BLAST
(Basic Local Alignment Search Tool)

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

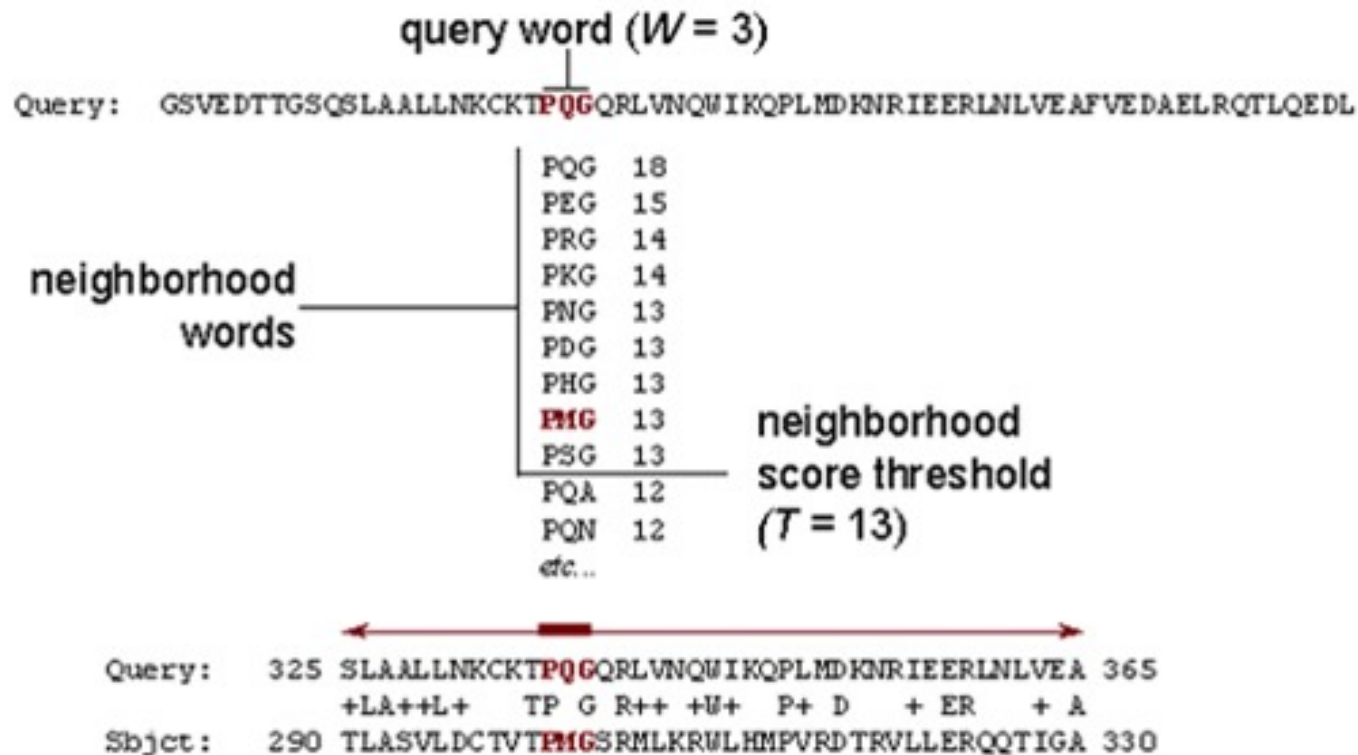## Stephen Altschul
NCBI, NLM, NIH
Bioinformatics
Verified email at nih.gov

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, SF Altschul, MS Boguski, JS Liu, AF Neuwald, JC Wootton	1963	1993

The BLAST Search Algorithm

query word (W = 3)

Query: GSVEDTTGSQSLAALLNKCKTPQGQRLVNQWIKQPLMDKNRIEERLNLVEAFVEDAELRQTLQEDL

PQG	18
PEG	15
PRG	14
PKG	14
PNG	13
PDG	13
PHG	13
PMG	13
PSG	13
PQA	12
PQN	12

neighborhood words

neighborhood score threshold (T = 13)

etc...

Query: 325 SLAALLNKCKTPQGQRLVNQWIKQPLMDKNRIEERLNLVEA 365
+LA++L+ TP G R++ +W+ P+ D + ER + A
Sbjct: 290 TLASVLDCTVTPMGSRMLKRWLHMPVRDTRVLLERQQTIGA 330

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.

Anatomy of a BLAST result

glycoprotein precursor [Junin virus]
Sequence ID: gb|ABI51595.1| Length: 485   Number of Matches: 1

Range 1: 1 to 241 GenPept   Graphics

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    MGQLISFFQEIPVFLQEALNIALVAVSLIAVIKGIINLYKSGLFQFIFFLFLAGRSCSDG   60
             MGQ ISF QEIP FLQEALNIALVAVSLIA+IKGI+NLYKSGLFQF  FL LAGRSC++
Sbjct   1    MGQFISFMQEIPTFLQEALNIALVAVSLIAIIKGIVNLYKSGLFQFFVFLALAGRSCTEE   60

Query   61   TFKIGLHTEFQSVTFTMQRLLANHSNELPSLCMLNNSFYYMKGGANIFLIRVSDVSVLMK   120
              FKIGLHTEFQ+V+F+M   L +N+ ++LP LC LN S  Y+KGG    F+I   D++VL+
Sbjct   61   AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLLCTLNKSHLYIKGGNASFMISFDDIAVLLP   120

Query   121  EYDVSVYEPEDLGNCLNKSDSSWAIHWFSIALGHDWLMDPPMLCRNKTKKEGSNIQFNIS   180
             +YDV +   P D+  C    D  W    WF  A+GHDW +DPP LCRN+TK  EG   Q N S
Sbjct   121  QYDVVIQHPADMSWCSKSDDQIWLSQWFMNAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS   180

Query   181  KADESRVYGKKIRNGMRHLFRGFYDPCEEGKVCYVTINQCGDPSSFEYCGTNYLSKCQFD   240
             K    +   Y KK + GM HL+R + D C  GK+C +         P+S+          +C  D
Sbjct   181  KTGVNENYAKKFKTGMHHLYREYPDSCLNGKLCLMK----AQPTSWPL-------QCPLD   229

Query   241  HVNTLHFLVRSK   252
             HVNTLHFL R K
Sbjct   230  HVNTLHFLTRGK   241
```

Anatomy of a BLAST result

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(710)	1e-86	Compositional matrix adjust.	137/252(54%)	168/252(66%)	11/252(4%)

```
Query  1    MGQLISFFQEIPVFLQEALNIALVAVSLIAVIKGIINLYKSGLFQFIFFLFLAGRSCSDG  60
            MGQ ISF QEIP FLQEALNIALVAVSLIA+IKGI+NLYKSGLFQF  FL LAGRSC++
Sbjct  1    MGQFISFMQEIPTFLQEALNIALVAVSLIAIIKGIVNLYKSGLFQFFVFLALAGRSCTEE  60

Query  61   TFKIGLHTEFQSVTFTMQRLLANHSNELPSLCMLNNSFYYMKGGANIFLIRVSDVSVLMK  120
             FKIGLHTEFQ+V+F+M  L +N+ ++LP LC LN S  Y+KGG    F+I    D++VL+
Sbjct  61   AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLLCTLNKSHLYIKGGNASFMISFDDIAVLLP  120

Query  121  EYDVSVYEPEDLGNCLNKSDSSWAIHWFSIALGHDWLMDPPMLCRNKTKKEGSNIQFNIS  180
            +YDV +   P D+  C    D  W    WF  A+GHDW +DPP LCRN+TK  EG    Q N S
Sbjct  121  QYDVVIQHPADMSWCSKSDDQIWLSQWFMNAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS  180

Query  181  KADESRVYGKKIRNGMRHLFRGFYDPCEEGKVCYVTINQCGDPSSFEYCGTNYLSKCQFD  240
            K    +   Y KK + GM HL+R + D C   GK+C +            P+S+          +C  D
Sbjct  181  KTGVNENYAKKFKTGMHHLYREYPDSCLNGKLCLMK----AQPTSWPL-------QCPLD  229

Query  241  HVNTLHFLVRSK    252
            HVNTLHFL R K
Sbjct  230  HVNTLHFLTRGK    241
```

Anatomy of a BLAST result

glycoprotein precursor [Junin virus]

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

Range 1: 1 to 241 GenPept   Graphics   ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
278 bits(710)	1e-86 **E value** Compositional matrix adjust.	137/252(54%)	168/252(66%)	11/252(4%)

```
Query   1    MGQLISFFQEIPVFLQEALNIALVAVSLIAVIKGIINLYKSGLFQFIFFLFLAGRSCSDG   60
             MGQ ISF QEIP FLQEALNIALVAVSLIA+IKGI+NLYKSGLFQF  FL LAGRSC++
Sbjct   1    MGQFISFMQEIPTFLQEALNIALVAVSLIAIIKGIVNLYKSGLFQFFVFLALAGRSCTEE   60

Query   61   TFKIGLHTEFQSVTFTMQRLLANHSNELPSLCMLNNSFYYMKGGANIFLIRVSDVSVLMK   120
              FKIGLHTEFQ+V+F+M  L +N+ ++LP LC LN S  Y+KGG     F+I   D++VL+
Sbjct   61   AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLLCTLNKSHLYIKGGNASFMISFDDIAVLLP   120

Query   121  EYDVSVYEPEDLGNCLNKSDSSWAIHWFSIALGHDWLMDPPMLCRNKTKKEGSNIQFNIS   180
             +YDV +   P D+  C     D  W    WF   A+GHDW +DPP LCRN+TK  EG   Q N S
Sbjct   121  QYDVVIQHPADMSWCSKSDDQIWLSQWFMNAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS   180

Query   181  KADESRVYGKKIRNGMRHLFRGFYDPCEEGKVCYVTINQCGDPSSFEYCGTNYLSKCQFD   240
             K    +   Y KK + GM HL+R + D C  GK+C +        P+S+          +C  D
Sbjct   181  KTGVNENYAKKFKTGMHHLYREYPDSCLNGKLCLMK----AQPTSWPL-------QCPLD   229

Query   241  HVNTLHFLVRSK   252
             HVNTLHFL R K
Sbjct   230  HVNTLHFLTRGK   241
```

Anatomy of a BLAST result

glycoprotein precursor [Junin virus]
Sequence ID: gb|ABI51595.1| Length: 485   Number of Matches: 1

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    MGQLISFFQEIPVFLQEALNIALVAVSLIAV...                      60
             MGQ ISF QEIP FLQEALNIALVAVSLIA+  IKGI+NLYKSGLFQF  FL LAGRSC++
Sbjct   1    MGQFISFMQEIPTFLQEALNIALVAVSLIAI  IKGIVNLKSGLFQEFVELALAGRSCTW  60

Query   61   TFKIGLHTEFQSVTFTMQRLLANHSNELPSI CMLNNGFYYMKGGANIFLIRVSDVSVLMK 120
              FKIGLHTEFQ+V+F+M  L +N+ ++LP IG     K G     D++VL+
Sbjct   61   AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLL CTLNKSHLYIKGGNASFMISFDDIAVLLP 120

Query   121  EYDVSVYEPEDLGNCLNKSDSSWAIHWFSIALGHDWLMDPPMLCRNKTKKEGSNIQFNIS 180
             +YDV +  P D+  C    D  W    WF  A+GHDW +DPP LCRN+TK EG    Q N S
Sbjct   121  QYDVVIQHPADMSWCSKSDDQIWLSQWFMNAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS 180

Query   181  KADESRVYGKKIRNGMRHLFRGFYDPCEEGKVCYVTINQCGDPSSFEYCGTNYLSKCQFD 240
             K   +  Y KK + GM HL+R + D C  GK+C +          P+S+       +C  D
Sbjct   181  KTGVNENYAKKFKTGMHHLYREYPDSCLNGKLCLMK----AQPTSWPL-------QCPLD 229

Query   241  HVNTLHFLVRSK   252
             HVNTLHFL R K
Sbjct   230  HVNTLHFLTRGK   241
```

number and % of exact matches, near matches, and no matches

Anatomy of a BLAST result

glycoprotein precursor [Junin virus]

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

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    MGQLISFFQEIPVFLQEALNIALVAVSLIAV...
             MGQ ISF QEIP FLQEALNIALVAVSLIA+
Sbjct   1    MGQFISFMQEIPTFLQEALNIALVAVSLIAI...

Query   61   TFKIGLHTEFQSVTFTMQRLLANHSNELPSI...
              FKIGLHTEFQ+V+F+M  L +N+ ++LP L
Sbjct   61   AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLL...

Query   121  EYDVSVYEPEDLGNCLNKSDSSWAIHWFSIALGHDWLMDPPMLCRNKTKKEGSNIQFNIS  180
             +YDV +  P D+  C    D  W   WF  A+GHDW +DPP LCRN+TK EG   Q N S
Sbjct   121  QYDVVIQHPADMSWCSKSDDQIWLSQWFMNAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS  180

Query   181  KADESRVYGKKIRNGMRHLFRGFYDPCEEGKVCYVTINQCGDPSSFEYCGTNYLSKCQFD  240
             K   +  Y KK + GM HL+R + D C  GK+C +          P+S+          +C  D
Sbjct   181  KTGVNENYAKKFKTGMHHLYREYPDSCLNGKLCLMK----AQPTSWPL-------QCPLD  229

Query   241  HVNTLHFLVRSK    252
             HVNTLHFL R K
Sbjct   230  HVNTLHFLTRGK    241
```

number and % of exact matches, near matches, and no matches

exact match

Anatomy of a BLAST result

glycoprotein precursor [Junin virus]
Sequence ID: gb|ABI51595.1| Length: 485 Number of Matches: 1

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    MGQLISFFQEIPVFLQEALNIALVAVSLIAV...                          60
             MGQ ISF QEIP FLQEALNIALVAVSLIA+ IKGI+NLYKSGLFQF   FL LAGRSC++
Sbjct   1    MGQFISFMQEIPTFLQEALNIALVAVSLIAI...IKGIVNLLKSGLFQFFVELALAGRSCTLE 60

Query   61   TFKIGLHTEFQSVTFTMQRLLANHSNELPSICMLNNGFYYMKGGANIFLIRVSDVSVLMK 120
              FKIGLHTEFQ+V+F+M  L +N+ ++LP L                     D++VL+
Sbjct   61   AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLLCTLNKSHLYIKGGNASFMISFDDIAVLLP 120

Query   121  EYDVSVYEPEDLGNCLNKSDSSWAIHWFSIALGHDWLMDPPMLCRNKTKKEGSNIQFNIS 180
             +YDV +  P D+  C    D  W    WF  A+GHDW +DPP LCRN+TK EG    Q N S
Sbjct   121  QYDVVIQHPADMSWCSKSDDQIWLSQWFMNAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS 180

Query   181  KADESRVYGKKIRNGMRHLFRGFYDPCEEGKVCYVTINQCGDPSSFEYCGTNYLSKCQFD 240
             K   +  Y KK + GM HL+R + D C  GK+C +              P+S+      +C  D
Sbjct   181  KTGVNENYAKKFKTGMHHLYREYPDSCLNGKLCLMK----AQPTSWPL-------QCPLD 229

Query   241  HVNTLHFLVRSK  252
             HVNTLHFL R K
Sbjct   230  HVNTLHFLTRGK  241
```

number and % of exact matches, near matches, and no matches

near match (positive)

Anatomy of a BLAST result

glycoprotein precursor [Junin virus]

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

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    MGQLISFFQEIPVFLQEALNIALVAVSLIAV                              60
            MGQ ISF QEIP FLQEALNIALVAVSLIA+
Sbjct  1    MGQFISFMQEIPTFLQEALNIALVAVSLIAI                              60

Query  61   TFKIGLHTEFQSVTFTMQRLLANHSNELPSI                              120
             FKIGLHTEFQ+V+F+M  L +N+ ++LP I
Sbjct  61   AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLL                              120

Query  121  EYDVSVYEPEDLGNCLNKSDSSWAIHWFSIALGHDWLMDPPMLCRNKTKKEGSNIQFNIS 180
            +YDV +  P D+  C    D  W   WF  A+GHDW +DPP LCRN+TK  EG    Q N S
Sbjct  121  QYDVVIQHPADMSWCSKSDDQIWLSQWFMNAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS  180

Query  181  KADESRVYGKKIRNGMRHLFRGFYDPCEEGKVCYVTINQCGDPSSFEYCGTNYLSKCQFD  240
            K    +  Y KK + GM HL+R + D C  GK+C +            P+S+       +C  D
Sbjct  181  KTGVNENYAKKFKTGMHHLYRTYPDSCLNGKLCLMK----AQPTSWPL-------QCPLD  229

Query  241  HVNTLHFLVRSK   252
            HVNTLHFL R K
Sbjct  230  HVNTLHFLTRGK   241
```

number and % of exact matches, near matches, and no matches

no match (gap)

Anatomy of a BLAST result

glycoprotein precursor [Junin virus]

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

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    MGQLISFFQEIPVFLQEALNIALVAVSLIAVIKGIINLYKSGLFQFIFFLFLAGRSCSDG    60
              MGQ ISF QEIP FLQEALNIALVAVSLIA+IKGI+NLYKSGLFQF  FL LAGRSC++
Sbjct    1    MGQFISFMQEIPTFLQEALNIALVAVSLIAIIKGIVNLYKSGLFQFFVFLALAGRSCTEE    60

Query   61    TFKIGLHTEFQSVTFTMQRLLANHSNELPSLCMLNNSFYYMKGGANIFLIRVSDVSVLMK   120
               FKIGLHTEFQ+V+F+M   L +N+ ++LP LC LN S  Y+KGG    F+I   D++VL+
Sbjct   61    AFKIGLHTEFQTVSFSMVGLFSNNPHDLPLLCTLNKSHLYIKGGNASFMISFDDIAVLLP   120

Query  121    EYDVSVYEPEDLGNCLNKSDSSWAIHWFSIALGHDWLMDPPMLCRNKTKKEGSNIQFNIS   180
              +YDV +   P D+  C     D W    WF  A+GHDW +DPP LCRN+TK  EG   Q N S
Sbjct  121    QYDVVIQHPADMSWCSKSDDQIWLSQWFMNAVGHDWHLDPPFLCRNRTKTEGFIFQVNTS   180

Query  181    KADESRVYGKKIRNGMRHLFRGFYDPCEEGKVCYVTINQCGDPSSFEYCGTNYLSKCQFD   240
              K    +  Y KK + GM HL+R + D C  GK+C +         P+S+          +C  D
Sbjct  181    KTGVNENYAKKFKTGMHHLYREYPDSCLNGKLCLMK----AQPTSWPL-------QCPLD   229

Query  241    HVNTLHFLVRSK    252
              HVNTLHFL R K
Sbjct  230    HVNTLHFLTRGK    241
```