BIOINFORMĀTIKAS RISINĀJUMI AUGU ĢENĒTIKĀ UN GENOMIKĀ

Dr. biol. Nils Rostoks
Latvijas Universitāte
Bioloģijas fakultāte

Semināra plāns

- Bioinformātika: biologa skatu punkts
- Augu genomika un bioinformātika
- Bioinformātika LU Bioloģijas fakultātes Augu molekulārās ģenētikas grupā

GOD... THE HUMAN GENOME CODE'S BEEN UNRAVELLED
DAMN HACKERS!!! NOW, I HAVE TO CHANGE THE PASSWORD

Bioinformātika un bioloģija

- Bioinformātika nav gluži “informātikas pielietojums bioloģijā”
- Tradicionāli bioinformātika nodarbojas ar ģenētiskās informācijas ieguvi, uzglabāšanu un analīzi
- Molekulārās bioloģijas un ģenētikas metodes aizvien plašāk izmanto arī citās bioloģijas zinātnēs, piemēram, botānikā, zooloģijā un ekoloģijā, tātad arī tur ir vajadzīga ne tikai statistika (biometrija), bet arī bioinformātika

Bioinformātika un biloģija

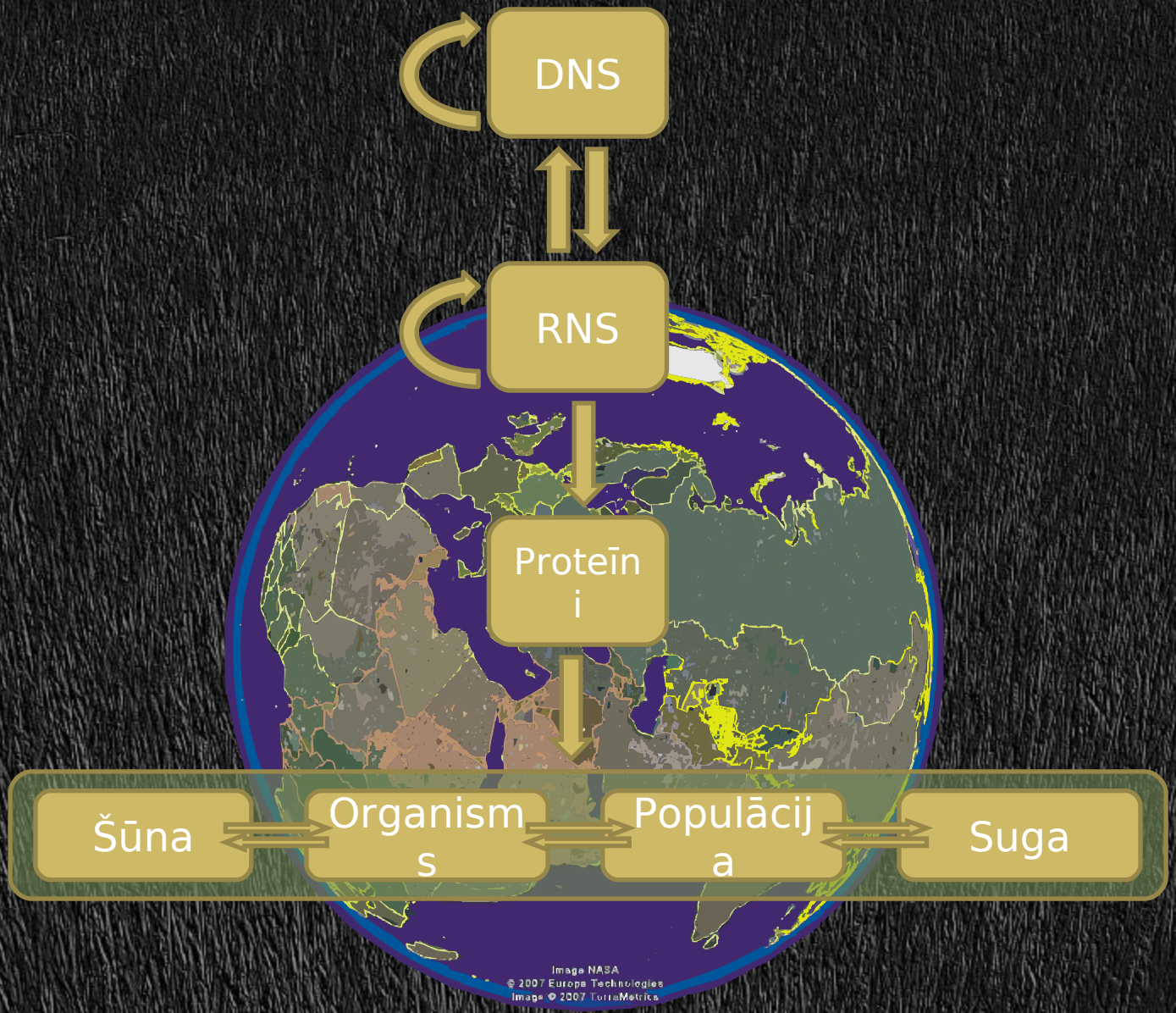
- Piemēram, priežu mežs Kolkā...

  Biometrija - saskaita priedes, nomēra to augstumu vai stumbra diametru

  Augu fizioloģija – audzē priežu stādus vidē ar kontrolētu sāls daudzumu un pēta to atbildes reakciju

  Bioinformātika - analizē priežu genomu un tā kodētos proteīnus

Bioloģiskās informācijas plūsma



Bioloģija, statistika, datorzinātnes, informāciju tehnoloģijas

1. Aprakstošā informācija par dzīvo dabu un vidi – ekoloģija, morfoloģija, taksonomija...
2. Eksperimentālā informācija par dzīvajiem organismiem - genomika, proteomika, transkriptomika...

Datu bāzes

IT infrastruktūra un DB programmēšana

Jaunu algoritmu izstrāde

Jaunu programmu izstrāde

Datu apstrādes un analīzes metodes – statistika, informātika, informāciju tehnoloģijas

Eksperimentālie dati un bioinformātika

- Eksperimentālā bioloģija ‘ražo’ datus, ko analizēt, uzglabāt un analizēt:

Genomika – DNS (genomu sekvences un to analīze)

Transkriptomika – RNS (gēnu ekspresija)

Proteomika – proteīni (proteīnu struktūra)

Metabolomika – šūnas metabolīti

- Bioinformātika apkopo un organizē šos datus, kā arī ļauj iegūt padziļinātas zināšanas par šo datu fundamentālo saistību

DNS sekvenēšana

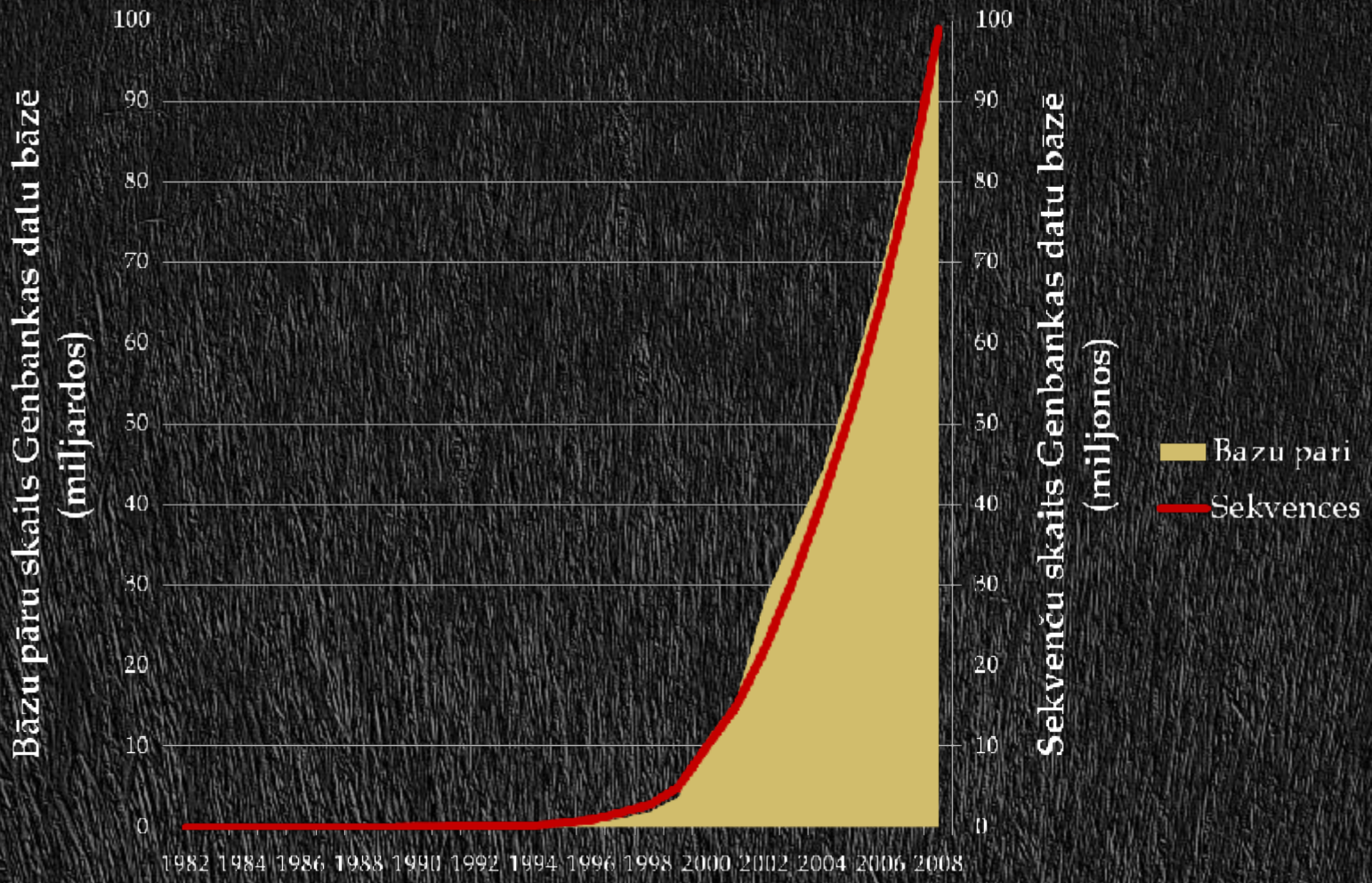
- Divas metodes 1975. - 1977. gadā (*Allan Maxam un Walter Gilbert, Frederick Sanger*)

“These chemical procedures ... soon allowed the entire sequence of the plasmid cloning vector pBR322 (4362 bp) to be worked out by a single scientist in only one year.”

Molecular Biology of the Gene IVth ed. 1987

DNS sekvence == DNS virkne

DNS sekvenču pieaugums GenBank



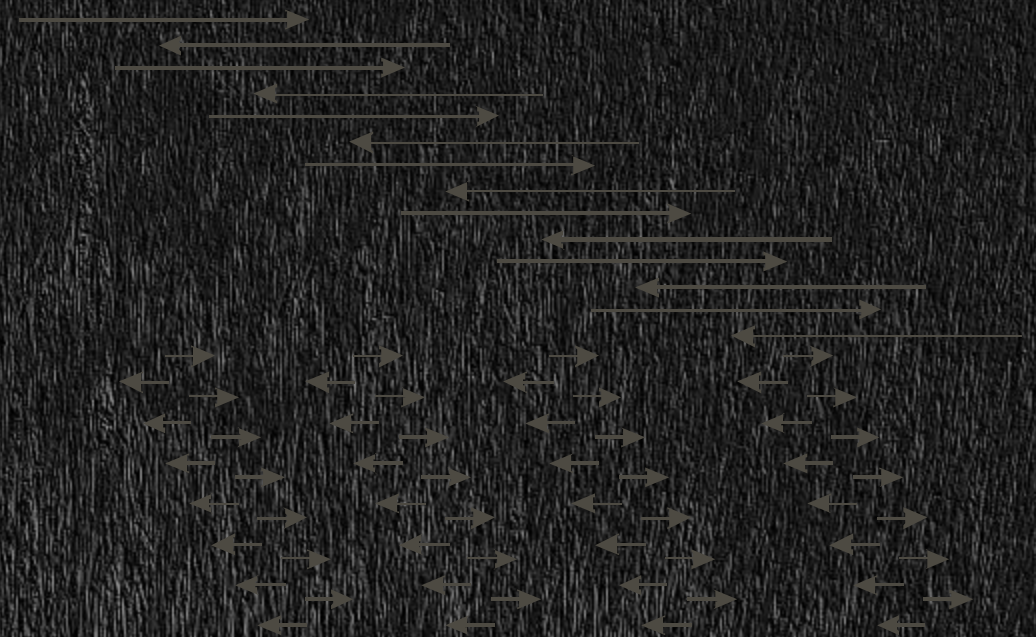
Genbankas attīstība 1982. - 2008. gada

Jaunas paaudzes sekvenēšanas tehnoloģijas

- Sangera sekvenēšana – industrijas standarts, pēc būtības nemainīgs no 1977. gada
- *Next generation sequencing* ir apkopojošs nosaukums vairākiem atšķirīgiem sekvenēšanas veidiem - 454 (Roche), ABI Solid, Illumina Solexa, Helicos HeliScope
- Pieļauj daudz straujāku genomu sekvenēšanu (un rada daudz vairāk datu)

Izaicinājumi bioinformātikai

- Izstrādāt jaunas metodes pilnu genoma sekvenču savākšanai
- Izlobīt jēgu no strauji pieaugošā genomu sekvenču daudzuma
- Salīdzināt daudzas genomu sekvences savā starpā (salīdzinošā genomika)

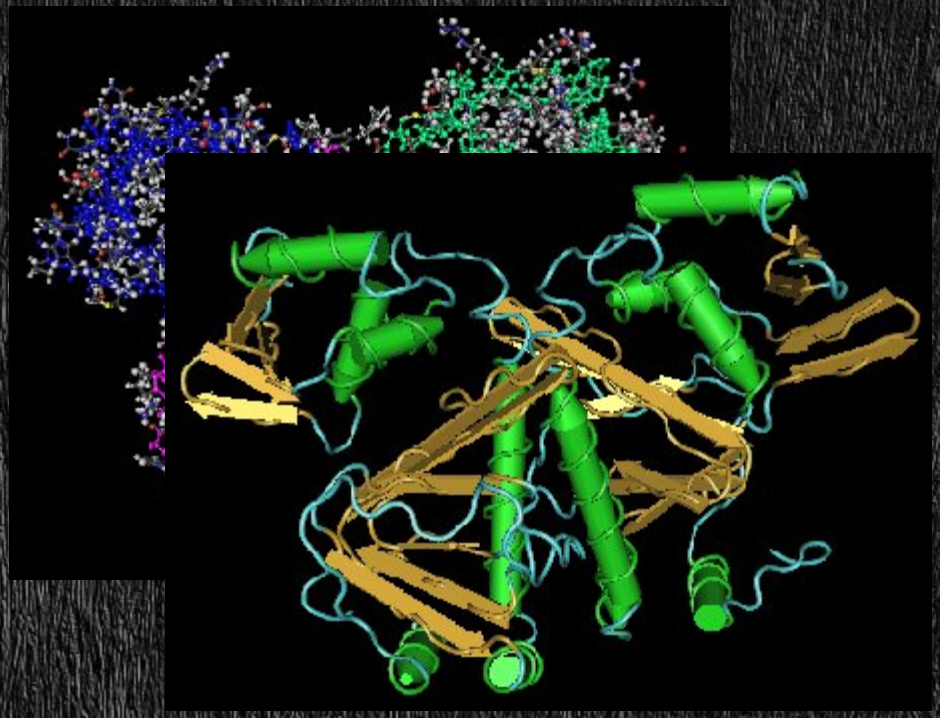


```
GAATTCCATTGCCAAA
 AATTCCATTGCCAAAT
  ATTCCATTGCCAAATT
   TTCCATTGCCAAATTC
    TCCATTGCCAAATTCG
     CCATTGCCAAATTCGC
      CATTGCCAAATTCGCA
       ATTGCCAAATTCGCAA
        TTGCCAAATTCGCAAA
         TGCCAAATTCGCAAAT
          GCCAAATTCGCAAATC
           CCAAATTCGCAAATCG
            CAAATTGGCAAATCGG

GAATTCCATTGCCAAATTGGCAAATCGG
```

Gēna X promoters
Gēns X

****
CCSTVNLVRPVNNIAHVNCGRCRTTLMYPHGAPSVKCAIC



Datu vizualizācija

Ģenētikas uzdevumi

- Noteikt pilnas genomu secības
- Identificēt visus gēnus un gēnu ekspresiju kontrolējošos elementus
- Noskaidrot visu gēnu regulāciju un funkcijas
- Izvērtēt ģenētisko daudzveidību, kas atrodama dažādu augu sugu populācijās
- Pētīt saikni starp alēlisko daudzveidību un gēnu funkciju

Augu pielietojums

- Pārtika
- Skābeklis un CO_2
- Ainava
- Enerģija
- Rūpniecības izejmateriāli

Production			
Agriculture 06 *		Pulpwd, Rnd & Split **	521,715
Cereals	2,227,980	Round Wood **	3,417,660
Meat	265,105	Log:Saw & Veneer **	970,481
Fruits & Vegetables	1,392,253	Sawn Wood **	415,553
Roots & Tubers	711,682	Wood Charcoal *	43,694
Pulses	61,706	Wood Fuel **	1,771,978
Oilseeds & Nuts *(oil eq.)*	146,353	Wood Pulp *	174,635
Sugar Crops	1,533,272	Wood-based Panels **	211,128
Tobacco	6,565	**Fisheries 04 ***	
Fibre Crops	28,782	Pelagic	40,664
Forestry 04		Freshwater	36,147
Indust.Roundwood **	1,645,681	Demersal	22,324
Paper, Paper Board *	354,489	Molluscs	16,786

* *1000Tonnes* ** *1000m³*

FAO www.fao.org

Augu īpatnības

- Ļoti atšķirīgi genomu izmēri, kas nav saistīti ar augu kompleksitāti
- Trīs genomi vienā šūnā – kodola, mitohondriju, hloroplastu
- Izplatīta poliploīdija
- Atšķirīgs pielietojums cilvēka vajadzībām – mehāniskās īpašības koksnei, barības vielu saturs un sastāvs lauksaimniecības augiem, sekundārie metabolīti medicīnas un garšaugiem

Augu genomi

- Astoņi pabeigti augu genomi

  Arabidopsis thaliana, *Glycine max*, *Medicago truncatula*, *Oryza sativa*, *Populus trichocarpa*, *Sorghum bicolor*, *Vitis vinifera*, *Zea mays*
- Daudzi genomu sekvenēšanas projekti

  Lotus, *Brachypodium*, kvieši, mieži, kartupeļi, tomāti
- Ģenētiskās kartes
- Gēnu resursi – expressed sequence tags

http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html

EST (gēnu) datubāze

dbEST: database of "Expressed Sequence Tags"

dbEST release 012210

Summary by Organism - January 22, 2010

Number of public entries: 64,679,625

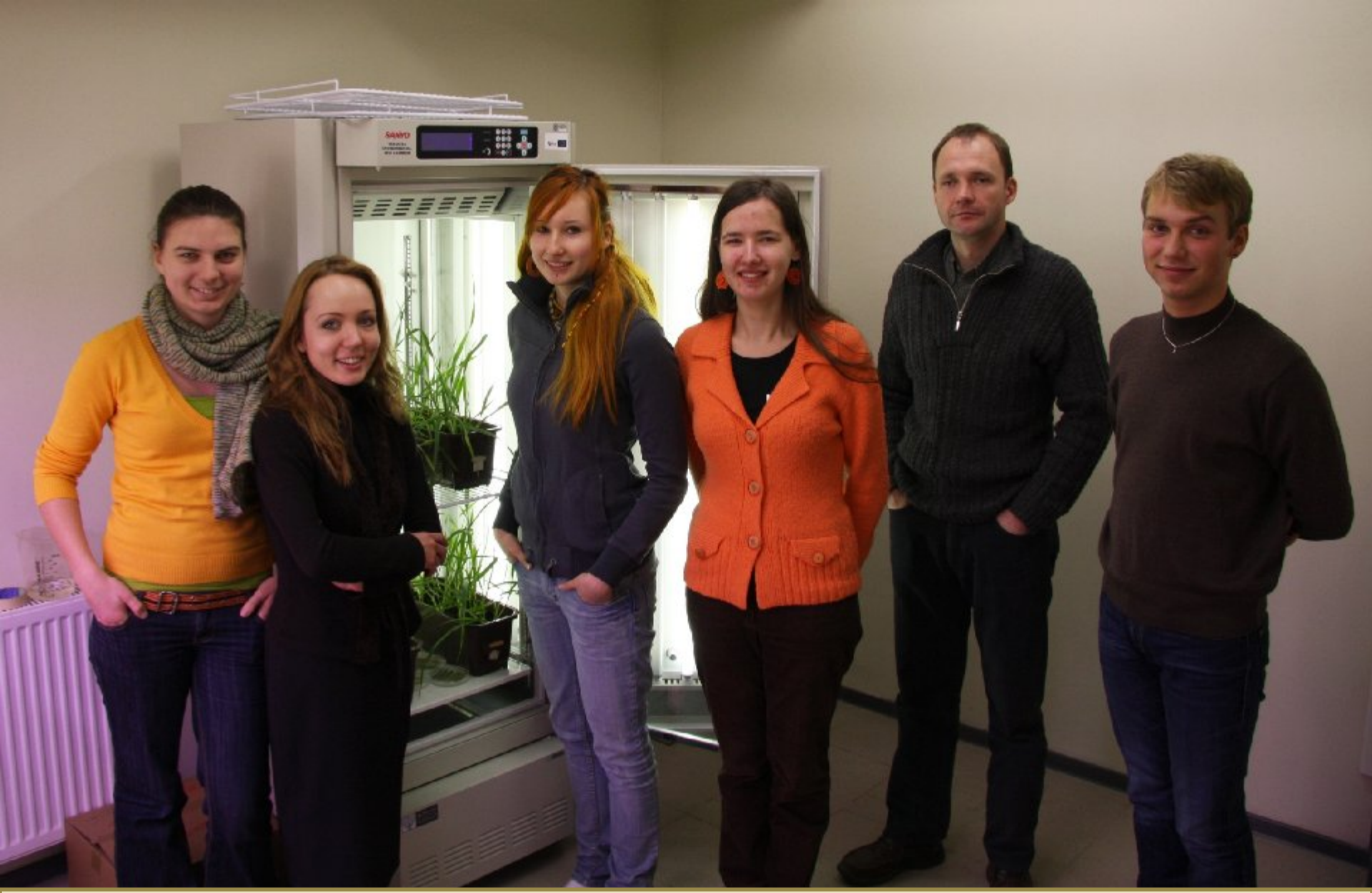
Homo sapiens (human)	8,300,249
Mus musculus + domesticus (mouse)	4,852,146
Zea mays (maize)	2,018,798
Bos taurus (cattle)	1,558,493
Sus scrofa (pig)	1,538,636
Arabidopsis thaliana (thale cress)	1,527,299
Danio rerio (zebrafish)	1,481,930
Glycine max (soybean)	1,422,982
Xenopus (Silurana) tropicalis (western clawed frog)	1,271,375
Oryza sativa (rice)	1,249,110
Ciona intestinalis	1,205,674
Triticum aestivum (wheat)	1,067,304
Rattus norvegicus + sp. (rat)	1,009,820

Bioinformātika Augu molekulārās ģenētikas grupā

- Nelieli sekvenēšanas projekti (viena vai vairāku gēnu salīdzinājums vienas sugas ietvaros)
- Genoma polimorfisma datu analīze
- Filoģenētiskā analīze
- Proteīnu struktūras paredzēšana, piemēram, transmembrānu domēni vai signālpeptīdi
- Iztiekam ar esošajām brīvpieejas programmām

Bioloģija un datorzinātnes

- Vēsturiski bioloģija ir aprakstoša zinātne, bet tas mainās un parādās nepieciešamība pēc datu bioinformātiskas analīzes
- Biologiem caurmērā ir vājas zināšanas programmēšanā, matemātikā un IT; ne īpaši labas statistikā
- Biologiem un datozinātniekiem ir savstarpēji papildinošas iemaņas – tātad potenciāls sadarbībai
-



Augu molekulārās ģenētikas grupa 2009
http://plantgenetics.lu.lv/