Shared bioinformatics database within Unipro UGENE

BOSC
July 11–12, 2014

by Ivan Protsyuk
UniPro LLC
Novosibirsk State University
• 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
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
Future Plans

- Performance improvements
- Distributed workflow system
- User requests
Contacts

UGENE:
• http://ugene.unipro.ru/shared-db
• ugene@unipro.ru

Ivan Protsyuk: iprotsyuk@unipro.ru
Graphical View