BLAST

Basic Local Alignment Search Tool

Jos sulla on jokin sekvenssi, saat äkkiä lisätietoja, kuten sen, onko se sekvenssi ennestään tunnettu

Kotitehtävä #8

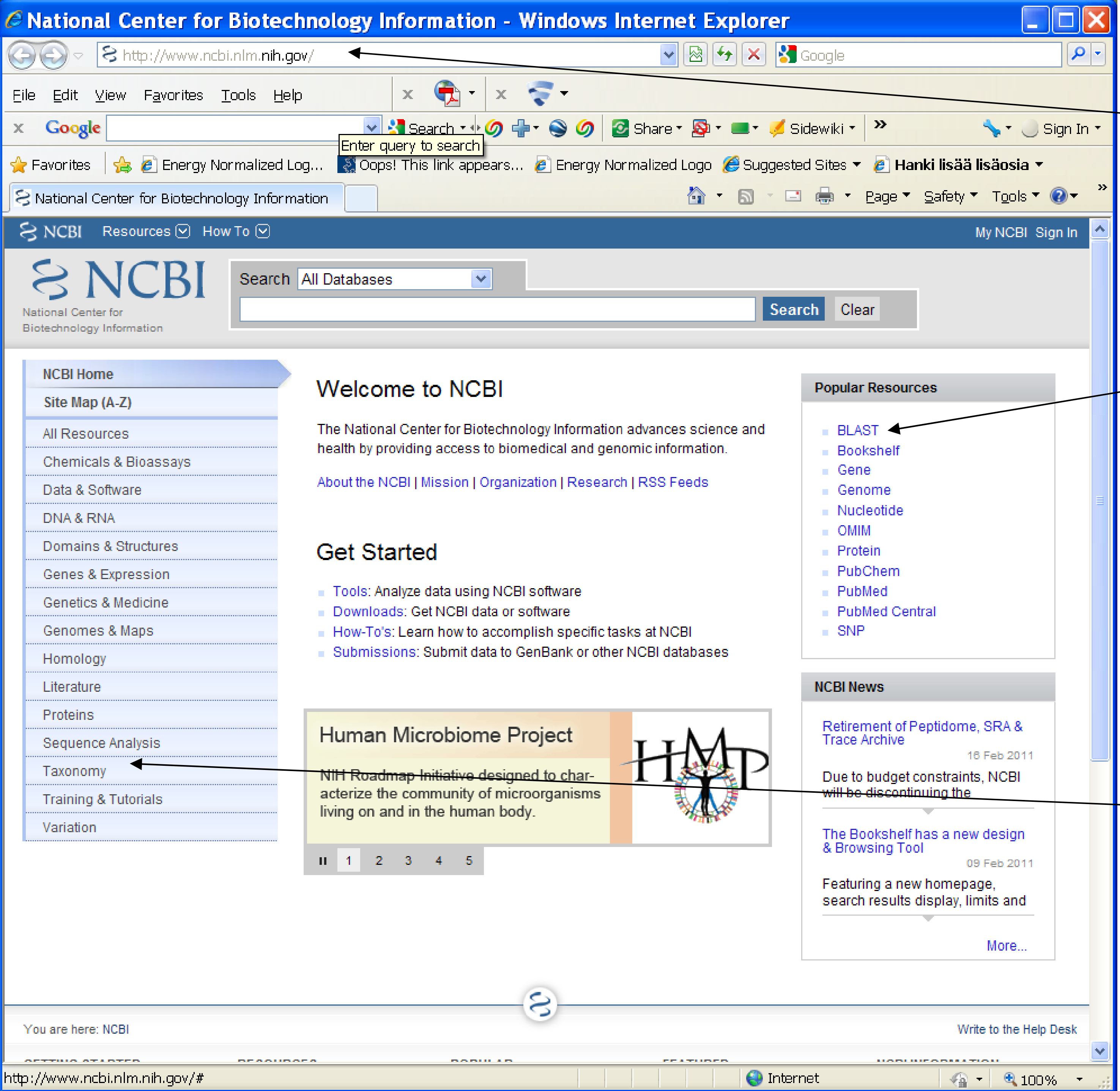
Kotitehtävä #7

Luennolla jaettiin paperi, jossa oli 20 DNA-sekvenssin pätkää (löytyy myös verkosta)

Valitse satunnaisesti yksi rivi, kopioi se tekstiksi

Mene sitten geenipankkiin ja blastaa

Seuraavassa on yksi polku malliksi



National Center for Biotechnology Information - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/

NCBI Home
Site Map (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | RSS Feeds

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How-To's: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

Human Microbiome Project

NIH Roadmap Initiative designed to characterize the community of microorganisms living on and in the human body.

Popular Resources

- BLAST
- Bookshelf
- Gene
- Genome
- Nucleotide
- OMIM
- Protein
- PubChem
- PubMed
- PubMed Central
- SNP

NCBI News

Retirement of Peptidome, SRA & Trace Archive
16 Feb 2011
Due to budget constraints, NCBI will be discontinuing the

The Bookshelf has a new design & Browsing Tool
09 Feb 2011
Featuring a new homepage, search results display, limits and

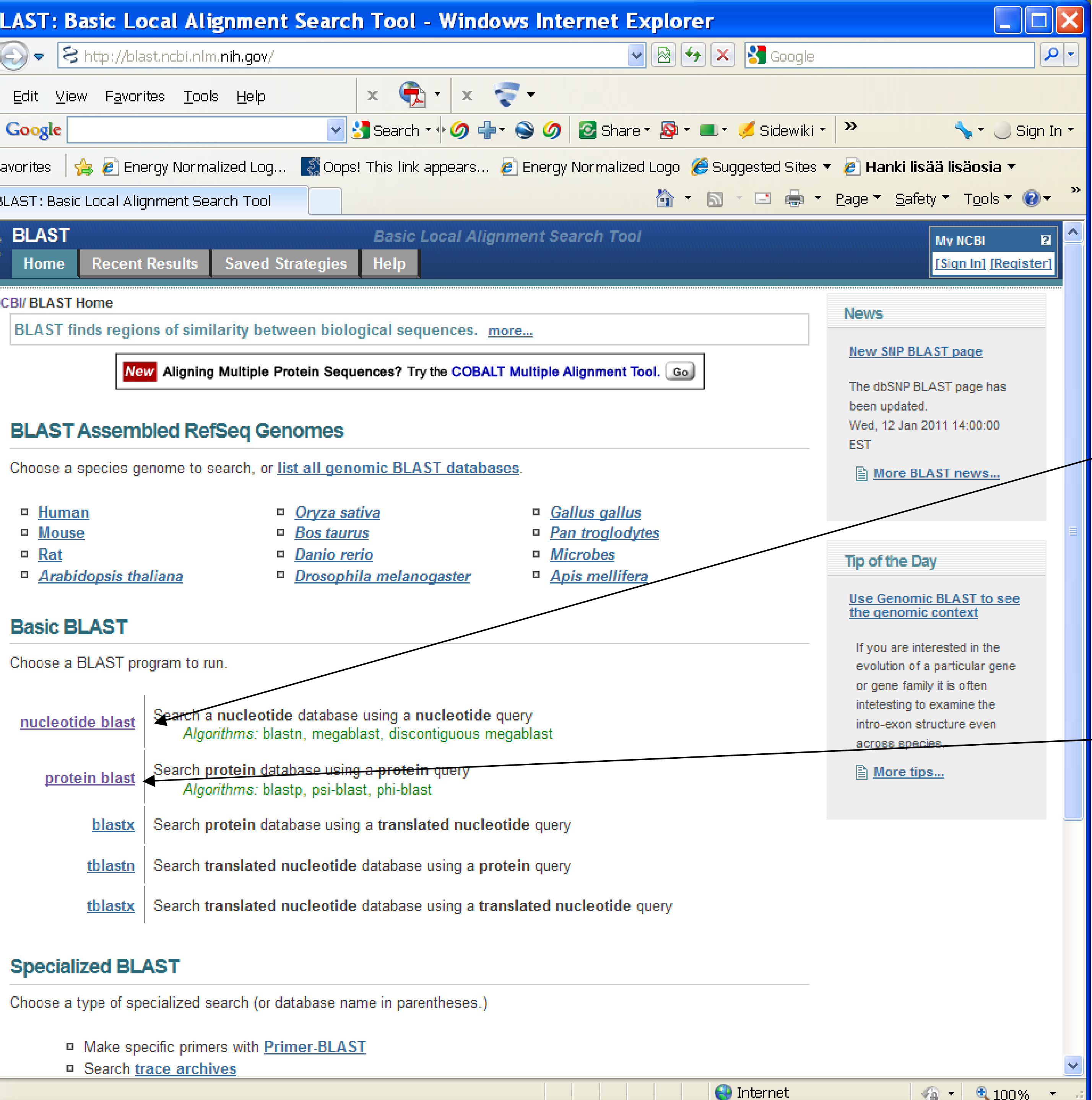
More...

You are here: NCBI
Write to the Help Desk

osoite

valikko: BLAST

taksonomiavalikko



BLAST *Basic Local Alignment Search Tool*

Home | Recent Results | Saved Strategies | Help

My NCBI
[Sign In] [Register]

CBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. more...

New Aligning Multiple Protein Sequences? Try the **COBALT Multiple Alignment Tool.** Go

BLAST Assembled RefSeq Genomes

Choose a species genome to search, or list all genomic BLAST databases.

- Human
- Mouse
- Rat
- *Arabidopsis thaliana*
- *Oryza sativa*
- *Bos taurus*
- *Danio rerio*
- *Drosophila melanogaster*
- *Gallus gallus*
- *Pan troglodytes*
- *Microbes*
- *Apis mellifera*

Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query
tblastx	Search **translated nucleotide** database using a **translated nucleotide** query

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with Primer-BLAST
- Search trace archives

News

New SNP BLAST page

The dbSNP BLAST page has been updated.
Wed, 12 Jan 2011 14:00:00 EST

More BLAST news...

Tip of the Day

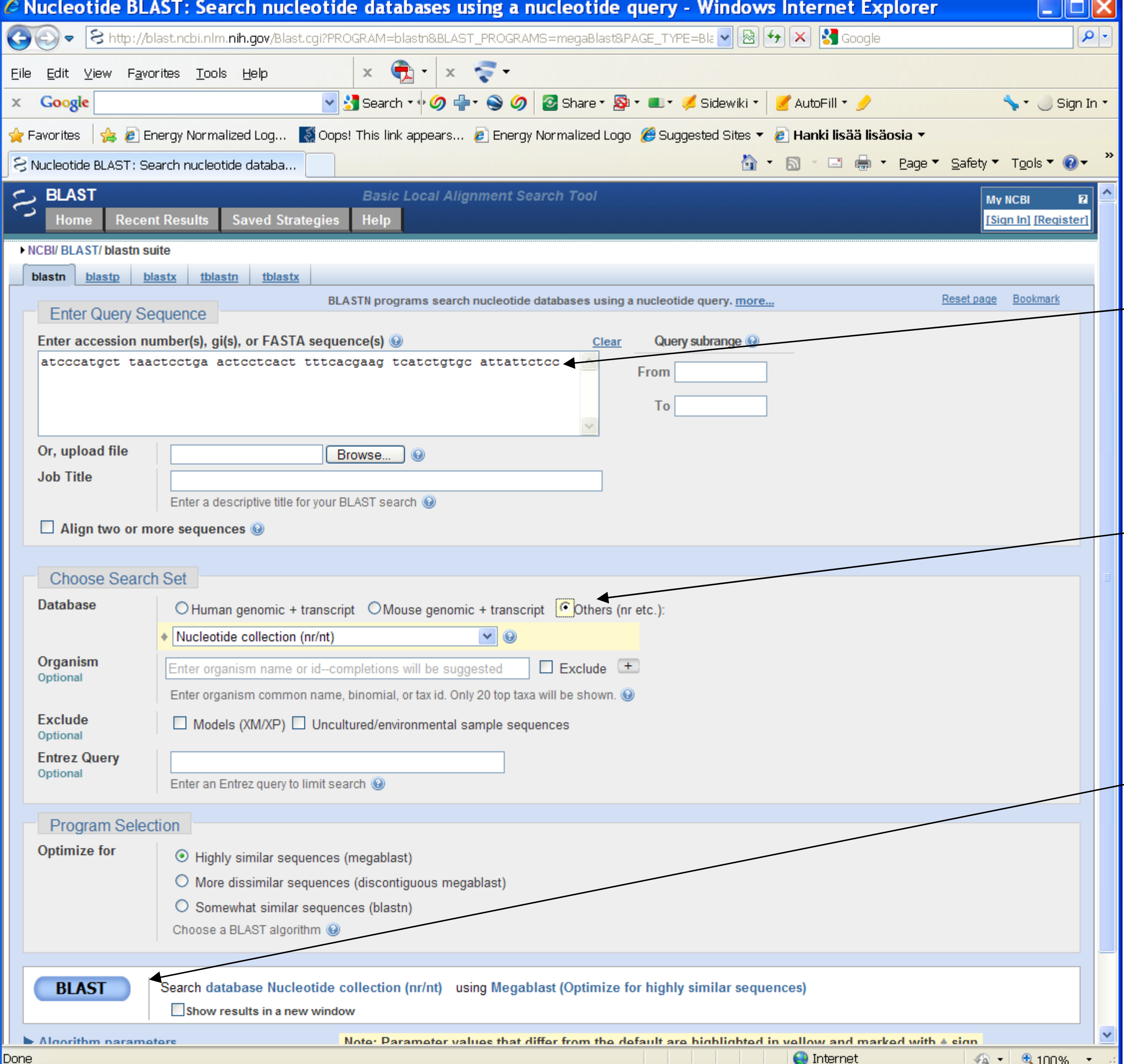
Use Genomic BLAST to see the genomic context

If you are interested in the evolution of a particular gene or gene family it is often intetesting to examine the intro-exon structure even across species.

More tips...

nukleotidiBLAST

proteiiniBLAST



BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
atcccatgct taactcctga actcctcact tttcacgaag tcatctgtgc attattctcc
```

Query subrange

From

To

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database ○ Human genomic + transcript ○ Mouse genomic + transcript ● Others (nr etc.):

Nucleotide collection (nr/nt)

Organism Optional: Enter organism name or id--completions will be suggested ☐ Exclude +

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Optional: ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query Optional

Enter an Entrez query to limit search

Program Selection

Optimize for

● Highly similar sequences (megablast)

○ More dissimilar sequences (discontiguous megablast)

○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

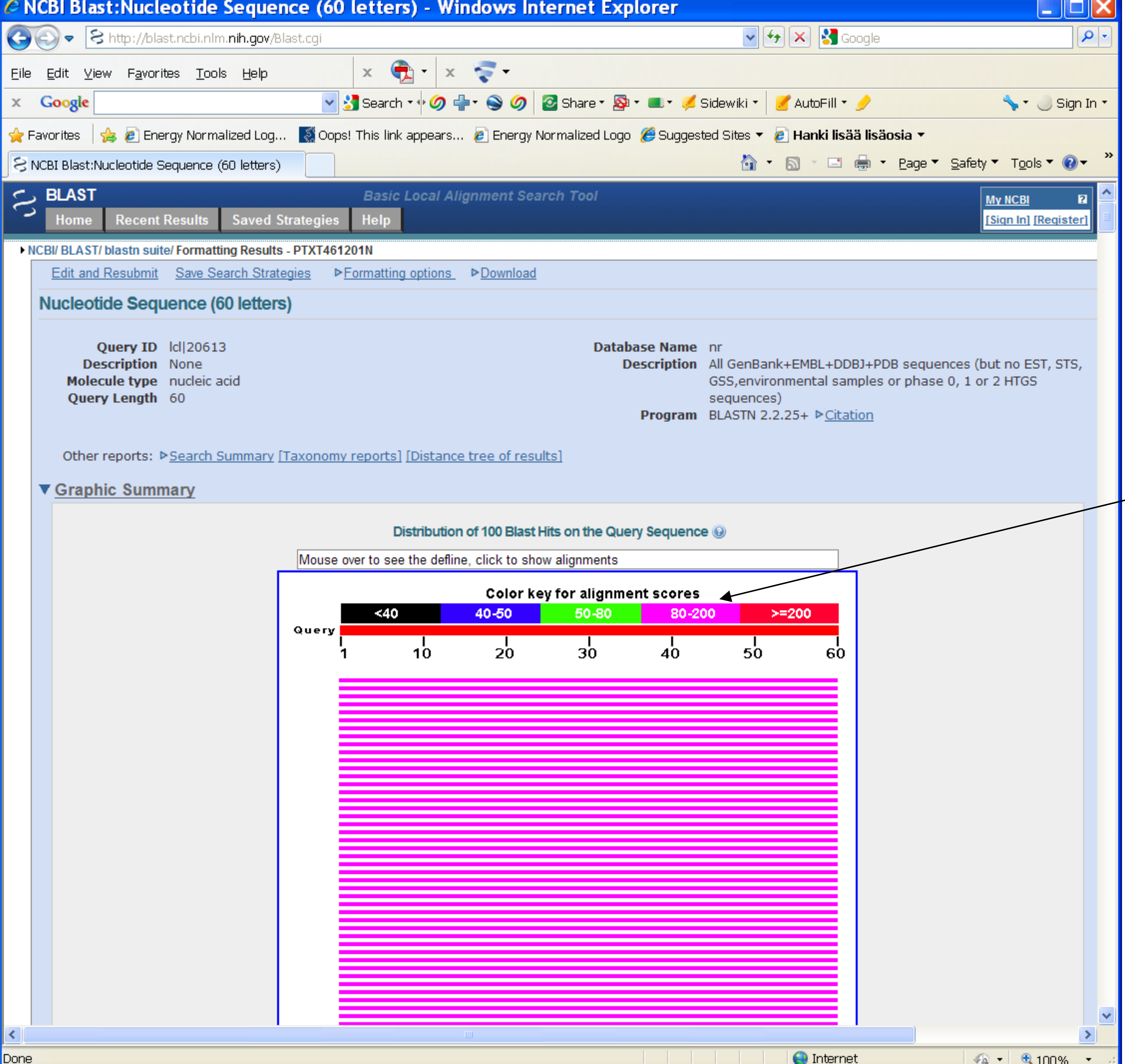
Algorithm parameters

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

liimattu sekvenssi

nappi ihmissekvenssien väistämiseksi

Go



Nucleotide Sequence (60 letters)

Query ID	lcl|20613	**Database Name**	nr
Description	None	**Description**	All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
Molecule type	nucleic acid	**Program**	BLASTN 2.2.25+ Citation
Query Length	60		

Other reports: Search Summary [Taxonomy reports] [Distance tree of results]

Graphic Summary

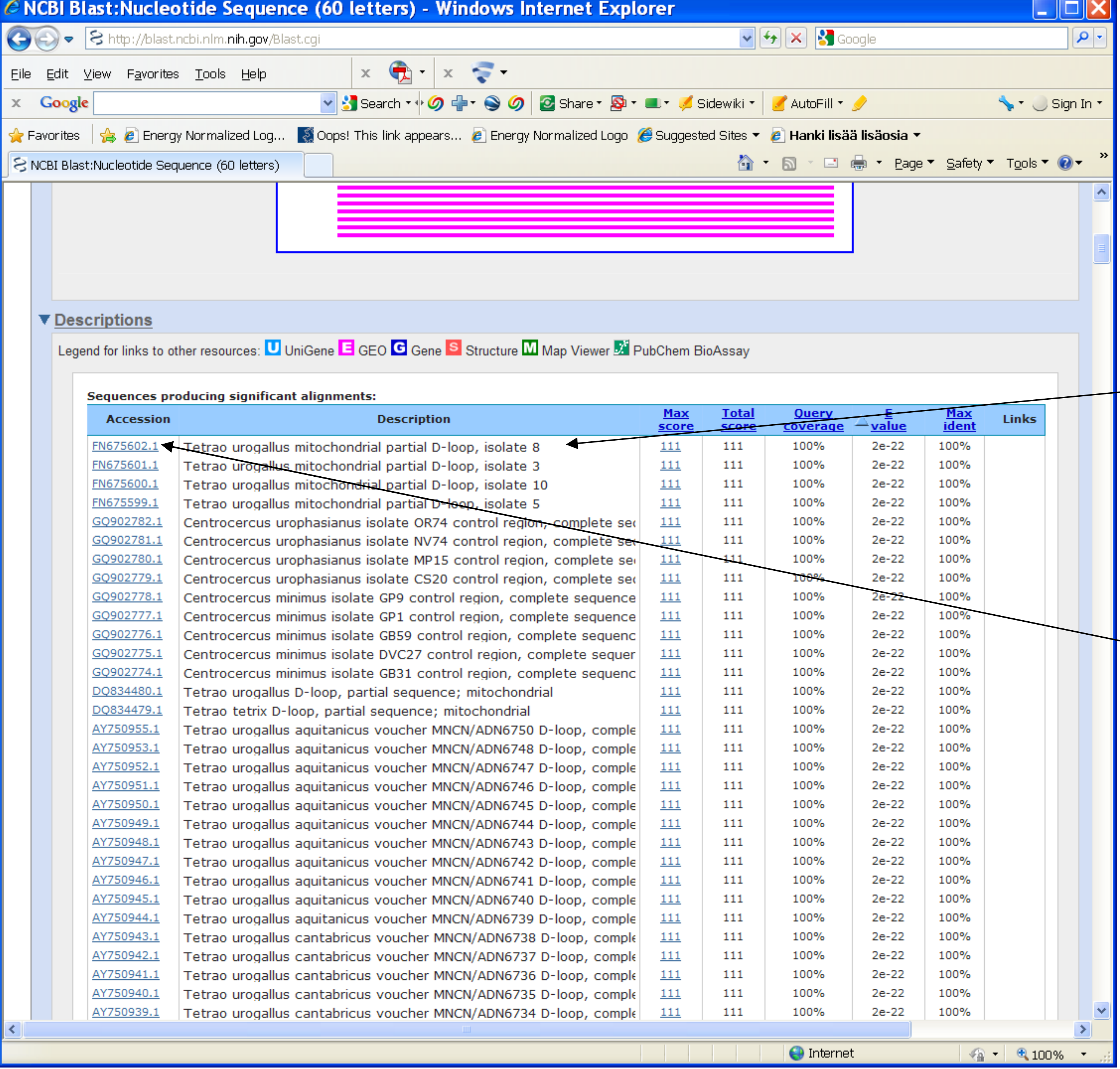
Distribution of 100 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments

Tulossivu

Graafisen osumataulun alku

laatu



▼ Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer PubChem BioAssay

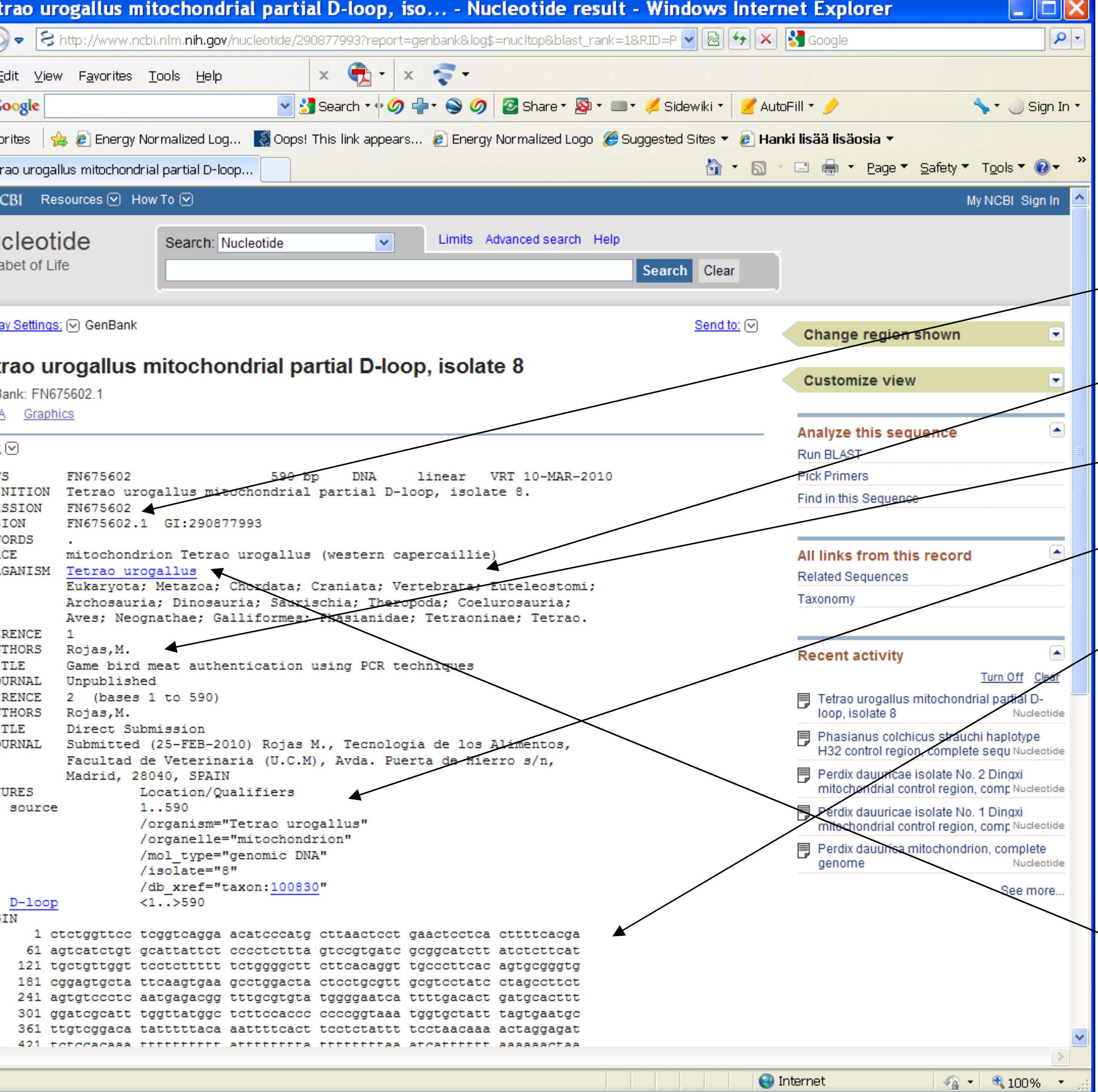
Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
FN675602.1	Tetrao urogallus mitochondrial partial D-loop, isolate 8	111	111	100%	2e-22	100%	
FN675601.1	Tetrao urogallus mitochondrial partial D-loop, isolate 3	111	111	100%	2e-22	100%	
FN675600.1	Tetrao urogallus mitochondrial partial D-loop, isolate 10	111	111	100%	2e-22	100%	
FN675599.1	Tetrao urogallus mitochondrial partial D-loop, isolate 5	111	111	100%	2e-22	100%	
GQ902782.1	Centrocercus urophasianus isolate OR74 control region, complete se	111	111	100%	2e-22	100%	
GQ902781.1	Centrocercus urophasianus isolate NV74 control region, complete se	111	111	100%	2e-22	100%	
GQ902780.1	Centrocercus urophasianus isolate MP15 control region, complete se	111	111	100%	2e-22	100%	
GQ902779.1	Centrocercus urophasianus isolate CS20 control region, complete se	111	111	100%	2e-22	100%	
GQ902778.1	Centrocercus minimus isolate GP9 control region, complete sequence	111	111	100%	2e-22	100%	
GQ902777.1	Centrocercus minimus isolate GP1 control region, complete sequence	111	111	100%	2e-22	100%	
GQ902776.1	Centrocercus minimus isolate GB59 control region, complete sequenc	111	111	100%	2e-22	100%	
GQ902775.1	Centrocercus minimus isolate DVC27 control region, complete seque	111	111	100%	2e-22	100%	
GQ902774.1	Centrocercus minimus isolate GB31 control region, complete sequenc	111	111	100%	2e-22	100%	
DQ834480.1	Tetrao urogallus D-loop, partial sequence; mitochondrial	111	111	100%	2e-22	100%	
DQ834479.1	Tetrao tetrix D-loop, partial sequence; mitochondrial	111	111	100%	2e-22	100%	
AY750955.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6750 D-loop, comple	111	111	100%	2e-22	100%	
AY750953.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6748 D-loop, comple	111	111	100%	2e-22	100%	
AY750952.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6747 D-loop, comple	111	111	100%	2e-22	100%	
AY750951.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6746 D-loop, comple	111	111	100%	2e-22	100%	
AY750950.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6745 D-loop, comple	111	111	100%	2e-22	100%	
AY750949.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6744 D-loop, comple	111	111	100%	2e-22	100%	
AY750948.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6743 D-loop, comple	111	111	100%	2e-22	100%	
AY750947.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6742 D-loop, comple	111	111	100%	2e-22	100%	
AY750946.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6741 D-loop, comple	111	111	100%	2e-22	100%	
AY750945.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6740 D-loop, comple	111	111	100%	2e-22	100%	
AY750944.1	Tetrao urogallus aquitanicus voucher MNCN/ADN6739 D-loop, comple	111	111	100%	2e-22	100%	
AY750943.1	Tetrao urogallus cantabricus voucher MNCN/ADN6738 D-loop, comple	111	111	100%	2e-22	100%	
AY750942.1	Tetrao urogallus cantabricus voucher MNCN/ADN6737 D-loop, comple	111	111	100%	2e-22	100%	
AY750941.1	Tetrao urogallus cantabricus voucher MNCN/ADN6736 D-loop, comple	111	111	100%	2e-22	100%	
AY750940.1	Tetrao urogallus cantabricus voucher MNCN/ADN6735 D-loop, comple	111	111	100%	2e-22	100%	
AY750939.1	Tetrao urogallus cantabricus voucher MNCN/ADN6734 D-loop, comple	111	111	100%	2e-22	100%	

tekstitulostus

paras osuma

tiedot



trao urogallus mitochondrial partial D-loop, isolate 8

ank: FN675602.1

A Graphics

```
S         FN675602                590 bp    DNA     linear   VRT 10-MAR-2010
NITION   Tetrao urogallus mitochondrial partial D-loop, isolate 8.
SSION    FN675602
ION      FN675602.1  GI:290877993
ORDS     .
CE       mitochondrion Tetrao urogallus (western capercaillie)
GANISM   Tetrao urogallus
         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
         Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria;
         Aves; Neognathae; Galliformes; Phasianidae; Tetraoninae; Tetrao.
RENCE    1
THORS    Rojas,M.
TLE      Game bird meat authentication using PCR techniques
URNAL    Unpublished
RENCE    2  (bases 1 to 590)
THORS    Rojas,M.
TLE      Direct Submission
URNAL    Submitted (25-FEB-2010) Rojas M., Tecnologia de los Alimentos,
         Facultad de Veterinaria (U.C.M), Avda. Puerta de Hierro s/n,
         Madrid, 28040, SPAIN
URES              Location/Qualifiers
 source           1..590
                  /organism="Tetrao urogallus"
                  /organelle="mitochondrion"
                  /mol_type="genomic DNA"
                  /isolate="8"
                  /db_xref="taxon:100830"
 D-loop           <1..>590
IN
    1 ctctggttcc tcggtcagga acatcccatg cttaactcct gaactcctca cttttcacga
   61 agtcatctgt gcattattct cccctcttta gtccgtgatc gcggcatctt atctcttcat
  121 tgctgttggt tcctcttttt tctggggctt cttcacaggt tgcccttcac agtgcgggtg
  181 cggagtgcta ttcaagtgaa gcctggacta ctcctgcgtt gcgtcctatc ctagccttct
  241 agtgtccctc aatgagacgg tttgcgtgta tggggaatca ttttgacact gatgcacttt
  301 ggatcgcatt tggttatggc tcttccaccc ccccggtaaa tggtgctatt tagtgaatgc
  361 ttgtcggaca tatttttaca aattttcact tcctctattt tcctaacaaa actaggagat
```

Datalehti

Accession number

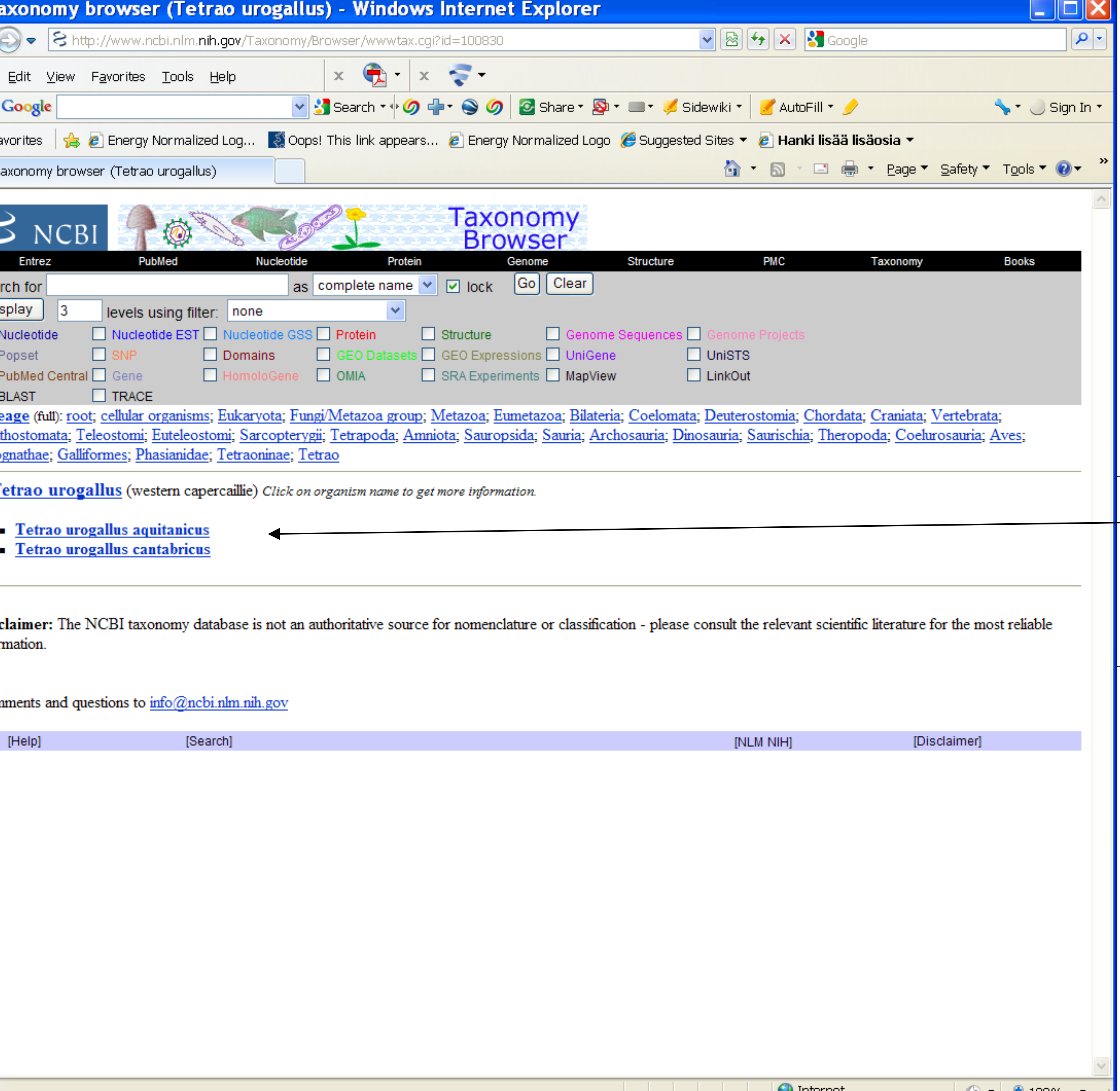
Systematiikka

Tekijä(t)

Tietoja

Sekvenssi kokonaan

ylöspäin taksonomiaan



rch for ☐ as complete name ☑ lock

3 levels using filter: none

Nucleotide ☐ Nucleotide EST ☐ Nucleotide GSS ☐ Protein ☐ Structure ☐ Genome Sequences ☐ Genome Projects

Popset ☐ SNP ☐ Domains ☐ GEO Datasets ☐ GEO Expressions ☐ UniGene ☐ UniSTS

PubMed Central ☐ Gene ☐ HomoloGene ☐ OMIA ☐ SRA Experiments ☐ MapView ☐ LinkOut

BLAST ☐ TRACE

eage (full): root; cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; thostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Sauropsida; Sauria; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Aves; gnathae; Galliformes; Phasianidae; Tetraoninae; Tetrao

etrao urogallus (western capercaillie) *Click on organism name to get more information.*

- **Tetrao urogallus aquitanicus**
- **Tetrao urogallus cantabricus**

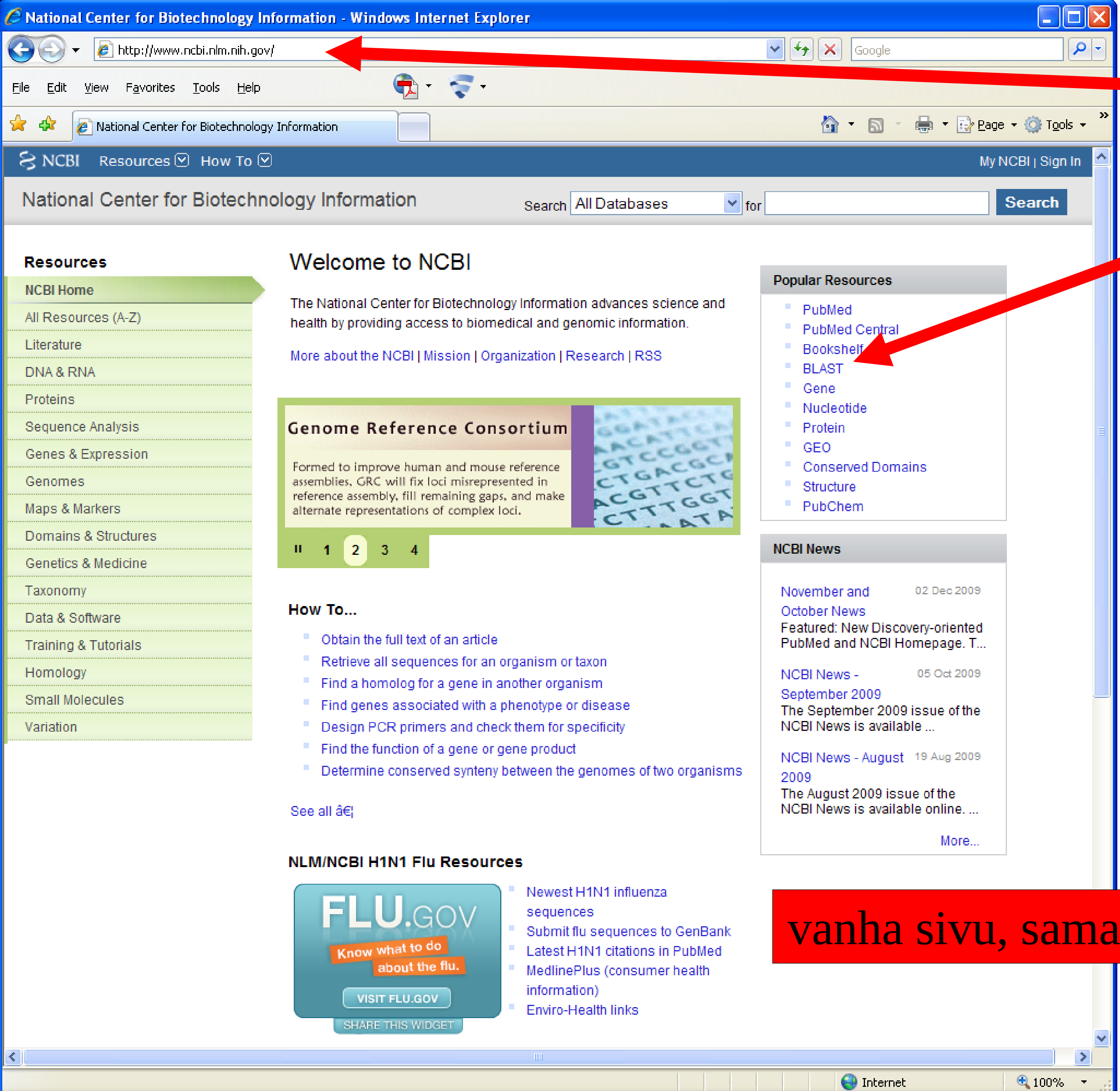
claimer: The NCBI taxonomy database is not an authoritative source for nomenclature or classification - please consult the relevant scientific literature for the most reliable rmation.

nments and questions to info@ncbi.nlm.nih.gov

Taksonomia-selain: haku nimillä

Muut tätä lajia koskevat tiedostot

Seuraavat slidet: ylimääräistä, voisi poistaakin mutten nyt viitsi



National Center for Biotechnology Information

Resources

- NCBI Home
- All Resources (A-Z)
- Literature
- DNA & RNA
- Proteins
- Sequence Analysis
- Genes & Expression
- Genomes
- Maps & Markers
- Domains & Structures
- Genetics & Medicine
- Taxonomy
- Data & Software
- Training & Tutorials
- Homology
- Small Molecules
- Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

More about the NCBI | Mission | Organization | Research | RSS

Genome Reference Consortium

Formed to improve human and mouse reference assemblies, GRC will fix loci misrepresented in reference assembly, fill remaining gaps, and make alternate representations of complex loci.

II 1 2 3 4

How To...

- Obtain the full text of an article
- Retrieve all sequences for an organism or taxon
- Find a homolog for a gene in another organism
- Find genes associated with a phenotype or disease
- Design PCR primers and check them for specificity
- Find the function of a gene or gene product
- Determine conserved synteny between the genomes of two organisms

See all â€¦

NLM/NCBI H1N1 Flu Resources

FLU.GOV Know what to do about the flu. VISIT FLU.GOV SHARE THIS WIDGET

- Newest H1N1 influenza sequences
- Submit flu sequences to GenBank
- Latest H1N1 citations in PubMed
- MedlinePlus (consumer health information)
- Enviro-Health links

Popular Resources

- PubMed
- PubMed Central
- Bookshelf
- BLAST
- Gene
- Nucleotide
- Protein
- GEO
- Conserved Domains
- Structure
- PubChem

NCBI News

November and October News 02 Dec 2009
Featured: New Discovery-oriented PubMed and NCBI Homepage. T...

NCBI News - September 2009 05 Oct 2009
The September 2009 issue of the NCBI News is available ...

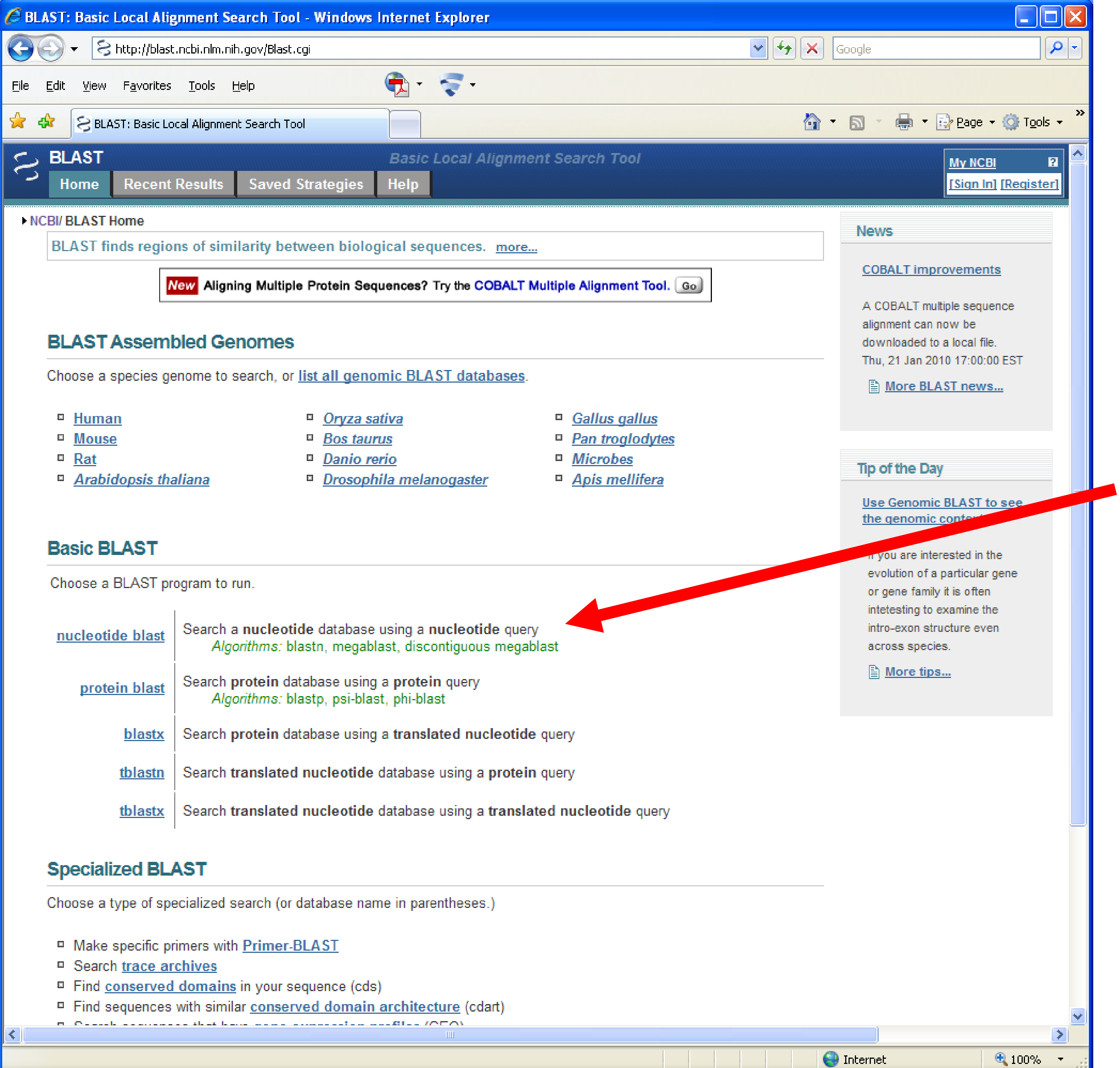
NCBI News - August 2009 19 Aug 2009
The August 2009 issue of the NCBI News is available online. ...

More...

osoite

valikko:
BLAST

vanha sivu, samat toiminnat



NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. more...

New Aligning Multiple Protein Sequences? Try the COBALT Multiple Alignment Tool. Go

BLAST Assembled Genomes

Choose a species genome to search, or list all genomic BLAST databases.

- Human
- Mouse
- Rat
- *Arabidopsis thaliana*
- *Oryza sativa*
- *Bos taurus*
- *Danio rerio*
- *Drosophila melanogaster*
- *Gallus gallus*
- *Pan troglodytes*
- *Microbes*
- *Apis mellifera*

Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query
tblastx	Search **translated nucleotide** database using a **translated nucleotide** query

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with Primer-BLAST
- Search trace archives
- Find conserved domains in your sequence (cds)
- Find sequences with similar conserved domain architecture (cdart)

News

COBALT improvements

A COBALT multiple sequence alignment can now be downloaded to a local file.
Thu, 21 Jan 2010 17:00:00 EST

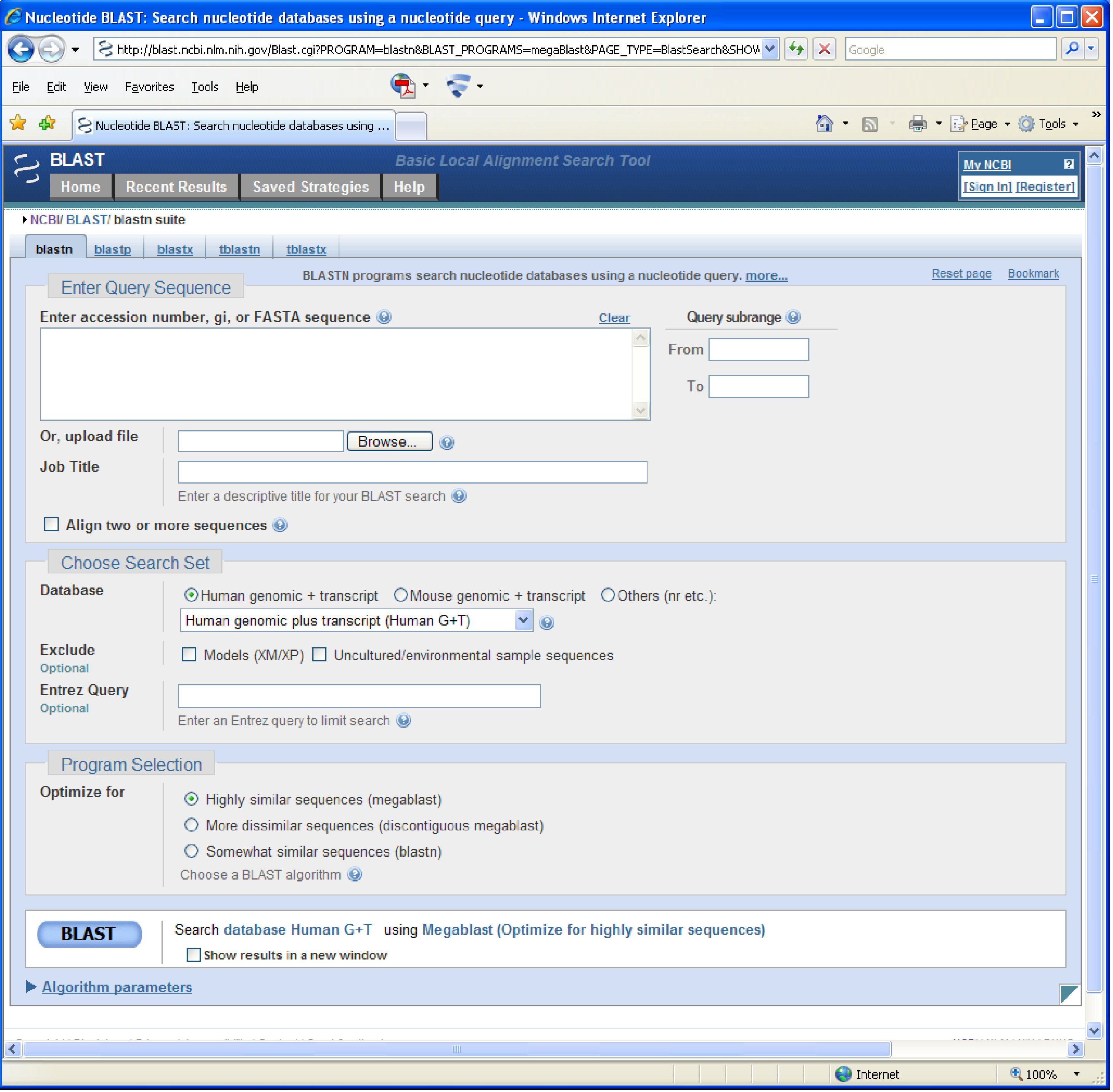
More BLAST news...

Tip of the Day

Use Genomic BLAST to see the genomic conte

If you are interested in the evolution of a particular gene or gene family it is often intetesting to examine the intro-exon structure even across species.

More tips...



Nucleotide BLAST: Search nucleotide databases using a nucleotide query - Windows Internet Explorer

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW

BLAST Basic Local Alignment Search Tool

NCBI/ BLAST/ blastn suite

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

Query subrange

From

To

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database ● Human genomic + transcript ○ Mouse genomic + transcript ○ Others (nr etc.):

Human genomic plus transcript (Human G+T)

Exclude Optional ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query Optional

Enter an Entrez query to limit search

Program Selection

Optimize for

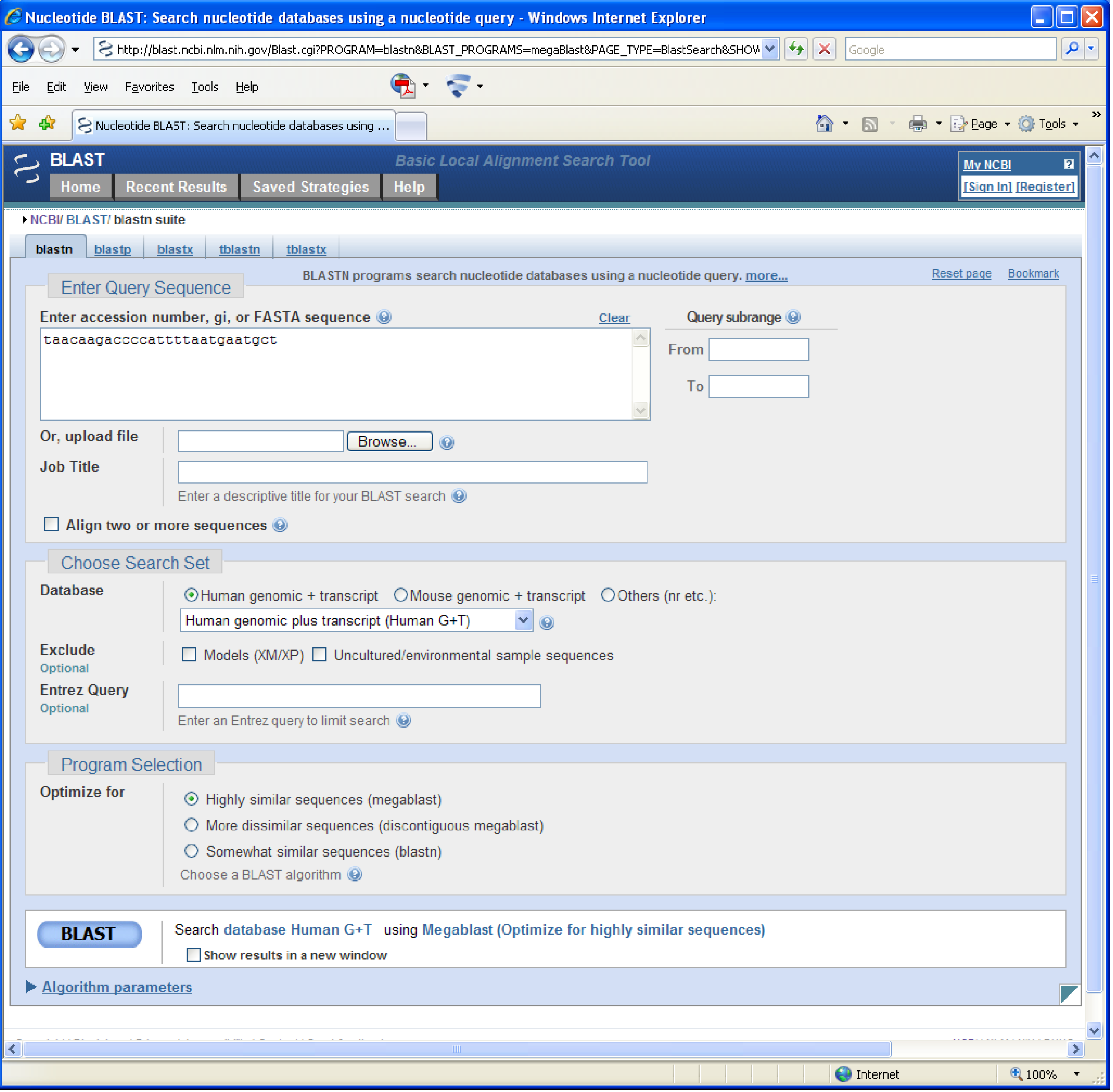
- ● Highly similar sequences (megablast)
- ○ More dissimilar sequences (discontiguous megablast)
- ○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search database Human G+T using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

Algorithm parameters



http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

```
taacaagaccccattttaatgaatgct
```

Query subrange

From

To

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database: ● Human genomic + transcript ○ Mouse genomic + transcript ○ Others (nr etc.):

Human genomic plus transcript (Human G+T)

Exclude (Optional): ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query (Optional)

Enter an Entrez query to limit search

Program Selection

Optimize for

● Highly similar sequences (megablast)

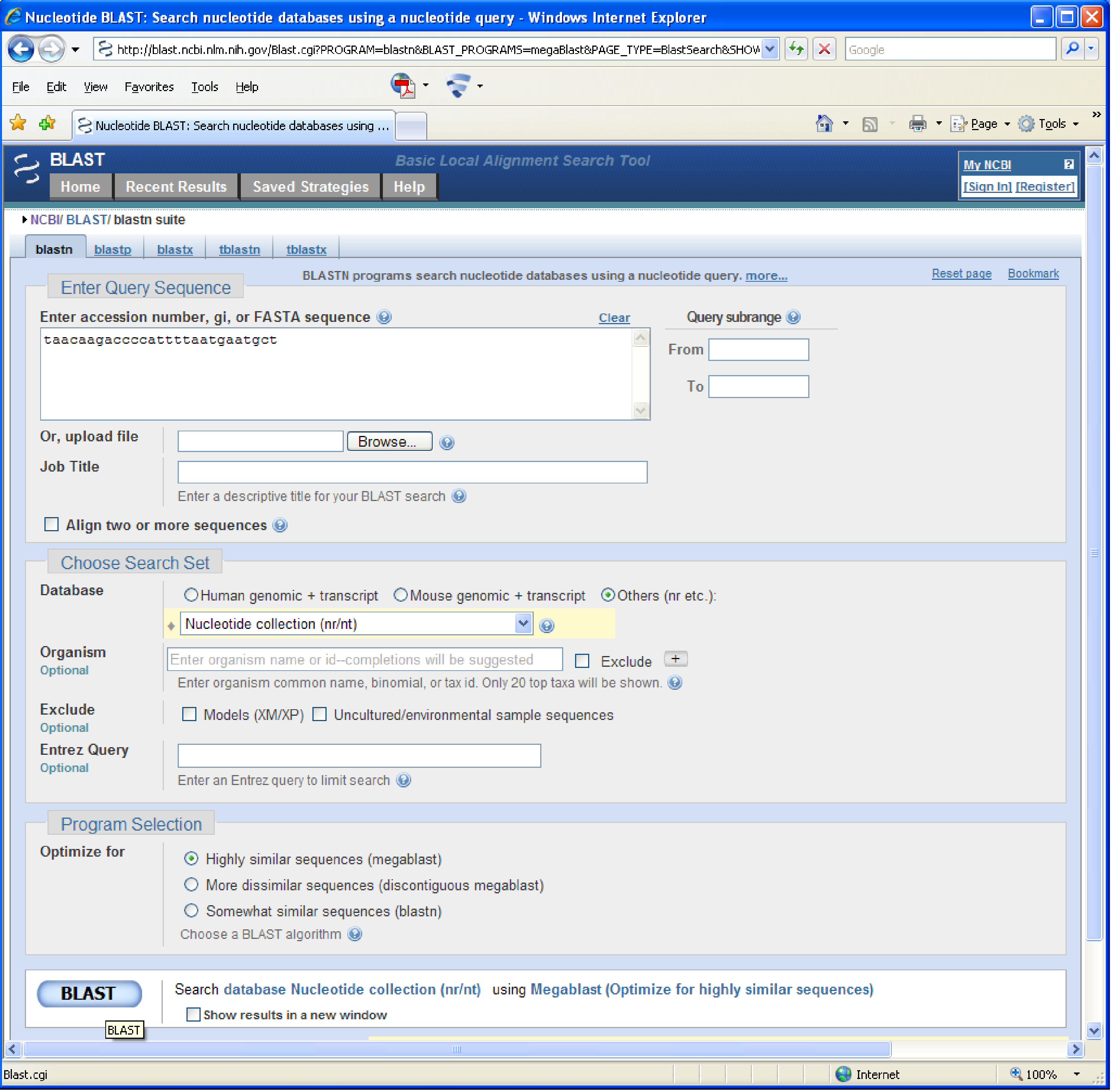
○ More dissimilar sequences (discontiguous megablast)

○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search database Human G+T using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window



BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

```
taacaagaccccattttaatgaatgct
```

Query subrange

From

To

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database: ○ Human genomic + transcript ○ Mouse genomic + transcript ● Others (nr etc.):

Nucleotide collection (nr/nt)

Organism
Optional

Enter organism name or id--completions will be suggested ☐ Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude
Optional

☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query
Optional

Enter an Entrez query to limit search

Program Selection

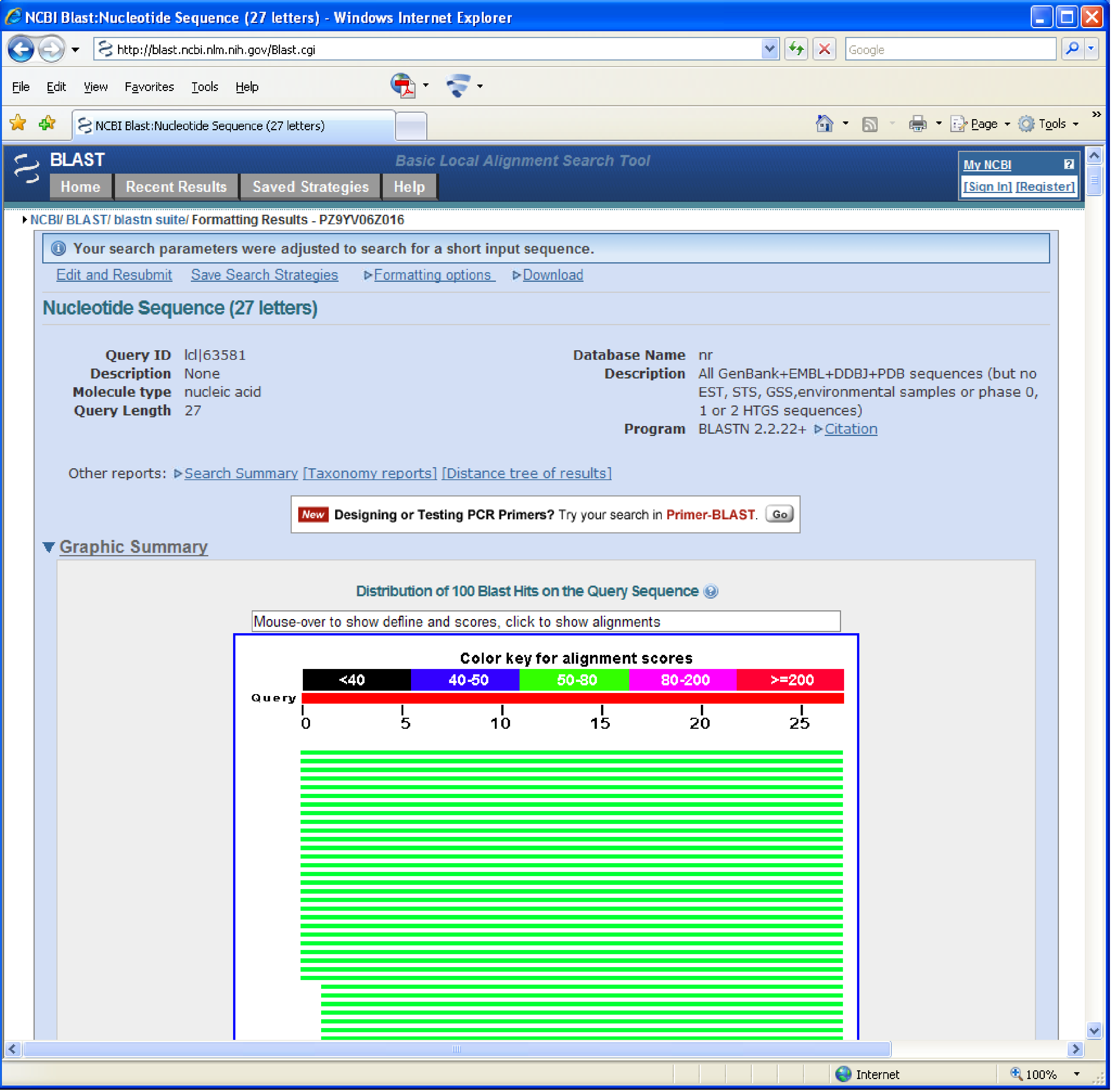
Optimize for

- ● Highly similar sequences (megablast)
- ○ More dissimilar sequences (discontiguous megablast)
- ○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window



NCBI/ BLAST/ blastn suite/ Formatting Results - PZ9YV06Z016

Your search parameters were adjusted to search for a short input sequence.

Edit and Resubmit   Save Search Strategies   ▷Formatting options   ▷Download

Nucleotide Sequence (27 letters)

Query ID lcl|63581
Description None
Molecule type nucleic acid
Query Length 27

Database Name nr
Description All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
Program BLASTN 2.2.22+ ▷Citation

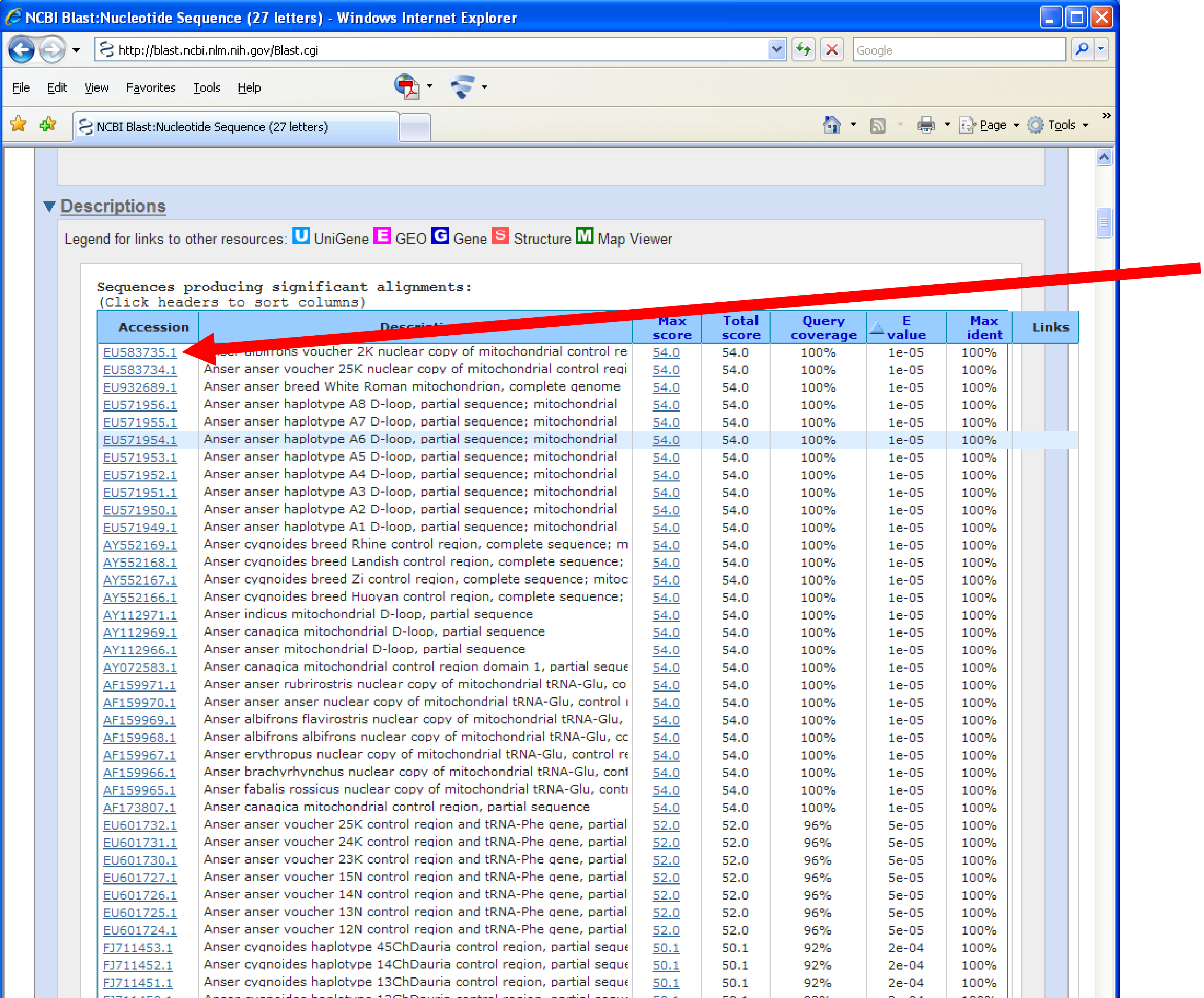
Other reports: ▷Search Summary [Taxonomy reports] [Distance tree of results]

New **Designing or Testing PCR Primers?** Try your search in **Primer-BLAST**. Go

▼ Graphic Summary

Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments

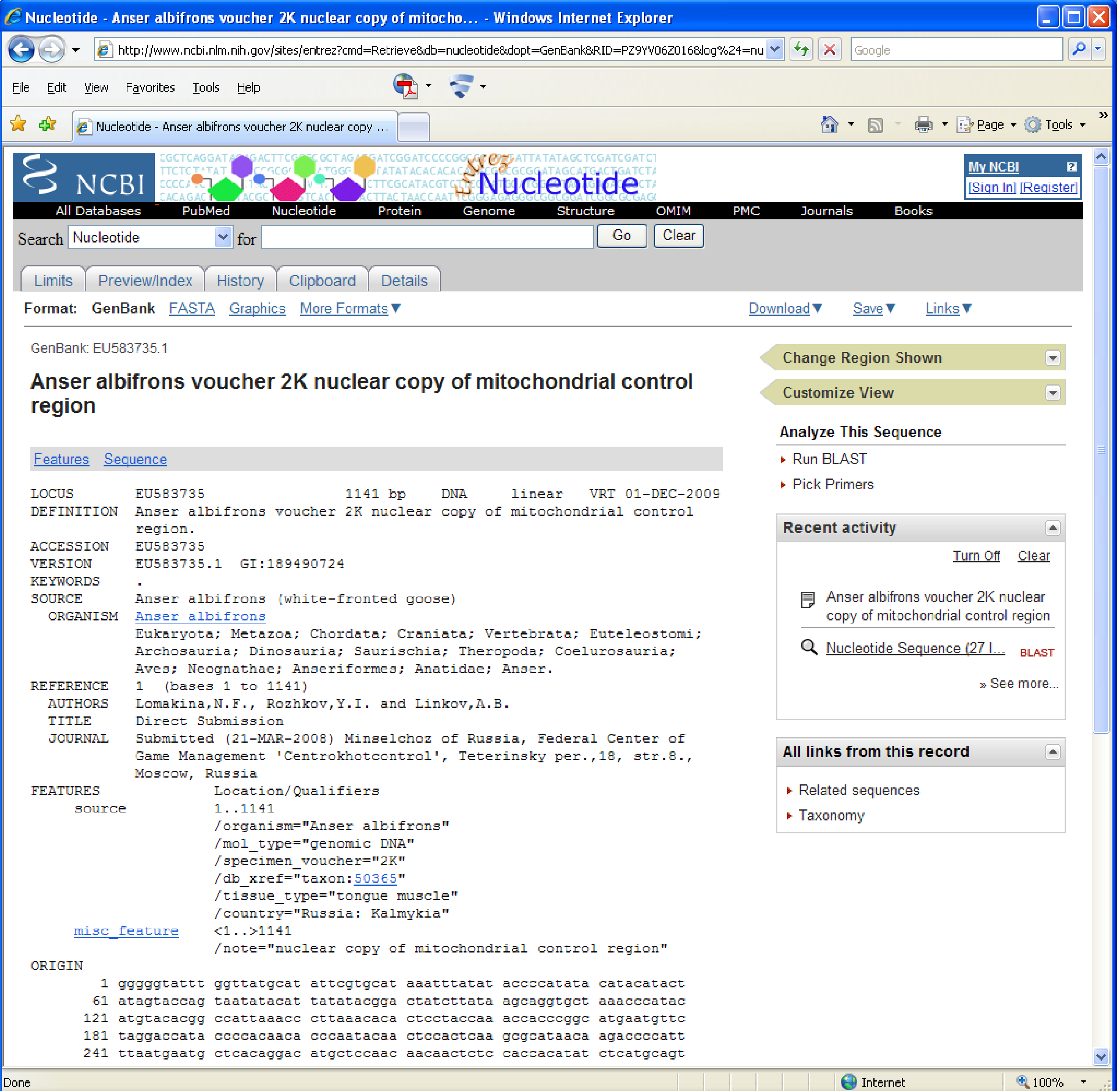


Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
EU583735.1	albifrons voucher 2K nuclear copy of mitochondrial control re	54.0	54.0	100%	1e-05	100%	
EU583734.1	Anser anser voucher 25K nuclear copy of mitochondrial control regi	54.0	54.0	100%	1e-05	100%	
EU932689.1	Anser anser breed White Roman mitochondrion, complete genome	54.0	54.0	100%	1e-05	100%	
EU571956.1	Anser anser haplotype A8 D-loop, partial sequence; mitochondrial	54.0	54.0	100%	1e-05	100%	
EU571955.1	Anser anser haplotype A7 D-loop, partial sequence; mitochondrial	54.0	54.0	100%	1e-05	100%	
EU571954.1	Anser anser haplotype A6 D-loop, partial sequence; mitochondrial	54.0	54.0	100%	1e-05	100%	
EU571953.1	Anser anser haplotype A5 D-loop, partial sequence; mitochondrial	54.0	54.0	100%	1e-05	100%	
EU571952.1	Anser anser haplotype A4 D-loop, partial sequence; mitochondrial	54.0	54.0	100%	1e-05	100%	
EU571951.1	Anser anser haplotype A3 D-loop, partial sequence; mitochondrial	54.0	54.0	100%	1e-05	100%	
EU571950.1	Anser anser haplotype A2 D-loop, partial sequence; mitochondrial	54.0	54.0	100%	1e-05	100%	
EU571949.1	Anser anser haplotype A1 D-loop, partial sequence; mitochondrial	54.0	54.0	100%	1e-05	100%	
AY552169.1	Anser cygnoides breed Rhine control region, complete sequence; m	54.0	54.0	100%	1e-05	100%	
AY552168.1	Anser cygnoides breed Landish control region, complete sequence;	54.0	54.0	100%	1e-05	100%	
AY552167.1	Anser cygnoides breed Zi control region, complete sequence; mitoc	54.0	54.0	100%	1e-05	100%	
AY552166.1	Anser cygnoides breed Huoyan control region, complete sequence;	54.0	54.0	100%	1e-05	100%	
AY112971.1	Anser indicus mitochondrial D-loop, partial sequence	54.0	54.0	100%	1e-05	100%	
AY112969.1	Anser canagica mitochondrial D-loop, partial sequence	54.0	54.0	100%	1e-05	100%	
AY112966.1	Anser anser mitochondrial D-loop, partial sequence	54.0	54.0	100%	1e-05	100%	
AY072583.1	Anser canagica mitochondrial control region domain 1, partial seque	54.0	54.0	100%	1e-05	100%	
AF159971.1	Anser anser rubrirostris nuclear copy of mitochondrial tRNA-Glu, co	54.0	54.0	100%	1e-05	100%	
AF159970.1	Anser anser anser nuclear copy of mitochondrial tRNA-Glu, control r	54.0	54.0	100%	1e-05	100%	
AF159969.1	Anser albifrons flavirostris nuclear copy of mitochondrial tRNA-Glu,	54.0	54.0	100%	1e-05	100%	
AF159968.1	Anser albifrons albifrons nuclear copy of mitochondrial tRNA-Glu, cc	54.0	54.0	100%	1e-05	100%	
AF159967.1	Anser erythropus nuclear copy of mitochondrial tRNA-Glu, control re	54.0	54.0	100%	1e-05	100%	
AF159966.1	Anser brachyrhynchus nuclear copy of mitochondrial tRNA-Glu, cont	54.0	54.0	100%	1e-05	100%	
AF159965.1	Anser fabalis rossicus nuclear copy of mitochondrial tRNA-Glu, conti	54.0	54.0	100%	1e-05	100%	
AF173807.1	Anser canagica mitochondrial control region, partial sequence	54.0	54.0	100%	1e-05	100%	
EU601732.1	Anser anser voucher 25K control region and tRNA-Phe gene, partial	52.0	52.0	96%	5e-05	100%	
EU601731.1	Anser anser voucher 24K control region and tRNA-Phe gene, partial	52.0	52.0	96%	5e-05	100%	
EU601730.1	Anser anser voucher 23K control region and tRNA-Phe gene, partial	52.0	52.0	96%	5e-05	100%	
EU601727.1	Anser anser voucher 15N control region and tRNA-Phe gene, partial	52.0	52.0	96%	5e-05	100%	
EU601726.1	Anser anser voucher 14N control region and tRNA-Phe gene, partial	52.0	52.0	96%	5e-05	100%	
EU601725.1	Anser anser voucher 13N control region and tRNA-Phe gene, partial	52.0	52.0	96%	5e-05	100%	
EU601724.1	Anser anser voucher 12N control region and tRNA-Phe gene, partial	52.0	52.0	96%	5e-05	100%	
FJ711453.1	Anser cygnoides haplotype 45ChDauria control region, partial seque	50.1	50.1	92%	2e-04	100%	
FJ711452.1	Anser cygnoides haplotype 14ChDauria control region, partial seque	50.1	50.1	92%	2e-04	100%	
FJ711451.1	Anser cygnoides haplotype 13ChDauria control region, partial seque	50.1	50.1	92%	2e-04	100%	



GenBank: EU583735.1

Anser albifrons voucher 2K nuclear copy of mitochondrial control region

```
LOCUS       EU583735                1141 bp    DNA     linear   VRT 01-DEC-2009
DEFINITION  Anser albifrons voucher 2K nuclear copy of mitochondrial control
            region.
ACCESSION   EU583735
VERSION     EU583735.1  GI:189490724
KEYWORDS    .
SOURCE      Anser albifrons (white-fronted goose)
  ORGANISM  Anser albifrons
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria;
            Aves; Neognathae; Anseriformes; Anatidae; Anser.
REFERENCE   1  (bases 1 to 1141)
  AUTHORS   Lomakina,N.F., Rozhkov,Y.I. and Linkov,A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2008) Minselchoz of Russia, Federal Center of
            Game Management 'Centrokhotcontrol', Teterinsky per.,18, str.8.,
            Moscow, Russia
FEATURES             Location/Qualifiers
     source          1..1141
                     /organism="Anser albifrons"
                     /mol_type="genomic DNA"
                     /specimen_voucher="2K"
                     /db_xref="taxon:50365"
                     /tissue_type="tongue muscle"
                     /country="Russia: Kalmykia"
     misc_feature    <1..>1141
                     /note="nuclear copy of mitochondrial control region"
ORIGIN
        1 gggggtattt ggttatgcat attcgtgcat aaatttatat accccatata catacatact
       61 atagtaccag taatatacat tatatacgga ctatcttata agcaggtgct aaacccatac
      121 atgtacacgg ccattaaacc cttaaacaca ctcctaccaa accacccggc atgaatgttc
      181 taggaccata ccccacaaca cccaatacaa ctccactcaa gcgcataaca agaccccatt
      241 ttaatgaatg ctcacaggac atgctccaac aacaactctc caccacatat ctcatgcagt
```

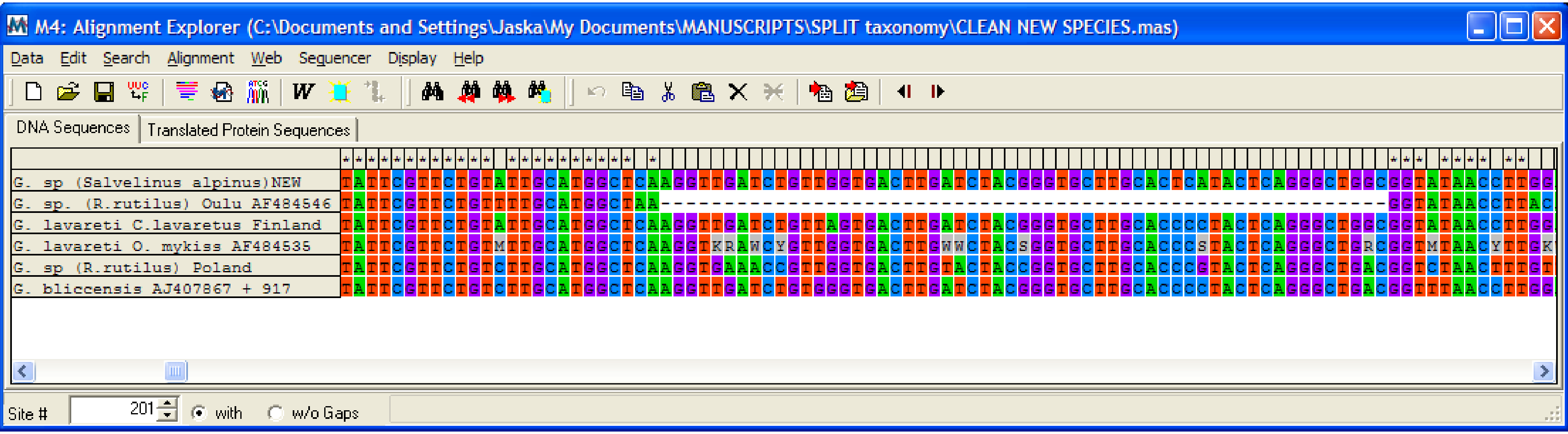
Tiedonhakua elävässä elämässä

BLAST

Basic Local Alignment Search Tool

Jos sulla on jokin sekvenssi, saat äkkiä lisätietoja, kuten sen, onko se sekvenssi ennestään tunnettu

Laitan tähän ihan oikean esimerkin



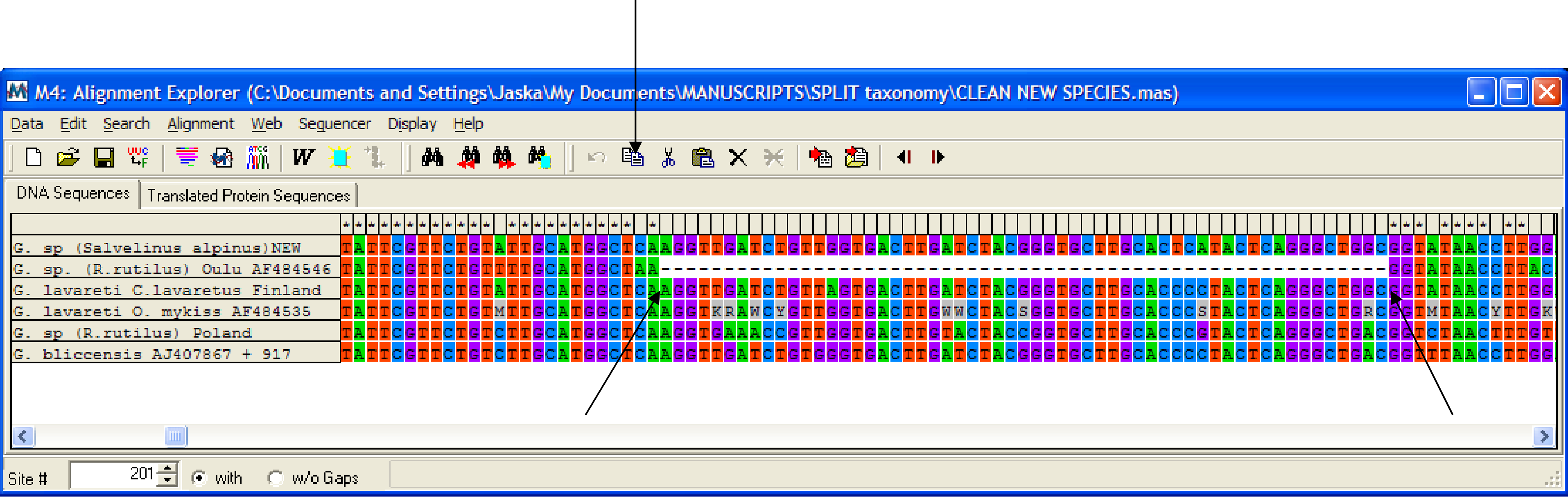
```
                                  ************ ********** *                                                         *** **** **
G. sp (Salvelinus alpinus)NEW     TATTCGTTCTGTATTGCATGGCTCAAGGTTGATCTGTTGGTGACTTGATCTACGGGTGCTTGCACTCATACTCAGGGCTGGCGGTATAACCTTGG
G. sp. (R.rutilus) Oulu AF484546  TATTCGTTCTGTTTTGCATGGCTAA---------------------------------------------------------GGTATAACCTTAC
G. lavareti C.lavaretus Finland   TATTCGTTCTGTATTGCATGGCTCAAGGTTGATCTGTTAGTGACTTGATCTACGGGTGCTTGCACCCCTACTCAGGGCTGGCGGTATAACCTTGG
G. lavareti O. mykiss AF484535    TATTCGTTCTGTMTTGCATGGCTCAAGGTKRAWCYGTTGGTGACTTGWWCTACSGGTGCTTGCACCCSTACTCAGGGCTGRCGGTMTAACYTTGK
G. sp (R.rutilus) Poland          TATTCGTTCTGTCTTGCATGGCTCAAGGTGAAACCGTTGGTGACTTGTACTACCGGTGCTTGCACCCGTACTCAGGGCTGACGGTCTAACTTTGT
G. bliccensis AJ407867 + 917      TATTCGTTCTGTCTTGCATGGCTCAAGGTTGATCTGTGGGTGACTTGATCTACGGGTGCTTGCACCCCTACTCAGGGCTGACGGTTTAACCTTGG
```

Site # 201 ● with ○ w/o Gaps

Tällainen MEGA4 näkymä ilmestyi kerran, kun tutkittiin muutamia kirjolohifarmin *Gyrodactylus*-loisia ja linjattiin ne lähimpien sukulaistensa kanssa. Selvä aukko linjauksessa.

Sehän nyt voi olla aukko (deleetio) tai lisäys (insertio), mutta kun lisättiin muita lajeja, se näytti lisäykseltä.

Tutkitaan ihan lyhyesti, mitä BLAST sanoisi.

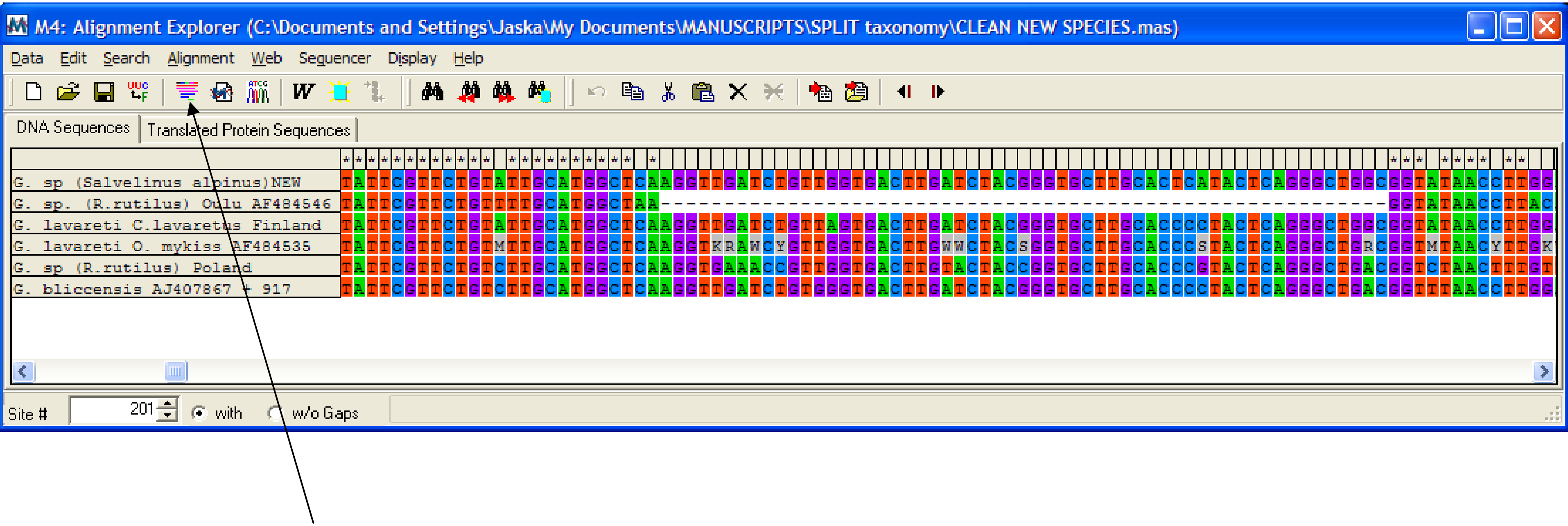


```
                                 ************ ********** *                                                         *** **** **
G. sp (Salvelinus alpinus)NEW    TATTCGTTCTGTATTGCATGGCTCAAGGTTGATCTGTTGGTGACTTGATCTACGGGTGCTTGCACTCATACTCAGGGCTGGCGGTATAACCTTGG
G. sp. (R.rutilus) Oulu AF484546 TATTCGTTCTGTTTTGCATGGCTAA---------------------------------------------------------GGTATAACCTTAC
G. lavareti C.lavaretus Finland  TATTCGTTCTGTATTGCATGGCTCAAGGTTGATCTGTTAGTGACTTGATCTACGGGTGCTTGCACCCCTACTCAGGGCTGGCGGTATAACCTTGG
G. lavareti O. mykiss AF484535   TATTCGTTCTGTMTTGCATGGCTCAAGGTKRAWCYGTTGGTGACTTGWWCTACSGGTGCTTGCACCCSTACTCAGGGCTGRCGGTMTAACYTTGK
G. sp (R.rutilus) Poland         TATTCGTTCTGTCTTGCATGGCTCAAGGTGAAACCGTTGGTGACTTGTACTACCGGTGCTTGCACCCGTACTCAGGGCTGACGGTCTAACTTTGT
G. bliccensis AJ407867 + 917     TATTCGTTCTGTCTTGCATGGCTCAAGGTTGATCTGTGGGTGACTTGATCTACGGGTGCTTGCACCCCTACTCAGGGCTGACGGTTTAACCTTGG
```

AAGGTTGATCTGTTAGTGACTTGATCTACGGGTGCTTGCAC CCCTACTCAGGGCTGGC

Merkkasin jakson (shift painettuna) ja kopioin sen (printtasin sen tuohon)

Sitten menen geenipankkiin



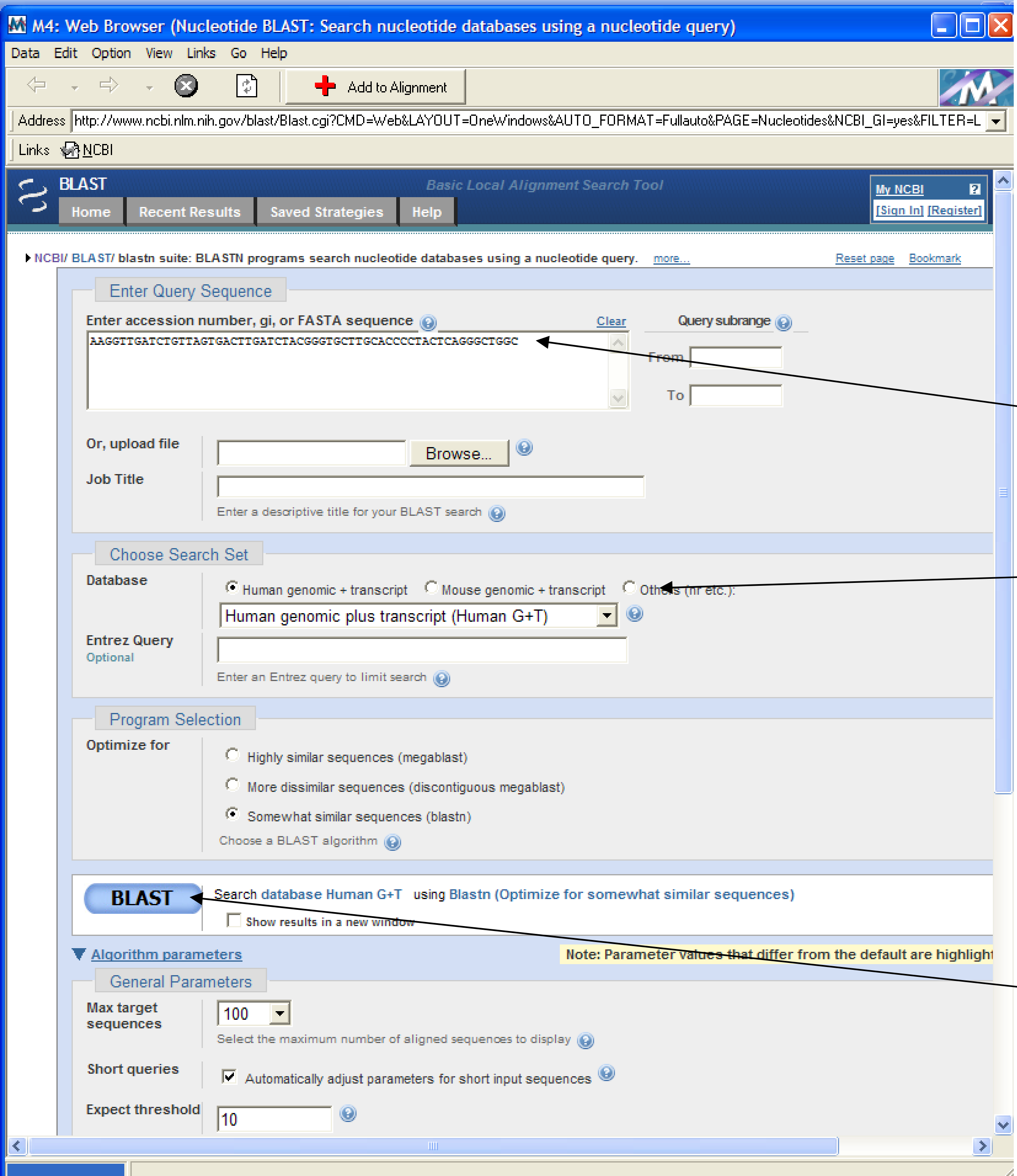
M4: Alignment Explorer (C:\Documents and Settings\Jaska\My Documents\MANUSCRIPTS\SPLIT taxonomy\CLEAN NEW SPECIES.mas)

```
                                 ************ ********** *                                                         *** **** **
G. sp (Salvelinus alpinus)NEW    TATTCGTTCTGTATTGCATGGCTCAAGGTTGATCTGTTGGTGACTTGATCTACGGGTGCTTGCACTCATACTCAGGGCTGGCGGTATAACCTTGG
G. sp. (R.rutilus) Oulu AF484546 TATTCGTTCTGTTTTGCATGGCTAA---------------------------------------------------------GGTATAACCTTAC
G. lavareti C.lavaretus Finland  TATTCGTTCTGTATTGCATGGCTCAAGGTTGATCTGTTAGTGACTTGATCTACGGGTGCTTGCACCCCTACTCAGGGCTGGCGGTATAACCTTGG
G. lavareti O. mykiss AF484535   TATTCGTTCTGTMTTGCATGGCTCAAGGTKRAWCYGTTGGTGACTTGWWCTACSGGTGCTTGCACCCSTACTCAGGGCTGRCGGTMTAACYTTGK
G. sp (R.rutilus) Poland         TATTCGTTCTGTCTTGCATGGCTCAAGGTGAAACCGTTGGTGACTTGTACTACCGGTGCTTGCACCCGTACTCAGGGCTGACGGTCTAACTTTGT
G. bliccensis AJ407867 + 917     TATTCGTTCTGTCTTGCATGGCTCAAGGTTGATCTGTGGGTGACTTGATCTACGGGTGCTTGCACCCCTACTCAGGGCTGACGGTTTAACCTTGG
```

Site # 201 ● with ○ w/o Gaps

Täältä MEGAsta pääsee suoraan geenipankkiin, kun on merkannut hakusekvenssinsä

Tuo ikoni heittää suoraan BLASTiin



M4: Web Browser (Nucleotide BLAST: Search nucleotide databases using a nucleotide query)

Address http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&LAYOUT=OneWindows&AUTO_FORMAT=Fullauto&PAGE=Nucleotides&NCBI_GI=yes&FILTER=L

NCBI/ BLAST/ blastn suite: BLASTN programs search nucleotide databases using a nucleotide query. more... Reset page Bookmark

Enter Query Sequence

Enter accession number, gi, or FASTA sequence Clear

```
AAGGTTGATCTGTTAGTGACTTGATCTACGGGTGCTTGCACCCCTACTCAGGGCTGGC
```

Query subrange

From

To

Or, upload file Browse...

Job Title

Enter a descriptive title for your BLAST search

Choose Search Set

Database ● Human genomic + transcript ○ Mouse genomic + transcript ○ Others (nr etc.):

Human genomic plus transcript (Human G+T)

Entrez Query Optional

Enter an Entrez query to limit search

Program Selection

Optimize for

○ Highly similar sequences (megablast)

○ More dissimilar sequences (discontiguous megablast)

● Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search database Human G+T using Blastn (Optimize for somewhat similar sequences)

☐ Show results in a new window

Algorithm parameters

Note: Parameter values that differ from the default are highlight

General Parameters

Max target sequences 100

Select the maximum number of aligned sequences to display

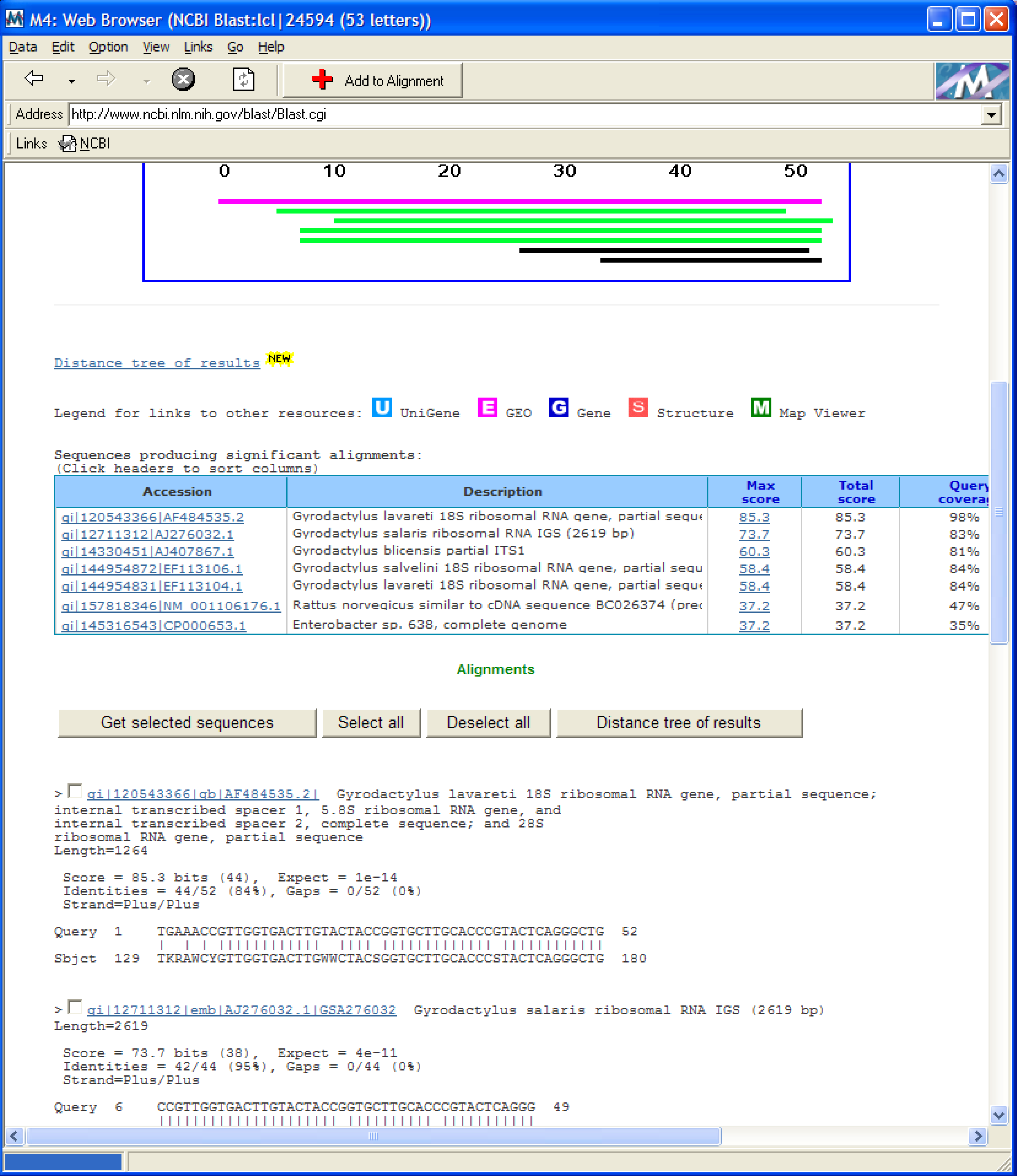
Short queries ☑ Automatically adjust parameters for short input sequences

Expect threshold 10

merkattu jakso

pitää muistaa merkitä tämä "muut", niin ei tarvitse odotella kaikkia ihmissekvenssejä

Lopuksi komennetaan toimi



Distance tree of results NEW

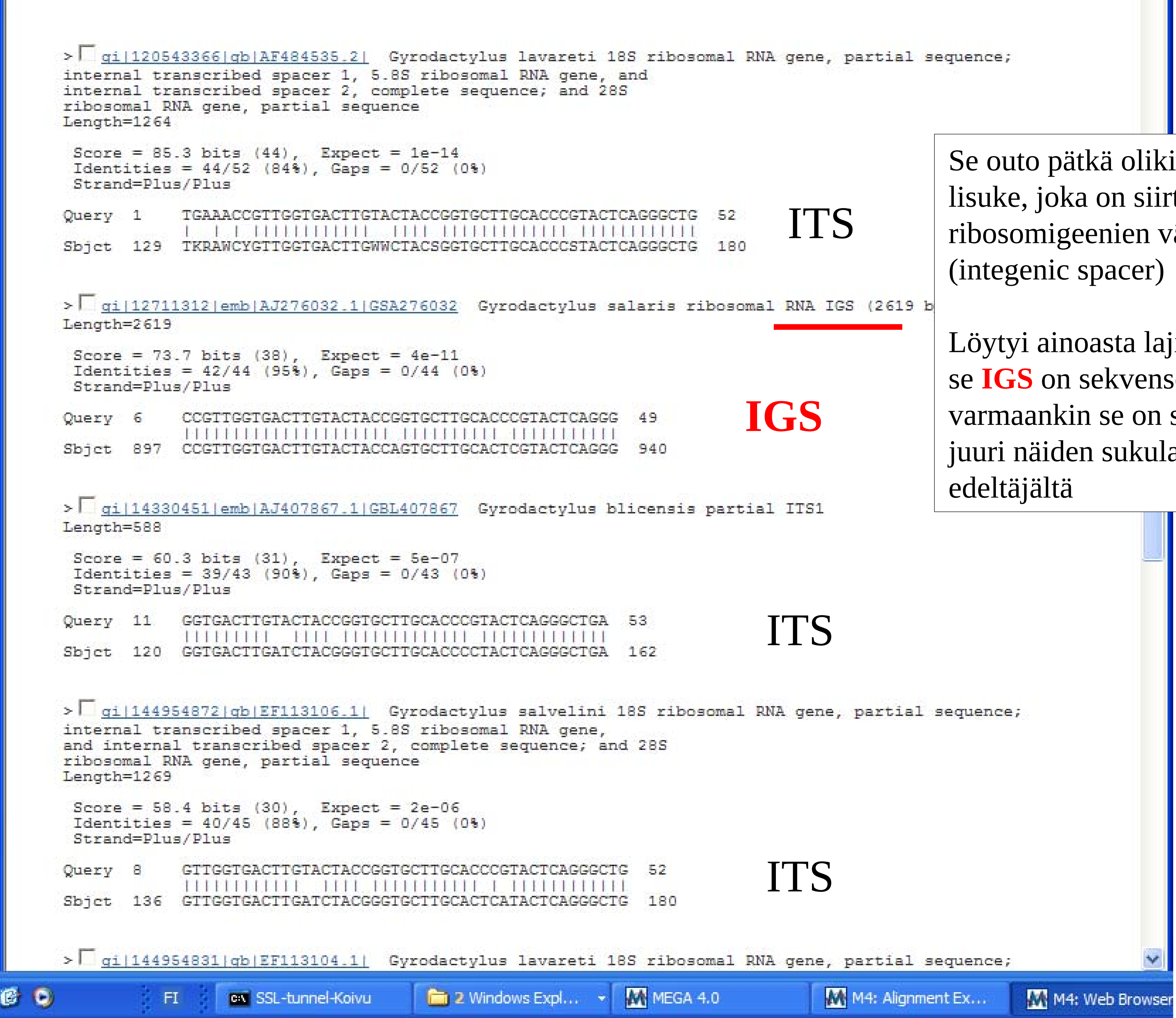
Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query covera
gi|120543366|AF484535.2	Gyrodactylus lavareti 18S ribosomal RNA gene, partial seque	85.3	85.3	98%
gi|12711312|AJ276032.1	Gyrodactylus salaris ribosomal RNA IGS (2619 bp)	73.7	73.7	83%
gi|14330451|AJ407867.1	Gyrodactylus blicensis partial ITS1	60.3	60.3	81%
gi|144954872|EF113106.1	Gyrodactylus salvelini 18S ribosomal RNA gene, partial sequ	58.4	58.4	84%
gi|144954831|EF113104.1	Gyrodactylus lavareti 18S ribosomal RNA gene, partial seque	58.4	58.4	84%
gi|157818346|NM_001106176.1	Rattus norvegicus similar to cDNA sequence BC026374 (prec	37.2	37.2	47%
gi|145316543|CP000653.1	Enterobacter sp. 638, complete genome	37.2	37.2	35%

Alignments

```
> ☐ gi|120543366|gb|AF484535.2|  Gyrodactylus lavareti 18S ribosomal RNA gene, partial sequence;
internal transcribed spacer 1, 5.8S ribosomal RNA gene, and
internal transcribed spacer 2, complete sequence; and 28S
ribosomal RNA gene, partial sequence
Length=1264

 Score = 85.3 bits (44),  Expect = 1e-14
 Identities = 44/52 (84%), Gaps = 0/52 (0%)
 Strand=Plus/Plus

Query  1    TGAAACCGTTGGTGACTTGTACTACCGGTGCTTGCACCCGTACTCAGGGCTG  52
            |  | | ||||||||||||  |||| ||||||||||||| ||||||||||||
Sbjct  129  TKRAWCYGTTGGTGACTTGWWCTACSGGTGCTTGCACCCSTACTCAGGGCTG  180


> ☐ gi|12711312|emb|AJ276032.1|GSA276032  Gyrodactylus salaris ribosomal RNA IGS (2619 bp)
Length=2619

 Score = 73.7 bits (38),  Expect = 4e-11
 Identities = 42/44 (95%), Gaps = 0/44 (0%)
 Strand=Plus/Plus

Query  6    CCGTTGGTGACTTGTACTACCGGTGCTTGCACCCGTACTCAGGG  49
            ||||||||||||||||||||| |||||||||| |||||||||||
```



```
> ☐ gi|120543366|gb|AF484535.2|  Gyrodactylus lavareti 18S ribosomal RNA gene, partial sequence;
internal transcribed spacer 1, 5.8S ribosomal RNA gene, and
internal transcribed spacer 2, complete sequence; and 28S
ribosomal RNA gene, partial sequence
Length=1264

 Score = 85.3 bits (44),  Expect = 1e-14
 Identities = 44/52 (84%), Gaps = 0/52 (0%)
 Strand=Plus/Plus

Query  1    TGAAACCGTTGGTGACTTGTACTACCGGTGCTTGCACCCGTACTCAGGGCTG  52
            |  | | ||||||||||||  |||| ||||||||||||| ||||||||||||
Sbjct  129  TKRAWCYGTTGGTGACTTGWWCTACSGGTGCTTGCACCCSTACTCAGGGCTG  180


> ☐ gi|12711312|emb|AJ276032.1|GSA276032  Gyrodactylus salaris ribosomal RNA IGS (2619 b
Length=2619

 Score = 73.7 bits (38),  Expect = 4e-11
 Identities = 42/44 (95%), Gaps = 0/44 (0%)
 Strand=Plus/Plus

Query  6    CCGTTGGTGACTTGTACTACCGGTGCTTGCACCCGTACTCAGGG  49
            ||||||||||||||||||||| |||||||||| |||||||||||
Sbjct  897  CCGTTGGTGACTTGTACTACCAGTGCTTGCACTCGTACTCAGGG  940


> ☐ gi|14330451|emb|AJ407867.1|GBL407867  Gyrodactylus blicensis partial ITS1
Length=588

 Score = 60.3 bits (31),  Expect = 5e-07
 Identities = 39/43 (90%), Gaps = 0/43 (0%)
 Strand=Plus/Plus

Query  11   GGTGACTTGTACTACCGGTGCTTGCACCCGTACTCAGGGCTGA  53
            |||||||||  |||| ||||||||||||| |||||||||||||
Sbjct  120  GGTGACTTGATCTACGGGTGCTTGCACCCCTACTCAGGGCTGA  162


> ☐ gi|144954872|gb|EF113106.1|  Gyrodactylus salvelini 18S ribosomal RNA gene, partial sequence;
internal transcribed spacer 1, 5.8S ribosomal RNA gene,
and internal transcribed spacer 2, complete sequence; and 28S
ribosomal RNA gene, partial sequence
Length=1269

 Score = 58.4 bits (30),  Expect = 2e-06
 Identities = 40/45 (88%), Gaps = 0/45 (0%)
 Strand=Plus/Plus

Query  8    GTTGGTGACTTGTACTACCGGTGCTTGCACCCGTACTCAGGGCTG  52
            ||||||||||||  |||| ||||||||||| | ||||||||||||
Sbjct  136  GTTGGTGACTTGATCTACGGGTGCTTGCACTCATACTCAGGGCTG  180


> ☐ gi|144954831|gb|EF113104.1|  Gyrodactylus lavareti 18S ribosomal RNA gene, partial sequence;
```

ITS

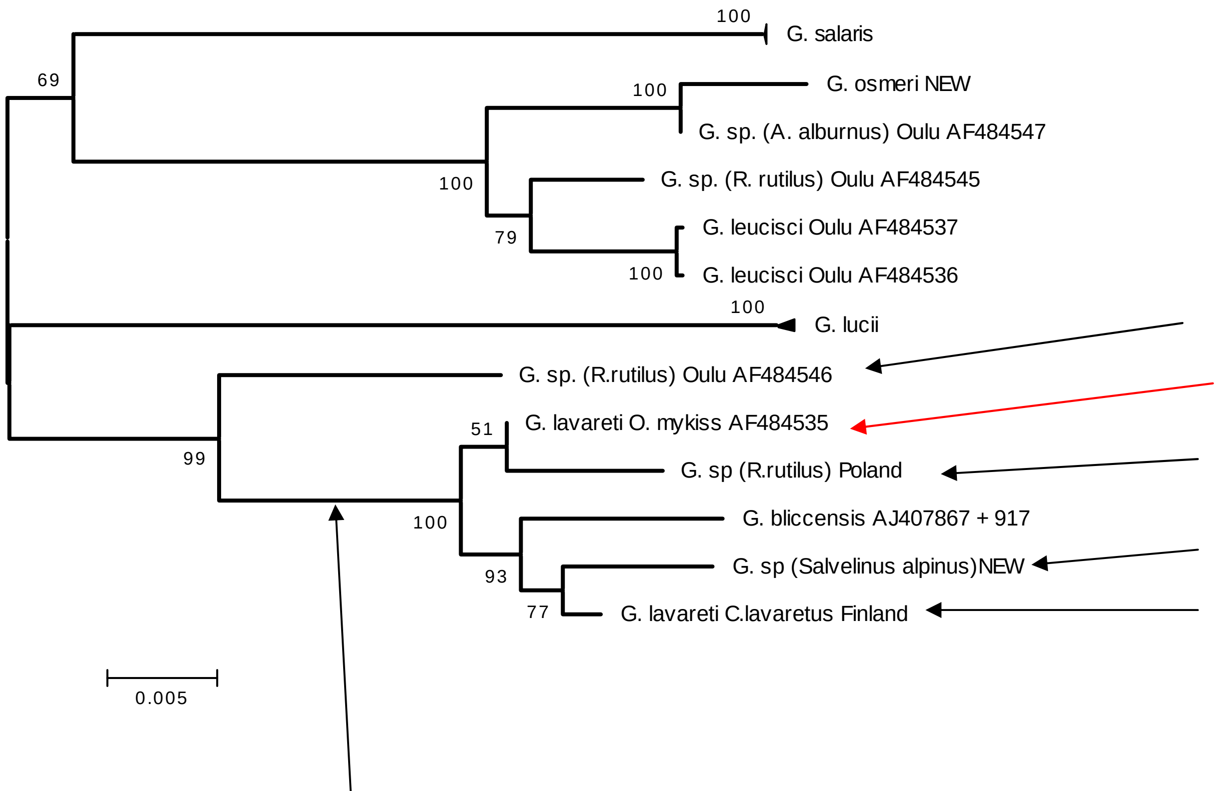
IGS

ITS

ITS

Se outo pätkä olikin insertio, lisuke, joka on siirtynyt ribosomigeenien välikkeestä (integenic spacer)

Löytyi ainoasta lajista, jolta se **IGS** on sekvensoitu, mutta varmaankin se on siirtynyt juuri näiden sukulaislajien edeltäjältä



Uusia lajeja:

G. ouluensis risteytymä kirjolohella

G. pomeraniae (ÄITI)

G. salvelini

G. lavareti (ISÄ)

Insertion (57 bp) tapahtumapaikka fylogeniassa

Sitten piirrettiin fylogenia (MEGAlla sekin) ja tiedettiin, missä insertio on tapahtunut. Risteytymä farmilla johti useiden lajien kuvailuun. Toisessa puun haarassa on monta nimetöntä lajia.