisation of repeats	the elminication
an and		genomes	Structural assignments to genes	and man second
和学校 计包括		(1.0 million – 3 billion bases each)	Genomic-scale censuses	24042031
E AMIES		5 childh buses caelly	(characterisation of protein content, metabolic	1111-12/111-12/11/1
e martir er			pathways)	Industry margin
States -			Linkage analysis relating specific genes to	
ADIGA			diseases	SEANCE
E P P C	Gene expression	largest: ~20 time point measurements	Correlating expression patterns Mapping expression data to sequence, structural	対応がななるが開
CALL CLEVE		for ~6,000 genes	and biochemical data	ASCINCIP
e qa ai	Other data			ayend (a)
AGVAVAS	Literature	million citations	Digital libraries for automated bibliographical	4 Counces
EAUER			searches	(为1)发生的支付用1
uniner	Metabolic pathways		Knowledge databases of data from literature	CARDING THE CO
	wietabolie paulways		Pathway simulations	1 Barris
A 12 4 10 24 5				102261010

General Types of .Informatics techniques. Databases Geometry - Building, Querying **Robotics** Object DB Graphics (Surfaces, **Text String Comparison** Volumes) **Text Search Comparison and 3D 1D Alignment** Matching Significance Statistics (Vision, recognition) **Physical Simulation Finding Patterns Newtonian Mechanics** - AI / Machine Learning **Electrostatics** Clustering **Numerical Algorithms** Datamining Simulation

**Bioinformatcs Tools and Services** Databases: text, sequence, structure Database annotation text searches Sequence similarity search tools Sequence and structure analysis tools **3D** structure visualization tools Structure prediction tools Phylogenetic analysis tools Metabolic analysis tools

![](_page_23_Figure_0.jpeg)

Sequence comparison: Gene sequences can be aligned to see similarities between gene from different sources 768 , TGTGTGCATTTAAGGGTGATAGTGTATTT 813 TTCTTTAACACCTC 87 TTGACAGGTACCCAACTGTGTGTGTGCTGAT TTGCTGGCCAAGGACTG 135 814 AGTGTTTGAGCCTCTGTTTGTGTGTGTAATTGAGTGTGCATGT 863 GTGGGAGTG 136 AAGGATC 172 GTGGCGG 864 AAATTGTGGAATGTGTGTATGCTCATAGCACTGAGTGAAAATAAAA 913 173 AAA.TATGGGATATGCATGTCGA...CACTGAGTG..AAGGCAAGATTAT 216

![](_page_25_Figure_0.jpeg)

and any of the particular international states of the second second second second second second second second s

![](_page_26_Figure_1.jpeg)

## **Comparative Sequence Analysis**

3-14:08 20 2

125 3 124 1

-----2011/12/2

Using the 'Experiments of Evolution' to Decode the Human Genome

![](_page_26_Figure_4.jpeg)

Database similarity searching: The BLAST program has been written to allow rapid comparison of a new gene sequence with the 100s of 1000s of gene sequences in data bases Sequences producing significant alignments: (bits) Value gnl PID e252316 (Z74911) ORF YOR003w [Saccharomyces cerevisiae] 7e-26 112 gi 603258 (U18795) Prblp: vacuolar protease B [Saccharomyces ce... 106 5e-24 gnl|PID|e264388 (X59720) YCR045c, len:491 [Saccharomyces cerevi... 69 7e-13 gn1|PID|e239708 (Z71514) ORF YNL238w [Saccharomyces cerevisiae] 0.66 30 gn1 PID e239572 (Z71603) ORF YNL327w [Saccharomyces cerevisiae] 29 1.1 gn1|PID|e239737 (Z71554) ORF YNL278w [Saccharomyces cerevisiae] 29 1.5 gnl PID e252316 (Z74911) ORF YOR003w [Saccharomyces cerevisiae] Length = 478112 bits (278), Expect = 7e-26 Score =Identities = 85/259 (32%), Positives = 117/259 (44%), Gaps = 32/259 (12%) Ouery: 2 OSVPWGISRVOAPAAHNRG-------LTGSGVKVAVLDTGIST-HPDLNIRGG-ASFV 50 G G GV VLDTGI T H D + PWG+ RV 174 EEAPWGLHRVSHREKPKYGODLEYLYEDAAGKGVTSYVLDTGIDTEHEDFEGRAEWGAVI Sbict: Ouery: 51 PGEPSTODGNGHGTHVAGTIAALNNSIGVLGVAPSAELYXXXXXXXXXXXXXXXXXXXXXXOGLE 110 Private D NGHGTH AG I + + GVA + ++ +G+ESbict: 234 PANDEASDLNGHGTHCAGIIGSKH----FGVAKNTKIVAVKVLRSNGEGTVSDVIKGIE 288

Dependence of proteins from different organisms can be a to see similarities and differences Dependence of proteins from different organisms can be a to see similarities and differences	(S. LAN	CHEFRES	CONSCIENCES (CA)	IDACIUM GERICACIZZA	SCARDENC.	
Description       Operations of proteins from different organisms can be a to see similarities and differences       109Y:C_sa     10       100Y:C_sa     10       11     10       11     10       11     10	19834	EAADTEN	BIGMAGNACINAC	nonnyennu zyn se zalei	<b>WASSER</b>	tere re portanisti
Sequences of proteins from different organisms can be a to see similarities and differences	in cimi	ient:	ce alignm	ne seguen		in electronet
Sequences of proteins from different organisms can be a to see similarities and differences	1	iont.	oc angini	ne sequen	IVIG	
Decide in the protein is from unretent of gams is called a construction of the protein is from unretent of gams is called a construction of the protein is from unretent of gams is called a construction of the protein of the protein is called a construction of the protein of the	anoc	n ho aliar	ont organisms of	protains from differe	auoncoa	ACCE
to see similarities and differences     109Y:C_ss   10   Ender Fill     109Y:C_10   PALDSLALDTIRGELLITLAELERILDAETIET-GISEG-HATLEGGVARGELWDVEG-HALGUET     EHEC_Sep0   232   LSNETVMNRIYCAIGTVHIDIHMLRNVKKDDINSGYHLFGGCQLIRNNTTLAYGSIVKINEDFYFTVST     Yp_Ysc0   226   VOLBDIPOTVME GRITHPICE KGLAVGOTLACO-THCYGEVNICINGGVGASGELWDVEGNEUVIRAQ     CL_Ysc0   322   ATPLPSATK VAEVARYSISVGETLAKGESVLIGGESVLOPGHPTIGVNILINAGKVGSSILARDEVLGINLAG     At_AGRO988p   16   ENMOPTAVCIGTSRMLVSGIMSLEGGATIALD-RKIGEPVNINNGRLAGETVLEDDDTFGVKILE     Ba_P11N   53   TILMVPNINSLVIGGRANKSVAOLQUGGGSTELD-RKVGEADDIVNNRUVARGEVVVDEKLGVTBE     Cc_F11N   23   APVPDVVNISLGVIGRANNSVAOLQUGGGSTELD-RKVGEADDIVNNRUVARGEVVVDENELGVTBE     Cc_F11N   25   SOVINAVIVEISVIGGSTINEVOQULRAGGAVPLD-AKTHDEWNILANNHFLARGEQUEVVDDRLGVTBE     Cj_F11N   18   EDILDITVDFVSELGTINNSVAELKEVGSVIDLE-RPAGE-SVELVINNRUVARGEVVDDRLGVTBE     Cj_F11N   18   EDILDITVDFVSELGGTINNSVAELKEVGSVIDLE-RPAGE-SVELVINNRUTGGELVRUDDRLGVTBE     Cj_F11N   18   EDILDITVDFVSELGGTINNSVAELKEVGSVIDLE-RPAGE-SVELVINNRUTGGELVRUDDKLGVETHS     Fa_BCC_Ba0   254	yneu	in be aligi	sin organisms ca	proteins nom unere	equences	THE PLACE
109Y:C_ss 10	and you a	and was placed	nd differences	to see similarities a	Internimente	minen er chinie
109Y:C_ss 10 109Y:C_10 PALOSIAIDITRCGEIRITAEIRRIDAGTIEVT-GISPG-NATICHGEVVXBCELVDVECRELOOTR EHEC_Sep0 232 LONETVMURIYCAIGTVHIDIHMIRVVKKDDIINSDGYHLEG-GCQLIRNNTTIAYGSIVKINEDFYFYJST Yp_Ysc0 236 VOLED LPOTIVMEIGRITIPIGEIKQLAVGDIACO-THCYC-EVNICLMOGVGRGSLERCDEKLVVRIAQ CC_Ysc0 302 ATPLESATKIVAEVARYSLSVGEFIKLGPGSVLOFDGYHLFIGVDIINACKVGRGNIIALODVLCIRVLE ALAGRO980P 116 EMMODIDVOIVJGTSRMLVSGIMSLEEGATIALD-RHIGEPUEINVGRATAGGEVVVLEDKLVVRIAQ CC_FIN 33 APVPVVVNISVLGTSRMLVSGIMSLEEGATIALD-RHIGEPUEINVGRATAGGEVVVLED	14 4 7 1	and the second the				THE FLORE STOR
109Y:C 10 PALDSLALDITLRCGELRLTLAELRLDAGTILEVT-GISPG-HATLCHGEQVVAEGELVDVEGRLGLOTTR EHEC_SepQ 232 LSNETVMIRITCA.GTYHIDIHMLRWYKNDDI.NSDGYHLFGGCQLIRNNTTLAYGSIYK.HNEDFYFTYST Yp_YacQ 266 VQLEBLPOTUME GENTLTLGAGTYLDAACQ-THCYGEVNICLAGGSVGRGSLACDEKLVVRIAQ Ct_YacQ 302 ATPLPSATKIVAEVARYSLSVGEFLKLGPGSVLQFDGVHPTLGVDIILMQSVGRGSLIALDDVVGIRVLE At_AGRC989p 116 EMIMDIFTDVQIVLGTSRMLVSGIMSLEEGATALD-RKIGEPVEINVMORRIARGEITVLEDDDTRFGVKLIE Ba_FIN 53 -TILWYVNITVEIGTSRNLVSGIMSLEEGATALD-RKVGEAIDIYVNNGRIARGEITVLEDDDTRFGVKLIE Cc_FIN 22 SQVNAVIVEISVLGRSILHVQQILAKGRGAVIPLD-AKTHDEVWILANNHPIARGEIGISDDRLIQVTME Cc_FIN 22 SQVNAVIVEISVLGRSILHVQQILAKGRGAVIPLD-AKTHDEVWILANNHPIARGEIGISDDRLIQVTME Cc_FIN 22 SQVNAVIVEISVLGRSILHVQQILAKGRGAVIPLD-AKTHDEVWILANNHPIARGEIGISDDRLIQVTME Cj_FIN 199 NLIMVYRDIPVKVR (GNIKMLKDVIVTADIGSVVELM-QLANDEVWILANNHPIARGEIGISDDRLIQVTME Cj_FIN 18 EDILDITVDFVSELGTTNMSVAELLKLEVGSVIDLE-KPAGESVELYINKRIFGKGEVMVYEKNLAIRINE EHEC_EPAO 254ELVVKIFFVLGKKINNIYEIDELCARRISLDD-SALDGSVELYINKRIFGKGEVMVYEKNLAIRINE EHEC_EPAO 254ELVVKIFFVLGKKINNIYEIDELCARRISLDD-SALDD-SVUCGIGELVLDDKLGVENTE Ro_HrcQ 276 VPLEQLEVPVHLEIAVMGMPLAELAALQQHVITLPVKIRDVSVRLVCHGQTLGGGLVAVGEQLGLQIAS St_SsaQ 245 VELEQIPQQUFEVGRASLEIGURQLKTRDDVIPVG-GCRPAEVTIFXNDRIIGGGELLACCMEFMVRITRI St_SpaO 225 PCINQIVEFVGKURVIXINTINTALELAKGQQUSIEF-TNAELNUTEINXNDKILGNGELVGMDDKLGVENTE Yp_FIN 57 SLFSRIPVTITLEVASVEIPLSELLTVINDSVIELD-KLAGEPLDIRVNDRILGGELLACCMEFWRITRI A_1100701 2KINTTYLGVGRANKTIGEILAMGIGSVEIDD-SVRLKKGVERELAKGELVIDGKUGVKIK LI_LIN0706 441 QILEDIPVIEFVGRAVKKETVFIE-DSTGNVVEITIGKGILVNDDHGIQURE LI_LIN0708 441 QILEDIPVIEFVGRAVKKETVFIEJSSLEVVFILVEVVILAINGEVITKKGELTKDGKUGVKIK LI_LIN0708 441 QILEDIPVIEFVGRAVKKEKFISSCEKDVIILK-ESTME-PLVALANGVTIGKGILVRDDHGGIQME AL-FIN 237 QVKR-SQYTLEARKKETUERTSRLVAGDVYFF0-LINQDDIGVEVSANGSKYNCEFGKSGDKUAVGVSQI TJ_FIN 255 ALLSGVSQNIVYEGAKKKEKFISSCEKDVIILK-SSTDPLVILADNGVIGKGILVRDDHGGIQME RM_HrcQ 279 EDLDVFKMUPFIALEKTISRLVAGDVYFF0-LINQDIG	83	EFFERE-FE	-EFFERE		10	109Y-C 55
EHEC_SepQ232LSNETWMNRTYCAIGTVHIDIHMLRNVKKDDIINSDGYHLFGEGQLIRNNTTIAYGSIVKINEDFYFTVSIYp_YacQ226VQLEDLPQTIVMEIGRLTJPLGEIKQLAVGQTIACQ-THCYGEVMICINGQSVGRGSILRCDEKLVVRIAQCt_YacQ302ATPLIPARK VAEVARYSLSVGFLKLGGCSVLQFDGVHPTLGVDIILMGRKVGRGNIIALQDVLGIRVLEAt_AGRC988p116EMIMDIPIDVQIVGTSRMLVSGLMSLEEGATIALD-RKIGEPVSIMVNGRRIARGEITVLEDDTRFGVKLEBa_F11N53-TTLNVPVNITVEIGTSRMLVSGLMSLEEGATIALD-RKVGEAIDIYVNNRLVARGEVVVDEKVGIRITNCt_F11N23SQVMAVVCISVLGRSINKIKOFLKPSKGSMIIL-SSKEPLNIFKNGHLIASGEIVVDG	TR 83	RLGLOITRLYTR	-HATLCHGEOVVAEGELVDVEG	ELRLTLAELRRLDAGTILEVT-GISPG-	10 PALDSLALD	109Y:C
Yp_YscQ 226 VOLEDLPQTLVMEIGRLTLPIGEIKQLAVGQTLACQ-THCYG-EVNICLMGQSVGRGSLLRCDEKLVVRLAQ CC_YscQ 302 ATPLPSATKIVAEVARYSLSVGEIKLGPGSVLQFDGVUPTIGVDILMGAKVGRGNILALQDVLGTRVLE At_AGRC988p 116 EMINDIPIDVQTLVGTSRMLVSGIMSLEEGATIALD-RKIGEPVEIMVMGRIARGETVLEDDTNFGVKLIE Ba_F11N 53 -TILMVPVNITVEIGTSRMLVSGIMSLEGATIALD-RKIGEPVEIMVMGRIARGETVLEDEKYGIRITM Ci_F11N 33 APVFDVFVNISAVLGRANMSVAQLLQLQGGSILELDRKVGEADDIYMMRUARGEVVVDERLGVTMTE CC_F11N 22 SOVMAVVETSVLLGRSILFMQQLLRNGGACAVIPLD-AKTHDEVVILANMFPIARGETVVLED	D- 305	-DFYFTVSIVCD-	-GCOLIRNNTTIAYGSIVKINE	TVHIDIHMLRNVKKDDIINSDGYHLFG-	232 LSNETVMNR	EHEC SepQ
Ct_yscQ302ATPLPSATKIVAEVARYSLSVGEFLKLGPGSVLQFDGVHPTLGVDIILNGAKVGRGNIIALQDVLGIRVLEAt_AGR0980116EMIMDIPIDVQIVLGTSRMLVSGLMSLEEGATIALD-RKIGEPVEINVNGRIARGSITVLEDDDTREGVKLEBa_FIIN53-TILNVPVNITVELGTSRKLKNFLKFSKGSMIIN-ESKE-PINNFMALIASGITVLEEDRKGIRITNCi_F1IN33APVFDVPVNISAVLGRANMSVAQLLQLGQGSILELD-RKVGEALDIYVNNRLVARGEVVVDERLGVTMECc_F1IN22SQVNAVNVEISVLGRSILPMQQLLANGRGAVTPLD-AKTHDEVWILANNHPIARGEIQISDRLAQVTMECj_F1IN18EDILDITVDFVSELGTINNSVAELLKLEVGSVTDL-AKTHDEVWILANNHPIARGEIQUSDNRGQUTECj_F1IN18EDILDITVDFVSELGTINNSVAELLKLEVGSVTDL-AKTHDEVWILANNHPIARGEIQUSDNRGQUTECj_F1IN18EDILDITVDFVSELGTINNSVAELLKLEVGSVTDL-AKTHDEVWILANNHVIGGEUGUNVEKNLAIRINEEHEC_Epa024ELPVKIEVLGKKINNIYEIDELCARRIISL-PESEKNIEIEVNAALTCYGELVEVDDRLGVENTSPa_pscQ233HELDQIPPVSFEVGRRTLDLHISTLOPGSLLDL-SALGC-EVRILANQCLGIGELVRLQDRLGVENTSSt_SsaQ245VELEQIPOQULFEVGRASLEIGQIRQLKTGDVIPVG-GCFAPEVTIRVNDRIIGQGELIACGNEPMVRITRSt_SsaQ245VELEQIPOQULFEVGRASLEIGQIRQLKTGDVIPVG-SVRLVHMANVLLGNGELVQNUDTLGVENTSYp_F1IN57SLFSRIPVTITEVLSVEPLSLLVNNDSVIELDNAECEPDDIVENSINFGQAUVVINDRLGVKIKSLi_Lin070128RQVDNIGVNIIVFCKKEMPVGIARLSIGSVIED-REPKDLVDIKVMGKLIAKGELVIIDGKNGVKIKSLi_Lin0703441QILEDIPVTLEVVFGTAKKKLEKFISWCEKDVIILK-ESMNEPLVIALMOVTIGKGILVRUDDHFGIRMSML_HrcQ279EDLDDVEZMIVFCGWPPPLAGILMQPGDVIPFL-REPLD-PVDIIANGCIGRGITVRGGVAALNINTSVp_F11M250KINTMAPUNTERKUPEGRWPPPLAGILMQPGDVIPFL-PLTELSULVGQPIFEAAPGERGGQAALNINTS<	LQ 299	-KLVVRIAQWGLQ	-EVNICLNGQSVGRGSLLRCDE	RLTLPLGEIKQLAVGQTLACQ-THCYG-	226 VQLEDLPQT	Yp_YscQ
At_AGRe988p   116   EMIMODYPIDVOLVUGTSKMLVSCIMSLEEGATLID-RKIGE-PVEIMVMGRTARGEIV/LEDDDTRFGVKIE     Ba_F11N   53   -TILNVPVNITVELGTSRIKIKDFIKFSKGSMLILD-RKIGE-PLNIFMNGHLIASGEIV/LEDKYGIRITN     Ci_F11N   33   APVFDVPVNISAVLGRANKSVALLLGGGSILELD-RKVGE-AZDIYVNNELVARGEVVVVDEKIGVTMTE     Cc_F11N   22   SOVNAVNVEISVLGRSILPMOQLLRMGRGAVIPLD-AKTHDEVWILANNHPIARGEIQISDDRIAIQVTM     Cj_F11N   199   NLIMDVRLPVKVRIGNKIMLLKDVITMDIGSVUELN-QLANDPLEILLGDKRIAVGEVVIVDCNFGVQITE     Cj_F11N   18   EDILIDTVDFVSELGTTNNSVAELLKLEVGSVIDLE-KPAGE-SVELYTNKRIFGKGEWVVEKNKIGVETHSI     Pa_PscQ   233   HEDQUPIPVSEVGRRTIDHATISTLQPGSLIDD-SALDG-EVRILANGRCIGIGEIVRLOPKIGVRVTR     Rs_HrcQ   276   VELGUEVPVHLELAVKMVPLAELBAMQQQLLSLPGGCFAP-EVTIRVNRIIGGGUELVGHOUKIGVRVTR     St_SsaQ   245   VELGUIVPVHLELAVKENPLAELALQPOHVTIPVKIRDV-SVRLVCHGQLANGECTGVRTR     Yp_F11N   57   SLFSRIPVTITLEVASVETPLSELTVNNDSVIELD-KLAGEPLDIRVNGRIGGGEVUNDTKGVRTR     Yp_F11N   57   SLFSRIPVTUTLEVASVETPLSELTVNNDSVIELD-KLAGEPLDIRVNGRMGVLGNGGAVVVNDEKYGLRTIN     Aa_F11N   30   HESDIPVCVVGRANKTIGELLANGOV PVG-DLKGEV-SUNGKINGTGKGILAKGEVVIIDENFGTVISE     Li_Lin0701   28   RQVDNIGVNITVRIGKKEMPVGDIALSIGDVIED-KVGFIVVKITSGKCIGTGEITKDGKWYGRIKTR     Li_Lin0703   441   DIEDDVENUVYGGSLK	373	-VLGIRVLEV	-GVDIILNGAKVGRGNIIALQD	RYSLSVGEFLKLGPGSVLQFDGVHPTL-	302 ATPLPSATK	Ct_YscQ
Ba_Flin53 -TILNVPUNTVELGTSRIKIRDFLKFSKGSMLILN-ESIKE-PINTFMGHLIASGEUVLEEKYGIRITMCi_Flin33 APVFDVPVNISAVLGRANMSVAQLLQLGQGSILIN-ESIKE-PAIDIYVNNRLVARGEVVVDERLGVTMTECc_Flin25 OVNAVIVEISVLGRSILPMOQLLRMGGAVTPLD-AKTHDEVVILANNGFIARGEVIDGRLATOTRCj_Fliy199 NLIMDVRLPVRVIGKKMLLKDVLTMDIGSVVELN-QLANDPLEILIGDKRIAYGEVVIDGNFGVQITECj_Flin18 EDILDITVDFVSEJGTTNNSVAELLKLEVGSVIDLE-KPAGESVELYINKRIFGKGEVMVYEKNLATRINEEHEC_Epa0254ELPVKIEFVIGKKIMLYPIDELCAKRITSL-PESEKNIEIRVNGALTGYGELVEVDDKLGVETHSPa_PacQ233 HELQQIPIPVSFEVGRRTLDIHTISTLOPGSLLDLD-SALDCEVTRILANORCLGIGELVRLDDKLGVETHSSL_SsaQ245 VELQIPOVLEEVGRASLEIGURQLKIGDVIPVG-GCFAP-EVTRVNDRIIGQELAGHEFMUTARSSL_SsaQ245 VELQIPOVLEEVGRASLEFVLYRKWTLAELEAAGQQOLLSIP-TNAELNVEIMANGVLGNEELVGMDTLGVETHSYp_Flin57 SLFSRIPVTITLEVASVEIPLSELLTVNNDSVIELD-KLAGEPLDIRVNGRILGNEELVGMDDTLGVETHSYp_Flin57 SLFSRIPVTITLEVASVEIPLSELLTVNNDSVIELD-KLAGEPLDIRVNGRILGREELVINGNA_F11N37 QHFSDIPVEVEVVVGRAANKTLGELLAMGIGSVIELD-KLAGEPLDIRVNGRILAKGELVIIDGKIGVKIKSLi_Lin07062KINHTIPLRIDFELGRTKQPVGSILDVKKGTVFRLE-DSTGHVVKITISGKCIGGELIKDGKIGVKIKSLi_Lin0708441 QILEDIPVLEVVFGARVKLEKFISKCEKDVIILK-ESMNEPLULAINGYIGKGILVRUDDHFGIQNTESLi_Lin0708441 QILEDIPVLEVVFGARVKLEKFISKCEKDVIILK-ESMNE-PLULAINGYGCIGRGDIVRIGDTLGRRIGML=F11M250250VKR-SOVTLERIKLEFTITRTISTIRLVADVTFFQ-DLKODDIGVEVSANGSKLIVCEFKSGDRXMVRVKNHg_F11M252AILS-IPLTISARLCEPEVPLRQLMQMQPGDVIPVH-LTEALSLVEGQPIFEAAPGERGGQALINITRVVp_F11M252AILS-IPLTISKRLEFTISKEEKDVIILK-SSIND-PVVLIANGCCIGRGDIVRIGDTLGRRIGML	ST 192	TRFGVKLIEVLST	-PVEIMVNGRRIARGEITVLEDDI	TSRMLVSGLMSLEEGATIALD-RKIGE-	116 EMIMDIPID	At_AGRc988p
Ci_FliN 33 APVFDVPVNILSAVLGRANMSVAQLLQLGQGSTLELD-RKVGEAIDIYVNNRLVARGEVVVVDERLGVTMTE Cc_FliN 22 SOVNAVNVEISVLGRSILPMOQLLRMGRAVIPLD-AKTHDEVWILANNHPIARGEIQISDDRLGVTMTE Cj_FliN 19 NLIMDVRLPVRVRIGNKKMLLKDVITMDIGSVVELN-QLANDPLEILGGKRIAYGEVVVVDGNFGVQITE Cj_FliN 18 EDILDITVDFVSELGTINMSVAELLKLEVGSVIDLE-KPAGESVELYINKRIFGKGEVMVYEKNLAIRINE EHEC_Epa0 254ELPVKIEFVLGKKIMNLYEIDELCARRISL-PESEKNIEIRVNSALTEYGELVEVDDRLGVRTR Rs_HrcQ 276 VPLEQLEVPVHELAVMGMPLAELAALQPQNVITHPVKIRDVSVRUCHQQTLGGGELVRLQDRLGVRTR St_SsaQ 245 VELEQIPQQVLFEVGRASLEIGQLRQLKTGDVLPVG-GCFAPEVTIRVNSALTEYGELVEVDDRLGVRTR St_Spa0 225 PGLNQLPVKLEFVLYRNVTLAELEAMGQQLISLP-TNAELNVEIMANSVLLGNGELVAVGEQLGLQTAS St_SsaQ 245 VELEQIPQQVLFEVGRASLEIGQLRQLKTGDVLPVG-GCFAPEVTIRVNDRILGQGELIACGNEFMVRITR St_Spa0 225 PGLNQLPVKLEFVLYRNVTLAELEAMGQQLISLP-TNAELNVEIMANSVLLGNGELVAVGEQLGLQTAS St_SsaQ 245 VELEQIPQQVLFEVGRASLEIGQLRQLKTGDVLPVG-GCFAPEVTIRVNDRILGQGELIACGNEFMVRITR St_Spa0 225 PGLNQLPVKLEFVLYRNVTLAELEAMGQQLISLP-TNAELNVEIMANSVLLGNGELVONDDTLGVRIHE YD_F1N 57 SLFSRIPVTITLEVASVEIPLSELITVNNDSVIELD-KLAGEPLDIRVNGIMFGQAEVVVINEKYGLRIN Aa_F1N 37 QHFSDIPVEVVVVGRANKTLGELLAMGIGSVFED-REPKDLVDIKVNGKLIAKGELVINDGKIGVKKE Li_Lin0706 2 KINHTIPLRIDFLGRTKQPVGSLLDVKKGTVFRLE-DSTGNVVKITISGKCIGQGELITKDGKNFVKITK Li_Lin0706 2 KINHTIPLRIDFLGRTKQPVGSLLDVKKGTVFRLE-DSTGNVVKITISGKCIGYGELITKDGKNFVKITK Li_Lin0706 2 KINHTIPLRIGEGRNPIPLGELRSAGEGHIFFLG-RFIQDPVDILANGQCIGRGDIVRUDDHFGIQNTES Li_Lin0706 2 KINHTIPLRIGEGRNPIPLGELRSAGEGHIFFLG-RFIQDPVDILANGQCIGRGDIVRUDDHFGIQNTES Li_Lin0706 2 KINHTAPLKGGRNPIPLGELRSAGEGHIFFLG-RFIQDPVDILANGQCIGRGSLYNCEGKSGDRYMVRVKN HP_F11M 253 KRIMTAQIPVVAELGTSELTIEFFLSLEVGDCITVFN-LTKALSLUFGQPIFEAAPGERGGQAALNITR VD_F11M 254 ALLS-FVTIRQLEKKDSLGVGDUPVH-LTKALSLUFGQPIFEAAPGERGGQAALNITR HP_F11M 255 KELMTAMPIVAEIGEVKLKVREILSLDKGDVINLE-SSLINK-DLTKVGTKKKFKCRMGLMGHKVSVQITEF TP_F11M 255 ALLSGVSVDMIVFLGAVELSLKEMDLDVGDTIRH-THVHCWDC-PFVVSLGRKKKYLASVGFQGYKLAQVGID Bb_F11M 255 ALLSGVSVDMIVFLGAVELSLKEMDL	NS 125	-KYGIRITNIKNS	-PLNIFMNGHLIASGEIVVLEE	TSRIKIKDFLKFSKGSMLILN-ESIKE-	53 -TILNVPVN	Ba_FliN
Cc_FliN22 SQVNAVVEISVLGRSILPMQQLLMMGRGAVIPLD-AKTHDEVWILANNHPIARGEIQISDDRIAUQVTRCj_FliY199 NLIMDVRLPVRVRIGNKKLLKDVLTMDIGSVVELN-QLANDPLEILIGDKRIAYGEVVIVGNFGVQITECj_FliN18 EDILDTVDFVSELGTIMSVAELLKLEVGSVIDLE-KPAGE-SVELVINKRIFGKGEVMVYEKNLAIRINEEHEC_Epa0254ELPVKIEFVLGKKIMNLYEIDELCAKRIISLL-PESEKNEIRVNGALTGYGELVEVDDKLGVETHSPa_PcQ233 HELDQ1PIPVSEVGRRTLDLHTSTLQPGSLLDLD-SALDG-EVRILANQRCLGIGELVRLQDRLGVRTRRs_HrcQ276 VPLEQLEVPVHELAVMGMPLAELAALQPQHVITLPVKIRDV-SVRLVCHQTGHGQLVAVGEQLGLQIASSt_SsaQ245 VELEQIPQQULFEVGRASLEIGQIRQLKTGDVLPVG-GCFAP-EVTIRVNDRIIGQEELVACOEFMURIRISt_Spa0225 PGLNQLPVKLEFVLYRKNVTLAELEAMGQQQLLSLP-TNAELNVEIMANGVLLGNGELVQMNDTLGVETHEIYp_FliN37 SLFSRIPVTITLEVASVEFPISELTVNNDSVFLD-KLAGEPLDIRVNGINFGQAEVVNNEKYGURINAa_F11N37 QHFSDIPVEVEVVVGRANKTIGELLAMGIGSVFETD-KLAGEPLDIRVNGINFGQAEVVNEKYGURINLi_Lin07062 KINHTPIRIPFEGRTKQPVGSLDVKKGTVFRLE-DSTGNVKITISGKCIGYGEILMDGKIGVKIKELi_Lin07062 KINHTPIRIPFEGRTKQPVGSLDVKKGTVFRLE-DSTGNVKITISGKCIGYGEILTNDGTIGRIGMEMLHrcQ279 EDIDDVENUVFGGRAVKLEKFISWCEMVILKGTVFRLE-DSTGNVKITISGKCIGYGEITMGGDTIGRIGMEMLHrcQ270 VKR-SOVTLEARIKLETITLRTISRLVAGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGD	KD 106	-RLGVTMTEIIKD	-AIDIYVNNRLVARGEVVVVDE	RANMSVAQLLQLGQGSILELD-RKVGE-	33 APVFDVPVN	Ci_FliN
Cj_Fliy199NLIMDVRLPVRVRLEVKKMLLKDVLTMDIGSVVELN-QLANDPLEILIGDKRIAYGEVVIVDGNFGVQITECj_FliN18EDILDITVDFVSELGTTNMSVAELLKLEVGSVVDLE-KPAGESVELYINKRIFGKGEVMVYEKNLAIRINEEHEC_Epa0254ELPVKIEFVJGKKIMNLYEDELCARRIISLI-PESEK-NTEIRVNGALTGYGELVEVDDKLGVRVTR.Rs_HrcQ276VPLEQLEVPVHELAVMGMPLAELAALQPQRVLTPVKIRDVSVRLVCHQTLGHGQLVAVGEQLGLQIAS.St_SsaQ245VELEQIPQVLFEVGRASLEIGQIRQLKTGDVLPVG-GGFAPEVTIRVNDRIIGQGELACGNEFMVRITR.St_SsaQ245VELEQIPQVVLEFVGRASLEIGQIRQLKTGDVLPVG-GGFAPEVTIRVNDRULGNGELVQMNDTLGVEIHELYp_FliN57SIFSRIPVTITLEVASVEIPLSELITVNIDSVIELD-KLAGEPLDIRVNGIMFGQAEVVVINEKYGURIKEYp_FliN73QHFSDIPVEVEVVVGRANKTIGELLAMGIGSVIEID-REPKDLVDIKVNGKLIAKGELVIIDGKIGUKIKELi_Lin070128RQVDNIGVILIVRLGKKEMPVGDIAELSIGOVLEVE-KKPGHKVEIFLDEKKVGIGGALIMDENFGUVISELi_Lin07062 KINHTPLRIDFELGRTKQPVGSLLDVKKGTVFRLE-DSTGNVVKITISGKCIGYGELITKDGKFGUVISELi_Lin07062 KINHTPLRIDFELGRTKQPVGSLLDVKKGTVFRLE-DSTGNVVKITISGKCIGYGEILTKDGKFGUVISEML_HrcQ279EDLDDVEIMUVFECGRWPIPLGELRSAGEGHIFELG-RPIQDPUTLANGQCIGRGDIVRIGDTLGIRLRGIAt_F1iM237QVKR-SQVTLEARIKLETTIRTISKLVAGOVIPFQ-DLKQDDIGVEVSANGSKLWICEFGKSGDKHAVGVSQTp_F1iM253KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKQAVQILDFBb_F1iM263EKLENTAMPUVAELGESKLKVREILSLDKGDVINE-SSLINK-DLTKVGTKEKFKCRMGLMGNKVSVQITEFTp_F1iM252ALSGVSVDMIVVLEGSLRLSVRDILGLEVGDIRLR-HVNUKGSLVGHGELVSIEDKLAAQITEFBb_F1iM253KRIMTAQIPVVAELGESKLKVREILSLDKGDVINE-SSLINK-DLTKVGSVGVGKCKLAAQILEF	DV 95	-RIAIQVTRAADV	-EVWILANNHPIARGEIQISDD	RSILPMQQLLRMGRGAVIPLD-AKTHD-	22 SQVNAVNVE	Cc_FliN
Cj_Film 18 EDILDIVDFVSELGTTMMSVAELKLEVGSVIDLE-KPAGESVELYINKRIFGKGEVMVYEKNLAIRINE EHEC_Epa0 254ELPVKIEFVLGKKIMMLYEIDELCAKRIISLL-PESEKNIEIRVNGALTGYGELVEVDDKLGVETHS Pa_PacQ 233 HELDQJPIPVSFEVGRTIDLHTISTLQPGSLIDLD-SALDGEVRILANQRCLGIGELVRLQDKLGVETHS St_SsaQ 245 VELEQIPQ0VLFEVGRASLEIGQLRQLKTGDVLPVG-GCFAPEVTIRVNDRIIGQGELACGNEFMVRITRI St_Spa0 225 FGLNQIPVKLEFVLYRKNVTLAELEAMGQQLISIP-TNAELNVEIMANGVLLGNGELVQMNDTLGVETHEI Yp_F1in 57 SLFSRIPVTITLEVASVEIPLSELITVNNDSVIELD-KLAGEPLDIRVNGIMFGQAEVVVINEKYGLRIN Aa_F1iN 37 QHFSDIPVEVEVVGRANKTIGELAMGIGSVIEID-KLAGEPLDIRVNGIMFGQAEVVVINEKYGLRIN Li_Lin0701 28 RQVDNIGVNLIVRLGKEMPVGDIAELSIGDVLEVE-KKPGHKVEIFLDEKKVGIGEAILMDEKNFVKITKS Li_Lin0706 2 KINHTIPLRIDFELGRTKQPVGSLLDVKKGTVFRLE-DSTGNVKITISGKCIGYGEILTKDGKMFVKITKS Li_Lin0708 441 QILEDIPVTLEVVFGTAKVKLEKFISWCEKDVIILK-ESMNEPLVJALNGQCIGRGDIVRIGDHFGIQMTES ML_HrcQ 279 EDLDDVEIMLVFECGRWPIPLGELRSAGEGHIFELG-RPIQDPVJILANGQCIGRGDIVRIGDTLGVRIRG At_F1IM 237 QVKR-SOVTLEARIKLEFIITIRTISRLVAGDVIPPQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVRVKN Hp_F1IM 252 AILS-TPITLSARLCEPEVPLRQLMQMQPGOVLPVH-LTEALSILVEGQIFFGKSGDKLAVQVSQI Tp_F1IM 253 KRIMTAQIPVVAELGTSEITIEEFISLEVGDCITID-KSVTDPLTVLVGDREKFLGQAGRVNRKAAVQIDD Bb_F1IM 253 KRIMTAQIPVVAELGTSEITIEEFISLEVGDCITID-KSVTDPLTVLVGDREKFLGQAGRVNRKVAVQIDD Bb_F1IM 254 ALLSGVSVDMIVFLGAVELSLKKVREILSLDKGDVLNLE-SSLINK-DITLKVGTKEKFKCRMGIMGHKVSVQITE Tp_F1IM 255 ALLSGVSVDMIVFLGAVELSLKKMDLDVGDIRILH-DTHVGD-PFVLSIGNKKKFLQPGGVGKKLAAQUSQI FD_F1IM 252 ALLSGVSVDMIVFLGAVELSLKMDLDVGDIRIN-FKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSI Bp_SctQ 231 TRIGELEPVQFE DTVSSPIDQLSALEPGVVIEP-VAVTD-ARLR.VHGQTVGYGGELVAVGEHLGVRIIR Ea_HrcQ 259 PQLASLPLSLEVRCDRTALTIGELQRUQAGSVVTID-NVTPCEAGLYHGDTLARGELVDVEGHLGVRIIR FALTCO 259 PQLASLPLSLEVRCDRTALTIGELQRUAGSVVTEND-NVTPCEAGLYHGDTLARGELVDVEGHLGVRIIR FALTCO 259 PQLASLPLSLEVRCDRTALTIGELQRUAGSVVTEND-NVTPCEAGLYHGDTLARGELVDVEGHLGVRIIR FALTCO 259 PQLASLPLSLEVRCDRTALTIGELQRUAGSVVTEND-NVTPCEAGLYHGDTLARGELVDVEGHLGVRIIR FALTCO	SK 272	-NFGVQITEIGSK	-PLEILIGDKRIAYGEVVIVDG	NKKMLLKDVLTMDIGSVVELN-QLAND-	199 NLIMDVRLP	Cj_FliY
EHEC_ppa0   254  ELPVKIEFVLGKKIMNLYEDELCARRIISLL-PESEKNIEIRVNGALTGYGELVEVDDKLGVEIHS     Pa_BcQ   233   HELDQIPIPVSFEVGRRILDLHI.STLQPGSLIDLD-SALDGEVRILANQRCLGIGELVRLQDRLGVRVTR     Rs_HrcQ   276   VPLEQLEVPVHLELAVMGMPLAELAALQPGRUTIPVKIRDVSVRLVCHGQTLGHGQLVAVGEQLGLQIAS     St_SsaQ   245   VELEQIPQQVLFEVGRASLEIGQLRQLKTGDVLPVG-GCFAPEVTIRVNDRIIGQGELIACGNEFMVRITRI     St_Spa0   225   PGLNQIPVKLEFVLYRKNVTLAELEAMGQQQLLSIP-TNAELNVEIMANGVLLGNGELVQMINDTLGVEIHEI     Yp_F11N   57   SLFSRIPVTITLEVASVEIPLSELITVINDSVIELD-KLAGEPLDIRVNGKLIAKGELVINDGKIGVRIKEI     Na_F11N   37   QHFSDIPVEVEVVGRANKTIGELLAMGIGSVIEID-REPKDIVDIKVNGKLIAKGELVINDGNFGVKIKEI     Li_Lin0706   2   KINHTIPLRIDFELGRTKQPVGSLLDVKKGTVFRLE-DSTGNVVKITISGKCIGYGEILTKDGNFGVKIKEI     Li_Lin0708   441   QILEDIPVTLEVVFGTAKVLEKFISWCEKDVIILK-ESMNEPLVLALNGVTIGKGILVRVDDHFGIQMTEI     ML_HrcQ   279   EDLDDVEIMVFECGRWPIPLGELRSAGGEHIFELG-RFIQDPVDILANGCCIGRGDIVRIGDTLGIRLRGI     At_F11M   237   QVKR-SQVTLEARIKLEFIIRTISRLVAGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVKNI     Hp_F11M   252   ALLSGVSVDMIVFLEARIKLEFISKECBOVIEFGSULVAGDIGVEVSANGSKLYNCEFGKSGDRYMVKNI     Vp_F11M   250   EEIMDCPVNFKVNLLEKDISLRDMMQPGDVPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVKNRI     Vp_F11M   2	DS 91	-NLAIRINEILDS	-SVELYINKRIFGKGEVMVYEK	TTNMSVAELLKLEVGSVIDLE-KPAGE-	18 EDILDITVD	Cj_FliN
Pa_PscQ233HELDQIFIPVSFEVGRATUDIATISTICPGSLIDLD-SALDG-EVALIANORCLGIGELVALODRLGVRVIRRs_HrcQ276VPLEQLEVPVHELAVMGMPLAELAALQPQHVITLPVKIRDV-SVRLVCHGQTLGHQLVAVGEQLGLQIASSt_SsaQ245VELEQIPVQVLFEVGRASLEIGQLRQLKTGDVLPVG-GCFAP-EVTTRVNDRILGOGELIACGNEFMVRITRISt_SpaO225PGLNQLFVKLEFVLYRKNVTLAELEAMGQQQLLSLP-TNAELNVEIMANGVLLGNGELVQMNDTLGVEIHEIYp_FliN57SLFSRIPVTITLEVASVEIPLSELLTVNNDSVIELD-KLAGEPLDIRVNGIMFGQAEVVVINEKYGLRIINAa_FliN37QHFSDIPVEVEVVVGRANKTLGELLAMGIGSVIEID-REPKDLVDIKVNGKLIAKGELVIIDGKIGVKIKEILi_Lin070128RQVDNIGVNLIVRLGKKEMPVGDIAELSIGDVLEVE-KKPGHKVEIFLDEKKGGIGEAILMDENFGIVISELi_Lin07062KINHTIPLRIDFELGRTKQPVGSLLDVKKGTVFRLE-DSTGNVVKITISGKCIGYGEILTKDGKIGVKIKEILi_Lin0707441QILEDIPVTLEVVFGTAKVKLEKFISWCEKDVIILK-ESMNEPLVLALNGVTGKGCIURKUDDHFGIQMTEIMLHrcQ279EDLDDVEIMUVFCGAKVKLEKFISWCEKDVIILK-ESMNEPLVLALNGVTGKGILVRUDDHFGIQMTEIMLHrcQ279EDLDDVEIMUVFCGRKVKLEKFISWCEKDVIILK-ESMNEPLVLALNGVTGKGILVRUDDKIGVKIKHp_F1iM252AILS-IPLTLSARLCEPEVPLRQLMQMQPGDVLPVH-LTEALSLLVEGQPIFEAAPGERGGQAALNITRIVp_F1iM250EEIMDCPVNFRVNLLEKDISLRDIMELQPGDIIPIE-MPEHATMFIEDLPTYRVKMGSEDKLAVQVSQITp_F1iM253KRIMTAQIPVAELGESELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKVAQVILDEBb_F1iM263EKLENTAMPIVAEIGEVKLKVREILSLOKGDUNLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITEITp_F1iM252ALLSGVSVDMIVFLGAVELSLKEMDLDVGDTIRLN-HAND-EVSVYVHKKKRYLASVGFQGYKLAQQILEERp_F1iM252ALLSGVSVDMIVFLGAVELSLKEMDLDLVGDITRLN-KIAND-EVSVYVHKKKRYLAS	SG 323	-KLGVEIHSWLSG	-NIEIRVNGALTGYGELVEVDD	KKIMNLYEIDELCAKRIISLL-PESEK-	254ELPVK	EHEC_Epa0
RS_HFQ   216   VELEQIPPOVLELAVMERE JARJAAUGPONTILEVANDVSVALVERGILGAGUAVAGEQUGLQIAS     St_SsaQ   245   VELEQIPPOVLEEVGRASLEIGQIRQLKTGDVLPVG-GCFAPEVTIRVMDRIIGQGELIACGNEFWRITR     St_SpaO   225   PGLNQLPVKLEFVLYRKNVTLAELEAMGQQULSLP-TNAELNVEIMANGVLLGNGELVQMNDTLGVEIHE     Yp_FliN   57   SLFSRIPVTITLEVASVEIPLSELLTVNNDSVIELD-KLAGEPLDIRVNGIMFGQAEVVVINEKYGLRINN     Aa_FliN   37   QHFSDIPVEVEVVVGRANKTIGELLAMGIGSVIEID-REPKDLVDIKVNGKLIAKGELVIIDGKIGVKIKE     Li_Lin0701   28   RQVDNIGVNLIVRLGKKEMPVGDIAELSIGDVLEVE-KKPGHKVEIFLDEKKVGIGEAILMDENFGIVISE     Li_Lin0706   2   KINHTIPLRIDFELGRTKQPVGSLUDVKGTVFRLE-DSTGNVVKITISGKCIGYGEILTKDGKMFVKITK     Li_Lin0708   441   QILEDIPVTLEVVFGTAKVKLEKFISWCEKDVIILK-ESMNEPLVLALMGVTIGKGILVRVDDHFGIQMTE     MLHrcQ   279   EDLDDVEIMIVFECGRWPIPLGELRSAGEGHIFELG-RPIQDPVDILANGQCIGRGDIVRIGDTLGIRLGIR     At_F11M   237   QVKR-SQVTLEARIKLETITRTISRLVAGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVRVKNI     Hp_F11M   250   EEIMDCPVNFRVNLLEKDISLRDIMELOPGDIIPIE-MPEHATMFTEDLPTYRVKNGRSEDKLAVQVSQI     Tp_F11M   253   KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKVAVQIDH     Bb_F11M   263   EKLENTAMPIVAEIGSVKLKVREILSLOKGDVINLE-SSLINK-DLTLKVGTKEKFKCRMGIMGNKVAVQIDH     Bb_F11M   263 <td>H 306</td> <td>-RLGVRVIRLFGH</td> <td>-EVRILANORCLGIGELVRLOD</td> <td>RRTLDLHTLSTLQPGSLLDLD-SALDG-</td> <td>233 HELDQLPIP</td> <td>Pa_PscQ</td>	H 306	-RLGVRVIRLFGH	-EVRILANORCLGIGELVRLOD	RRTLDLHTLSTLQPGSLLDLD-SALDG-	233 HELDQLPIP	Pa_PscQ
St_Stad215VERGUTOW HED GRADELGOURSELGOURSELGOURSENGED FUNCTION FUNCTION FUNCTION FUNCTIONSt_Spao225PGLNQLPVKLEFVLYRKNVLAELEAMGQQULSLP-TNAELNVEIMANGVLGNGELVQMNDKIGVRIFYp_Flin57SLFSRIPVTITLEVASVETPLSELLTVNNDSVIELD-KLAGEPLDIRVNGINFGQAEVVVINEKIGVRIKENAa_Flin37QHFSDIPVEVEVVVGRANKTIGELLAMGIGSVIELD-KLAGEPLDIRVNGKLIAKGELVINDGKIGVRIKELi_Lin070128RQVDNIGVNIIVRLGKKEMPVGDIAELSIGDVLEVE-KKPGHKVEIFLDEKKVGIGEAILMDENFGIVISELi_Lin07062KINHTIPLRIDFELGRTKQPVGSLLDVKKGTVFRLE-DSTGNVVKITISGKCIGYGEILTKDGKMFVKITKLi_Lin0708441QILEDIPVTLEVVFGTAKVKLEKFISWCEKDVIILK-ESMNEPLVLALNGVTIGKGILVRVDDHFGIQMTEML_HrcQ279EDLDDVEIMUVFECGRWPIPIGELRSAGEGHIFELG-RPIQDPVDILANGQCIGRGDIVRIGDTLGIRLRGIAt_Flin237QVKR-SQVTLEARIKLETLILRTISRLVAGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVRVKNIHp_Flin252AILS-IPLTISARLCEPEVPLRQIMQQPGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVRVKNIHp_Flin253KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKFKFLGQAGRVNRKLAVQVSQITEITp_Flin253KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKFKFLGQAGRVNRKLAVQVSQITEITp_Flin252CKLSTVDMDVVAEVGSLRLSVRDILGLRVGDIIRLH-DTHVGD-PFVLSIGNRKKFLCQPGVVGKKLAQIENGSf_Spa0219FNYDDINVKDFILLEKNMTINELKMYVENELFFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGTEISSIBp_SctQ231TRIGELEPVVFEIDTVSLPIDQLSALEPGYVENELFFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGTEISSIBp_SctQ231TRIGELEPVVFEIDTVSLPIDQLSALEPGYVENELFFKP-DDIVKHVNIKVNGSLVGHGELVAVGEHLGLQLTQCongeneus259PQLASL	C 319	-QLGLQIASIGKH	-SVRLVCHGQILGHGQLVAVGE	MGMPLAELAALQPQHVLILPVKIKDV-	245 VELEQUEVP	RS_HrcQ
Sc_paceSignature <t< td=""><td>SE 298</td><td>-TLOVETHEWLSE</td><td>-NVETMANGVILGNCELVOMND</td><td>PKNUTIAFIFAMCOOOLISID-TNAFI-</td><td>225 PCLNOLPVK</td><td>St_SbaQ</td></t<>	SE 298	-TLOVETHEWLSE	-NVETMANGVILGNCELVOMND	PKNUTIAFIFAMCOOOLISID-TNAFI-	225 PCLNOLPVK	St_SbaQ
Aa_Flin37QHFSDIPVEVEVVVGRANKTLGELLAMGIGSVIEID-REPKDLVDIKVNGKLIAKGELVIIDGKIGVKIKELi_Lin070128RQVDNIGVNLIVRLGKKEMPVGDIAELSIGDVLEVE-KKPGHKVEIFLDEKKVGIGEAILMDENFGIVISELi_Lin07062KINHTIPLRIDFELGRTKQPVGSLLDVKKGTVFRLE-DSTGNVVKITISGKCIGYGEILTKDGKMFVKITK.Li_Lin0708441QILEDIPVTLEVVFGTAKVKLEKFISWCEKDVIILK-ESMNEPLVLALNGVTIGKGILVRVDDHFGIQMTEM1_HrcQ279EDLDDVEIMLVFECGRWPIPLGELRSAGEGHIFELG-RPIQDPVDILANGQCIGRGDIVRIGDTLGIRLRGIAt_Flim237QVKR-SQVTLEARIKLETLTLRTISRLVAGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVRVKNIHp_Flim252AILS-IPLTLSARLCEPEVPLRQLMQMQPGDVLPVH-LTEALSLLVEGQPIFEAAPGERGGQAALNITRIVp_Flim250EEIMDCPVNFRVNLLEKDISLRDIMELQPGDIIPIE-MPEHATMFIEDLPTYRVKMGRSEDKLAVQVSQITp_Flim253KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKQAVQILDIBb_Flim263EKLENTAMPLVAEIGEVKLKVREILSLDKGDVINLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITEITp_Flim252ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLH-DTHVGD-PFVLSIGNRKKFLCQPGVVGKKIAAQILEIRp_Flim252ALLSGVSVDMIVFIGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYHKKKRYLASVGFQGYRKTQIKEVSf_Spa0219FNVDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSIBp_SctQ231TRIGELEPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGGGELVAVGEHLGVITQConsensus/800hpl.hpl.hpl.hpl.hpl.hpl.hpl.hpl.hpl.hpl.	50 130	-KYGLETININSO	-PLDTRVNGIMEGOAEVVVINE	SVETPLSELLTVNNDSVIELD-KLAGE-	57 SLESRIPVT	Yp FliN
Li_Lin070128RQVDNIGVNLIVRLGKKEMPVGDIAELSIGDVLEVE-KKPGHKVEIFLDEKKVGIGEAILMDENFGIVISELi_Lin07062KINHTIPLRIDFELGRTKQPVGSLLDVKKGTVFRLE-DSTGNVVKITISGKCIGYGEILTKDGKMFVKITK.Li_Lin0708441QILEDIPVTLEVVFGTAKVKLEKFISWCEKDVIILK-ESMNEPLVLALNGVTIGKGILVRVDDHFGIQMTEM1_HrcQ279EDLDDVEIMLVFECGRWPIPLGELRSAGEGHIFELG-RPIQDPVDILANGQCIGRGDIVRIGDTLGIRLRGIAt_F1iM237QVKR-SQVTLEARIKLETLTLRTISRLVAGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVRVKNIHp_F1iM252AILS-IPLTLSARLCEPEVPLRQLMQMQPGDVLPVH-LTEALSLLVEGQPIFEAAPGERGGQAALNITRIVp_F1iM250EEIMDCPVNFRVNLLEKDISLRDIMELQPGDIIPIE-MPEHATMFIEDLPTYRVKMGRSEDKLAVQVSQITp_F1iM253KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKQAVQILDIBb_F1iM263EKLENTAMPLVAEIGEVKLKVREILSLDKGDVINLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITEITp_F1iM252DKLSTVDMDVVAEVGSLRLSVRDILGLRVGDIIRLH-DTHVGD-PFVLSIGNRKKFLC0PGVVGKKIAAQILEIRp_F1iM252ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYVHKKKRYLASVGFQGYRKT1QIKEVSf_Spa0219FNYDDINVKVDFILLEKNMTINELKMVVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSIBp_SctQ231TRIGELELPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGGELVAVGEHLGVRIIRNEa_HrcQ259PQLASLPLSLEVRCDRTALTIGELQRLQAGSVVTLD-NVTPGEGCLYHGDTLIARGELVDVEGHLGVRIConsensus (803b b b b b b b b b b b b b b b b b b b	KE 110	-KIGVKIKEVVKE	-LVDIKVNGKLIAKGELVIIDG	RANKTLGELLAMGIGSVIEID-REPKD-	37 OHESDIPVE	Aa FliN
Li_Lin07062KINHTIPLRIDFELGRTKQPVGSLLDVKKGTVFRLE-DSTGNVVKITISGKCIGYGEILTKDGKMFVKITK.Li_Lin0708441QILEDIPVTLEVVFGTAKVKLEKFISWCEKDVIILK-ESMNEPLVLALNGVTIGKGILVRVDDHFGIQMTE.M1_HrcQ279EDLDDVEIMLVFECGRWPIPLGELRSAGEGHIFELG-RPIQDPVDILANGQCIGRGDIVRIGDTLGIRLRGIAt_F1iM237QVKR-SQVTLEARIKLETITLRTISRLVAGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVRVKNIHp_F1iM252AILS-IPLTISARLCEPEVPLRQLMQMQPGDVLPVH-LTEALSLLVEGQPIFEAAPGERGGQAALNLTRIVp_F1iM250EEIMDCPVNFRVNLLEKDISLRDIMELQPGDIIPIE-MPEHATMFIEDLPTYRVKMGRSEDKLAVQVSQITp_F1iM253KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKQAVQILDIBb_F1iM263EKLENTAMPLVAEIGEVKLKVREILSLDKGDVLNLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITEITp_F1iM252DKLSTVDMDVVAEVGSLRLSVRDILGLRVGDIIRLH-DTHVGD-PFVLSIGNRKKFLCQPGVGKKLAQULEIRp_F1iM252ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYVHKKKRYLASVGFQGYRKT1QIKEVSf_Spa0219FNYDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSIBp_SctQ231TRIGELEPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVHGQTVGYGELVAVGEHLGVRIIRNEa_HrcQ259PQLASLPLSLEVRCDRTALIGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGVRIIRNEa_HrcQ259PQLASLPLSLEVRCDRTALIGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGVLIQUConcensus(808b al b al b an blackbal black	99	-NFGIVISEID	-KVEIFLDEKKVGIGEAILMDE	KKEMPVGDIAELSIGDVLEVE-KKPGH-	28 ROVDNIGVN	Li Lin0701
Li_Lin0708441QILEDIPVTLEVVFGTAKVKLEKFISWCEKDVIILK-ESMNEPLVLALNGVTIGKGILVRVDDHFGIQMTEMI_HrcQ279EDLDDVEIMLVFECGRWPIPLGELRSAGEGHIFELG-RPIQDPVDILANGQCIGRGDIVRIGDTLGIRLRGIAt_Flim237QVKR-SQVTLEARIKLETLTLRTISRLVAGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVRVKNIHp_Flim252AILS-IPLTLSARICEPEVPLRQLMQMQPGDVLPVH-LTEALSLVEGQPIFEAAPGERGGQAALNLTRIVp_Flim250EEIMDCPVNFRVNLLEKDISLRDIMELQPGDIIPIE-MPEHATMFIEDLPTYRVKMGRSEDKLAVQVSQLTp_Flim253KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKQAVQILDIBb_Flim263EKLENTAMPLVAEIGEVKLKVREILSLDKGDVINLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITEITp_Flim252DKLSTVDMDVVAEVGSLRLSVRDILGLRVGDIIRLH-DTHVGD-PFVLSIGNRKKFLCQPGVVGKKIAAQILEIRp_Flim252ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYVHKKKRYLASVGFQGYRKTIQIKEVSf_Spa0219FNYDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSIBp_SctQ231TRIGELEPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGYGELVAVGEHLGVRIINEa_HrcQ259PQLASLPLSLEVRCDRTALTLGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGVRIIN	EG 75	-KMFVKITKLGEG	-VVKITISGKCIGYGEILTKDG	RTKOPVGSLLDVKKGTVFRLE-DSTGN-	2 KINHTIPLR	Li_Lin0706
M1_HrcQ279EDLDDVEIMLVFECGRWPIPLGELRSAGEGHIFELG-RPIQDPVDILANGQCIGRGDIVRIGDTLGIRLRGIAt_Flim237QVKR-SQVTLEARIKLETLTLRTISRLVAGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVRVKNIHp_Flim252AILS-IPLTLSARLCEPEVPLRQLMQMQPGDVLPVH-LTEALSLLVEGQPIFEAAPGERGGQAALNLTRIVp_Flim250EEIMDCPVNFRVNLLEKDISLRDIMELQPGDIIPIE-MPEHATMFIEDLPTYRVKMGRSEDKLAVQVSQITp_Flim253KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKQAVQILDIBb_Flim263EKLENTAMPLVAEIGEVKLKVREILSLDKGDVLNLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITEITp_Flim252DKLSTVDMDVVAEVGSLRLSVRDILGLRVGDIIRLH-DTHVGD-PFVLSIGNRKKFLCQPGVVGKKIAAQILEIRp_Flim252ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYVHKKKRYLASVGFQGYRKTQIKEVSf_Spa0219FNYDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSIBp_SctQ231TRIGELELPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGYGELVAVGEHLGVRIIRNEa_HrcQ259PQLASLPLSLEVRCDRTALTLGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGVRIIRN	R- 513	-HFGIQMTELVR-	-PLVLALNGVTIGKGILVRVDD	TAKVKLEKFISWCEKDVIILK-ESMNE-	441 QILEDIPVT	Li_Lin0708
At_FliM237QVKR-SQVTLEARIKLETLTLRTISRLVAGDVIPFQ-DLKQDDIGVEVSANGSKLYNCEFGKSGDRYMVRVKNHp_FliM252AILS-IPLTLSARLCEPEVPLRQLMQMQPGDVLPVH-LTEALSLLVEGQPIFEAAPGERGGQAALNLTRVp_FliM250EEIMDCPVNFRVNLLEKDISLRDLMELQPGDIIPIE-MPEHATMFIEDLPTYRVKMGRSEDKLAVQVSQTp_FliM253KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKQAVQILDBb_FliM263EKLENTAMPLVAEIGEVKLKVREILSLDKGDVINLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITEITp_FliM252DKLSTVDMDVVAEVGSLRLSVRDILGLRVGDIIRLH-DTHVGD-PFVLSIGNRKKFLCQPGVVGKKLAQILEIRp_FliM252ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYVHKKKRYLASVGFQGYRKTQIKEVSf_Spa0219FNYDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSIBp_SctQ231TRIGELELPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGYGELVAVGEHLGVRIIREa_HrcQ259PQLASLPLSLEVRCDRTALTIGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGVRIIREa_HrcQ259PQLASLPLSLEVRCDRTALTIGELQRLQAGSVVTLD-NVTPG-EAGLYHGDTLIARGELVDVEGHLGVRIIR	GC 352	-TLGIRLRGRLGC	-PVDILANGOCIGRGDIVRIGD	RWPIPLGELRSAGEGHIFELG-RPIQD-	279 EDLDDVEIM	M1_HrcQ
Hp_FliM252AILS-IPLTLSARLCEPEVPLRQLMQMQPGDVLPVH-LTEALSLLVEGQPIFEAAPGERGGQAALNLTRVp_FliM250EEIMDCPVNFRVNLLEKDISLRDLMELQPGDIIPIE-MPEHATMFIEDLPTYRVKMGRSEDKLAVQVSQTp_FliM253KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKQAVQILDBb_FliM263EKLENTAMPLVAEIGEVKLKVREILSLDKGDVLNLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITEITp_FliM252DKLSTVDMDVVAEVGSLRLSVRDILGLRVGDIIRLH-DTHVGD-PFVLSIGNRKKFLCQPGVVGKKIAAQILEIRp_FliM252ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYVHKKKRYLASVGFQGYRKTQIKEVSf_Spa0219FNYDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSUBp_SctQ231TRIGELELPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGYGELVAVGEHLGVRIIREa_HrcQ259PQLASLPLSLEVRCDRTALTIGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGLQLTQUConserpsusADLLLLLDLLDLLDLLDLLDLLDLLDLDLLDLDLLDLDLDLDLDLDLDLDLDLDLDLDLDLDLDLDL <tr< td=""><td>ST 311</td><td>-RYMVRVKNNVST</td><td>IGVEVSANGSKLYNCEFGKSGD</td><td>LETLTLRTISRLVAGDVIPFQ-DLKQDD</td><td>237 QVKR-SQVT</td><td>At_FliM</td></tr<>	ST 311	-RYMVRVKNNVST	IGVEVSANGSKLYNCEFGKSGD	LETLTLRTISRLVAGDVIPFQ-DLKQDD	237 QVKR-SQVT	At_FliM
Vp_Flim   250   EEIMDCPVNFRVNLLEKDISLRDIMELQPGDIIPIE-MPEHATMFIEDLPTYRVKMGRSEDKLAVQVSQ     Tp_Flim   253   KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKQAVQILD     Bb_Flim   263   EKLENTAMPIVAEIGEVKLKVREILSLDKGDVINLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITE     Tp_Flim   252   DKLSTVDMDVVAEVGSLRLSVRDILGLRVGDIIRLH-DTHVGD-PFVLSIGNRKKFLCQPGVVGKKIAAQILEN     Rp_Flim   252   ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYVHKKKRYLASVGFQGYRKTIQIKEN     Sf_Spa0   219   FNYDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSN     Bp_SctQ   231   TRIGELELPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGYGELVAVGEHLGVRIIRN     Ea_HrcQ   259   PQLASLPLSLEVRCDRTALTIGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGLQLTQD	VR 322	-QAALNLTRRHVR	SLLVEGQPIFEAAPGERGG	EPEVPLRQLMQMQPGDVLPVH-LTEAL-	252 AILS-IPLT	Hp_FliM
Tp_F1im   253 KRIMTAQIPVVAELGTSELTIEEFLSLEVGDCITLD-KSVTDPLTVLVGDKPKFLGQAGRVNRKQAVQILD     Bb_F1iM   263 EKLENTAMPLVAEIGEVKLKVREILSLDKGDVINLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITEI     Tp_F1iM   252 DKLSTVDMDVVAEVGSLRLSVRDILGLRVGDIIRLH-DTHVGD-PFVLSIGNRKKFLCQPGVVGKKIAAQILEI     Rp_F1iM   252 ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYVHKKKRYLASVGFQGYRKTIQIKEV     Sf_Spa0   219 FNYDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSI     Bp_SctQ   231 TRIGELELPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGYGELVAVGEHLGVRIIN     Ea_HrcQ   259 PQLASLPLSLEVRCDRTALTLGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGLQLTQI	ER 321	-KLAVQVSQEIER	TMFIEDLPTYRVKMGRSED	EKDISLRDIMELQPGDIIPIE-MPEHA-	250 EEIMDCPVN	Vp_FliM
Bb_F11M   263 EKLENTAMPLVAEIGEVKLKVREILSLDKGDVLNLE-SSLINK-DLTLKVGTKEKFKCRMGLMGNKVSVQITE     Tp_F1iM   252 DKLSTVDMDVVAEVGSLRLSVRDILGLRVGDIIRLH-DTHVGD-PFVLSIGNRKKFLCQPGVVGKKIAAQILE     Rp_F1iM   252 ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYVHKKKRYLASVGFQGYRKTQIKEV     Sf_Spa0   219 FNYDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSI     Bp_SctQ   231 TRIGELELPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGYGELVAVGEHLGVRIIRN     Ea_HrcQ   259 PQLASLPLSLEVRCDRTALTLGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGLQLTQU	IR 326	-KQAVQILDHDIR	-PLTVLVGDKPKFLGQAGRVNR	TSELTIEEFLSLEVGDCITLD-KSVTD-	253 KRIMTAQIP	Tp_F11M
Tp_FILM   252   DKLSTVDMDVVAEVGSLKLSVKDILGLKVGDILKLH-DIHVGD-PFVLSIGNKKKFLCOPGVGKKIAAGILE     Rp_F1M   252   ALLSGVSVDMIVFLGAVELSLKEMLDLDVGDTIRLN-KIANDEVSVYVHKKKRYLASVGFQGYRKTIQIKE     Sf_Spa0   219   FNYDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSI     Bp_SctQ   231   TRIGELELPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGYGELVAVGEHLGVRIIR     Ea_HrcQ   259   PQLASLPLSLEVRCDRTALTLGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGLQLTQI     Consensus (80%   h. pl. h. l. h. lsp. ph. l. pl.bph. cp.ll. l.   h. pl. h. l. h. lsp. ph. l. pl.bph. cp.ll. l.	GD 337	-KVSVQITEKIGD	-DLTLKVGTKEKFKCRMGLMGN	EVKLKVREILSLDKGDVLNLE-SSLINK	263 EKLENTAMP	Bb_FliM
Sf_Spa0   219   FNYDDINVKVDFILLEKNMTINELKMYVENELFKFP-DDIVKHVNIKVNGSLVGHGELVSIEDGYGIEISSI     Bp_SctQ   231   TRIGELELPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGYGELVAVGEHLGVRIIRI     Ea_HrcQ   259   PQLASLPLSLEVRCDRTALTLGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGLQLTQL     Consensus (80%   h. pl. h. l. h. l.sp. pb. l. pl.bph. cp.l.l.   h.pl.	VC 225	PETTOTENTYC	-PFVLSIGNKKKFLCQPGVVGK	AVEL CI VEMI DI DUCOTTOTAL VIANO	252 DKLSTVDMD	Pp_F11M
Bp_SctQ 231 TRIGELELPVQFEIDTVSLPIDQLSALEPGYVIELP-VAVTD-ARLRLVVHGQTVGYGELVAVGEHLGVRIIRI Ea_HrcQ 259 PQLASLPLSLEVRCDRTALTLGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGLQLTQU Consensus (80% h pl h l h lsp ph l plbph spll l	IS 325	CACTELCOM	- HUNTKUNGELUGHGELUGTED	EVENATINET PMANE PEPE	210 ENVDETNUE	Sf Space
Ea_HrcQ 259 PQLASLPLSLEVRCDRTALTLGELQRLQAGSVVTLD-NVTPGEAGLYHGDTLIARGELVDVEGHLGLQLTQL Consensus/80% h pl h l h lsp ph l plbph spll l	HR 305	-HLGVRITEMAHR	ARLRIVVHGOTVGYGELVAVCE	TVSLPIDOLSALEPCVUTFLP-VAVTD-	231 TRICELELP	Bp_Sct0
Concency /80% h n h h h h n h h n h h n h h n h	LT 332	-HLGLOLTOLLLT	-EAGLYHGDTLIARGELVDVEG	RTALTLGELORLOAGSVVTLD-NVTPG-	259 POLASLPLS	Ea HrcO
CONSCIPTION 1111011111111111111111111111111111111		.pbslpl.p.h.	hpl. hssp. hhbtphh	p.pb.l.pldpbspll.l.	%h.pl.h.	Consensus/80

Phylogeny inference: Analysis of sequences allows evolutionary relationships to be determined E.coli **C.botulinum C.cadavers** C.butyricum **B.subtilis B.cereus** NUMBER OF THE OWNER OF STREET

![](_page_30_Figure_0.jpeg)

"...UNDERSTAND and organise the information..."

(但在前時間之下)已成而是 EDECTORE MILLER 国际国自机组织和 CONTRACTOR OF SOME PARTIES. entra drie dece 回日日 MILIA 25 200 2013232 2012月1月日日(Clean)

![](_page_31_Figure_2.jpeg)

**Challenges in bioinformatics**  Explosion of information Need for faster, automated analysis to process large amounts of data Need for integration between different types of information (sequences, literature, annotations, protein levels, RNA levels etc...) Need for "smarter" software to identify interesting relationships in very large data sets Lack of "bioinformaticians" Software needs to be easier to access, use an understand Biologists need to learn about the software, its limitations, and how to interpret its results

![](_page_33_Picture_0.jpeg)

in silico Biology in vivo, in vitro, ..... in silico Bioinformatics is not just using a computer to store data, or speed up biology With bioinformatics, you do biological hypothesis testing on a computer

![](_page_33_Picture_2.jpeg)

Major Application I: Designing Drugs Understanding How Structures Bind Other Molecules (Function) **Designing Inhibitors Docking, Structure Modeling** (From left to right, figures adapted from Olsen Group Docking Page at Scripps, Dyson NMR Group Web page at Scripps, and from Computational Chemistry Page at Cornell Theory Center).

![](_page_34_Picture_1.jpeg)

![](_page_35_Figure_0.jpeg)

![](_page_36_Picture_0.jpeg)

# **Pharmaceutical Motivation**

Speed up drug discovery pipeline Drug Discovery 10,000 candidates 10 preclinical 5 clinical 1 New Drug oximate cost for a new drug = 

Precise, predictive model of transcription initiation and termination: ability to predict where and when transcription will occur in a genome Precise, predictive model of RNA splicing/alternative splicing: ability to predict the splicing pattern of any primary transcript in any tissue Precise, quantitative models of signal transduction pathways: ability to predict cellular responses to external Determining effective protein:DNA, protein:RNA and protein:protein recognition codes Accurate ab initio protein structure prediction

Rational design of small molecule inhibitors of proteins Mechanistic understanding of protein evolution: understanding exactly how new protein functions evolve Mechanistic understanding of speciation: molecular details of how speciation occurs Continued development of effective gene ontologies systematic ways to describe the functions of any gene or protein Education: development of appropriate bioinformatics curricula for secondary, undergraduate and graduate education

![](_page_40_Picture_0.jpeg)