

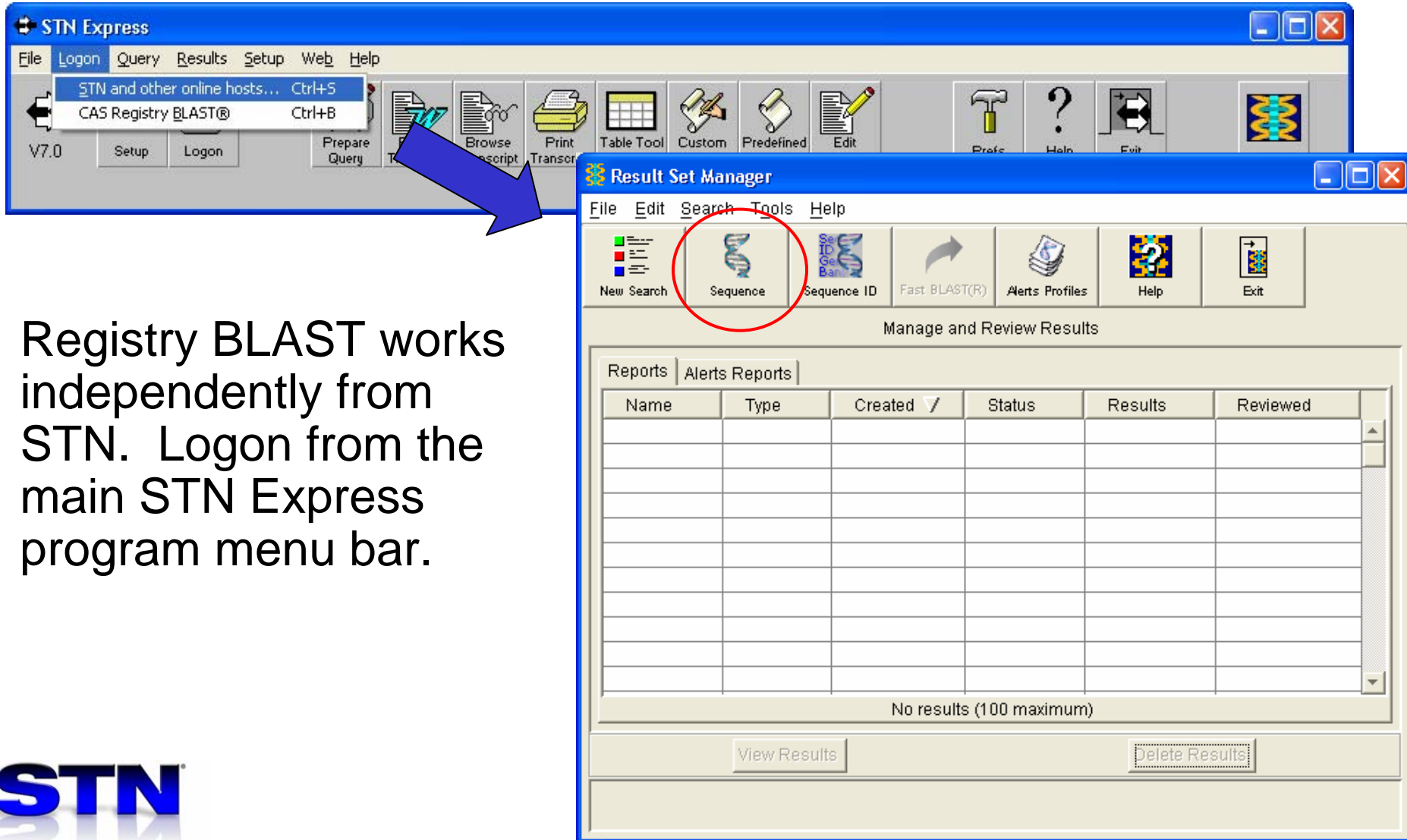
# STN<sup>®</sup>

Olennaiset kohdat  
biosekvenssihauissa

*Registry, DGENE, PCTGEN, USGENE*

(osia Robert Austinin esityksestä)

# CAS Registry BLAST





The image shows two overlapping windows from the STN Express software. The top window is the main application, with a menu bar including 'File', 'Logon', 'Query', 'Results', 'Setup', 'Web', and 'Help'. The 'Query' menu is open, showing options like 'STN and other online hosts...' and 'CAS Registry BLAST@'. A blue arrow points from the 'CAS Registry BLAST@' option to the 'Result Set Manager' window below. The 'Result Set Manager' window has a menu bar with 'File', 'Edit', 'Search', 'Tools', and 'Help'. The 'Search' menu is open, and the 'Sequence' option is circled in red. Below the menu bar are icons for 'New Search', 'Sequence', 'Sequence ID', 'Fast-BLAST(R)', 'Alerts Profiles', 'Help', and 'Exit'. The main area of the window is titled 'Manage and Review Results' and contains a table with columns for 'Name', 'Type', 'Created', 'Status', 'Results', and 'Reviewed'. The table is currently empty, and a message at the bottom states 'No results (100 maximum)'. At the bottom of the window are buttons for 'View Results' and 'Delete Results'.

Registry BLAST works independently from STN. Logon from the main STN Express program menu bar.

**Similar Sequences** [minimize] [maximize] [close]

File Edit Help

Result Name (optional):  

Sequence to be searched:   

```
mkfylvrhvk tqwnlegrfq gangdspile eaieeleelg qylssihfda
vyssdlgrar dtvnilndan scpkeihytp qlrewalgtl egckiatmqa
iyprqmtafy qnplqfkhdn fgaeslyqtt hrvesflrsl asknydkvli
vghganltas irsllgyqyg slhykdkldn asltiethd fkdfncltwn
dksylrgevkh mth
```

Example #1: a nucleotide sequence

```
tcccctctcc ctatgcaggt tccctgggtgc aggcagcgct gactctgcca
gactcattat cagcaaactt gggagaaact gtcaggataa cctgggtctgg
aggtagctat gattatggtt ggcaccagca gaaatcacct ggcagtgccc
ctgtcactgt gatctatagc aacaacaaga gaccctcgga catccctca
```

Example #2: a protein sequence

```
TVDQHLGSHLVEALYSVWVHEAKGLPRAAAGAPGVRAELWLDGALLARTAPRAGPG
QLFWAERFHFEALPPARRLSRLRLGPGSAVLGRVALALEELDAPRAPAAGLERWF
PLLGAPAGAALRARIRARRLRVLPSEYKELAEFLTFHYARLCGALEPALPA
```

Protein Query

**BLASTp**

Compare a protein sequence against a protein database.



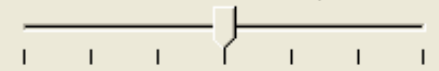
tBLA

**BLASTp Settings - AAM99639**

BLASTp Settings - AAM99639

Additional Option Presets

Search Sensitivity



Fewer Answers → More Answers

 Show Additional Options

Basic Options

 Low Complexity Filtering

Query G

Max No

OK

Back

**BLASTp Database Subsets**

Select Protein Subsets:

 All Referenced Sequences Referenced in at least one Patent document Referenced in a non-patent document Only retrieve sequences with 50 residues or less

OK

Back

Cancel

**Result Set Manager** [Window Controls]

File Edit Search Tools Help

New Search Sequence Sequence ID Fast BLAST(R) Alerts Profiles Help Exit

Manage and Review Results

Reports Alerts Reports

Name	Type	Created ▾	Status	Results	Reviewed
<b>AAM99639</b>	<b>BLASTp</b>	<b>2004-03-25 06:12 PM</b>	<b>Complete</b>	<b>221</b>	

1 result (100 maximum)

View Results Delete Results



CAS Registry BLAST(R) Report - AAM00620

File Edit View Search Tools

Unique Sequences: 221

Alignment Scores

<40 40-50 =200

Alignment Summary

1 54 213

Alignment Details

<input type="checkbox"/>	<input checked="" type="checkbox"/>	439	e-122	<a href="#">(479125-93-4) Protein (Streptococcus agalactiae strain ATCC12403 clone FR2824074-SEQID-2011)</a>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	439	e-122	<a href="#">(440171-08-4) Protein (Streptococcus agalactiae strain 2603V/R clone WO0234771-SEQID-8556)</a>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	439	e-122	<a href="#">(422580-99-2) Protein (Streptococcus agalactiae strain 2603V/R clone WO0234771-SEQID-1064)</a>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	248	3e-65	<a href="#">(422581-00-8) Protein (Streptococcus agalactiae strain 2603V/R clone WO0234771-SEQID-1066)</a>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	235	3e-61	<a href="#">(437961-34-7) Essential protein (Streptococcus pneumoniae clone WO0170955-SEQID-13306)</a>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	235	4e-61	<a href="#">(660059-83-6) Protein (Streptococcus pneumoniae strain 14453 clone US6699703-SEQID-5205 open reading frame-encoded)</a>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	234	6e-61	<a href="#">(463393-17-1) Protein (Streptococcus pneumoniae strain JNR.7/87 clone WO02077021-SEQID-1944)</a>

Get STN Data Cancel



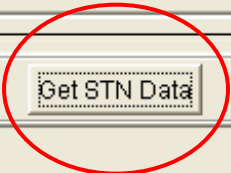
# CAS Registry BLAST(R) Report

File Edit View Search Tools

Select All Ctrl+A

Unselect All

Invert Selection



# CAS Registry BLAST

## Alignment Details

439 e-122 [\(479125-93-4\) Protein \(Streptococcus agalactiae strain ATCC12403 clone FR2824074-SEQID-2011\)](#)

Length = 217

Score = 439 Expect = e-122

Identities = 213/213 (100%) Positives = 213/213 (100%)

Query: 1 MKFYLVRRHGKTQWNLEGRFQGAM

MKFYLVRHGKTQWNLEGRFQGAM

Subject: 5 MKFYLVRRHGKTQWNLEGRFQGAM

This is the best REGISTRY answer. Even though the answer is longer than the query, this does not affect the score (100% identity).

Query: 56 LGRARDTVNILDANSPKEIHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFY 110

LGRARDTVNILDANSPKEIHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFY

Subject: 60 LGRARDTVNILDANSPKEIHY

Note that this is a different best answer from those found in both PCTGEN and DGENE

Query: 111 QNPLQFKHDMFGAESLYQTTHRV

QNPLQFKHDMFGAESLYQTTHRVESFLRSLASKNYDKVLIVGHGANLTASIRSL

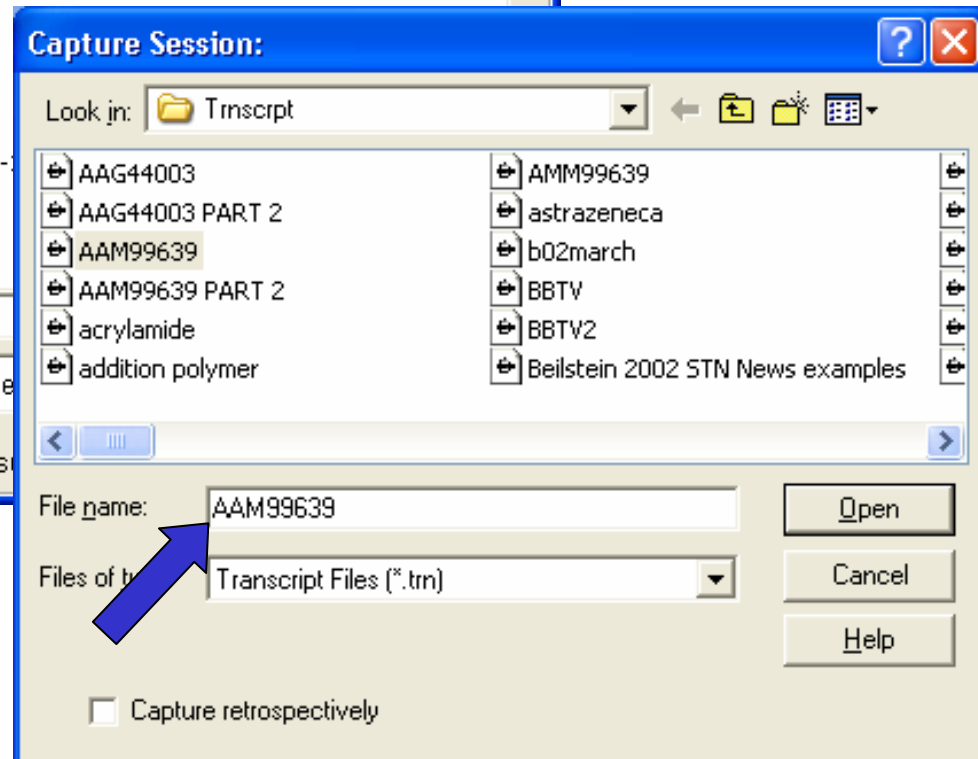
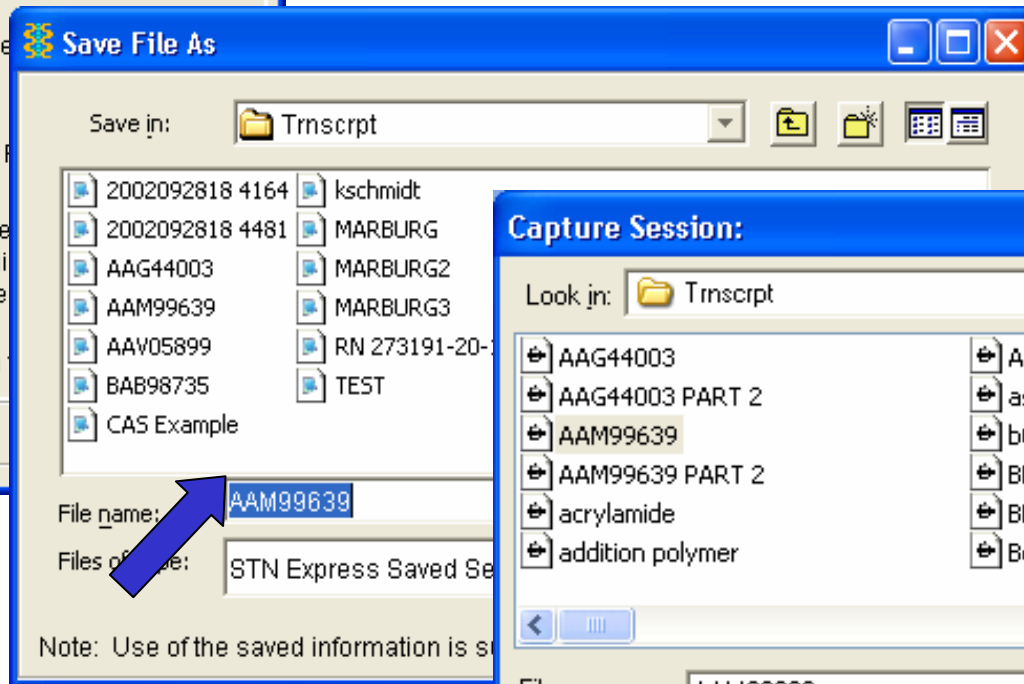
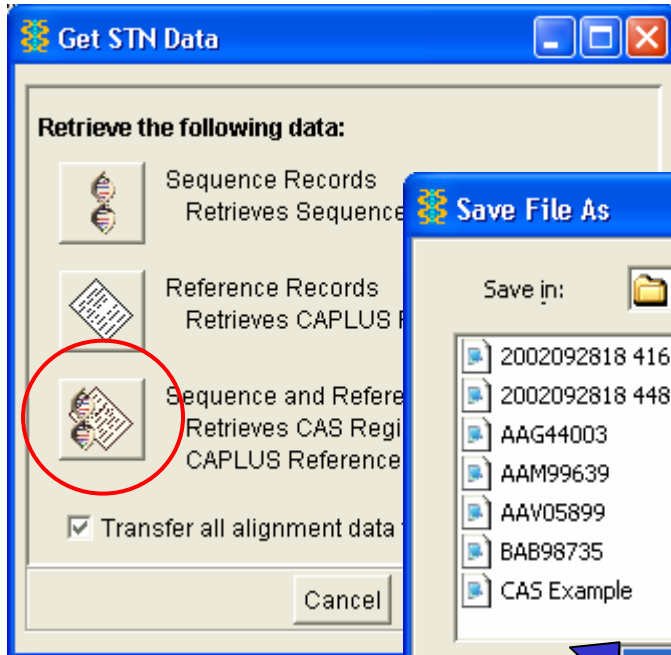
Subject: 115 QNPLQFKHDMFGAESLYQTTHRVESFLRSLASKNYDKVLIVGHGANLTASIRSL 169

Query: 166 GYQYGSLSHYKDKLDNASLTIETHDFKDFNCLTWNKSYLRQEVKMT 213

GYQYGSLSHYKDKLDNASLTIETHDFKDFNCLTWNKSYLRQEVKMT

Subject: 170 GYQYGSLSHYKDKLDNASLTIETHDFKDFNCLTWNKSYLRQEVKMT 217

# CAS Registry BLAST





# CAS Registry BLAST

=> FILE REGISTRY

The STN Express script runs automatically to create an L-number in REGISTRY (L25)

```
=> QUE (479125-93-4 OR 440171-08-4 OR 422580-99-2 OR . . .)
L6  QUE (479125-93-4 OR 440171-08-4 OR 422580-99-2 OR . . .)/RN
. . .
```

```
=> QUE (612128-61-7)/RN
L24  QUE (612128-61-7)/RN
```

```
=> S L6 OR L7 OR L8 OR L9 OR L10 OR L11 OR L12 OR L13 OR L14 OR L15 OR
L16 OR L17 OR L18 OR L19 OR L20 OR L21 OR L22 OR L23 OR L24 OR L24
```

```
L25      253 L6 OR L7 OR L8 OR L9 OR L10 OR L11 OR L12 OR L13 OR L14
          OR L15 OR L16 OR L17 OR L18 OR L19 OR L20 OR L21 OR L22
          OR L23 OR L24
```

=> FILE HCAPLUS

Crossover results to HCAPLUS

```
=> S L25
L26      135 L25
```

```
=> S L26 AND P/DT
L27      122 L26 AND P/DT
```

Limit to HCAPLUS patent records with P/DT

## =&gt; D BIB HITRN 1-

L27 ANSWER 1 OF 122 HCAPLUS COPYRIGHT 2004 ACS on STN  
 AN 2004:181844 CAPLUS  
 TI Methods of testing for bronchial asthma or chronic obstructive . . .  
 IN Ohtani, Noriko; Sugita, Yuji; Yamaya, Mutsuo; Kubo, Hiroshi; . . .  
 PA Genox Research, Inc., Japan

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
EP 1394274	A2	20040303	EP 2003-254857	20030804

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,  
 IE, SI, LT, LV, FI, RO, MK, CY, AL, TR, BG, CZ, EE, HU, SK

PRAI JP 2002-229312 A 20020806 . . .  
 IT 663637-03-4 663642-53-3  
 RL: ANT (Analyte); BSU (Biological study, unclassified); DGN (Diagnostic use); PRP (Properties); THU (Therapeutic use); ANST (Analytical study); BIOL (Biological study); USES (Uses)  
 (amino acid sequence; methods of testing for bronchial asthma . . .

REGISTRY BLAST alignments can be merged in with the bibliographic results using a post processing feature. RNs must be included when displaying CAPLUS records for this feature to work.

# Post-process Registry alignments

The screenshot displays the STN Express software interface. The main window shows a menu bar with 'File', 'Logon', 'Query', 'Results', 'Setup', 'Web', and 'Help'. Below the menu bar is a toolbar with icons for 'Table Tool', 'Custom Report', 'Predefined Report', 'Edit Transcript', 'Prefs', 'Help', 'Exit', and 'BLAST'. A dropdown menu is open under 'Results', showing options like 'Browse Transcript...', 'Print Transcript...', 'Edit Transcript...', 'Export Transcript...', 'Accounting...', 'Table Tool...', 'Report Tool...', 'BLAST® Report with...', 'Predefined Reports', 'R-group Analysis', and 'Open...'. A blue arrow points to the 'BLAST® Report with...' option.

The 'Report Tool' dialog box is open, showing a list of 'Field Name' options on the left and a 'Selected Fields' list on the right. The 'Selected Fields' list includes:
 

- Inventor
- Patent Assignee
- Patent Information
- Application Information
- Priority Application Information
- Other Source
- Cross Reference
- Description
- Patent Sequence Locator
- DGENE BLAST® Alignment**
- BLAST® Alignment Data**
- Controlled or Index Terms

 A blue arrow points to the 'BLAST® Alignment Data' field. Below the 'Selected Fields' list are buttons for 'Change Order', 'Format', and 'Courier New, 14 pts'. At the bottom of the dialog box are buttons for 'Save Template', 'Finish', '< Back', 'Next >', and 'Cancel'.

L33 ANSWER 2 OF 49 CAPLUS COPYRIGHT 20  
AN 2004:176539 CAPLUS  
DN 140:176343  
TI Nucleic acid and amino acid sequences relating to Streptococcus . . .  
IN Doucette-stamm, Lynn; Bush, David; Zeng, Qiandong; . . .  
PA Genome Therapeutics Corporation, USA

This is an example of a CAPLUS record with a merged REGISTRY BLAST alignment.

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI	US 6699703	B1	20040302	US 2000-583110	20000526
PRAI	US 1997-51553P	P	19970702		
	US 1998-85131P	P	19980512		
	US 1998-107433	A2	19980630		

IT 660059-83-6 660040-29-9 660049-22-9  
RL: BSU (Biological study, unclassified)  
660059-83-6  
Length = 206      Score = 235      Expect =  
Score = 235      Expect = 4e-61  
Identities = 120/204 (58%)      Positives = 141/204 (69%)

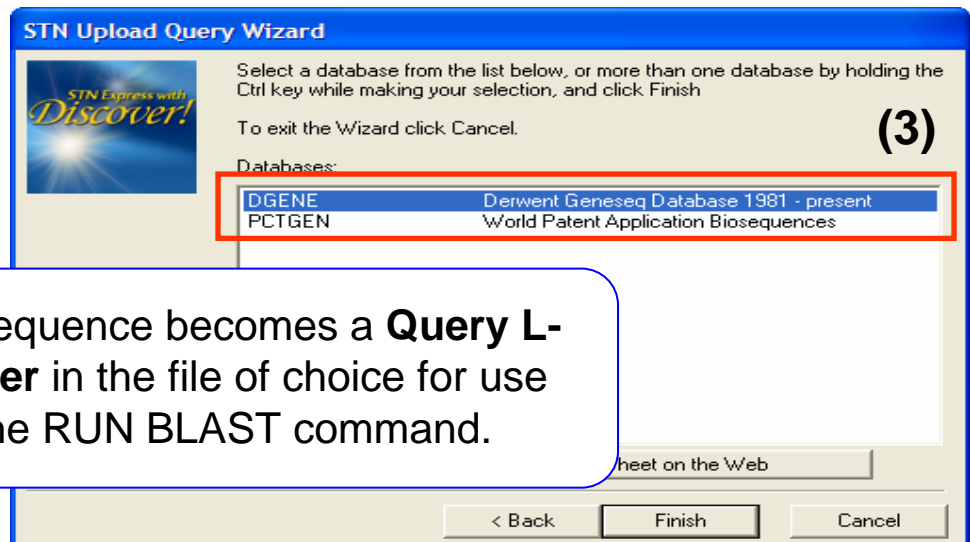
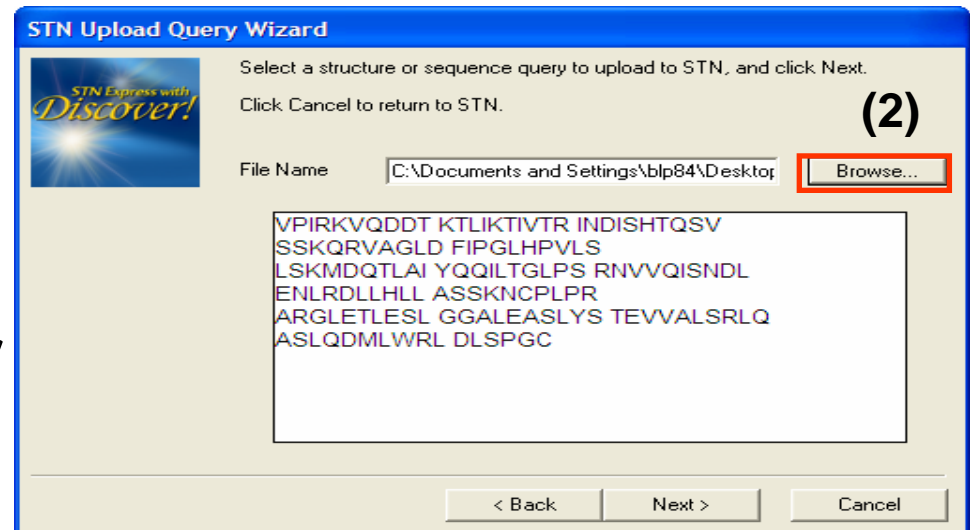
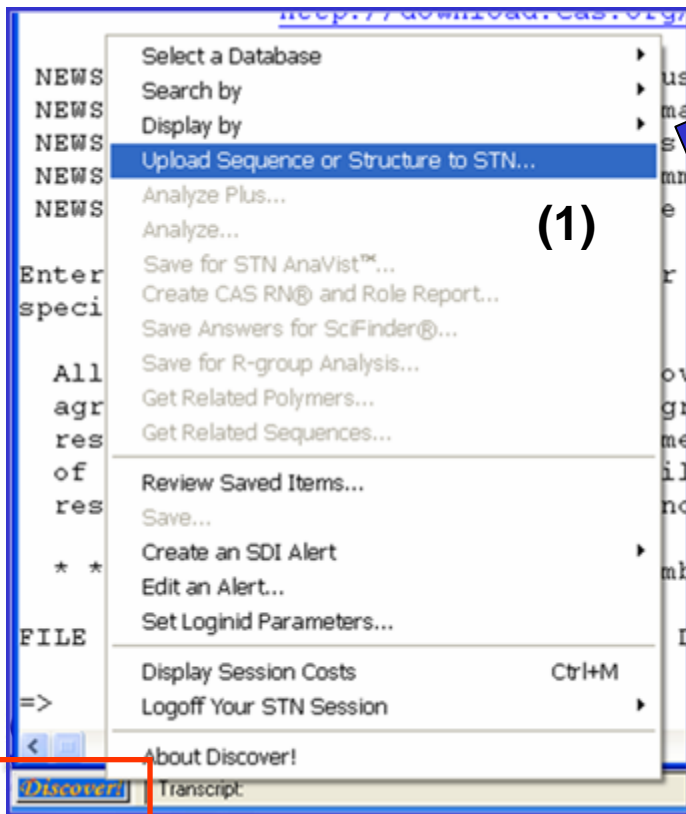
This sequence was uniquely retrieved by REGISTRY BLAST, because the patent has not yet been indexed into DGENE.

Query: 1 MKFYLVRHGKTQWNLEGRFQGANSDPLLEEAIIEELEELGQYLSSIHFDAVYSSD 55  
MK Y VRHG+T WN EGRFQGA+GDSPLL E+IE L+ LGQYL I FD +YSSD  
Subject: 1 MKLYFVRHGRTLWNQEGRFQASGDSPLLPESIETLKRLLGQYLKEIPFDQIYSSD 55  
Query: 56 LGRARDTVNILNDANSCPKIHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFY 110  
L RA + I+ P + P LREW LG LEG KIAT++AIYP+Q+ AF  
Subject: 56 LPRAVKSAEIIQSOLYTPCSLEIVPNLREWQLGKLEGLKIATLEAIYPQQIQAFR 110  
Query: 111 QNPLQFKHDMFGAESLYQTTHRVESFLRSLASKNYDKVLIVGHGANLTASIRSL 165  
N QF MFGAESLY TT R F++SL +++LIVGHGANLTAS+R+LL  
Subject: 111 SNLAQFDTRMFGAESLYSTTQRTIQFIKSLKDSPAERILIVGHGANLTASLR 165  
Query: 166 GYQYGSLSHYKDKLDNASLTIIETHDFKDFNCLTWNDKSY 204  
GY+ L L NASLTIIETHDF+ F TWND SY  
Subject: 166 GYKEPLLKDGGLANASLTIIETHDFETFTLNTWNDTSY 204

# DGENE & PCTGEN

## STN Express Upload Sequence Query Wizard

- (1) Click **Discover!** and **Upload...**
- (2) Browse, select & **Upload Query**.
- (3) Choose the STN file of interest.




The sequence becomes a **Query L-number** in the file of choice for use with the RUN BLAST command.

# DGENE & PCTGEN

## Uploading The Query

=> FILE DGENE  
=> FILE PCTGEN

UPLOAD SUCCESSFULLY COMPLETED  
L1 GENERATED



=> D L1 LQUE

L1 ANSWER 1 PCTGEN COPYRIGHT 2004 WIPO on STN  
LQUE MKFYLVVRHGKTQWNLEGRFQGANGLSPLLEEAIIEELEELGQYLSSIHFDVYSSDLGRAR  
DTVNILNDANSCPKIHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFYQNPLQFKHDM  
FGAESLYQTTHRVESFLRSLASKNYD  
ASLTIIETHD FKDFNCLTWNDKSYL

After logging on to STN the sequence query is UPLOADED to DGENE, PCTGEN

The process for *query upload* using STN Express (8.x) is on the previous slide. . .

Verify the uploaded query with **D LQUE**

# DGENE, PCTGEN & USGENE - BLAST search<sup>15</sup>

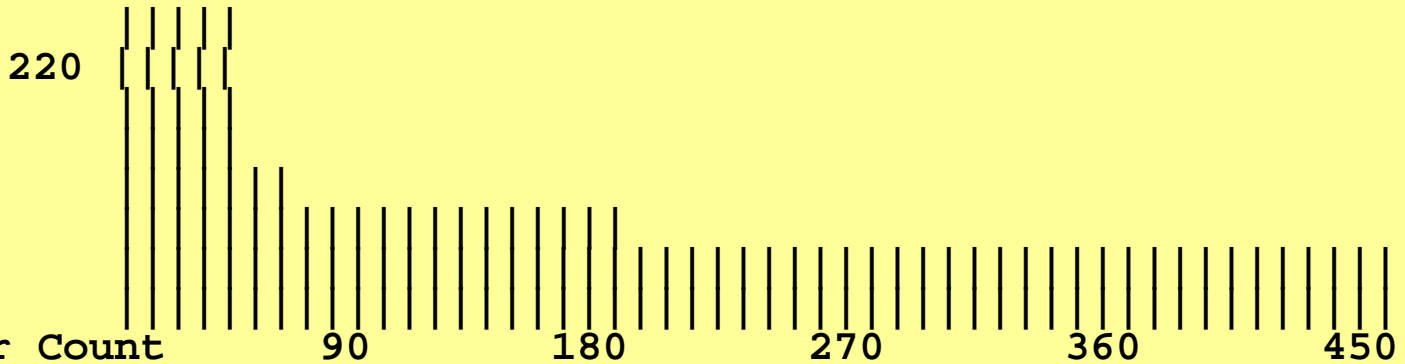
=> FILE DGENE

=> RUN BLAST L1/SQP -F F

RUN BLAST SQP for a BLASTP peptide similarity search. Reuse the PCTGEN sequence query

BLAST Version 2.2

. . . .



HOW MANY ANSWERS WOULD YOU LIKE TO KEEP ? (ALL) OR ? : **ALL**

L4 RUN STATEMENT CREATED

L4 171 MKFYLVVRHGKTQWNLEGRFQGAN GDSPLLEEAI EEELEELGQYLSSIHFDA  
VYSSDLGRARDTVNINLDANSCPK EIHYPQLREWALGTLEGCKIATMQA  
IYPRQMTAFYQNPLOQFKHDMFGAESLYQTTHR VESFLRSLASKNYDKVLI  
VGHGANLTASIRSL LGYQYGSLHYKDKLDNASLTI IETHDFKDFNCLTWN  
DKSYLRQEVKMTH/SQP.-F F

Answer set arranged by accession number; to sort by descending similarity score, enter at an arrow prompt (=>) "sor score d".

=> **SOR SCORE D**

PROCESSING COMPLETED FOR L3

L5 171 SOR L4 SCORE D

SORT the answer set by descending similarity SCORE

=> D SCORE TRIAL ALIGN

Maksuton selailuformaatti

=> D BIB ALIGN

Hyvä maksullinen muoto,  
koska mukana alignment

L5 ANSWER 1 OF 171 DGENE COPYRIGHT  
AN ABP29690 Protein DGENE  
TI New Streptococcus protein for the treatment or prevention of infection or  
disease caused by Streptococcus bacteria, such as meningitis, and for  
detecting a compound that binds to the protein  
IN Telford J; Massignani V; Margarit Ros Y I; Grandi G; Fraser C; Tettelin H  
PA (CHIR-N) CHIRON SPA (GENO-N) INST GENOMIC RES.  
PI WO 2002034771 A2 20020502 999p  
AI WO 2001-GB4789 20011029  
PRAI GB 2000-26333 20001027  
OS 2002-352536 [38]  
CR N-PSDB: ABN70321  
DESC Streptococcus polypeptide SEQ ID NO 8556.

BLASTALIGN

Query = 213 letters

Length = 213

Score = 439 bits (1130), Expect = e-128

Identities = 213/213 (100%), Positives = 213/213 (100%)

Query: 1 MKFYLVVRHGKTQWNLEGRFQGANGDSPLEEAIEEELGQYLSSIHFDAVYSSDLGRAR  
MKFYLVVRHGKTQWNLEGRFQGANGDSPLEEAIEEELGQYLSSIHFDAVYSSDLGRAR  
Sbjct: 1 MKFYLVVRHGKTQWNLEGRFQGANGDSPLEEAIEEELGQYLSSIHFDAVYSSDLGRAR  
Query: 61 DTVNILDANDANCPKEIHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFYQNPLQFKHDM  
DTVNILDANDANCPKEIHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFYQNPLQFKHDM  
Sbjct: 61 DTVNILDANDANCPKEIHYTPQLREWALGTLEGCKIATMQAIYPRQMTAFYQNPLQFKHDM  
Query: 121 FGAESLYQTTHRVESFLRSLASKNYDKVLIVGHGANLTASIRSLGQYQYGSLSHYKDKLDN  
FGAESLYQTTHRVESFLRSLASKNYDKVLIVGHGANLTASIRSLGQYQYGSLSHYKDKLDN  
Sbjct: 121 FGAESLYQTTHRVESFLRSLASKNYDKVLIVGHGANLTASIRSLGQYQYGSLSHYKDKLDN  
Query: 181 ASLTIIETHDFKDFNCLTWNDKSYLRQEVKMT 213  
ASLTIIETHDFKDFNCLTWNDKSYLRQEVKMT  
Sbjct: 181 ASLTIIETHDFKDFNCLTWNDKSYLRQEVKMT 213