



New percent identity feature for sorting BLAST answer sets in DGENE, USGENE and PCTGEN

A new sequence sorting feature for BLAST answer sets from DGENE, USGENE and PCTGEN has been available since November 30, 2009. This new sorting option is a great support for reviewing large sequence answer sets and makes the identification of relevant sequences much more efficient and reliable.

The enhanced sorting functionality allows you to sort a BLAST answer set according to local percent identity (IDENT). Additional sorting with this BLAST parameter is especially useful to identify short, highly similar sequences with a low overall homology score.

The new percent identity field IDENT is available for sorting and display purposes. Ideally it can be used in combination with the conventional homology sorting parameter SCORE. Using both parameters simultaneously for sorting generates answer sets which are easier to review. For this kind of dual

sorting either of the two BLAST parameters can be specified to be the premier sorting parameter. On top of this, customers have the choice to group sequence records into invention families using the FSORT command. When FSORT is applied to an answer set previously sorted by IDENT and SCORE, the sorting order is retained for each family (see example below). Thus the new sorting option adds additional flexibility to the display of BLAST answer sets.

The new sorting feature is available for all online and batch BLAST search results, and also for BLAST alert results in all three databases.

=> SOR L2 IDENT D SCORE D

L3 348 SOR L2 IDENT D SCORE D

BLAST answer set L2 is sorted by
 1. descending percent identity (IDENT)
 2. descending similarity score (SCORE)

=> D IDENT SCORE 1-8

```
L3 ANSWER 1 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 100%
SCORE 496 100% of query self score 496

L3 ANSWER 2 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 100%
SCORE 496 100% of query self score 496

L3 ANSWER 3 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 100%
SCORE 496 100% of query self score 496

L3 ANSWER 4 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 100%
SCORE 496 100% of query self score 496

L3 ANSWER 5 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 100%
SCORE 496 100% of query self score 496

L3 ANSWER 6 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 100%
SCORE 60 12% of query self score 496

L3 ANSWER 7 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 100%
SCORE 56 11% of query self score 496

L3 ANSWER 8 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 100%
SCORE 56 11% of query self score 496
```

TRIAL ALIGN SCORE is a free of charge display format which is very useful for a fast relevance check; it provides the alignment together with the DGENE title (TI), description (DESC) and keyword (KW) field.

=> D TRIAL ALIGN SCORE 1-

```
L3 ANSWER 1 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
AN ASS06733 cDNA DGENE
TI Providing to a mammal a mammalian Interleukin (IL)-1 receptor antagonist
(Ra) peptide or polypeptide, by introducing to cells or tissue of the
mammal, a composition comprising a recombinant self-complimentary adeno-
associated viral vector.
DESC Horse Interleukin-1 receptor antagonist coding sequence, SEQ ID 11.
KW ss; gene; IL1RA; Interleukin-1 receptor antagonist; antagonist;
therapeutic; gene therapy; protein therapy; prophylactic to disease; drug
delivery; vector; arthritis; antiarthritic; antiinflammatory;
osteoarthritis; rheumatoid arthritis; immunosuppressive; injury;
vulnerary; joint disease; periarticular joint disease;
musculoskeletal-gen.; neuromuscular disease; neuroprotective; autoimmune
disease; immunosuppressive.
```

```
SQL 534
BLASTALIGN
Query = 250 letters
Length = 534
Score = 496 bits (250), Expect = e-145
Identities = 250/250 (100%)
Strand = Plus / Plus
```

← percent identity

```
Query: 1 ggaagagaccctgcaagatgcaagccttcagaatctgggatgtaaccagaagaccttct
|
Sbjct: 86 ggaagagaccctgcaagatgcaagccttcagaatctgggatgtaaccagaagaccttct

Query: 61 acatgaggaataaccaactagttgctggatacttgcaagaatcaataactaaattacaag
|
Sbjct: 146 acatgaggaataaccaactagttgctggatacttgcaagaatcaataactaaattacaag

Query: 121 agaagatagatgtggtgccattgagcctgatgctctattcctgggactccatgggagga
|
Sbjct: 206 agaagatagatgtggtgccattgagcctgatgctctattcctgggactccatgggagga

Query: 181 agctgtgcctggcctgtgtcaagtctggtgatgagattaggtccaattggaggcagtta
|
Sbjct: 266 agctgtgcctggcctgtgtcaagtctggtgatgagattaggtccaattggaggcagtta

Query: 241 acatcactga 250
|
Sbjct: 326 acatcactga 335
SCORE 496 100% of query self score 496
```

oooooooooooooooo

```

L3 ANSWER 8 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
AN AAC63779 DNA DGENE
TI Method for predicting the risk of interstitial lung disease, comprising
detecting an interleukin-1 receptor antagonist allele and tumor necrosis
alpha allele or an allele in linkage disequilibrium with either of these
alleles
DESC Human IL-1RN gene Sacman assay probe 1.
KW Human; IL-1RN; interleukin-1 receptor; cytostatic; antiinflammatory;
immunosuppressive; dermatological; antimicrobial; antiarthritic; IL-1
receptor antagonist; tumour necrosis factor alpha antagonist interstitial
lung disease; interstitial pneumonia; pulmonary fibrosis; rheumatoid
arthritis; systemic lupus erythmatosis; Sjogren's syndrome; systemic
sclerosis; dermatomyocitis; chromosome 2; probe; ss.
SQL 28
BLASTALIGN
Query = 250 letters
Length = 28
Score = 56.0 bits (28), Expect = 6e-14
Identities = 28/28 (100%)
Strand = Plus / Plus

Query: 72 aaccaactagttgctggatacttgcaag 99
          |||
Sbjct: 1 aaccaactagttgctggatacttgcaag 28
SCORE 56 11% of query self score 496

oooooooooooooooooooo
    
```

ANSWER 8 refers to a short sequence fragment with a high percent identity (100%) but a low homology score (11%).

The same sequence would be ANSWER 248 in an answer set sorted by SCORE only.

=> FSORT L3

```

SEL L3 1- PN,APPS
L4 SEL L4 1- PN APPS : 638 TERMS
L4 348 FSO L4
      62 Multi-record Families  Answers
          Family 1             Answ
          Family 2             Answ
          Family 3             Answers 12-13
          Family 4             Answers 14-15
          Family 5             Answers 16-25
oooooooo
          Family 58            Answers 279-281
          Family 59            Answ
          Family 60            Answ
          Family 61            Answ
          Family 62            Answ
      54 Individual Records  Answ
      0 Non-patent Records
    
```

The family sort command FSORT groups sequence records into invention families according to common publication, application and/or priority numbers.

348 records of answer set L3 are grouped into: 62 families comprising multiple records 54 families with just one record

=> DISPLAY PFAM

```

ENTER (L4) OR L#:L4
ENTER PATENT FAMILY NUMBER OR RANGE (1):1
ENTER ANSWER NUMBER OR RANGE (1):1-
ENTER DISPLAY FORMAT (BIB):IDENT SCORE
YOU HAVE REQUESTED DATA FROM 8 ANSWERS - CONTINUE

L4 ANSWER 1 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 100%
SCORE 496 100% of query self score 496

L4 ANSWER 2 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 90%
SCORE 305 61% of query self score 496

L4 ANSWER 3 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 88%
SCORE 264 53% of query self score 496

L4 ANSWER 4 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 88%
SCORE 258 52% of query self score 496

L4 ANSWER 5 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 86%
SCORE 222 44% of query self score 496

L4 ANSWER 6 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN
IDENT 85%
SCORE 107 21% of query self score 496

oooooooooooooooooooo
    
```

With DISPLAY PFAM you can specify which families you want to see and how many records per family. All members of FAMILY 1 can be displayed as indicated.

FAMILY 1 is sorted by
 1. descending percent identity (IDENT)
 2. descending similarity score (SCORE)

=> DISPLAY PFAM

ENTER (L4) OR L#:.
 ENTER PATENT FAMILY NUMBER OR RANGE (1):1-
 ENTER ANSWER NUMBER OR RANGE (1):.
 ENTER DISPLAY FORMAT (BIB):**TRIAL ALIGN SCORE**

The best %-identity hit of each family can be displayed in the free of charge formats **TRIAL ALIGN SCORE**.

```
L4 ANSWER 1 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN FAMILY 1
AN ASS06733 cDNA DGENE
TI Providing to a mammal a mammalian Interleukin (IL)-1 Receptor antagonist
(Ra) peptide or polypeptide, by introducing to cells or tissue of the
mammal, a composition comprising a recombinant self-complimentary adeno-
associated viral vector.
DESC Horse Interleukin-1 receptor antagonist coding sequence, SEQ ID 11.
KW ss; gene; IL1RA; Interleukin-1 receptor antagonist; antagonist;
therapeutic; gene therapy; protein therapy; prophylactic to disease; drug
delivery; vector; arthritis; antiarthritic; antiinflammatory;
osteoarthritis; rheumatoid arthritis; immunosuppressive; injury;
vulnerable; joint disease; periarticular joint disease;
musculoskeletal-gen.; neuromuscular disease; neuroprotective; autoimmune
disease; immunosuppressive.
SQL 534
BLASTALIGN
Query = 250 letters
Length = 534
Score = 496 bits (250), Expect = e-145
Identities = 250/250 (100%)
Strand = Plus / Plus

Query: 1 ggaagagaccctgcaagatgcaagccttcagaatctgggatgttaaccagaagaccttct
|
Sbjct: 86 ggaagagaccctgcaagatgcaagccttcagaatctgggatgttaaccagaagaccttct

Query: 61 acatgaggaataaccaactagttgctggatacttgcaagaatcaataactaaattacaag
|
Sbjct: 146 acatgaggaataaccaactagttgctggatacttgcaagaatcaataactaaattacaag

Query: 121 agaagatagatgtggtgccattgagcctgatgctctattcctgggactccatgggagga
|
Sbjct: 206 agaagatagatgtggtgccattgagcctgatgctctattcctgggactccatgggagga

Query: 181 agctgtgcctggcctgtgtcaagtctggtgatgagattagggtccaattggaggcagtta
|
Sbjct: 266 agctgtgcctggcctgtgtcaagtctggtgatgagattagggtccaattggaggcagtta

Query: 241 acatcactga 250
|
Sbjct: 326 acatcactga 335
SCORE 496 100% of query self score 496
```

oooooooooooooooo

```
L4 ANSWER 14 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN FAMILY 4
AN ADW78370 DNA DGENE
TI Determining risk of undergoing myocardial infarction related to
atherosclerosis using expression product of gene in subject involves
measuring an inflammation parameter in sample e.g. circulating body
fluid/whole blood sample of leukocyte.
DESC Hybridizing sequence of long probe oligonucleotide, SEQ ID 60.
KW Myocardial infarction; atherosclerosis; inflammation; probe; ss.
SQL 38
BLASTALIGN
Query = 250 letters
Length = 38
Score = 60.0 bits (30), Expect = 6e-15
Identities = 30/30 (100%)
Strand = Plus / Plus

Query: 32 aatctgggatgttaaccagaagaccttcta 61
|
Sbjct: 1 aatctgggatgttaaccagaagaccttcta 30
SCORE 60 12% of query self score 496
```

oooooooooooooooo

```
L4 ANSWER 66 OF 348 DGENE COPYRIGHT 2009 THOMSON REUTERS on STN FAMILY 10
AN AOG43604 DNA DGENE
TI Prognosing a rheumatoid arthritis phenotype in a subject by obtaining
outcomes for one or more single nucleotide polymorphism (SNP) variables
and one or more clinical variables and using the obtained outcomes to
prognose the phenotype.
DESC Human IL1RN2 gene SNP detecting probe, SEQ ID 61.
KW snp detection; therapeutic; prognosis; rheumatoid arthritis;
antiarthritic; antirheumatic; Interleukin 1 receptor antagonist allele 2;
IL1RN2; ss; probe.
SQL 25
BLASTALIGN
Query = 250 letters
Length = 25
Score = 50.1 bits (25), Expect = 3e-12
Identities = 25/25 (100%)
Strand = Plus / Plus

Query: 74 ccaactagttgctggatacttgcaa 98
          |||
Sbjct: 1 ccaactagttgctggatacttgcaa 25
SCORE 50 10% of query self score 496

oooooooooooooooooooo
```

FIZ Karlsruhe
STN Europe
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen, Germany

Phone: +49 7247 808 555

Fax: +49 7247 808 259

helpdesk@fiz-karlsruhe.de

www.fiz-karlsruhe.de

November 2009