Pairwise Alignment

Mark Voorhies

3/27/2012

æ

Making a file executable:

chmod "a+x" pydotter.py

Handling file/directory names with spaces:

cd My $\ Directory \ with \ Spaces$

or

 $cd ~~"My_Directory_with_Spaces"$

Killing a process on OS X:

- Try ctrl-c
- If that doesn't work:
 - **ps** -awx | grep name_of_process
 - First column in ps output is PID (process ID)
 - kill PID
 - If that doesn't work:
 - kill -KILL PID
- On Linux:

ps -ealf | grep name_of_process

FASTA files

>Name Free-form annotation MGCLLIMKEGGPGRKHKLIVMLYLDENQ EHELPIMTRAPPEDINADNAMACHINEW NQEDLYMNILKHGPPGEDEDRKHEDEDG

伺 ト く ヨ ト く ヨ ト

э

FASTA files

>Name Free-form annotation MGCLLIMKEGGPGRKHKLIVMLYLDENQ EHELPIMTRAPPEDINADNAMACHINEW NQEDLYMNILKHGPPGEDEDRKHEDEDG

• Dotplots: unbiased plot of all possible ungapped alignments of two sequences.



What do we mean by best?



What do we mean by best?



• Residues with equivalent functional roles are paired

What do we mean by best?



- Residues with equivalent functional roles are paired
- Residues that derive from the same position in the common ancestor are paired (homology)



What do we mean by best?

- Residues with equivalent functional roles are paired
- Residues that derive from the same position in the common ancestor are paired (homology)
- The sequence alignment maximizes a similarity function

Frequency of residue *i*:

P.

Frequency of residue *i*:

pi

Frequency of residue *i* aligned to residue *j*:

qij

Frequency of residue *i*:

pi

Frequency of residue *i* aligned to residue *j*:

qij

Expected frequency if i and j are independent:

p_ip_j

Frequency of residue *i*:

pi

Frequency of residue *i* aligned to residue *j*:

qij

Expected frequency if i and j are independent:

p_ip_j

Ratio of observed to expected frequency:

<u>qij</u> pipj

Frequency of residue *i*:

pi

Frequency of residue *i* aligned to residue *j*:

qij

Expected frequency if i and j are independent:

p_ip_j

Ratio of observed to expected frequency:

q_{ij} p_ipj

Log odds (LOD) score:

$$s(i,j) = \log \frac{q_{ij}}{p_i p_j}$$

• PAM1 matrix originally calculated from manual alignments of highly conserved sequences (myoglobin, cytochrome C, etc.)

- PAM1 matrix originally calculated from manual alignments of highly conserved sequences (myoglobin, cytochrome C, etc.)
- We can think of a PAM matrix as evolving a sequence by one unit of time.

- PAM1 matrix originally calculated from manual alignments of highly conserved sequences (myoglobin, cytochrome C, etc.)
- We can think of a PAM matrix as evolving a sequence by one unit of time.
- If evolution is uniform over time, then PAM matrices for larger evolutionary steps can be generated by multiplying PAM1 by itself (so, higher numbered PAM matrices represent greater evolutionary distances).

- PAM1 matrix originally calculated from manual alignments of highly conserved sequences (myoglobin, cytochrome C, etc.)
- We can think of a PAM matrix as evolving a sequence by one unit of time.
- If evolution is uniform over time, then PAM matrices for larger evolutionary steps can be generated by multiplying PAM1 by itself (so, higher numbered PAM matrices represent greater evolutionary distances).
- The BLOSUM matrices were determined from automatically generated ungapped alignments. Higher numbered BLOSUM matrices correspond to *smaller* evolutionary distances. BLOSUM62 is the default matrix for BLAST.

BLOSUM80



▲御 ▶ ▲ 臣 ▶

문 🛌 문

BLOSUM62



▲ □ ► < □ ►</p>

< ≣ >

æ

BLOSUM45



▲母 ▶ ▲ 臣 ▶

< ≣⇒

æ

In log space, multiplication and division become addition and subtraction:

$$log(xy) = log(x) + log(y)$$
$$log(x/y) = log(x) - log(y)$$

Therefore, exponentiation becomes multiplication:

$$\log(x^y) = y \log(x)$$

Also, we can change of the base of a logarithm like so:

$$\log_A(x) = \log(x) / \log(A)$$

Log odds (LOD) score:

$$s(i,j) = \log \frac{q_{ij}}{p_i p_j}$$

æ

Log odds (LOD) score:

$$s(i,j) = \log \frac{q_{ij}}{p_i p_j}$$

Multiplying independent probabilities is equivalent to adding independent log probabilities.

Log odds (LOD) score:

$$s(i,j) = \log \frac{q_{ij}}{p_i p_j}$$

Multiplying independent probabilities is equivalent to adding independent log probabilities.

Therefore, for an ungapped alignment can be scored as:

$$S(x,y) = \log \prod_{i}^{N} \frac{q_{x_i y_i}}{p_{x_i} p_{y_i}} = \sum_{i}^{N} s(x_i, y_i)$$

Log odds (LOD) score:

$$s(i,j) = \log \frac{q_{ij}}{p_i p_j}$$

Multiplying independent probabilities is equivalent to adding independent log probabilities.

Therefore, for an ungapped alignment can be scored as:

$$S(x,y) = \log \prod_{i}^{N} \frac{q_{x_iy_i}}{p_{x_i}p_{y_i}} = \sum_{i}^{N} s(x_i,y_i)$$

What about gaps?

Log odds (LOD) score:

$$s(i,j) = \log \frac{q_{ij}}{p_i p_j}$$

Multiplying independent probabilities is equivalent to adding independent log probabilities.

Therefore, for an ungapped alignment can be scored as:

$$S(x,y) = \log \prod_{i}^{N} \frac{q_{x_iy_i}}{p_{x_i}p_{y_i}} = \sum_{i}^{N} s(x_i,y_i)$$

What about gaps?

- Probability of an insertion/deletion event (gap opening, G)
- Length distribution of insertions/deletions (gap extension, E)

Log odds (LOD) score:

$$s(i,j) = \log \frac{q_{ij}}{p_i p_j}$$

Multiplying independent probabilities is equivalent to adding independent log probabilities.

Therefore, for an ungapped alignment can be scored as:

$$S(x,y) = \log \prod_{i}^{N} \frac{q_{x_iy_i}}{p_{x_i}p_{y_i}} = \sum_{i}^{N} s(x_i,y_i)$$

What about gaps?

- Probability of an insertion/deletion event (gap opening, G)
- Length distribution of insertions/deletions (gap extension, E)

$$S_{gapped}(x,y) = S(x,y) + \sum_{i}^{gaps} (G + E * L_i)$$

Log odds (LOD) score:

$$s(i,j) = \log \frac{q_{ij}}{p_i p_j}$$

Multiplying independent probabilities is equivalent to adding independent log probabilities.

Therefore, for an ungapped alignment can be scored as:

$$S(x,y) = \log \prod_{i}^{N} \frac{q_{x_iy_i}}{p_{x_i}p_{y_i}} = \sum_{i}^{N} s(x_i,y_i)$$

What about gaps?

- Probability of an insertion/deletion event (gap opening, G)
- Length distribution of insertions/deletions (gap extension, *E*)

$$S_{gapped}(x,y) = S(x,y) + \sum_{i}^{s-r} (G + E * L_i)$$

We find an optimal alignment by finding x and y that maximize S.



@▶ ∢ ≣▶

æ



⊡ ► < ≣ ►

글▶ 글



▲□ ▶ ▲ 三 ▶ ▲

글▶ 글



□ ▶ ▲ 臣 ▶ ▲ 臣 ▶

æ

Binomial formula:

$$\binom{k}{r} = \frac{k!}{(k-r)!r!}$$
$$\binom{2n}{n} = \frac{(2n)!}{n!n!}$$

Stirling's approximation:

$$x! \approx \sqrt{2\pi} \left(x^{x+\frac{1}{2}} \right) e^{-x}$$
$$\binom{2n}{n} \approx \frac{2^{2n}}{\sqrt{\pi n}}$$

Scoring an alignment quickly

$$\frac{2^{2n}}{\sqrt{\pi n}}$$
 is too expensive.

æ

⊡ ► < ≣

$$\frac{2^{2n}}{\sqrt{\pi n}}$$
 is too expensive.

$$S_{gapped}(x,y) = S(x,y) + \sum_{i}^{gaps} (G + E * L_i)$$

æ

⊡ ► < ≣

$$\frac{2^{2n}}{\sqrt{\pi n}}$$
 is too expensive.

$$S_{gapped}(x,y) = S(x,y) + \sum_{i}^{gaps} (G + E * L_i)$$

The best alignment of any pair of subsequences is independent of the global alignment.

Dynamic Programming



æ

-

▲ □ ▶ ▲ 三









문 문 문



/⊒ > < ∃ >

- ∢ ≣ ▶

э

- DOTTER: $O(n^2)$
- Exhaustive search: $\frac{2^{2n}}{\sqrt{\pi n}}$
- Dynamic programming: $O(n^2)$ to $O(n^3)$

A∄ ▶ ∢ ∃=

Setting gap penalties in CLUSTALX

ОК	
Slow-Accurate v	
Slow/Accurate Pairwise Parameters Fast/Approx Pe < >	
Gap Opening [0-100]: 0	
Gap Extend [0-100]: 0	
Protein Weight Matrix	
BLOSUM 30 PAM 350 • Gonnet 250	
O Identity matrix O User defined	
Load protein matrix:	
DNA Weight Matrix	
IUB CLUSTALW(1.6) User defined	
Load DNA matrix:	

\odot	Alignment Parameter	rs 🛛 🖉 🛇 🛇	0
ОК			
	Multiple Param	eters	
Gap Openir	ng [0-100]:	0	
Gap Extens	sion [0-100]:	0	
Delay Diver	gent Sequences(%):	30	
DNA Transit	tion Weight[0-1]	0.5	
Use Negati	ve Matrix	Off v	
	Protein Weight	Matrix	
	M series 🔵 PAM serie	es 📀 User defined	
 Gonnet 	t series 🔵 Identity n	natrix	
Load pro	otein matrix:		
	DNA Weight M	atrix	
• IUB		1.6) 🕖 User defined	
Load DN	A matrix:		

			Jalview 2.3
<u>F</u> ile Tools Help Window			
/home/mvoorhie/Projects/Courses/P	racticalBioinformatics/gui/webr 🗗 🔂		/home/mvoorhie/Projects/Courses/PracticalBioinformatics/gui/web 🖬 🗹 🗵
File Edit Select View Format	Colour Calculate WebService		File Edit Select View Format Colour Calculate WebService
gi 28904 amb CA439602.2 /2-508 1 M <mark>B</mark> T R <mark>B</mark> L	ALALLAAVLLL <mark>O</mark> TVL <mark>P</mark> AASEAE <mark>GLVR</mark> IAL85	-	0126904/em8/C4A39402.1/2-308 1 MC T GLAVAL VAVL U T VL PAA E A E QLVE IA 34
gi 264604 g6 A4431096.2 /2-385 1 · · · · M K	WLLLLSLVVLS <mark>E</mark> CLVKV <mark>P</mark> ·····LV <mark>R</mark> KKS26		0/1264604/ge/a4431096.1/2-385 1 M - WL - L - LL S LV - VL E C - LV VP 20
gr]26904[emb)CA439602.2]/2-508 36 K K R P I D	RNSRVATGLSGGEEOPLLSGANPLRSEEE70		97124904/mm5/C4439662.21/2-308 35 - KKRP DRN SRVATG - LSBCEEDPLL - SCA - NP 64
gr]264604[g6]A4432096.2]/2-385 27 L <mark>B</mark> QNL I	KNGKLKDFLKTHKHNP-ASKYFPEAAALI50		97124964/ms6/C4439662.21/2-385 21 LVRKKSL - BONLIKNCKL KD - FLKTH KHNP 48
gr[26904]emb(CAA39602.2]/2-508 71 GD VAL	KNYMNADYFDEIGYGTPPOKFTVIFDTGE105		gv]28904∥emb/CAAS9602.2/2-508 65 L - 8 - · SEEE - · - GD IVAL KNYMNADYFEEIGVGT 92
gr[264604]gb]A4A32096.2]/2-385 61 GD - E PL	ENYLDTEYFGTIGIGTPADDFTVIFDTGE94		gv]264604/ge AAA31096.1/2-385 49 KSEYFEAAALIGD - EPLENYLD TEYFETIGIGT 81
gr 28904 emb CA439602.21/2-508 106 88 LWV P	SAKCYFSIACYLHSRYKAGASSTYKKNOK 140		9/28904/emb/CAA39602.2/2-508 93 PPOK KFTVIFDTGSSNLWVPSAKCYFSIACYLHSR 125
gr 264604 g6 44431096.21/2-385 95 88 LWV P	Svycs-SlacsdhnofnPddsstfeatso 128		9/264604/g6/AAA31096.1/3-385 82 PADOFTVIFDTGSSNLWVPSVYCS-ELACSDHNO.114
gr 28904 amb CA439602.2 /2-508 141 PAATOY	GTGSIAGYFSEDSYTVGDLVVKDOEFIEA175		pil26904 jemb jCAA39602.2/2-508 127 ¥KAGASSTYKKNGKPAAIOYGTGSIASYFSEDSV 150
gr 264604 gr A4A31096.1 /2-385 129 ELSTY	GTGSMTGILGYDTVOVGGISDTNOIFGLS163		pil264604 jej jAA432096.2/2-385 115 FNPDDSSTFEATSQELSITYGTGSMTGILGYDTV 148
g#[28904]emb[CAA39602.2]/2-508 176 TK <mark>EPG</mark>	T FLVAK FDG ILGLGFKEISVGKAV PVWYK 210		gri26904 jem b/CAA39602.2/2-508 161 TVGDLVVKDD1F-1EATKEFGITFLV-AKFDGIL 192
g#[264604]g#[A4A31096.2]/2-385 164 ET <mark>EPG</mark> S	FLYYAPFDG ILGLAYPSISASGAT PVFON 190		gri264604 jg6j4A432096.2/2-385 149 0VGGISDTND1FGLSET-EFG-SFLVYAFFDGIL 180
gr]26904 emb CAA39602.2 /2-508 211 MIEDGL	VSDPVFSFWLNRHVDEGEGGEIIFGGMDP245		9/12/90/1/em/b/CA439602.21/2-509 199 GLOFKEISV-GKAVFVWYKMIEGGLVS-DFVFEF224
gr]264604 g6 A4432096.2 /2-385 199 LWDQGL	VSQDLFSVYLSSNDDSSSVVLLGGIDS231		9/12/9/0/1/ge/a/a/31096.1/3-385 181 GLAVESISASG-ATFVFDNLWDDGLVSOD-LFSV212
gr 28904 emb CAA39602.21/2-508 246 K HYVS E	HTYVPVTOK GYWOFDMGDVLVGGKSTGFC200		gv128904/jem/b/CAA59602.2/2-508 225 WLNRHVDEGEGGEIIFGGMDPKHVVSEHTVVFVT258
gr 264604 ge A4A32096.21/2-385 232 S WYTS S	LNWVPVSVEGYWOITLDSITMOG-ETIAC265		gv1264604/g6/A431096.1/2-383 213 YLSSN-DDS-GSVVLLGGIDSSYYTGSLNWVPVE244
gr 28904 emb CA439602.2 /2-508 281 A <mark>333C</mark> AA	I AD SGTS LLAGPTAILTEINEK I GAAGVV 315		9128904(emb)(24439602.2)/2-508 259.0 K 8 YWO F DMGD - V L VGG K 8 T 8 F C A 6 6 C A A I A D 8 C 231
gr 264604 ge A4A31096.1 /2-385 266 S <mark>355C</mark> OA	I VD TGTS LL TGPT SALANI Q SD I GAS 297		91264604(g6)(44431096.3)/3-385 245 V L 0 YWO I T L - D S I T MDG L T I A - C S 6 6 C O A I VD I G 276
gv 28904 emb CA439602.2 /2-508 316 SQECKT gv 264604 ge A4A31096.1 /2-385	IVSQY <mark>Q</mark> QQILDLLLAETQ <mark>P</mark> KKICSQV <mark>Q</mark> LC350		gyi26904/jem/s/CAA39602.2/2-508 292 TSLLAGPTAILTELEEKIGAAGVVSDECKTIVD0325 gyi264604/g/sjaa432096.3/2-385 277 TSLLTSPTSAIA-H-I
gv]28904 emb[CAA39602.2 /2-508 851 TFD <mark>G</mark> TR gv]264604 g6 A4A32096.1 /2-385 ·····	GVSAGIRSVVDDE <mark>P</mark> VKSN <mark>G</mark> LRAD P MCSAC385		9/12/90/(em/s)/CAA39602.21/2-50/8 326 YBOOILDLLLAETORKKICSDVGLCTFDBTRGVS 359 p/12/9/64/9/g6)/44/31096.1/2-385 293
gr]26904 emb CAA39602.2 /2-508 386 EMAVVW	MONOLAONKTODLILDYVNOLCNRL <mark>P</mark> SPM 420		9/12/80/4/em/b/CA439602.2/2-508 560 AG I R VVDDE RVK ENGL R AD PMC S AC EMAVVWAQ 595
gr]264604 ge A4A32096.2 /2-385 298	En S <mark>y</mark> 301		9/12/9/6/4/31096.1/2-385 296 A
gr]26904]emb)CAA39602.21/2-508 421.01 SAVD	COSLOSMPDIEFTIGGKKFALKPEIYILK455		9/128904/emb/C4439602.2/2-508 894 NQLAQNKTQDLILDYVNQLCNRLP8PMGEA4VDC 427
gr]264604]g6]A4432096.21/2-385 802.0 EMVI S	CSSIDSLPDIVFTINGVQYPLSPSAYILQ335		9/1264604/g6/44431096.1/3-385 306
gr]28904 emb)CA439602.2 /2-508 456 ∨ <mark>0</mark> E <mark>G</mark> AA	AQ <mark>CISGFTAMDIPPPRSPLWILGDVFMGP</mark> 490		9/12/90/1/em/(CAAS9602.2)/2-508 428 65 L 6 SM PD I E FT I 66 KK FAL K FE E YI L K V6 E 6 A A 461
gr]264604 ge]A4A32096.2 /2-385 387 DDD	- SCT <mark>SGFEGMDVP</mark> TSS <mark>GELWILGDVFIR</mark> 0367		9/12/#/60/1/ge(A4A312096.3)/3-385 313 - 1 PD I VI T I N6 VO YPL S SAYILO - DD S - 340
gv 28904 emb CA439602.2 /2-508 491 YHTVFD	Y <mark>GKLRIGFAKAA 508</mark>		9/12/90/(em/b)CAA39602.2)/2-308 462 A 0 C I SC FT A MO I PPP RCP LWIL COVPMCPYHTVF 455
gv 264604 g6 A4A31096.1 /2-385 368 YYTVFD	Rann evglapva 385		gr]26/60/gb[A4431096.1]/3-385 341 - C T SC FEGMOVF TS SC ELWIL COVFIRG YTVF 372
Sequence 2 ID: gi[164604[gb]AAA31096.1]			Sequence 2 ID: gij164604 gb AAA31096.1 Residue: ASN (47)

· • ×

æ

- 《圖》 《문》 《문》

Jalview 2.3														
ile Tools Help Wind	ow													
/home/mvoorhie/Projec	ts/Courses/Practica	Bioinformatics	/gui/webr 🖬 🗹	X	1	home/mv	oorhie/P	rojects	/Courses	/Practica	IBioinforma	tics/gui/web	್ ಡಿ	\boxtimes
File Edit Select Vie	w Format Colour	Calculate We	eb Service		F	File Edit	Select	View	Format	Colour	Calculate	Web Service		
p]28904 emb CA439602.2 /2-508 p]264604 g6 A4432096.2 /2-385	1 M <mark>G</mark> TR <mark>G</mark> LALALL 1 MKWLLLL	AAVELE <mark>D</mark> TVE <mark>P</mark> A SEVVES E CEVKV	ASEAE <mark>GLVR</mark> IAL 85 P LVRKKS 26	-		26904 emb (264604 gé .4	AA39602 1 4431096 1	12-508 2-385	1 MG T 1 M · · · K	UALALLA VL-L-LL	AAVLLLOTVI SLVVI	PAASEAEGLV	VP 20	-
y]26904 emb CAA39602.2 /2-508 y]264604 g6 AAA31096.2 /2-385	36 K <mark>KRPIDRN</mark> SRV 27 L <mark>R</mark> ONLIKNGKL	AT <mark>GL</mark> S <mark>GG</mark> EEQ <mark>P</mark> L KDFLKTHKHN <mark>P</mark> -	L <mark>SGANP</mark> LRSEEE70 ASK <mark>Y</mark> FPEAAALI60			28904 emb 6 264604 g8 A	AA39662.1 4431096.1	V2-508 2-385	35 21 	election	RVAT <mark>G-L</mark> SBC ▶ Edit	EEQPLL - BOA	• NP 64	
p]28904 emb CAA39602.21/2-508 p]264604 g6 A4A32096.21/2-385	71 GOIVAL K <mark>NYM</mark> N 61 GO-E <mark>PLENYL</mark> O	A DYFGEIGVGTP T EYFGTIGIGTP	POKETVIECTOS 105 AQDETVIEDTOS 94			28904 emb 6 264604 g8 A	4439602.2 4432096.2]	V2-508 2-385	65 49 A S K YI	SEEE Peaaal	Output to Create Se	Textbox quence Featu	re 1	
p]28904 emb CA439602.2 /2-508 p]264604 g6 A4432096.2 /2-385	106 SNLWYPSAKCY 95 SNLWYPSVYCS	FSIACYLHSRYK - Slacsdhnofn	AGASSTYKKNGK 140 PODSSTFEATSQ 128			28904 emb 0 264604 gá A	34439602.2 4437096.1]	V2-508 2-385	98 <mark>PPO</mark> K 82 <mark>PAD</mark> D	TVIEDTO	Group SSNLWVPSV	YCS - BLACSD	▶ 26 NO 114	
p 28904 emb CA439602.2 /2-508 p 264604 g6 A4A31096.1 /2-385	141 PAAIQYGTGSI 129 ELSITYGTGSM	AGYFSEDSVTVS TGILGYDTVOVS	DLVVKDOEFIEA175 GISDTNOIFGLS163			28904 emb 0 264604 gá A	AA39602.2 4431096 1]	V2-508 2-385	127 YK AG 115 E N P DI	N <mark>SSTY</mark> KKI D <mark>SSTF</mark> EA'	NG K PAAIOYO TSQELSITYO	TGSIAGYFSEL TGSMTGILGY	5V 160 TV 148	
p]28904 emb CA439602.2 /2-508 p]264604 g6 A4A32096.2 /2-385	176 TK <mark>EPG</mark> IT <mark>FLVA</mark> 164 ET <mark>EPG</mark> SFL <mark>YY</mark> A	K FDG ILGLGF KE PFDG ILGLA <mark>YP</mark> \$	I SVGKAVPVWYK 210 I SASGAT PVFON 198			26904 emb 0 264604 gb A	4432662-2 4432696-2]	V2-508 2-385	161 TVGD 149 QVGG	VVKDOE SDTNOI	F - I E A TKEPC FGL S E T - EPC	ITFLV-AKFDO -SFLYYA <mark>P</mark> FDO	IL 192	
p]28904 emb CAA39602.2 /2-508 p]264604 g6 A4A32096.2 /2-385	211 MIEDGLVSDPV 199 LWDQGLVSQDL	FSFWLNRHVDE <mark>G</mark> FSVYLSSNDDS-	EGGEIIFGGMDP245 - GSVVLLGGIDS231			28904 emb 0 264604 g8 A	(4439662.2 4432096.2]	V2-508 2-385	193 GLBF 181 GLAY	CEISV-G SISASG	KAV <mark>P</mark> VWYKM - AT <mark>P</mark> VFDNLV	EDGLVS-DPVI VDDGLVSQD-LI	SF 224 SV 212	
p]28904 emb CA439602.2]/2-508 p]264604 g6 44432096.2]/2-385	246 KHYVGEHTYVP 282 SYYTGSLNWVP	VTOKGYWOFDMG VSVEGYWOITLD	DVL VGG K S <mark>TGFC</mark> 200 SITMDG - ETIAC 265			28904 emb 0 264604 g8 A	24439602.2 4431096.1	V2-508 2-385	225 WLNR 218 YL S SI	HVDEGEGG N-DDS-G	SVVLLGGID	KHYVBEHTYVI SSYYTGSLNWVI	VT 258 VS 244	
p]28904 emb CA439602.2 /2-508 p]264604 g6 A4A31096.2 /2-385	281 A <mark>BBC</mark> AAIA <mark>D 8</mark> 6 266 S <mark>BBC</mark> OAIVD TG	TSLLA <mark>BPT</mark> ATIT TSLLT <mark>GPT</mark> SATA	E INEKIGAAG VV 315 NIQ SDIGA S 297			28904 emb 0 264604 gá A	04439602.2 4437096.1	V2-508 2-385	259 Q K <mark>G Y</mark> 245 V E <mark>G Y</mark>	OFDMGD	- VLVGGKST SITMOGETI/	FCABBCAALAT	25G 291 2TG 276	
p]28904 emb}CA439602.2 /2-508 p]264604 g6 A4A31096.2 /2-385	S16 SQECKTIVSQY	GQQILDLLLAET	Ф <mark>Р</mark> ККІС 8 Ф V <mark>G</mark> LC 850			26904 emb 0 264604 gá A	AA39602.1 4431096.1]	V2-508 2-385	292 TSLL 277 TSLL	GPTAII GPTSAI	TEINEK IGA/ A N I	GVVSDECKTI	/ 5 0 825 - 292	
p]28904 emb CA439602.2 /2-508 p]264604 g6 A4431096.2 /2-385	851 TED <mark>G</mark> TR <mark>G</mark> VSA <mark>G</mark>	IRSVVDDE <mark>P</mark> VKS	N <mark>G</mark> LRAD <mark>P</mark> MCSAC385			26904 emb)(264604 gá A	4432096 1	V2-508 2-385	326 Y <mark>8</mark> 00 298		ETQ <mark>P</mark> KK <mark>I</mark> CSC	OVELCTEDETR	VS 359	
p]28904 emb CAA39602.2 /2-508 p]264604 g6 A4A32096.2 /2-385	386 EMAVVWMONOL 298	AQNKTODLILDY	VNOLCNRL <mark>P</mark> S <mark>PM</mark> 420 ENS <mark>Y</mark> 301			28904 emb 0 264604 g8 A	44336662 2 4432096 2]	V2-508 2-385	960 AG I R 296 A • • •	VVDDEP E	VKSNGLRAD - NSYG	MCSACEMAVV	MAQ 393	
p]28904]emb CAA39602.2]/2-508 p]264604 g6 A4A32096.2]/2-385	421 GESAVDCGSLG 802 GEMVISCS <mark>SI</mark> D	SMPDIEFTIGGK SLPDIVFTINGV	K F A L K <mark>P</mark> E E <mark>Y I L</mark> K 455 Q <mark>Y P L S P</mark> S A <mark>Y I L Q</mark> 336			28904 emb 0 264604 gb A	4439602.2 4432096.2]	V2-508 1-385	994 N.Q.L.A. 306	ONK TOD L	LDYVNQL <mark>C</mark>	IRL <mark>PSP</mark> M <mark>G</mark> ESA	D C 427	
p 28904 emb CA439602.2 /2-508 p 264604 g6 A4431096.1 /2-385	456 VEEGAAAQCI 887 DDD SC T S	GFTAMDIPPPRB GFEGMDVPTSSB	PLWILGOVFMGP 490 ELWILGOVFIRQ 367			28904 emb 0 264604 gá A	AA33602.2 AA32096.1]	V2-508 2-385	428 <mark>68 6</mark> 313 - 6 -	PDIVE	T I GG K K F AL F T I NG VO <mark>Y P L</mark> 1	PEETILKVGE PSATILQ-DD	AA 461 28 - 340	
p]28904 emb}CA439602.2 /2-508 p]264604 g6 A4431096.2 /2-385	491 YHTVFDYGKLR 368 yytvfdrann	I <mark>G</mark> FAKAA V <mark>G</mark> LA <mark>P</mark> VA	508 385			2 <i>8</i> 904 emb 0 264604 gá A	AA39602.2 4432096-1]	V2+308 2-385	462 A 0 C 341 C T	G F T AMD G F E GMD	I <mark>PPP</mark> R <mark>gp</mark> lw V <mark>P</mark> TSS <mark>g</mark> elw	ILGDVEM <mark>GPYH</mark> ILGDVEIRQ <mark>YY</mark>	VF 495 VF 372	
				•										¥
Sequence 2 ID: gi[164604]gb	AAA31096.1				s	equence 1	D: gi[1890	4 emb	CAA39602.3	Residue:	: LYS (36)			

. . . .

æ

<ロト <部ト < 注ト < 注ト

Mark Voorhies Pairwise Alignment

			Jalview 2.3		
ile Tools Help Wind	ow				
/home/mvoorhie/Project	s/Courses/PracticalBioinformati	cs/gui/webr 🖬 🛛 🛛	/home/mvoorhie/Project	s/Courses/PracticalBioinform	iatics/gui/web 🗗 🗹 🔀
File Edit Select View	r Format Colour Calculate V	Veb Service	File Edit Select View	v Format Colour Calculate	Web Service
p]28904 emb CAA39602.2]/2-508 p]264604 g6 AAA31096.1]/2-385	1 M <mark>6</mark> TR <mark>6</mark> LALALLAAVLLL <mark>0</mark> TVL <mark>P</mark> 1 ····MKWLLLLSLVVLS E CLVK	AASEAEGLVRIAL85 VPLVRKKS26	gr[26904]emb]CAA39602.21/2-508 gr[264604]g6]AAA31096.11/2-385	1 MGTRGLALALLAAVLLUOT 1 M WAL-L-LLSLV	VLPAASEAEGLVRIA 34
p]26904 emb CAA39602.2]/2-508 p]264604 g6 A4A31096.2]/2-385	36 K K R P I D RN S RVATGL S GG E E O P 27 L RONL I KNGKL K D F L K TH KHN P	LLSGANPLRSEEE70 - ASKYFPEAAALI60	7/120904/emb/C443960221/2-508 7/1264604/g8/44431096.1/2-385	35 L - KKRPIDENSRVATG - LS 21 LVRKKSL - BONLIKNGKL -	GGEEQPLL - SGA - NP 64 KD FLKTHKHNP 48
p]28904 emb CAA39602.21/2-508 p]264604 g6 AAA32096.21/2-385	71 GOIVALKNYMNAQYFGEIGVGT 61 GO-EPLENYLDTEYFGTIGIGT	PROKETVIEDTGS 105 PAQOETVIEDTGS 94	pr[28904]em8[C4489602.2]/2-508 pr[264604]g8[A4A32096.2]/2-385	65 L - B SEEE GD IVALK 49 ASKYFPEAAALIGD - EPLE	NYMNAQYFGEIGVGT 92 Nyldteyfgtigigt 81
p]28904 emb)CA439602.2 /2-508 p]264604 g6 A4A31096.1 /2-385	106 SNLWYPSAKCYFSIACYLHSRY 35 SNLWYPSVYCS-SLACSDHNOF	KAGASSTYKKNGK140 N <mark>P</mark> DD SSTF EATSQ128	9/[28904]em6]CAA39602.2]/2-508 9/[264604]96]AAA31096.1]/2-385	98 PPOKETVIEDTGSSNLWVP 82 PAODETVIEDTGSSNLWVP	SAKCYFSIACYLHSR 126 Syycs - Slacsdhng 114
p]28904 emb CA459602.2 /2-508 p]264604 g6 A4A31096.2 /2-385	141 PAAIQYGTGSIAGYFSEDSVTV 129 ELSITYGTGSMTGILGYDTVOV	GOLVVKDOEFIEA175 GGISDTNOIFGLS168	gil28904[em6]CAA39602.2]/2-509 gil264604[g6]AA431096.1]/2-385	127 YKAGASSTYKKNGKPAAIQ 115 FNPDDSSTFEATSQELSIT	YGTGSIAGYFSEDSV 160 YGTGSMTBILGYDTV 148
p]28904 emb CAA39602.2 /2-508 p]264604 g6 AA431096.2 /2-385	176 TK <mark>EPG</mark> IT <mark>FLVAKFDGILGLGF</mark> K 164 ET <mark>EPG</mark> SFL <mark>YYAP</mark> FDGILGLA <mark>YP</mark>	Create New Sequ	ence Feature(s)	EF - TEATKE I FOLSET - E	PGITFLV-AKFDGIL 192 PG-SFLVYAPFDGIL 180
a]26904 emb CAA39602.2 /2-508 a]264604 g6 A4A32096.2 /2-385	211 MIEDGLVSOPVFSFWLNRHVDE 199 LWDDGLVSCOLFSVYLSSNDDS	B ? Name: P	ositive Patch	GKAVPVWYK G-ATPVFDN	MIEDGLVS-DPVFSF224 LWDDGLVSOD-LFSV212
p 28904 emb CAA39602.21/2-508 p 264604 g6 A4A32096.21/2-385	246 K HYVGEHTYVPVTOKGYWDEDM 282 Syytgsliwyvpvsvegyw0itl	B (Colour:	GGEIIFGGM GSVVLLGGI	DPKHYVSEHTYVPVT 258 DSSYYTSSLNWVPVS 244
p 28904 emb CAA39602.21/2-508 p 264604 g6 AAA32096.21/2-385	281 AGGCAATADSGTSLLAGPTATT 265 SGGCOATVDTGTSLLTGPTSAT	T A Description:	Run of three positive residues ap penalty	aligned with no g	TOF CABOCAATAD SG 291
a 18904 amb CAA39602.2 /2-508 a 164604 g6 AAA31096.1 /2-383	S16 SQECKTIVSQY <mark>0</mark> 00 ILDLLLAE	1		ITEINEK IG IA - N - I	AAGVVSDECKTIVS0 825
a 28904 amb CAA39602.2 /2-508 a 264604 g6 AAA31096.2 /2-385	851 TED <mark>G</mark> TR <mark>G</mark> VSA <mark>G</mark> IRSVVDDE <mark>P</mark> VK	۹ ـــــ	OK Cancel	AETOPKKIC	SOVOLCTEDOTROVS 359
a 26904 emb CA439602.2 /2-508 a 264604 g6 A4432096.2 /2-385	386 EMAVVWMONQLAONKTODLILD 298	YVNOLCNRL <mark>P</mark> S <mark>PM</mark> 420 ENS <mark>Y</mark> 301	gv[28904]emb}CAA39602.232-508 gv[264604]g8]AA431096.232-385	360 AGIRSVVDDEPVKSNOLRA 236 A · · · S · · · · E · · NSYG · · ·	DPMCSACEMAVVWMQ 393
p 28904 emb CAA39602.21/2-508 p 264604 g6 A4A32096.21/2-385	421 SESAVOCGSLGSMPDIEFTIGG 802 SEMVISCSSIDSLPDIVFTING	KK FALKPEEYILK455 VQ <mark>YPLSP</mark> SA <mark>YIL</mark> Q336	o 28904 emb C44396022 2-508 o 264604 g6 A4A310961 2-385	S94 NOLAONKTODLILDYVNOL 306	CNRLPSPMGESAVDC 427
28904 emb CA439602.2 /2-508 264604 g6 A4A32096.2 /2-385	456 VOEGAAAQCISGFTAMDIPPPR 887 DDD SCTSGFEGMDVPTSS	SPLWILGDVFMGP 490 GELWILGDVFIRQ 367	p/[28904]emb]CAA39602.2]/2-508 p/[264604]g6]AAA31096.1]/2-385	428 GSLGSMPDIEFTIGGKKFA 313 - SL PDIVFTINGVOYP	LKPEEVILKVGEGAA 461 LSPSAVILQ-DDQS-340
p]28904 amb CAA39602.2 /2-508 p]264604 g6 AAA31096.1 /2-385	491 YHTVFDYGKLRIGFAKAA 368 yytvfdrann <mark>e</mark> vglapva	508 385	gi[26904]emb]CAA39602.2]/2-509 gi[264604]g6]AAA31096.2]/2-385	462 ADCISCETAMDIPPPROPL 341 - CTSCECMDVPTSSCEL	WILSDVFMGPYHTVF 495 WILSDVFIROYYTVF 372
			-		*
Sequence 2 ID: gi[164604[gb]	AAA31096.1		Sequence 2 ID: gi[164604[gb]	AAA31096.1 Residue: LYS (25)	

. . .

æ

- 《圖》 《문》 《문》

				Jalview 2.3				
ile Tools Help Wind	ow							
/home/mvoorhie/Projec	ts/Courses/PracticalB	ioinformatics/gui/webr p	* 🗗 🛛	/home/mvoorhie/Proje	cts/Courses/Prad	icalBioinformatics/gu	i/web ¤ ⊠	\boxtimes
File Edit Select Viev	v Format Colour C	alculate Web Service		File Edit Select Vi	ew Format Colo	ur Calculate WebS	service	
v 28904 emb CAA39602.2 /2-508 v 264604 g6 AAA31096.1 /2-385	1 M <mark>6</mark> TR <mark>6</mark> LALALLAA 1 MKWLLLLSL	VLLLOTVLPAASEAEGLV <mark>R</mark> IA VVLSECLVKVPLV <mark>R</mark> KK	L 35 S 26	Fetch Sequence(s) Add Sequences		LAAVLLLOTVLPAASE	AEGLVRIA 34 C - LVKVP 20	
i]26904 emb CAA39602.2 /2-508 i]264604 g6 A4A31096.1 /2-385	36 K K R <mark>P I</mark> D <mark>RN</mark> S RVA T 27 L <mark>R</mark> ONL I KNG K L K D	GLSGGEEOPLLSGANPLRSEE Flkthkhn <mark>p - As</mark> kyf <mark>p</mark> eaaal	E 70 I 60	Reload Save	Ctrl-S	SRVATG-LSGGEEQPL NLIKNGKLKDF	L - <mark>SC</mark> A - <mark>N P</mark> 64 L K T H K H N P 48	
28904 emb CAA39602.2 /2-508 264604 g6 A4A32096.2 /2-385	71 GD IVALKNYMNAD 61 GD - EPLENYLD TE	YFGEIGVSTPPOKFTVIFDTG YFGTIGISTPAQOFTVIFDTG	5 105 5 94	Save As Output to Textbox	CtrH-Shitt-S	GDIVALKNYMNAQY	FGTIGIGT 81	
28904 emb CAA39602.2 /2-508 264604 g6 AAA31096.2 /2-385	106 SNLWVPSAKCYFS 95 SNLWVPSVYCS-5	IACYLHSRYKAGASSTYKKNG LACSDHNQFNPDDSSTFEATS	K 140 Q 128	Page Setup Print	Ctrl-P	TGSSNLWVPSAKCYFS TGSSNLWVPSVYCS-S	LACYLESR 126	
28904 emb CAA39602.2 2-508 264604 g0 AAA31096.1 2-385	141 PAAIQYGTGSIAG 129 ELSITYGTGSMTG	YFSEDSVTVGDLVVKDOEFIE ILGYDTVOVGGISDTNOIFGL	A 175 S 163	Export Image Export Features		KNGKPAAIOYGTGSIA ATSQELSITYGTGSMT	GYFSEDSV 160 GILGYDTV 148	
9 28904 emb CAA39602.2 /2-508 9 264604 g6 AAA31096.2 /2-385	176 TK <mark>EPG</mark> ITFLVAK 164 ET <mark>EPG</mark> SFLYVAPF	DGILGLGFKEISVGKAVPVWY DGILGLA <mark>VP</mark> SISASGATPVFD	'K 210 N 190	Export Annotations Load Associated Tree		EF - LEATKEPG ITELN	YA <mark>PFDGIL</mark> 192	
28904 emb CAA39602.2 /2-508 264604 g6 A4A32096.2 /2-385	211 MIE <mark>DGLVS</mark> OPVFS 199 LWDOGLVSODLFS	FWLNRHVDEGEGGEIIFGGMD Vylssndds-Ssvyllggid	P 245 S 231	Load Features / Anno <u>C</u> lose	ctrl-w	G ATPVEDNLWDOGLA	/S - DPVFSF 224 /SOD - LFSV 212	
28904 emb CAA39602.2 /2-508 264604 g6 A4A32096.2 /2-385	246 KHYVGEHTYVPVT 282 Syytgslnwvpvs	OKGYWOFDMODYLYGGKSTOF VEGYWOITLDSITMDG-ETIA	C 200 C 265	gi[28904]em6]C44396022[/2-5: gi[264604]g8]A44310961[/2-38	225 WLNRHVDEG 213 YLSSN - DDS	EGGEIIFGGMDPKHYVG GSVVLLGGIDSSYYTG	EHTYVPVT 258 Slnwvpvs 244	
28904 emb CA439602.2 /2-508 264604 g0 A4431096.1 /2-385	281 A <mark>B G C</mark> A I A <mark>D S G T S</mark> 265 S <mark>G G C O A I V D T G T S</mark>	LLA <mark>BPT</mark> ALITEINEKIGAAGV LLT <mark>BPT</mark> SALANIQSDIGAS -	/∨315 - 297	gi[28904]em6]CAA39602.2]/2-50 gi[264604]g6]AA431096.1]/2-30	259 Q K G Y W O F D M 245 V E G Y W O I T L	GD - VLVGG KSTOFCAGO - DSITMOGETIA - CSGO	CAATADSG 291	
28904 emb CA439602.2 /2-508 264604 g6 A4431096.1 /2-385	516 SQECKTIVSQY <mark>6</mark> 0	QILDLLLAETQ <mark>P</mark> KKICSQV <mark>G</mark> L	C 850	gi[28904]em6]CAA39602-2]/2-5/ gi[264604]g6]AA431096-1]/2-38	292 TSLLAGPTA 277 TSLLTGPTS	IITEINEK IGAAGVVS AlaNI	ЕСКТІV <mark>S</mark> 0 325 <mark>S</mark> - 292	
28904 emb CAA39602.2 /2-508 264604 g6 AAA31096.2 /2-385	851 TFD <mark>G</mark> TR <mark>G</mark> VSA <mark>G</mark> IR	SVVDDE <mark>P</mark> VKSN <mark>G</mark> LRAD <mark>P</mark> MCSA	C 885	gr[26904]em6}CAA39602-21/2-5i gr[264604]g8]A4431096-11/2-38	325 Y <mark>0</mark> 00 I LD L L 298 D	LAETO <mark>P</mark> KKICSOV <mark>B</mark> LCT	FD <mark>G</mark> TR <mark>G</mark> ∨S 359 295	
28904 emb CAA39602.2 /2-508 264604 g6 A4A32096.2 /2-385	386 EMAVVWMONOLAO 298	NKTODLILDYVNOLCNRL <mark>P</mark> S <mark>P</mark>	M 420 Y 301	gr[28904]em6}CAA396022]/2-50 gr[264604]g8]A44310963[/2-38	960 AG I R SVVDD 296 A · · · S · · · ·	EPVKSNOLRADPMCSAC ENSYG	EMAVVWMQ 393 EM-V 305	
28904 emb CAA39602.2]/2-508 264604 g6 A4A32096.2]/2-385	421 GESAVOCGSLGSM 302 GEMVISCSSIDSL	PDIEFTIGGKKFALKPEEYIL PDIVFTINGVOVPLSPSAYIL	K 455 Q 336	gr[28904]em6[C4439602.2]/2-51 gr[264604]g8[A4431096.1]/2-38	394 NQLAQNKTQ 306	DLILDYVNQL <mark>O</mark> NRL <mark>PS</mark> F	MGESAVDC 427	
28904 emb CA439602.2 /2-508 264604 g6 A4A31096.2 /2-385	456 V <mark>SES</mark> AAAQC I <mark>SG</mark> F 887 DDD SCT <mark>SG</mark> F	TAMDIPPPRSPLWILGDVFMG EgmdvpTSSGELWILGDVFIR	P 490 Q 867	gi[28904]em6]C4A89602.2]/2-51 gi[264604]g6]A4A33096.3]/2-38	428 BSLG SMPD 313 - SL PD	EFTIGGKKFALKPEETI VFTINGVOYPLSPSATI	LKVEEGAA 461 LQ-DDOS-340	
s 28904 emb CAA39602.2 /2-508 s 264604 g6 AAA31096.1 /2-385	491 YHTVFDYGKLRIG 368 yytvfdrann Kvg	FAKAA LA <mark>P</mark> VA	508 385	gv[28904]em6}CAA39602.2]/2-51 gv[264604]g6]A4A31096.1]/2-38	462 ADC I SGF TA 341 CTSGF E	M <mark>D I P P P R G P L WIL GD V P</mark> MD V P T S S G E L WIL GD V P	MGPYHTVF 495 IROYYTVF 972	
			-					
Sequence 1 ID: gi[18904]emb	CAA39602.1 Residue: LE	EU (35)		Sequence 2 ID: gi[164604]	b AAA31096.1 Resid	ue: LYS (25)		

. . .

æ

- 4 聞 と 4 臣 と 4 臣 と

				Jalview 2.3
ile Tools Help Windo	w			
/home/mvoorhie/Project	s/Courses/Practica	Bioinformatics/gui/webr	• 🛛 🖾	/home/mvoorhie/Projects/Courses/PracticalBioinformatics/gui/web 🖬 🗹 🔀
File Edit Select View	r Format Colour	Calculate Web Service		File Edit Select View Format Colour Calculate WebService
a)28904 emb CAA39602.2 /2-508 a)264604 g6 AAA32096.2 /2-385	1 M <mark>8</mark> TR <mark>8</mark> LALALLA 1 · · · · MKWLLLLS	AVELE <mark>DTVEPAA</mark> SEAEGEV <mark>R</mark> Sevvestereveve	I A L 35 K K S 26	p]2696/em6/c445%c2]/2-510 1 MG T GU A LALLAAVLLUTVLEAASEA00UV 1A 94 p]26660/g6]44431096]/2-535 1 M - WL-LLSLV-VV-VLSEC - LVVVF 20 □
128904 emb CA439602.2 /2-508 1264604 g6 A4431096.2 /2-385	36 K K R P I D RN S RV A 27 L R ONL I KNG K L P	(TGL SGGEEOPLLSGANPLRS DFLKTHKHNP-ASKYFPEAA	EEE 70 ALI 60	9126064 - m5/2453062 2/2-307 35 L - 14536 10 M SRVATO - L SOOE E 0 / LL - SOA - NH 64 9266604 944432095 2/2 325 22 L V 23235 L - SON L I KNGKL KD FLKEHKHR 48
p]28904 emb CAA39602.2]/2-508 p]264604 g6 A4A32096.2]/2-385	71. GOIVALKNYMNA 61. GO-EPLENYLD1	OYFGEIGVGTPPOKETVIFO EYFGTIGIGTPAQOETVIFO	TG S 105 TG S 94	2250/14m3/2455602 2/2-500 65 L
p 28904 emb CAA39602.2 /2-508 p 264604 g6 AAA31096.2 /2-385	106 SNLWYPSAKCYF 95 SNLWYPSVYCS	SIACYLHSRYKAGASSTYKK SLACSDHNOFNPDDSSTFEA	N <mark>G</mark> K 140 T S Q 128	Format: Jalview @ GFF
p 28904 emb CAA39602.2 /2-508 p 264604 g6 AAA31096.1 /2-385	141 PAATOYGTGSIA 129 ELSITYGTGSM1	GYFSEDSVTVGDLVVKDOEF Gilgydtvovggisdtnoif	IEA 175 GLS 163	to File to Textbox Close
a]28904 amb CAA39602.2 /2-508 a]264604 g6 AAA32096.2 /2-385	176 TK <mark>EPG</mark> IT <mark>FLVA</mark> # 164 ET <mark>EPG</mark> SFLYYAF	FDGILGLGFKEISVGKAVPV FDGILGLAVPSISASGATPV	WYK 210 F DN 198	GEF-IEATKEPGITELV-AKEDGIL 192 GIFGLSEN-EPG-SFLVVAPFDGIL 180
a]26904 emb CA439602.2 /2-508 a]264604 g6 A4432096.2 /2-385	211 MIEOGLVSDPVF 199 LWDQGLVSQDLF	SFWLNRHVDEBEGGEIIFGG SVYLSSNDDS-GSVVLLGG	MD P 245	212694-m6/C44CMC21/2-500 199 GLGFKEISV-GKAVPVWYKMIEDGLVS-DPVFSF224 2126604/m6/C44CMC20451/2-505 181 GLAVPSISASG-ATPVFDNLWDGGLVSOD-LFSV212
e]28904]emb CAA39602.21/2-508 e]264604]g6 AAA32096.21/2-385	246 KHYVGEHTYVP) 282 SyytgslnwvP)	TOKOVNOFOMOOVLVGGKST SVEGVNOITLOSITMOG-ET	G F C 200 I AC 265	225 WLNRHVDEGEGGEIIFGGMDPKHYVGEHTYVF 26660/glavdaos 1/2500 218 WLSSN - DDS - GSVVLLGGIDSS YYTGSLNWVPVS 244
p 28904 emb CA439602.2 /2-508 p 264604 26 44432096.2 /2-385	281 ABGCAALADSG 265 SBGCOALVDTG	SLLAGPTATITEINEKIGAA SLLTGPTSATANIOSDIGAS	GVV315 297	259 0 K G YWD FOMGD - VL VGC K B TOF CABGC AA I AD GG 291 2 K K K K G G A A I AD G 292 245 VE G YWD I TL - D S I TMDG E T I A - C S G C O A I VD TG 276
a 28904 amb CA439602.2 /2-508 a 264604 g6 A4A31096.1 /2-385	316 SQECKTIVSQY	QQILDLLLAETQ <mark>B</mark> KKICSQV	GLC 350	2012001/cm0/CAA30021/2500 292 TSLLAGPTALITEINEKIGAAGVVSDECKTIVG0 825 214460/cjaaa30061/2585 277 TELTGPTSALA-NI
a]28904 emb CAA39602.2 /2-508 a]264604 g6 AAA31096.2 /2-385	851 TED <mark>G</mark> TR <mark>G</mark> VSA <mark>G</mark>	RSVVDDE <mark>P</mark> VKSN <mark>G</mark> LRAD <mark>P</mark> MC	S A C 385	201200 4 m05/CAA.9402 3/2-500 5/264604 m05/CAA.9402 3/2-500 229 229 229 229 229 229 229 229 229 229
p]26904 emb CAA39602.2 /2-508 p]264604 g6 A4A32096.2 /2-385	386 EMAVVWMONOLA 298 · · · · · · · · · · · · · ·	ONKTODLILDYVNOLCNRL <mark>P</mark>	S PM 420 N S Y 301	20120040005/CAACXCC 1/2-500 5/264604/c644/310961/2-505 296 A · · · · · · · · · · · · · · · · · ·
p]28904 emb CAA39602.21/2-508 p]264604 g6 A4A32096.21/2-385	421 GESAVOCGSLG 302 GEMVISCSSID	MPDIEFTIGGKKFALKPEEY LPDIVFTINGVOVPLSPSAY	ILK 455 ILQ 336	02690 / km/s/CAA59502 3/2-505 5/26460 / joj AA431096 1/2 505 306
p]28904 emb}C4439602.2 /2-508 p]264604 g6 44432096.2 /2-385	456 V <mark>8 E B</mark> AAAQC I <mark>S</mark> 887 DDD S C T S C	FTAMDIPPPRSPLWILGDVF	MG P 490 I R Q 367	22560 4 g m () C 4 55602 2 1/2 505 2 26 4 g d 4 4 5 3 0 4 5 1 0 5
p 28904 emb C4439602.2 /2-508 p 264604 g6 444331096.2 /2-385	491 YHTVFDYGKLRI 368 YYTVFDRANNEY	G FAKAA G LAPVA	508 385	2 100 1 mil A 3002 1/2 502 462 A 0 C I SOFT AMD I PPEROPLWIL COVEMOPYHTYF 450 2 100 1 mil A 3002 1/2 50 341 - CT SOFE GMOVE TSSOELWIL COVE I ROVYTYF 572
Sequence 2 ID: gi[164604[gb]/	AAA31096.1 Residue:	ILE (79)		Sequence 2 ID: gi[164604]gb[AAA31096.1] Residue: PHE (210)

æ

<ロト <部ト < 注ト < 注ト

	Jaiview 2.3
<u>F</u> ile Tools Help Window	
/home/mvoorhie/Projects/Courses/PracticalBioinformatics/gui/webr 🖬 🖾	/home/mvoorhie/Projects/Courses/PracticalBioinformatics/gui/web 🖬 🗹
File Edit Select View Format Colour Calculate Web Service	File Edit Select View Format Colour Calculate WebService
Fetch Sequence(s) AAVLLETVLEASEAEGLVEIALB3 Add Sequences SLVVESCLVKVE Beload SLVVESCLVKVE	9/16/60/19/144310961//.387 1 M - WH - 1 - 110 12 - 12 - 12 - 12 - 12 - 12
Save Ctri-S LKDFLKTHKHNP-ASKYFPEAAALI60	25 L - KKRPIONSRVATO - LSOGE EOPLL - SCA - NP 64 g/26/60/g6/44432096.1/2-385 21 LVRKKSL - RONLIKNGKL KD - FLKTHKHNP 48
Save As Correction 100 PPG K FTVIFDTG 105 Output to Textbox PTEXF6 TIG IG TPAD OF TVIFDTG 94	CSL-COLVALX VYNA DYFGEIGVGT 92 971264604/g6/AAA310961/g-383 49 AS YFFEAAALIGD - EFLENYLD TEYFG TIGIGT 91
Page Setup Print Ctri-P S-BLACSDHNOENPODSSTFEATS0128	98 PPOKETVIEDTGSSNLWVPSAKGYFSIACYLHSR 126 grifefeorigelarasione 1/0-385 82 PADDETVIEDTGSSNLWVPSVYGS-SLACYLHSR 126
Export Image IAGYFSEDSUTVODLVVKODEFIEA175 Export Features MTCILGYDTVOVGEISDTNDIFELS183	212 TKAGA STYKKIGKEAA IOYOTOSIA STYKKIGKEAA IOYOTOSIA STYSEDSV 160 9/1244604/g6/jada10961/J-385 115 EN PODESTEA TSOELSITYOTOSMITSILOYDTV 148
Load Associated Tree AFFDG ILGLGFKEI BYGKAVFVWYK 210 AFFDG ILGLAVFSI BASGAT FYFDN 199	0/264604/g6/44432096.2/2-385 149.0V060 SDTN0 F6LSET - EPG - SFLVYAPFDG1L 180
Load Features / Annotations Close Ctrl-W LFBVYLSSNDDSGSVVLLGG1DS231	199 OL OF KEISV-OKAV PVWYKMIE OC V O DPVF BF 224 g/264604/g/444310961/0-385 181 CLAIP SI BASG - AT PVFDNLWDOCLVSOD - LF BV 212
9]24904/hem5/CA419602.21/2-50/9 246 K HYVE IN TYVE VE OK BYWE FOKOOV LVGC K STOFC 200 9]264604/g6/A4432096.21/2-385 292 STYTESLNWVFVE VE DYWE IT LOSITINGE - EI I AC 255	225 WLNRHYDEBECGEIIFCGMDRKHYDEGFTBYPYT 256 pilcefeorigilaaastoe 1ja 385 223 WLNRHYDEBECGEIIFCGMDRKHYVGEHTBYPYT 254
9/28904/em/b/CA499602.2/2-508 281 A BGC A A LAD 10 13 13 LL ABPT ALL TEINEK IOA AG VV 315 9/264604/gé/A4431096.1/2-383 266 8 GGC A LAD 10 TEIS LL TEPT SA LAN LO SD IGA 8 237	259 DK B VWC FDMCD . V L VGC K TOF G A GGC AA AD BC 231 prize#corige[A4A431096 1/13-385 245 VE DVWC I T L - DS T MO DE I I A - DS B GC A I VD I C 276
gy 28904 em/b CA4439602.2]/2-508 316 SQECKTIVSQY <mark>G</mark> QQILDLLLAETQ <mark>P</mark> KKICSQV <mark>G</mark> LC350 gy 264604]ge A44431096.3]/2-383	232 TELLABPTALITEINEKIGAAGVYSDECKTIVEO.825 pyl64604[g6]A4A31096.1]/3-885 277 TELLTEPTSALA-N-N
gy]28904 em/b CA439602.2 2-508 851 TFD©TR©LVSA©TTRSVVDDE₽VKSN©LRAD₽MCSAC885 gy]264604 g6 A4A312046.3 2-305	p/264604/gela4432096.2/2-385 238 244 245 245 245 245 245 245 245 245 245
g/1/28904/pm/b/CAA39602.2/2-308 385 EMAVVWMONOLAONKTODLILDYVNOLCNRL <mark>E</mark> S <mark>EM</mark> 420 g/1/24604/g6/AA431096.2/2-385 298ENS <mark>W</mark> 901	900 AG (R VVDD FVK NG L RAD MC SAC MAVVWK) 393 pile#60/gej44431096.1/1-385 236 A
gn[2690 4]em (b) (24439662,2) (2-50) 422 00 SAVD CCSLOSMPDIEFTIGG KKFALKFEIVTL k455 gn[26460 4]g6[44432096 2] (2-385 50 0 CEM VISCSEIDEL PDIVFTING VQ ¥PL SP SA ¥ILQ 336	21264604/gb[4A4A31096.1/1-385 306 427
gy]28904[em/b)CA439602.2]/2-508 455 VGEBAAACIISOFTAMDIPPPRDPLWILGOVFMGP490 gy]264604[gé A4432096.2]/2-383 887 DDD • • • • SGT <mark>SGFEGMDVP</mark> TSS <mark>GELWILGOVFI</mark> RQ367	428 GOLGENPOILEFILGEKKFALKREEVILKVGEGAA461 gri264604/gejaaa320963/3-383 - LPDIVITINGVOVFLSESAVILO-DDDS-340
© 29904 emb CAA59602.2 /2-308 491. <mark>YHTYFDYÖ</mark> KL <mark>E</mark> IGFAKAA 508 ⊚1264604 g6 AAA31096.1 2-383 368 <mark>WYTYFD</mark> RANN E VGLA P VA 385	462 ADD ISCETANDIFERRETWILGDVENSEPHEVE 455 9/1244604/g6/4443109611/1-385 941 GTEGEEEMOVETSSGEWILGDVERRUVY 75 372
	· · · · · · · · · · · · · · · · · · ·
Sequence 2 ID: gi[164604]gb[AAA31096.1] Residue: LYS (2)	Sequence 2 ID: gi[164604 gb AAA31096.1 Residue: GLY (80)

008

	Jalview 2.3
Eile Tools Help Window	
/home/mvoorhie/Projects/Courses/PracticalBioinformatics/gui/webr 🖬 🖾	/home/mvoorhie/Projects/Courses/PracticalBioinformatics/gui/web 🕫 🗹
File Edit Select View Format Colour Calculate Web Service	File Edit Select View Format Colour Calculate WebService
pi28904(emb(cAASB602.2) pi28904(emb(cAASB602.2) pi284604(pie)AAA32046.2)/ Expand Views x x x x x x x x x x x x x x x x x x x	1 Maireau Al All AAVI LOTVE AAA TAKEUVE I A 34 p/26-604/g6/A4432096 2/0-385 1 M
g128904[emb(CA439602.2] g1264604[g6[A4432096.2]/ Show ▶ 1 <mark>P</mark> - ASK FP EAAALI 60	SSL-KKRPIORNSRVATO-LSDGEEOPLL-SCA-NP64 gviz64604/g¢j444320963j42-385 21.LV3KKSL-RONLIKNGKLKDFLKDHKHMP48
g120904[em6(CA409602.2] p120904[em6(CA409602.2] p1204604[g6]A4432096.2]/ □ Show Annotations 3TEAODETV1FDT06.34	collocation (collocation) collocation (collo
p]26904/em6/CA439602.2 ¹² Show Sequence Features p]264604/g0/A4432096.2 / Feature Settings PENPOD SSTEEATS0.128	pland result and resul
p]28904/emb/p24439602.2 p]264604/p6/44432096.2 // Dise4604/p6/44432096.2 // Openview Window	127 WKAGASSYKKINGKEAALOVGTGSLASVESEDSV 160 gyl64604ggla444310961jQ-385 115 KNPDDSSTEEATSQELSITYGTGSMTGLLGVDTV148
9 28904 emb(54459602.2)2/309 1/5 TKEPGTTFLVAKFDBTLBLBFKETSVGKAVPVWYK210 9 264604 g6 44431096.2)2-305 164 ET <mark>EPG</mark> SFLYYAFFDGILGLAVFSISASGATPVFDN190	161 TVGOLVVKDCF-TAAKEPGIFLV-AKEOGIL 192 pi26#604/gb 444310961/JC-885 149 QVGGISDTNQIFGLSET-EPG-SFLWYAPEDGIL 190
g)[26904]emb)CAA39602.2]/2-308 211 MIEDGUVGDFVFSFWUNRHVDEGEGGEIIFGGMDF245 g)[264604]g6]A4A31096.1]/2-385 199 LWDDGUVGQDFSVYUSSNDDS6SVVULGGIDS231	199 DL&AKEI OV-OKAVPVWYKMIEDOLVO-DEVFEF 224 prize#604/ge/444310961/4-385 181 DLAFPSI SASG ATPVFDNLWDOGLVOOD-LFSV 212
p]26904jemb/CAA39602.21/2-509 246 K HTVG IIITYV PVTOK GYWOPDMOOVL VGG K STG FC 200 p]264604jgb]A4A31096.1 / 2-385 232 SYYTG SLNWV PVSV EGYWOITLOSITMOG - ETIAC 255	225 WLAIRH VOEGET FOGMDEX HYVGEHTYVFVT 258 271264604/26/444310961/2-383 213 VLSSN - DDS - GSVVLLGGIDSSWYTGSLNWVFVZ 244
p]26904jemb/CAA39602.2]/2-508 281 ASGC A I ADSC TSL L AGPTATITE IN EXICA AGV V 315 p]264604jej44431096.1/2-385 266 S <mark>GG</mark> C A I VDTG TSL L TGPTSA I AN I Q SD I GA S 237	259 0 K 5 W 0 F 0 M 6 D V U 6 6 K 8 T 6 F 0 A 6 6 6 A A A D 8 6 291 prize#ee/ge/aaa31096 1/2-083 245 V 6 6 W 0 T L - 0 S T M 0 6 F T A - 0 S 6 6 6 A V 0 T 6 276
p 26904 emb C44536602.2 2-508 S16 SQ ECKTIVSQ Y <mark>6</mark> Q Q ILDLLLA ETQ <mark>P</mark> KKICSQ V <mark>6</mark> LC 350 p 264604 ge 44432096.2 2-385	232 TELLAGPTALITEINEKIGAAGVVSDECKTIVB0 825 gr]264604[gb]44431096.1[J-385 277 TELLTBPTSALA
p)28904)em/b)CAA39602.2)/2-508 851. TFD <mark>G</mark> TR <mark>G</mark> VSA <mark>G</mark> IRSVVDDE <mark>P</mark> VKSN <mark>G</mark> LRAD P MCSAC385 p)2164604)g6 AA4312096.2)/2-365	9/16/60/1g6/44431096.1/12-385 238 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
p 26904 emb CAA59602.2 /2-506 366 EMAVVWMONOLAONKTOOLILDYVNOLCNRL <mark>R</mark> S <mark>RM</mark> 420 p 264604 ge A4A31096.1 /2-385 236	prize#ev4ge/44431096.1/2-385 236 4 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
p]2/8904/em/b/CA439602.21/2-509 421 50 SAVOCOLUSIANPDIEFTISSKK FALKFETYLK 455 p)264604/gé)44431096.21/2-385 802 50 MVISCSIDELPDIVFTINGVCYPLSPSAVILQ 336	212-00-10-0012-002-002-002-002-002-002-0
p]26904 wmb CAA59602.2 /2-509 456 V <mark>6</mark> EGAAAQCI <mark>BG</mark> FTAMDIPPPRB DLWILGDVFMGF 490 p]264604 ge 44A31096.2 /2-385 887 DDDSCTBGFEGMDVPTSS <mark>6ELWILGDVFI</mark> RQ367	21264664/gela4A31096.1/2.385 313 - L - POIVFINGVVVVVLSFSAVILO-DDDS-340
9/28904/emb/CAA59602.2/2-309 491 YHTVFDYGKLEIGFAKAA 508 9/264604/g6/AAA31096.1/2-383 368 YYTVFDRANNEVGLAPVA 383	462 AOC 1 86 F AMD I P P R 67 W L 60 V F M V F 495 gr[26#604]gb]A4A310961][2-385 341 C T & 6 F 60M0 V F 1 S S 6 E LWI L 60 V F I R 9 V T V F 372
	•
Dedneuce I ID: Bilitoaneleurolcywoacost I kesigne: CED (II)	pequence 2 in: gite+ee+ign/wexprosert

. . . .

æ

- 4 聞 と 4 臣 と 4 臣 と

Mark Voorhies Pairwise Alignment

jalview 2.3								
<u>F</u> ile Tools Help Window								
/home/mvoorhie/Projects/Courses/PracticalBioinformatics/qui/webr 🖬 🗖 🚺 / /home/mvoorhie/Projects/Courses/PracticalBioinformatics/qui/web 🛱 🏹								
File Edit Select View Format Colour Calculate Web Service File Edit Select View Format Colour Calculate Web Service								
9126904/800%/CAA996022/2-509 1 MGTRGLALALAAVLLLTAVLLTAAVLL 91264604/g6/AA431096.1/2-385 1	V <mark>R</mark> IAL 35 VRX 8 26	9/26/60/96/4432096.1/2-385 1 M ML 2 LS LV LC LV VP 20						
gizeeo4jembic4 Sequence Feature Settings 🛛 🗖 🗹	SEEE 70	gelessed = mskradower 2/2-508 S5 L - KK RP 10 RNSRVATG - LSDGEE0 PLU - SGA - NP 64						
gr/264604[g6]44. Feature Settings DAS Settings	AAL I 60	WIZE#E0#IgelA4A31096.1/2-385 21 UVIZERSL - BONLIKNGKL KD FLKTHKHNP 48						
gi20904(emb)Cr Feature Type Colour Display gi264604(gi)A4 Positive Patch V	2 TG S 105 2 TG S 94	9/1264604/g6/AAA31096.1/3.383 49 AS YF PEAAALIGD - EPLENYLD TEYFG TIGIGT 81						
gr]28904[emb)Cc gr]264604[gb]44	(NGK 140 NT SQ 128	prize#cerigeia4431096.1/3-385 82 PPD KFTVIFDTCSSNLWVP3AKCYF5LACYLHSR 126 gri2e#cerigeia44431096.1/3-385 82 PADDETVIFDTCSSNLWVP3VYDS-BLACSDHNO 114						
gr]28904[emb]Cc gr]264604[g6]A4	IEA175 GL S 169	9/264604/g6/A4A32096.1/2-385 115 FNPODESTFEATSCELS TYCTOSHTELCONTYLIE						
gi]28904[emb]C4 gi]264604[gi]AA	WYK 210	0/164604/g6/444310961/12385 149 0 VGG SDTNO I FOLSET - EPG - GFL YVAPFDG I 1 80						
g)28904[emb]Cr gi]264604[g6]A4	MD P 245 3 1 D S 281	199 OL OF KEI V. OKAV PVWYKM E OCLVO. DE VE F 224 @1264604/g6/444319961/1388 181 OLA PSI BASG - ATP VED NLWDOOLVOOD - LF SV 212						
g)28904[emb]2* g)264604[g6]44	1 AC 265	225 WLNRHVDEGEDGEIIFGGMDFKHVVGEHTVPVT 28 grij64604/g6j444310961j1385 213 VLSSN DDS S SVVLLGGIDSSWTTSSLNWVPVE 244						
g)28904[emb]2- g)264604[g6]44	3 · · · 297	259 OK B WO FOMOD V U VOK K TOF CABOCAA LADBO 291 gri26+60+ig6ja4431096 1/1-395 245 V B WWO I T L-B S T THO C F I LA - C S C C A I V D TO 276						
g)28904[emb]0- g)264604[g6]44	/GLC 350	232 SLLAGPTA11TEINEKIGAAGVVSDECKTVK0 825 gr[264604/gg]a44310961/13-385 277 SLLTGPTSATA						
p)28904[emb[Gr g)264604[g6]A4. Optimise Order	5 S A C 385	9/16/60/1g6/44431096.1/1.385 233 0 235						
gi26904[emb[Ci gi264604[g6]A4 OK Cancel Load Colours Save Colours	5 PM 420 EN S 7 301	900 AG R VVDD KVK NGL RAD KC SAC MAY VWMC 999 9/1264604/ge/44431996.1/1.385 256 A						
9]264604[g6]A4A31096.2 [2-385 802 8 EMVISESIDEL POIVETINEVOVELSES	AYILQ 336	394 NOLAONKTODI I LOYVNOL ONEL PRAMEE AVDC 427 9/126/60/lg6/44431096.1/2.385 306 8 10 312						
gy]28904 em/b)CA439602.2)/2-508 456 VGEGAAAQCISCFTAMDIPPPRCPLWILCO gy]264604 g6 44A431096.2 /2-385 837 DDD SCTSFFEGMDVPTSSGELWILCO	VFMGP 490 VFIR0 367	428 GBLGSMPDIEFTIGGKKFALKPEETILKVGEGAA461 gri264604/gejaaa330963/3-385 313 - 51 PDIVETINGVOYPLSPSAVILO-ODOS- 340						
©)28904 em/b)CA499602.2]/2-509 491, ¥HTVFDYGKLEIGFAXAA ©)264604]gé A4A31096.1]/2-363 368 ¥YTVFDRANN¥VGLAPVA	508 385	462 ÅDØ I BOFT ÅND I FPERDELWILDDVF WOPVH TVF 495 gr[64604g6[A4431096.1]4-385 941 - CT BOFE BMOVF TS SE ELWILDDVF ROPVH TVF 372						
Sequence 2 ID: gi[164604[gb]AAA31096.1]		Sequence 2 ID: gi[164604[gb]AAA31096.1] Residue: ASP (118)						

æ

'≣ ▶

・日・ ・ ヨ・・

To The Help Window To The Help Window The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Service The Call Select View Format Colour Calculate Web Serv		Jalview 2.3
Improvember bioler opiests/Courses/Practicalization/ormatics/gui/web n° (*) Dimension/ormatics/gui/web n° (*) Dimension/ormatics/gui/web n° (*) Improvember bioler opiests/Courses/Practicalization/ormatics/gui/web n° (*) Dimension/ormatics/gui/web n° (*) Dimension/ormatics/gui/web n° (*) Dimension/ormatics/gui/web n° (*) Improvember biol/regists/Courses/Practicalization/ormatics/gui/web n° (*) Dimension/ormatics/gui/web n° (*) Dimension/ormatics/gui/w	ile Tools Help Window	
III Edit Select View Format Colour Calculate Web Service File Select View<	/home/mvoorhie/Projects/Courses/PracticalBioinformatics/gui/webr	אר ס' ס' 🔀 🛛 /home/mvoorhie/Projects/Courses/PracticalBioinformatics/gui/web ס' ס' 🔀
Depresentation Deprese	File Edit Select View Format Colour Calculate WebService	e File Edit Select View Format Colour Calculate Web Service
Development	p]26904 emb CA4596022 /2-509 1 M <mark>6</mark> TR <mark>6</mark> LALLLAAVULL <mark>0</mark> TVL <mark>P</mark> A4SEAE <mark>CLV</mark> p]264604 gé A4432096.1/2-385 1 ····MKWLLLSLVVLS∎CLVKV <mark>P</mark> ····LVL	VIII18135 A 1 MITTER A A
Description	p]26904 emb/CA439602.2 /2-308 36 <mark> </mark>	.RSEEE70 5AAALI60 0/1264604/06/444310961/0-385 21.LV:KK1SL-MONLIKNGKLKDELK 0+KH4H 48
Bardy-Bardy-Bardy Control 100 Bardy-Bardy-Bardy Control 100 Bardy-Bardy Bardy-Bardy Bardy-Bardy-Bardy-Bardy-Bardy-Bardy-Bardy-Bardy Bar	p]26904]emb/CA439602.21/2-508 71.500 IVALKNYMINACYFGEIGVGTPPCKFTVIFD p]264604]gé A4A32096.21/2-385 61.500EPLENYLDTEYFGTIGIGTPACOFTVIFD	ID TO 5 105
Basel-Machinet State Basel-Mac	p]26904]em b)CA439602.21/2-508 106 SNLWYPSAKCYFSIACYLNSRYKAGASSTYKK p)264604]gé(A4A31096.21/2-383 95 SNLWYPSYYCS-SLACSDHNQENPDDSSTFEA	KKNOK 140 EA T SQ 128 prizefeorigejaAA31096 J[J-365 02 PADDF TVI FDTG S SN LWVP SAKOVF SI AGVLHSR 126 EA T SQ 128
Development	p 28904 em/b C4439602.2 2-503 141 <mark>P</mark> AATQY <mark>GTG5</mark> IAGYFSEDSVTVGCLVVKD <mark>0</mark> EF p 264604 gé A4A31096.2 2-383 129 ELSITY <mark>GTG5MTGILG</mark> YDTVQVGGISDTN <mark>0</mark> IF	EFIEA175 127 KAGASSTYKKINGKEAAIOYGTOSIACHESEDSV160 9/12/4/60/1g6/AAA310961/0-385 115 FNPDDstFEATSGELSITYGTOSMTGILGYDTV148
Development	p)28904 emb)CA439602.2]2-303 176 TKEPGITFLVA KFDGILGLGFKEIÖVGKA VPV p)264604]g6 A4432096.2]2-303 164 IT <mark>PG</mark> SFL <mark>YYA PFDGILGLA VP</mark> SISASGA TPV	PVWYK 210 161 TVGOL VYKODEF - LEAKKEPG I TPLV - AKKDGIL 192 (FVFDN 198) pj264604jg6j44431096.1j0.385 149 QVGB I SD TNG I FGL SE T - EPG - SFLWYAPFDGIL 180
Development 200 Total Participation State Participatin State Participation State Partinant Participatin State	p]26904 pmb/CA439602.2/2-308 22.1 MIEDGLVCDPVF5FWLNRHVDEGEGGEIIFGG p]264604 g6 A4432096.2/2-385 199LWDCGLVCCDLF5VYLSSNDDSGSVVLLGG	GGMD P 245 199 GL DF K E I SV - G K AV P VWYKM I E COLVE - DF VF B F 224 GG I D S 231 pilzereovigela4A31096.1/0.385 181 GL AM P SI SA SG - A T P VED N LWDD GL V G OD - LF SV 212
Development 2000 Development Development <tdd< td=""><td>p 2/8904 em/b CA439602.21/2-508 246 KHYVGEHTYVFVT0 KGYVGFDMGCVLVGGKS p 264604 g6 A4A32096.21/2-385 232 SYYTGSLNWVFVSVEGYVG1 TLDSITMOG-ET</td><td>CSTOP C 200 - ET I AC 265 p)[264604/g6]A4431096.1]/3.385 213 YL SSN - DOS - C SVVLL CC I D SS Y TOSLNINY PV 244</td></tdd<>	p 2/8904 em/b CA439602.21/2-508 246 KHYVGEHTYVFVT0 KGYVGFDMGCVLVGGKS p 264604 g6 A4A32096.21/2-385 232 SYYTGSLNWVFVSVEGYVG1 TLDSITMOG-ET	CSTOP C 200 - ET I AC 265 p)[264604/g6]A4431096.1]/3.385 213 YL SSN - DOS - C SVVLL CC I D SS Y TOSLNINY PV 244
2004/max.040000 [1:0.300 50 C K L L V K V K O L L L L L E T C K K L C V K L SO 2004/max.040000 [1:0.300 50 C K L V K V K V K V K V K V K V K V K V K	p)28904 pm/b)CA439602:2)2-508: 281.A <mark>BGC</mark> AAIAD <mark>SGTSLLABP</mark> AIITEINEKIGAA p)264604 g6)A4A32096.2)2-383: 266.8 <mark>BGC</mark> QAIVDTGTSLLTGPT <mark>SAI</mark> ANIQSDIGAS	A A G VV 915 A S · · · 297 9/264604/g6(A4431096.1/2-385 245 VE G YMG ITL · D S I THO D E I A · C S D C O A I VD TO 276
provinewscalescol	p 28904 em/b CA439602.2 2-308 31.6 SQECKTIVSQY <mark>B</mark> QQILDLLLAETQ <mark>P</mark> KKICSQV p 264604 g6 A4A31096.2 2-383	30 VGL C 850 292 19 L A GPT AL I TE I HEK I GAAG VVS HECKT I V 0 825 9/1244604/g6/A4431096.1/2-385 277 18 L I TOP 1 SAUA - H - I
2004/mar.2008/01/2013/02 S00 Million Control Link Contre Link Control Link Control Link Control Link Control	p)2890.4 xm/b)CA439602.2 /2-30.9 351. TFD <mark>G</mark> TR <mark>G</mark> VSA <mark>G</mark> IRSVVDDE <mark>P</mark> VKSN <mark>G</mark> LRAD <mark>P</mark> MC p)26460.4 g6)A4431.096.1 /2-385	BAC S AC 385 pil/24/04/04/05/06/01/04/05/04/01/04/05/04/01/04/05/04/01/04/05/04/01/04/05/04/01/04/05/04/01/04/05/04/01/04/05/04/04/02/06/01/04/05/04/04/02/06/01/04/05/04/04/02/05/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/02/04/04/04/04/04/04/04/04/04/04/04/04/04/
Developmenter 2000 Developme	p]26904]pmb/CA439602.2]2-308 386 EMAVVWMONQLAQNKTODLILDYVNQLCNRL <mark>E</mark> p]264604]g6]A4432096.2]2-385 298 - E	St. E S PM 420 S60 AG ()R VYDD EF VK NG L RAD MC SAC EMAN VWMC 355 - E N S W 301 pi/264604/gb/j4A431096.1/(J-385 256 A · · E · · · E · · · E · · E · · E · · · E · · E · · E · · E · · · · E · · · E · · · · E · · · · E · · · · E · · · · E · · · · E · · · · · · E · · · · E · · · · E · · · · E · · · · E · · · · E · · · · E ·
2004/ms/2004/m	p)2690-4)em 6(CAA3960 2.2)/2-508 421.01 SAVOCOBLOCMPDIUFII0OK KIALKINIU p)26460-4)g6(A4A32096.2)/2-385 802.02 EMVISOS 10 LPDIVFTINOVOYPLSPSAY	II Y I L K 455 SA X I L Q 396 gvizereorijeje Arazine 1/1-385 306
2894/emipCv69602 1/2 309 491 mm VPB VB VB VB C L 06 FA C MA 508 2894/emipCv69602 1/2 309 491 mm VPB VB	p)2890.4 em.b)CA439602.2 2-30.8 456 V <mark>BEG</mark> AAAC <mark>CISGFTAMDIPPPRGPLWILGDVF</mark> p)264604 g6)44432096.2 2-385 887 DDDSC <mark>TEGFEGMDVP</mark> TSS <mark>GELWILGD</mark> VF	SVFMGF 490 pil/second/pil/second
requence 2 (D. g)[64664/gb]AAA31096.1] Residue: ASP (91) Sequence 2 (D. g)[64664/gb]AAA31096.1]	») 1390 4) om b) (AAA39600 2.2) (2.50 8 491. ¥HTVFDYCKLEIGFAK AA 9) 126460 4) gó (AAA31096 1) (2.383 388 ¥¥TVFD RANNEVGLAPVA	508 pt/2604/emp(24432006.1/2.302 462.ADC 1.5CF TAMD (PPR 67.UW) LGDVFMGP (WTVF 455 9/264604/g6/44431096.1/2.385 341 - CT8CF 60MDVF T5 58 6 LWI LGDVF I R0 WTVF 972
equence 2 ID: gi[164604]gb]AAA31096.1] Residue: ASP [9]] Sequence 2 ID: gi[164604]gb]AAA31096.1]		x
	Sequence 2 ID: gi[164604[gb]AAA31096.1] Residue: ASP (91)	Sequence 2 ID: gi 164604 gb AAA31096.1

. . . .

æ

<ロト <部ト < 注ト < 注ト

Mark Voorhies Pairwise Alignment

List all differences between two text files # (empty output for identity) diff HvSs.gap0.0.both.aln HvSs.gap0.0.mult.aln # Report only whether the files differ # (empty output for identity) diff -q HvSs.gap0.0.both.aln HvSs.gap0.0.mult.aln (*NIX = *BSD, OS X, Solaris, Linux, Windows with Cygwin, ...)

- Use a text or sequence editor to create a spliced variant of HvPhytepsin that can be aligned to the full HsSaposinC sequence
- Find the GenBank entries for HvPhytepsin and SsPepsinogen (tip: use the identifiers from the FASTA files) and find the corresponding transcript sequences.
 - How easy is it to align the proteins vs. the transcripts?
 - Can you tell if you are getting equivalent results from the two alignments?
- Try repeating this exercise for a pair of sequences where genomic sequence is available; *e.g.*, *A. nidulans* VosA (ABQ18268.1), and *H. capsulatum* Ryp2 (ACB59236.1).