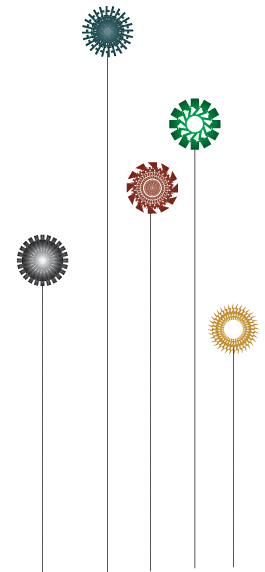


# Shared bioinformatics database within Unipro UGENE

BOSC  
July 11–12, 2014

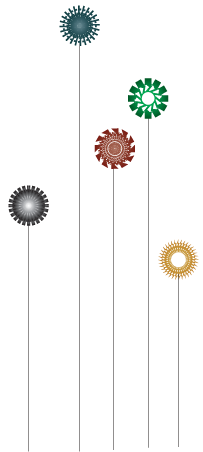
by Ivan Protsyuk

UniPro LLC  
Novosibirsk State University



# Agenda

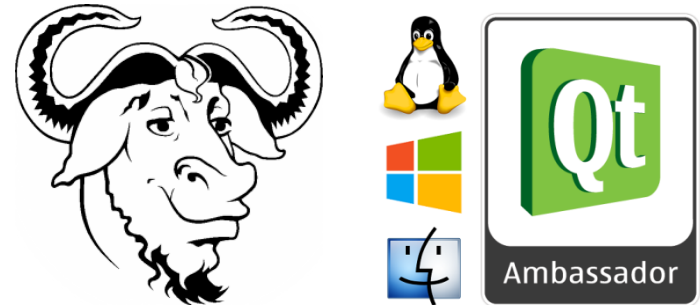
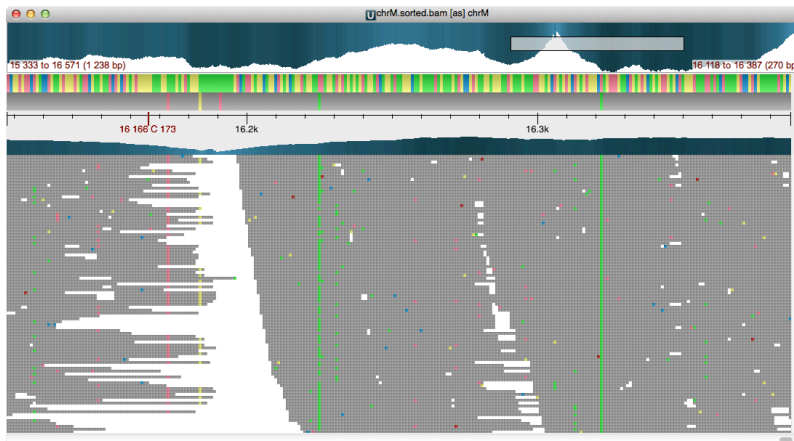
- About UGENE
- Shared databases in UGENE
  - generally
  - technically
  - numerically
- Prospects



# About UGENE

Dozens of algorithms:

- BLAST
- Multiple and pairwise alignment
- Short reads alignment
- And other...



- Visualization capabilities
- > 20 data formats
- Remote databases
- NGS framework



# Data Storage

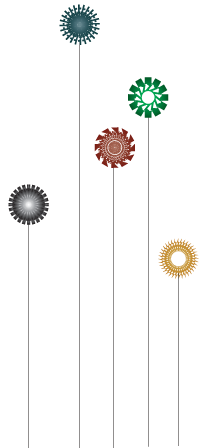
- Data can be big
- Simultaneous access



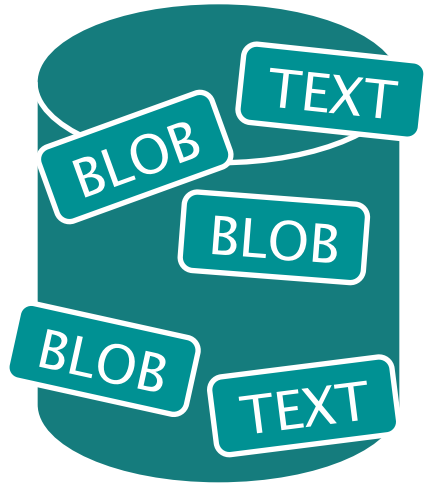
# Shared Bioinformatics Database



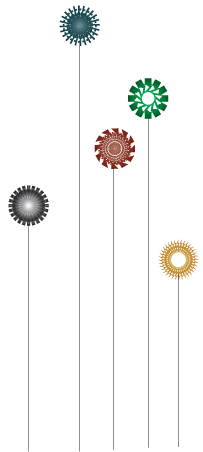
- Manipulations with shared objects
- Using folders to organize data
- Easy import/export of shared data
- Synchronization between users



# Existing Approaches



VS

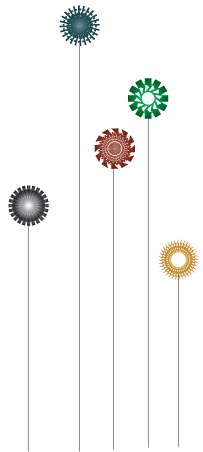


# UGENE Data Model



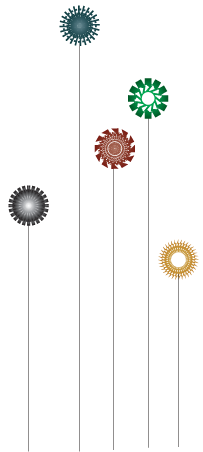
- Sequences
- Annotations
- Multiple alignments
- Genome assemblies

- Phylogenetic trees
- Chromatograms
- Protein tertiary structure
- Weight matrices
- Text



# Storage Capabilities

- Millions of biological objects
- NGS data supported
- Import 100 MBp ~ 1 min, but instant access later on



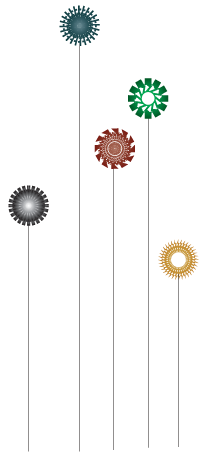
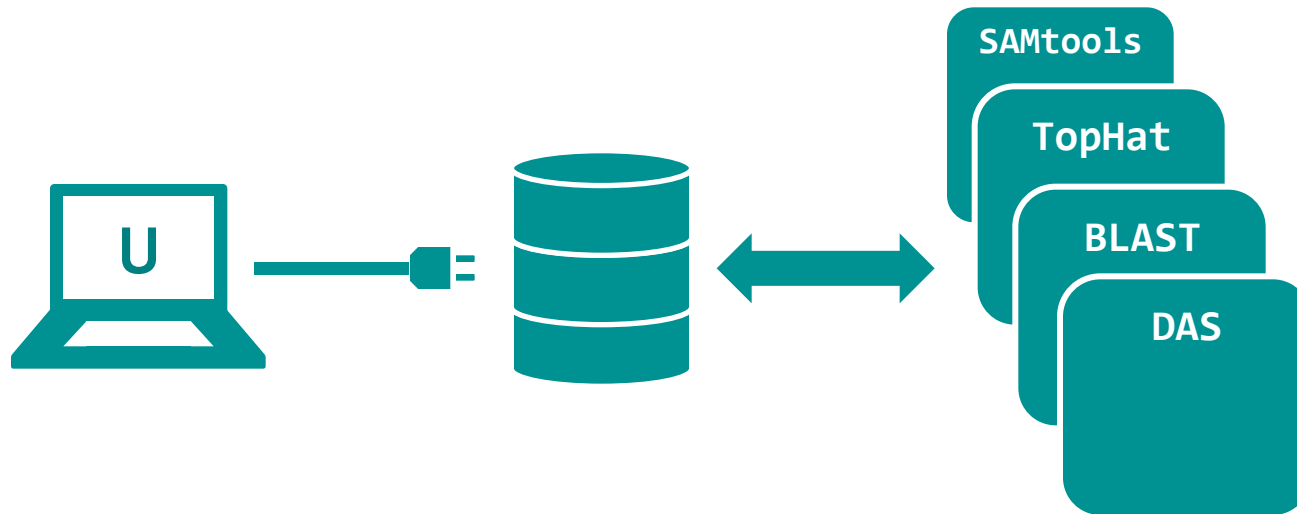


# Public Storage for Everyone



# Future Plans

- Performance improvements
- Distributed workflow system
- User requests



# Contacts

UGENE:

- <http://ugene.unipro.ru/shared-db>
- [ugene@unipro.ru](mailto:ugene@unipro.ru)



Ivan Protsyuk: [iprotsyuk@unipro.ru](mailto:iprotsyuk@unipro.ru)

# Graphical View

The screenshot displays the UGENE software interface. The main window shows a graphical view of a DNA sequence for NC\_016856. The sequence is displayed in a multi-line format with various annotations above and below it. The top track shows the sequence coordinates from 200k to 4.870265. Below the sequence, there are tracks for misc\_annotations, mobile\_element, ncRNA, rRNA, and tRNA. The sequence itself is shown in a multi-line format with a highlighted region in green. Below the sequence, there is a table of features.

Name	Value
Auto-annotations [Lab 42   NC_0168...	
NC_016856 features [Lab 42]	
CDS (0, 5315)	
gene (0, 5516)	
gene	190..255
gene	337..2799
gene	2801..3730
gene	3734..5020
gene	complement(5114..5887)
gene	complement(5966..7396)
gene	7665..8618
gene	8729..9319
gene	complement(9376..9942)
gene	complement(10092..10805)