Bioinformatiikan tehtäviä

Tehtävä 1:

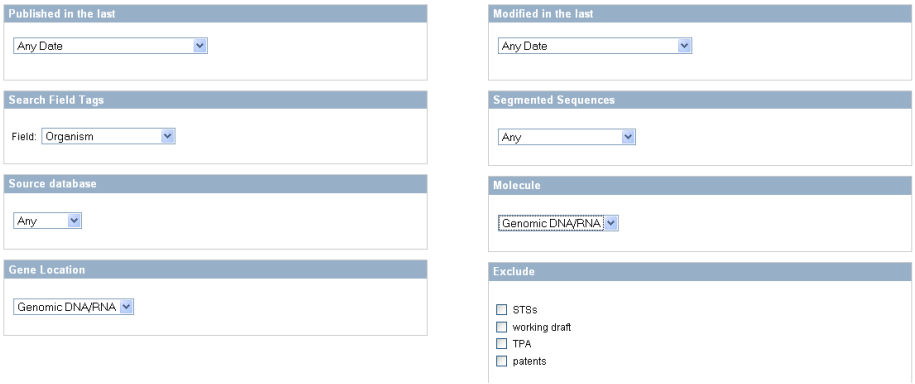
- In 2007 Lehtokari et al. published a joint Finnish-Belgian study in the journal “Neuromuscular disorders”. What was this study dealing about?

  - Entrez → Pubmed, advanced search:
    - (Lehtokari[Author]) AND "Neuromuscular disorders : NMD"[Journal]
    - AND “2007”
  - → Cap disease caused by heterozygous deletion of the beta-tropomyosin gene TPM2.

Tehtävä 2:

- How many genomic nucleic acid sequences from *Agrobacterium Tumefaciens* can be found from NCBI Nucleotide database? (Tip: Use search limits!)

  - Entrez → Nucleotide
    - search field: *Agrobacterium Tumefaciens*
    - limits:



  - → 500 kpl

Tehtävä 3:

- Synaptobrevin 1 is a small protein that plays an essential role in the mechanism of exocytosis of synaptic neurotransmitter-containing vesicles. Synaptobrevin 1 is also called VAMP-1. What is the AA-sequence of this protein in the rat? What does abbreviation “VAMP” stands for?
  - Entrez → protein
    - search field: Synaptobrevin 1 AND rat AND Vamp-1
  - → vesicle-associated membrane protein 1 [Rattus norvegicus]

    118 aa protein

    NP_037222.2 GI:76443677

```
ORIGIN
        1 msapaqppae gtegaapggg ppgpppntts nrrlqqtqaq veevvdimrv nvdkvlerdq
       61 klselddrad alqagasvfe ssaaklkrky wwknckmmim lgaicaiivv viviyift
//
```

Tehtävä 4:

- You got a message from your collaborator, with the text: "AAT74529 – check it out A.S.A.P. May be disease related /Y". Your collaborator is unfortunately is out of town for the coming weeks. You will just have to do this yourself. What is it? What human disease is it related to?

  - Entrez: AAT74529 → 1 hit in Protein database
  - → toxopain-2 [Toxoplasma gondii].
    - **Parasite**
      Toxoplasmosis, caused by the protozoan parasite *Toxoplasma gondii*, is one of the most common parasitic infections of man and other warm-blooded animals. It has been found worldwide from Alaska to Australia. Nearly one-third of humanity has been exposed to this parasite. In most adults it does not cause serious illness, but it can cause blindness and mental retardation in congenitally infected children and devastating disease in immunocompromised individuals.

Tehtävä 5:

- You have made a cDNA library derived from an unknown bacterium. At random you selected a clone and determined the sequence. This sequence is as follows:

  ATATTTTAGATCAGAATTTACAACCGCT
  TCCGATTGGGGT TCCTGGAGAGTTAC
  ATATTAGCAGTGTGGGGTTAGCGCGGG
  GTTATCTAAATCGTCTGGAATTAACTCA
  AGAAAAATTTAATTCTAACCCTTTTAATT
  CGGGTATTTTATATAAAACGGGGAATTT
  AGTTCGCTATCTTCCAGAGGGCAATATT
  GAAAATTTAGGGCGCATTGATAATCAGG
  TGAAGCTGAGAGGATTACGGATTGAAT
  TAGGAAAATAGAAGCAGAGAGACACA

- Can you check from which bacterium this sequence is derived, in which gene this sequence belongs and how long is the complete gene.
- Who first discovered this gene?

  - BLAST (nucleotide) → modify sequence to Fasta (">" at beginning)
  - → gb|AY034601.1| Microcystis aeruginosa PCC 7813 microcystin synthetase (mcyB)
    - gene, complete cds
    - Length=6381
    - → link to Nucleotide database: BOTES E. discovered

Tehtävä 6:

- You determined the sequence of a RT-PCR-product derived from the bacterium *E. coli*. Check with BLAST from which gene you have isolated a part.
  The sequence is:

  ATTCGTTCCAACACTTTTGTTGCGGAA
  CTGAAAGGCAAACAGCCAGGCGAAGT
  TGAAGTGCCGGTTATTGGCGGTCACTC
  TGGTGTTACCATTCTGCCGCTGCTGTC
  ACAGGTTCCTGGCGTTAGTTTTACCGA
  GCAGGAAGTGGCTGATCTGACCAAAC
  GTATCCAGAACGCAGGTACTGAAGTGG
  TTGAAGCGAAAGCCGGTGGTGGGTCTG
  CAACCCTGTCTATGGGCCAGGCAGCTG
  CACGTTTTGGTCTGTCTCTGGTTCGTGC
  ACTGCAGGGCG

- BLAST (nucleotide) → modify sequence to Fasta (“>” at beginning)

DQ389057.1	Escherichia coli isolate T51-urine malate dehydrogenase (mdh) gene, m
DQ441026.1	Escherichia coli isolate P31-urine malate dehydrogenase (mdh) gene, m
AY698683.1	Escherichia coli strain 05482 malate dehydrogenase (mdh) gene, partia
AY698665.1	Escherichia coli strain 395-2 malate dehydrogenase (mdh) gene, partial
AY973411.1	Escherichia coli serotype O26:H11 clone 2 Mdh (mdh) gene, partial cds
AY973410.1	Escherichia coli serotype O26:H11 clone 1 Mdh (mdh) gene, partial cds
AJ875445.1	Escherichia coli partial arcA gene for negative response regulator, strai
AJ287820.1	Escherichia coli partial mdh gene for malatdehydrogenase, strain IHIT0
AJ287818.1	Escherichia coli partial mdh gene for malatdehydrogenase, strain 413/8
AJ287817.1	Escherichia coli partial mdh gene for malatdehydrogenase, strain IHIT2
AJ287816.1	Escherichia coli partial mdh gene for malatdehydrogenase, strain RW22

Tehtävä 7:

- Out of the compost tank from your home, you isolated an unknown bacterium. In laboratory you extracted and sequenced a part of its 16S ribosomal RNA. Can you, based on this sequence, determine to which genus this bacterium belongs?

  The sequence is:

  AAGTGCTTCGAAAAATGGTAAACTTGAGTGCAGAAGAGGAAAGTTGAACTCCATGTGTAGCGGTGGAATG

  - BLAST (nucleotide) → modify sequence to Fasta (“>” at beginning)
  - Genus obviously is *Lactobacillus*

GU138609.1	Lactobacillus paralimentarius IMAU:10281 16S ribosomal RNA gene, par	119
GU138607.1	Lactobacillus paralimentarius IMAU:10279 16S ribosomal RNA gene, par	119
GU138543.1	Lactobacillus mindensis IMAU:10215 16S ribosomal RNA gene, partial se	119
GU138542.1	Lactobacillus mindensis IMAU:10214 16S ribosomal RNA gene, partial se	119
GU138541.1	Lactobacillus crustorum IMAU:10213 16S ribosomal RNA gene, partial s	119
GU138540.1	Lactobacillus paralimentarius IMAU:10212 16S ribosomal RNA gene, par	119
GU138539.1	Lactobacillus crustorum IMAU:10211 16S ribosomal RNA gene, partial s	119
GU138530.1	Lactobacillus paralimentarius IMAU:10202 16S ribosomal RNA gene, par	119
GU138528.1	Lactobacillus paralimentarius IMAU:10200 16S ribosomal RNA gene, par	119
GU138527.1	Lactobacillus mindensis IMAU:10199 16S ribosomal RNA gene, partial se	119
GU138525.1	Lactobacillus mindensis IMAU:10197 16S ribosomal RNA gene, partial se	119
GU138523.1	Lactobacillus paralimentarius IMAU:10195 16S ribosomal RNA gene, par	119
GU138520.1	Lactobacillus paralimentarius IMAU:10192 16S ribosomal RNA gene, par	119
GU138513.1	Lactobacillus paralimentarius IMAU:10185 16S ribosomal RNA gene, par	119
FJ982854.1	Uncultured Lactobacillus sp. clone OHW16 16S ribosomal RNA gene, par	119

Tehtävä 8:

- Mitä domeeneja löydät humaani synaptotagmin 1 proteiinista? Löytyykö kyseisestä domeenista kuva PDB-tietokannasta ja millä koodilla?

  - Uniprot: synaptotagmin 1 AND human → sekvenssi →PFAM on kuva, 2dmg (PDB)
  - Uniprot: synaptotagmin 1 AND human → sekvenssi →PROSITE ei ole kuvaa