BLAST (Basic Local Alignment Search Tool)

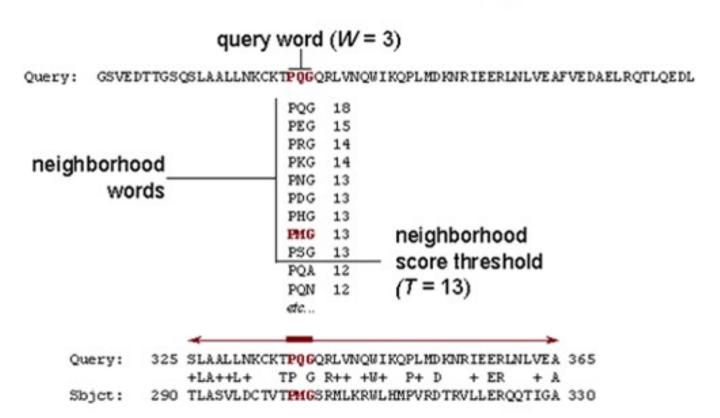
BLAST is the primary method to find sequences in modern sequence data bases



Stephen Altschul NCBI, NLM, NIH Bioinformatics Verified email at nih.gov

Title 1–20	Cited by	Year
Basic local alignment search tool SF Altschul, W Gish, W Miller, EW Myers, DJ Lipman Journal of molecular biology 215 (3), 403-410	54882	1990
Gapped BLAST and PSI-BLAST: a new generation of protein database search programs SF Altschul, TL Madden, AA Schäffer, J Zhang, Z Zhang, W Miller, Nucleic acids research 25 (17), 3389-3402	54781	1997
Identification of FAP locus genes from chromosome 5q21. KW Kinzler, MC Nilbert, LK Su, B Vogelstein, TM Bryan, DB Levy, Science (New York, NY) 253 (5020), 661	2060	1991
Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment CE Lawrence, SE Altschul, MS Boguski, JS Liu, AE Neuwald, JC Wootton	1963	1993

The BLAST Search Algorithm



High-scoring Segment Pair (HSP)

Primary BLAST quality metric: E value

The Expectation value or E value represents the number of different alignments with scores equivalent to or better than the one observed that are expected to occur in a database search by chance.

The lower the E value, the more significant the score and the alignment.

glycoprotein precursor [Junin virus]

Range :	1: 1 to	241 GenPept Graphics		▼ Next Ma	tch 🔺 Pro	evious Match
Score		Expect Method	Identities	Positives	Gaps	
278 bits	s(710)	1e-86 Compositional matrix adjust.	137/252(54%)	168/252(66%)	11/252(4%)
Query	1	MGQLISFFQEIPVFLQEALNIALVAVSLIZ				60
Sbjct	1	MGQFISFMQEIPTFLQEALNIALVAVSLI				60
Query	61	TFKIGLHTEFQSVTFTMQRLLANHSNELPS FKIGLHTEFQ+V+F+M L +N+ ++LP			/SVLMK -+VL+	120
Sbjct	61	AFKIGLHTEFQTVSFSMVGLFSNNPHDLP				120
Query	121	EYDVSVYEPEDLGNCLNKSDSSWAIHWFS: +YDV + P D+ C D W WF	IALGHDWLMDPPM A+GHDW +DPP		QFNIS O N S	180
Sbjct	121	QYDVVIQHPADMSWCSKSDDQIWLSQWFMI			~	180
Query	181	KADESRVYGKKIRNGMRHLFRGFYDPCEEC		PSSFEYCGTNYLS P+S+	SKCQFD +C D	240
Sbjct	181	KTGVNENYAKKFKTGMHHLYREYPDSCLNO		PTSWPL	-QCPLD	229
Query	241	HVNTLHFLVRSK 252 HVNTLHFL R K				
Sbjct	230	HVNTLHFLTRGK 241				

glycoprotein precursor [Junin virus]

Sequence ID: gb|ABI51595.1| Length: 485 Number of Matches: 1

sequence we found (subject sequence)

Range 1: 1 to 241 GenPept Graphics

Score		Expect	Method			Identities	Positives	Gaps	
278 bits	s(710)	1e-86	Composi	tional matri	x adjust.	137/252(54%	%) 168/252(66%)	11/252(4%)
Query	1	_		~			SGLFQFIFFLFLAG SGLFQF FL LAG		60
Sbjct	1	-		~			SGLFQFFVFLALAG		60
Query	61					SLCMLNNSFYY LC LN S Y	MKGGANIFLIRVSD +KGG F+I D	VSVLMK ++VL+	120
Sbjct	61		~				IKGGNASFMISFDD	IAVLLP	120
Query	121				WAIHWFS		PMLCRNKTKKEGSN P LCRN+TK EG	IQFNIS Q N S	180
Sbjct	121	QYDVVI	QHPADMS	WCSKSDDQI	WLSQWFMI	NAVGHDWHLDP	PFLCRNRTKTEGFI	FQVNTS	180
Query	181	KADESR K +		NGMRHLFRG		~	GDPSSFEYCGTNYL P+S+	SKCQFD +C D	240
Sbjct	181	KTGVNE					AQPTSWPL	-QCPLD	229
Query	241		FLVRSK FL R K	252					
Sbjct	230	HVNTLH	FLTRGK	241					

glycoprotein precursor [Junin virus]

Range 1: 1 to 241 GenPept Graphics

Sequence ID: gb|ABI51595.1| Length: 485 Number of Matches: 1

Kange 1	1 60	241 <u>001</u>	перс оп	артноз	2					INCAL MIC	CCII = FI	Svious
Score		Expect	E val	ш			Identiti		Positi		Gaps	
278 bits	(710)	1e-86	chmydi	y C	matrix a	djust.	137/252	(54%)	168/2	52(66%)	11/252(4%)
Query	1									FFLFLAGF FL LAGF		60
Sbjct	1	-	~	_					-	VFLALAGE		60
Query	61		TEFQSVT							FLIRVSDV F+I D+	SVLMK +VL+	120
Sbjct	61		~							FMISFDDI	- —	120
Query	121	EYDVSVY +YDV +			SDSSWA:		[ALGHDW] A+GHDW			rkkegsni rk eg	QFNIS O N S	180
Sbjct	121	QYDVVIQ								rktegfif	~	180
Query	181	KADESRV						~	PSSFE	YCGTNYLS	KCQFD +C D	240
Sbjct	181	KTGVNE								L		229
Query	241	HVNTLHI HVNTLHI		252								
Sbjct	230	HVNTLHI		241								

▼ Next Match A Previous Match

glycoprotein precursor [Junin virus]

Range 1: 1 to 241 GenPept Graphics

Sequence ID: gb|ABI51595.1| Length: 485 Number of Matches: 1

itulige .		242 Genrepe Grapmes	TTEXE Place II
Score		Expect Method Identities Positiv	es Gaps
278 bits	s(710)) 1e-86 Compositional matrix adjust. 137/252(54%) 168/25	2(66%) 11/252(4%)
Query	1	MGQLISFFQEIPVFLQEALNIALVAVSLIAV number and MGQ ISF QEIP FLQEALNIALVAVSLIA+	% of exact
Sbjct	1	MGQFISFMQEIPTFLQEALNIALVAVSLIAI matches, ne	ar matches,
Query	61	TFKIGLHTEFQSVTFTMQRLLANHSNELPSI AND NO MATC	irvsdvsvlmk 120 hesd++vl+
Sbjct	61	AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLLCTLNKSHLYIKGGNASF.	
Query	121	EYDVSVYEPEDLGNCLNKSDSSWAIHWFSIALGHDWLMDPPMLCRNKT +YDV + P D+ C D W WF A+GHDW +DPP LCRN+T	
Sbjct	121	QYDVVIQHPADMSWCSKSDDQIWLSQWFMNAVGHDWHLDPPFLCRNRT	~
Query	181	KADESRVYGKKIRNGMRHLFRGFYDPCEEGKVCYVTINQCGDPSSFEY K + Y KK + GM HL+R + D C GK+C + P+S+	CGTNYLSKCQFD 240 +C D
Sbjct	181	KTGVNENYAKKFKTGMHHLYREYPDSCLNGKLCLMKAQPTSWPL	QCPLD 229
Query	241	HVNTLHFLVRSK 252 HVNTLHFL R K	
Sbjct	230	HVNTLHFLTRGK 241	

▼ Next Match A Previous Match

glycoprotein precursor [Junin virus]

Range 1: 1 to	241 GenPept Graphics	▼ Next Match ▲ Previous Match
Score	Expect Method	Identities Positives Gaps
278 bits(710	1e-86 Compositional matrix adjust.	137/252(54%) 168/252(66%) 11/252(4%)
Query 1	MGQLISFFQEIPVFLQEALNIALVAVSLI	number and % of exact
Sbjct 1	MGQFISFMQEIPTFLQEALNIALVAVSLI	
Query 61	TFKIGLHTEFQSVTFTMQRLLANHSNELPS FKIGLHTEFQ+V+F+M L +N+ ++LP	i and no matches
Sbjct 61		LLCTLNKSHLYIKGGNASFMISFDDIAVLLP 120
Query 121	EYDVSVYEPEDLGNCLNKSDSSWAIHWFS: +YDV + P D+ C D W WF	IALGHDWLMDPPMLCRNKTKKEGSNIQFNIS 180 A+GHDW +DPP LCRN+TK EG ON S
Sbjct 121		NAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS 180
Query 181	KADESRVYGKKIRNGMRHLFRGFYDPCEE	GKVCYVTINQCGDPSSFEYCGTNYLSKCQFD 240 GK+C + P+S+ +C D
Sbjct 181	KTGVNENYAKKFKTGMANLYREYPDSCLN	GKLCLMKAQPTSWPLQCPLD 229
Query 241	HVNTLHFLVRSK 252 exact	match
Sbjct 230	HVNTLHFLTRGK 241	

glycoprotein precursor [Junin virus]

Range 1: 1 t	to 241 GenPept Graphics	▼ Next Match ▲ Previous Match
Score	Expect Method Id	lentities Positives Gaps
278 bits(710	0) 1e-86 Compositional matrix adjust. 13	37/252(54%) 168/252(66%) 11/252(4%)
Query 1	MGQLISFFQEIPVFLQEALNIALVAVSLIAV	number and % of exact
Sbjct 1	MGQFISFMQEIPTFLQEALNIALVAVSLIAII	matches, near matches,
Query 61	TFKIGLHTEFQSVTFTMQRLLANHSNELPSIC FKIGLHTEFQ+V+F+M L +N+ ++LP IC	and no matches
Sbjct 61	AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLLC	
Query 121		LGHDWLMDPPMLCRNKTKKEGSNIQFNIS 180 -GHDW +DPP LCRN+TK EG ON S
Sbjct 121		~
Query 181	KADESRVYGKKIRNGMRHLFRGFYDPCEEGKV K + Y KK + GM HL+R + D C GK+	
Sbjct 181	KTGVNENYAKKFKTGMHHLYREYPDSCLNGKL	CCLMKAQPTSWPLQCPLD 229
Query 241	HVNTLHFLVRSK 252 near m	atch (positive)
Sbjct 230	HVNTLHFLTRGK 241	

glycoprotein precursor [Junin virus]

Range 1: 1 to	241 GenPept Graphics	▼ Next Match ▲ Previous Match
Score	Expect Method	Identities Positives Gaps
278 bits(710)	1e-86 Compositional matrix adjust.	137/252(54%) 168/252(66%) 11/252(4%)
Query 1	MGQLISFFQEIPVFLQEALNIALVAVSLIZ	number and % of exact
Sbjct 1	MGQFISFMQEIPTFLQEALNIALVAVSLI	
Query 61	TFKIGLHTEFQSVTFTMQRLLANHSNELPS FKIGLHTEFQ+V+F+M L +N+ ++LP	i and no matches
Sbjct 61		LLCTLNKSHLYIKGGNASFMISFDDIAVLLP 120
Query 121	EYDVSVYEPEDLGNCLNKSDSSWAIHWFS: +YDV + P D+ C D W WF	IALGHDWLMDPPMLCRNKTKKEGSNIQFNIS 180 A+GHDW +DPP LCRN+TK EG O N S
Sbjct 121		NAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS 180
Query 181		GKVCYVTINQCGDPSSFEYCGTNYLSKCQFD 240 GK+C + P+S+ +C D
Sbjct 181	KTGVNENYAKKFKTGMHHLYRTYPDSCLN	GKLCLMKAQPTSWPLQCPLD 229
Query 241	HVNTLHFLVRSK 252 no m	atch (gap)
Sbjct 230	HVNTLHFLTRGK 241	

glycoprotein precursor [Junin virus]

Range 1: 1 to 241 GenPept Graphics

Sequence ID: gb|ABI51595.1| Length: 485 Number of Matches: 1

Rain	ge I. I to	241 Genrept Graphics
Sco	re	Expect Method Identities Positives Gaps
278	bits(710)	1e-86 Compositional matrix adjust. 137/252(54%) 168/252(66%) 11/252(4%)
Quer	ry 1	MGQLISFFQEIPVFLQEALNIALVAVSLIAVIKGIINLYKSGLFQFIFFLFLAGRSCSDG 60 MGQ ISF QEIP FLQEALNIALVAVSLIA+IKGI+NLYKSGLFQF FL LAGRSC++
Sbjo	t 1	MGQFISFMQEIPTFLQEALNIALVAVSLIAIIKGIVNLYKSGLFQFFVFLALAGRSCTEE 60
Quer	y 61	TFKIGLHTEFQSVTFTMQRLLANHSNELPSLCMLNNSFYYMKGGANIFLIRVSDVSVLMK 120 FKIGLHTEFQ+V+F+M L +N+ ++LP LC LN S Y+KGG F+I D++VL+
Sbjo	t 61	AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLLCTLNKSHLYIKGGNASFMISFDDIAVLLP 120
Quer	y 121	EYDVSVYEPEDLGNCLNKSDSSWAIHWFSIALGHDWLMDPPMLCRNKTKKEGSNIQFNIS 180 +YDV + P D+ C D W WF A+GHDW +DPP LCRN+TK EG O N S
Sbjo	t 121	QYDVVIQHPADMSWCSKSDDQIWLSQWFMNAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS 180
Quer	y 181	KADESRVYGKKIRNGMRHLFRGFYDPCEEGKVCYVTINQCGDPSSFEYCGTNYLSKCQFD 240 K + Y KK + GM HL+R + D C GK+C + P+S+ +C D
Sbjo	t 181	KTGVNENYAKKFKTGMHHLYREYPDSCLNGKLCLMKAQPTSWPLQCPLD 229
Quer	y 241	HVNTLHFLVRSK 252 HVNTLHFL R K
Sbjo	et 230	HVNTLHFLTRGK 241

▼ Next Match A Previous Match