



## UGENE Quick Start Guide



# Quick Start Guide

This document contains a quick introduction to UGENE. For more detailed information, you can find the UGENE User Manual and other special manuals in project website: <http://ugene.unipro.ru>.

## What is UGENE?

### What you can do in UGENE?

- View, edit and annotate DNA, RNA and protein sequences
- View, edit and align multiple sequence alignments
- View, align and assemble short reads
- 3D structures and surface algorithms
- Workflow Designer: pipelines and repeatable experiments

## What is UGENE?

**Unipro UGENE** is a free cross-platform genome analysis suite. It is distributed under the terms of the GNU General Public License. To learn more about UGENE visit [UGENE website](#). It works on Windows, Mac OS X or Linux and requires only a few clicks to install.

### Key Features

- Powerful viewers and editors for DNA and protein sequences, multiple sequence alignments, chromatograms, 3D-structures, short read assemblies and phylogenetic trees.
- Support for dozens data formats, set of tools for importing and exporting data.
- Visual designer, runner and debugger for computational workflows from over 100 embedded algorithmic elements.
- Reusable computational workflows that can be shared, published and read as plain text by everyone.
- Full featured command line only version for remote desktops.
- Automatic detection and optimization for multi-core hardware, GPGPU, special instruction sets available on user's PC
- Setting up UGENE package user automatically gets dozens of tools installed and configured for immediate use

### Tools provided with UGENE installation

- BLAST, Smith-Waterman - popular basic local sequence alignment tools.
- HMMER2/HMMER3 - sequence analysis using profile hidden Markov Models constructed from multiple sequence alignments.
- ClustalW, ClustalO, Muscle, K-Align, Mafft, T-Coffee multiple sequence alignment algorithms.
- Repeat finding and sequence comparison algorithms with a convenient dot-plot based results visualization.
- Open reading frames visualization and export.
- Restriction sites visualization tool with included version of REBASE database.
- Transcription factor binding sites analysis with SITECON, weight-matrix based algorithms and included version of JASPAR database.
- Short read aligners: BWA, Bowtie, Bowtie2, UGENE Aligner.
- GC-content, AT-content graphs.
- Mr bayes and Phylip tools for phylogenetic trees construction.
- ... and dozens of other popular and highly cited tools.

### Hardware requirements

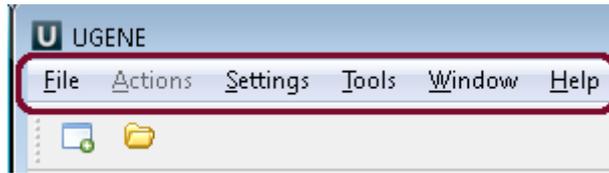
- UGENE runs smoothly on almost all hardware available: single core Pentium III or higher computer with 200Mb of free disk size and 64Mb memory is enough.
- Use UGENE on Mac, Windows or Linux operating systems. Both 32 and 64 bit versions are available.
- No 3rd-party software is needed to be installed to use UGENE!

### Customization and cooperation

- UGENE is opensource software: it's easy to customize UGENE, add new features and tools!
- UGENE team welcomes all developers and researchers to participate in joined projects to improve UGENE or create customizations needed by user!

## What you can do in UGENE?

When open UGENE for the first time, there is an empty UGENE window with the main menu bar on the top of the window:

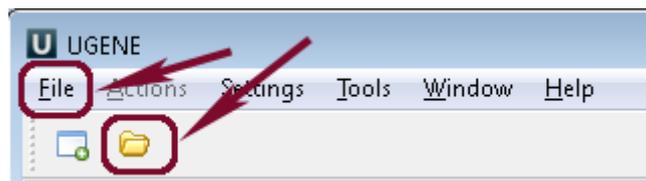


Using these menus you can run many algorithms, configure different settings and get help:

Menu	Description
<b>File</b>	A set of project/file level operations.  <b>Example:</b> create, open, etc. a project; open a document; access remote database to download a file.
<b>Actions</b>	Various actions associated with the active window.  <b>Example:</b> export, remove, edit, analyze a sequence using different plugins (for the <i>Sequence View</i> ); edit, align, change the consensus mode (for the <i>Alignment Editor</i> ).
<b>Settings</b>	Application, plugins and tools settings.
<b>Tools</b>	Various tools, independent of an active window. This menu is extended by different plugins.  <b>Example:</b> HMMER2 / HMMER3 tools, SITECON, Workflow Designer.
<b>Window</b>	A list of active windows and basic manipulations with the windows.  <b>Example:</b> close all windows, tile windows, select next window.
<b>Help</b>	Application help and check for updates.

The menus can be dynamically populated with new actions added by plugins.

You can select and analyze different files such as sequences, multiple sequence alignments, short reads assemblies and etc. To open a file click on the *File*-->*Open* or click on the *Open* icon and choose the file:



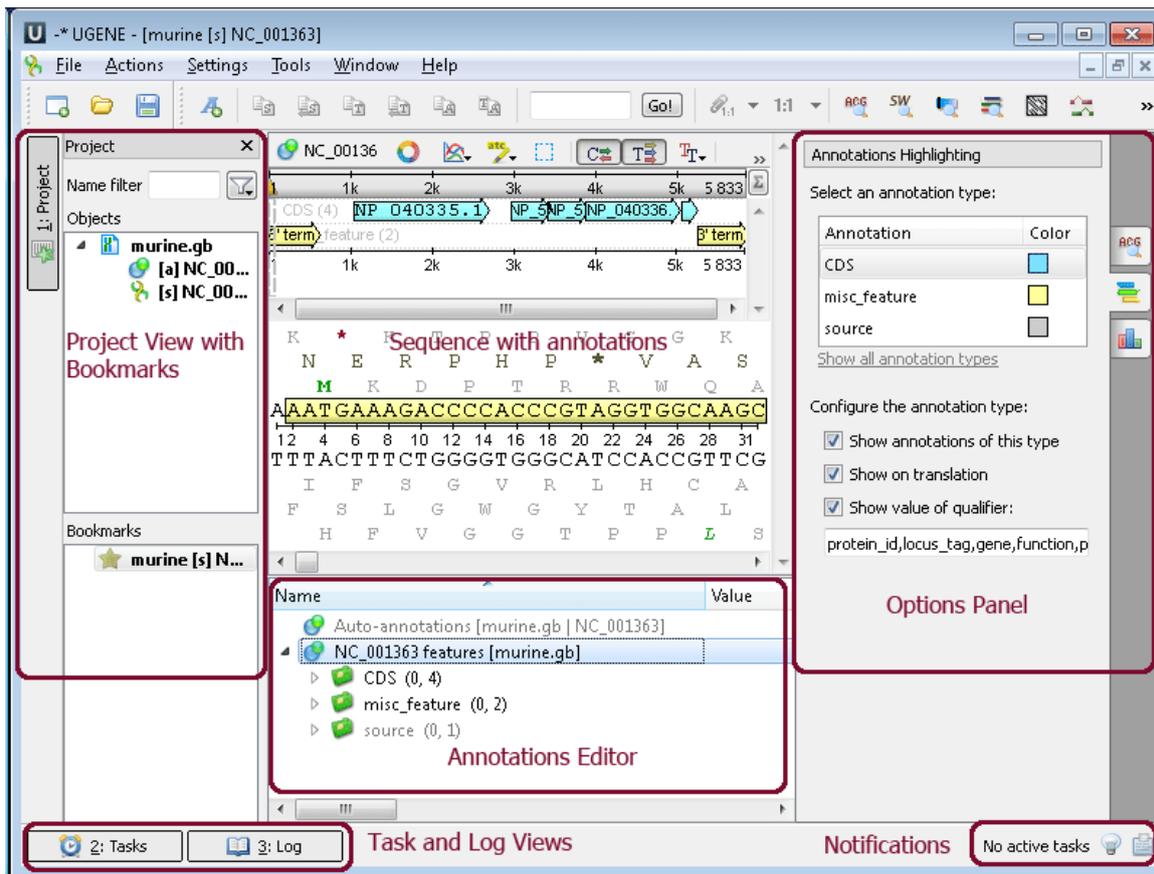
You can use UGENE sample files. There are the following sample data: ABIF, ACE, Assembly, CLUSTALW, EMBL, FASTA, FASTQ, Genbank, GFF, HMM, MMDB, MSF, Newick, PDB, Raw, SCF, Stockholm, Swiss-Prot. These files are located in the main UGENE directory: *ugene/data/samples*.

- View, edit and annotate DNA, RNA and protein sequences
- View, edit and align multiple sequence alignments
- View, align and assemble short reads
- 3D structures and surface algorithms
- Workflow Designer: pipelines and repeatable experiments

## View, edit and annotate DNA, RNA and protein sequences

**DNA, RNA or protein sequences (*Sequence View*).** The *Sequence View* is one of the major object views in UGENE aimed to visualize and edit DNA, RNA or protein sequences along with their properties like annotations, chromatograms, 3D models, statistical data, etc. For each le

UGENE analyzes the file content and automatically opens the most appropriate view. To activate the *Sequence View* open any file with at least one sequence. For example you can use the `$ugene/data/samples/Genbank/murine.gb` file provided with UGENE. After opening the file in UGENE the *Sequence View* window appears:



*Sequence View* - an object view aimed to visualize DNA, RNA or protein sequences along with their properties like annotations, chromatograms, 3D models, statistical data, etc.

*Project View* - a visual component used to manage active project and bookmarks.

*Annotations Editor* - contains tools to manipulate annotations for a sequence.

*Options Panel* - it is the panel with different information tabs and tabs with settings for Sequence View.

*Task View* - shows active tasks, for example, algorithms computations.

*Log View* - shows the program log information.

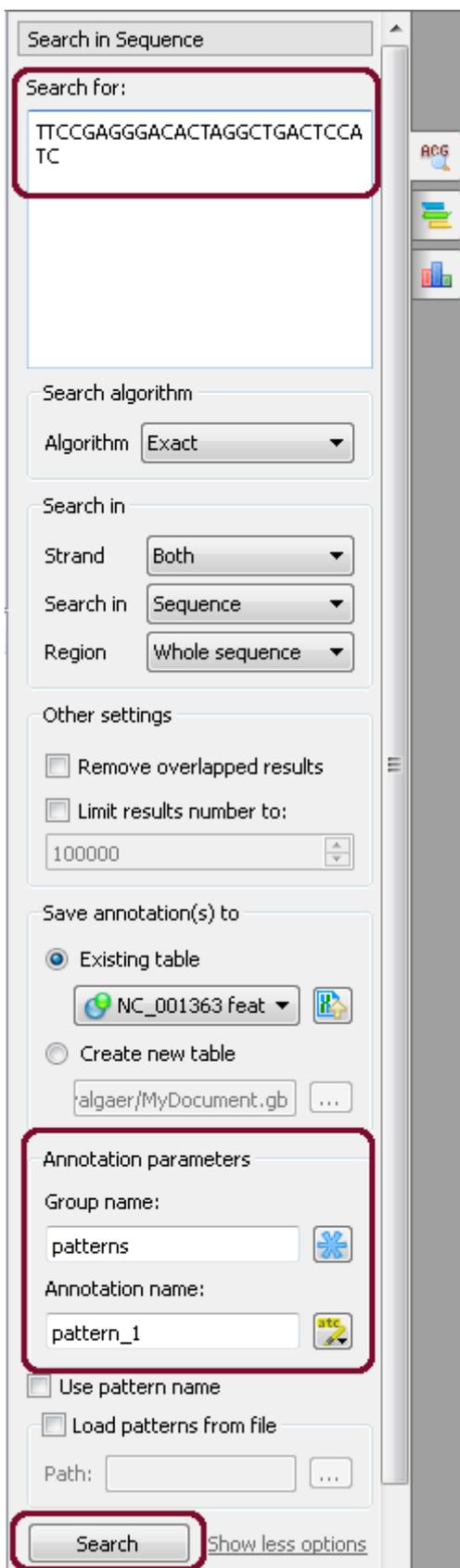
*Notifications* - shows notifications for task reports.

After the view is opened you can see a set of new buttons in the toolbar area. The actions provided by these buttons are available for all sequences opened in the view. These actions also available from the context menu. Many instruments and algorithms are available:

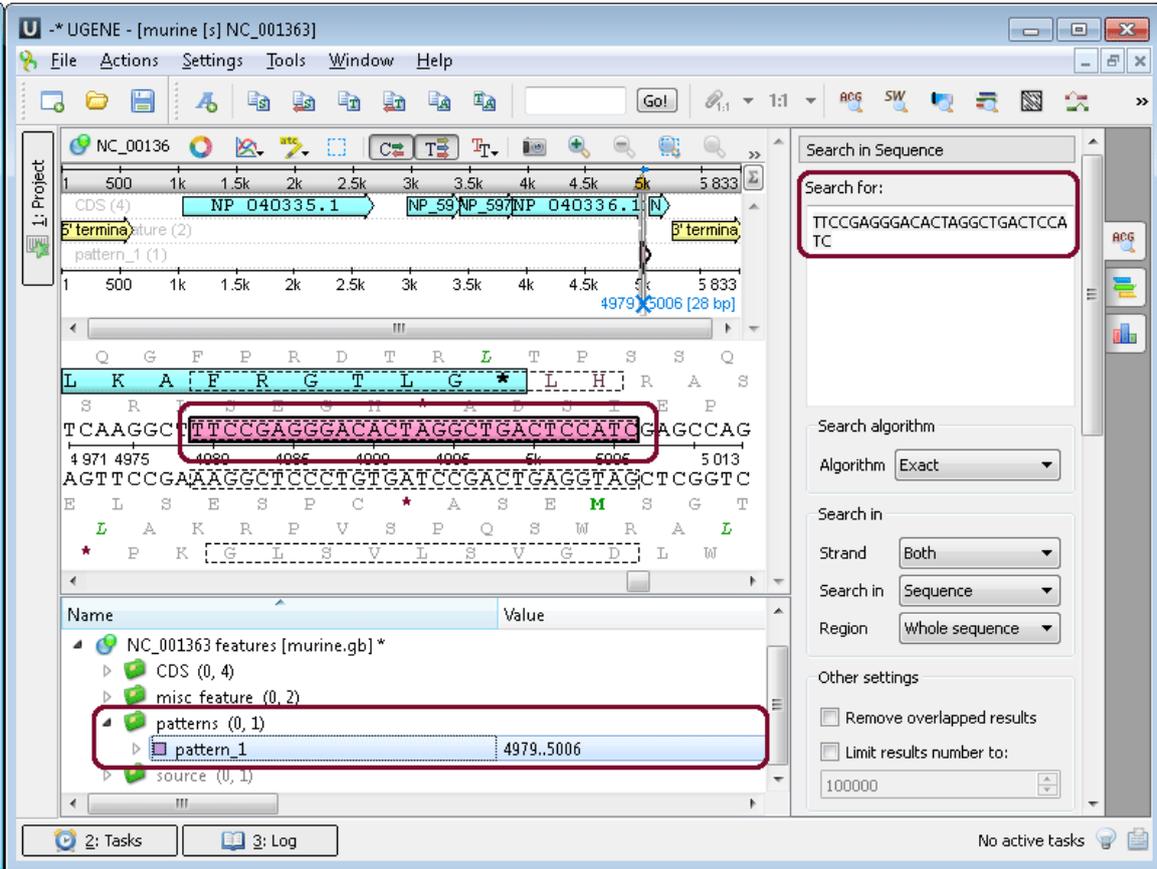
- Find pattern
- Find ORFs
- Find annotated regions
- Build dotplot
- Find repeats
- Find tandems
- Find restriction sites
- Primer 3
- Find high DNA flexibility regions
- ...

**Example 1:** Finding patterns in your sequence. Do the following steps:

Open the `ugene/data/samples/murine.gb` by *File*→*Open* menu, for example. Sequence View with murine.gb opens. Select the *Search in Sequence* tab of the *Options Panel*. Click *Show more options* and more options appear. Insert, for example "TTCCGAGGGGACACTAGGCTGACTCCATC" pattern into *Search for:* field and choose annotation parameters. For example as in the picture below:

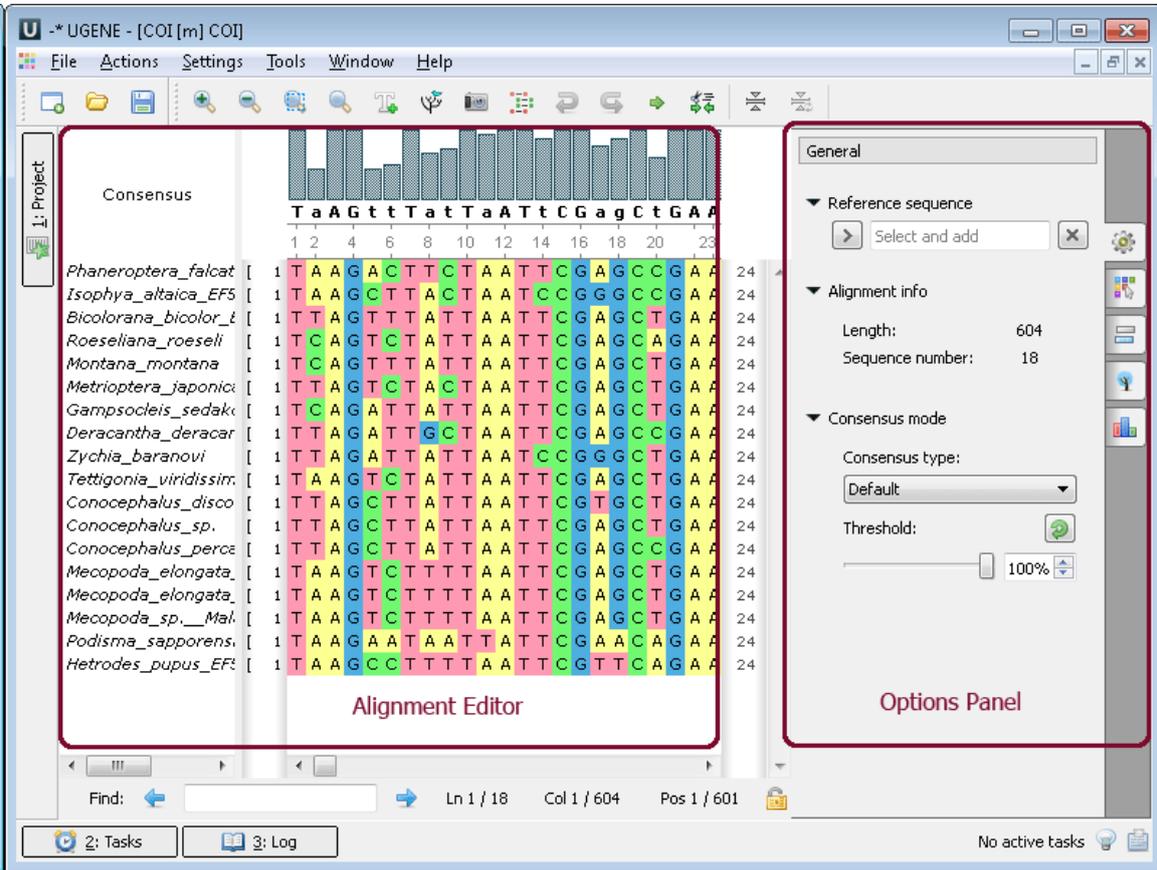


After that click the *Search* button. If the pattern there is or there are in the sequence it appears as new annotation(s):



### View, edit and align multiple sequence alignments

**Multiple sequence alignment (MSA Editor).** The *Alignment Editor* is a powerful tool for visualization and editing DNA, RNA or protein multiple sequence alignments. To activate the *Alignment Editor* open any alignment file. For example you can use the \$ugene/data/samples/C LUSTALW/COI.aln file provided with UGENE. After opening the file in UGENE the *Alignment Editor* window appears:



The editor supports different multiple sequence alignment (MSA) formats, such as ClustalW, MSF and Stockholm. The editor provides interactive visual representation which includes:

- Navigation through an alignment;
- Optional coloring schemes (for example Clustal, Jalview like, etc.);
- Flexible zooming for large alignments;
- Export publication-ready images of alignment;
- Several consensus calculation algorithms.

Using the *Alignment Editor* you can:

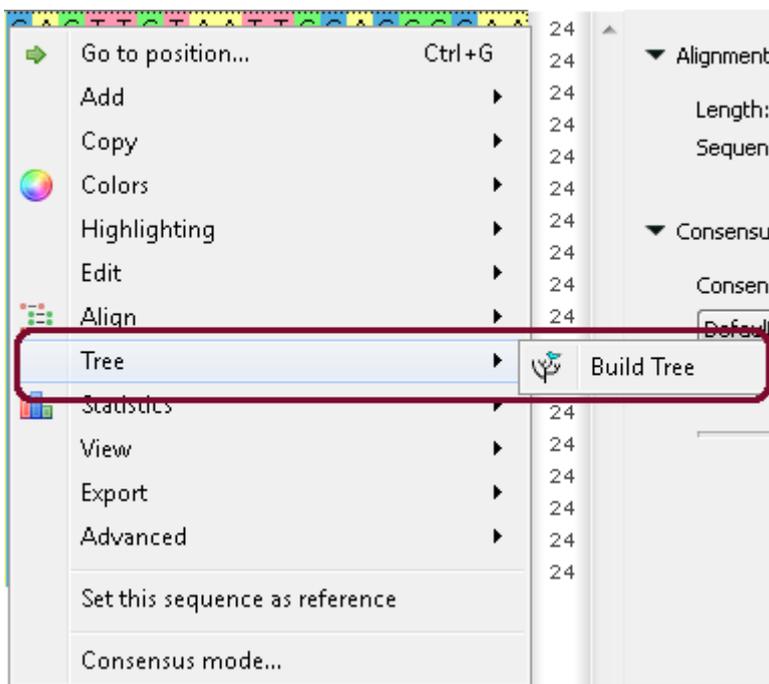
- Perform multiple sequence alignment using integrated MUSCLE and KAlign algorithms;
- Edit an alignment: delete/copy/paste symbols, sequences and subalignments;
- Build phylogenetic trees;
- Generate grid profiles;
- Build Hidden Markov Model profiles to use with HMM2/HMM3 tools.

**Example 2:** Build a tree from your alignment. You can do this by three different ways:

a. From the toolbar. Click to the tree icon:



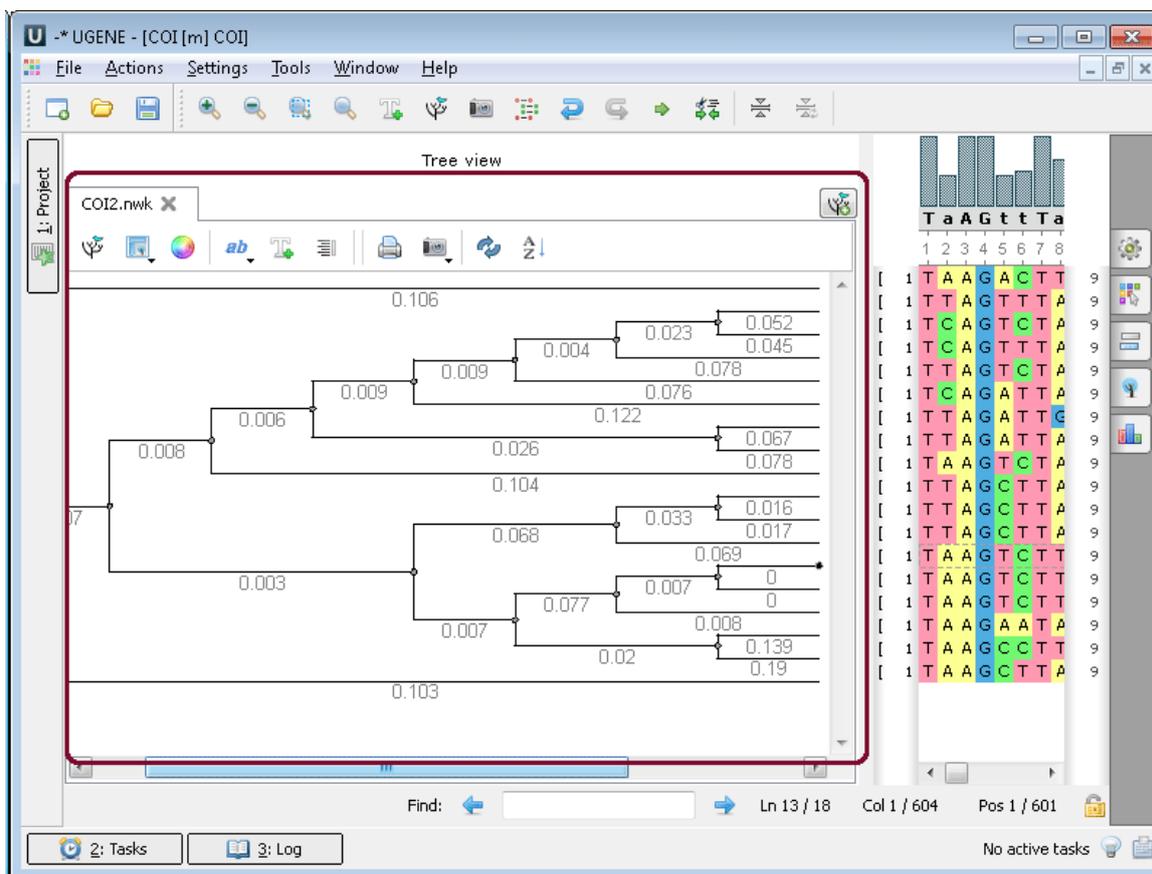
b. From the context menu:



c. From the Options Panel:

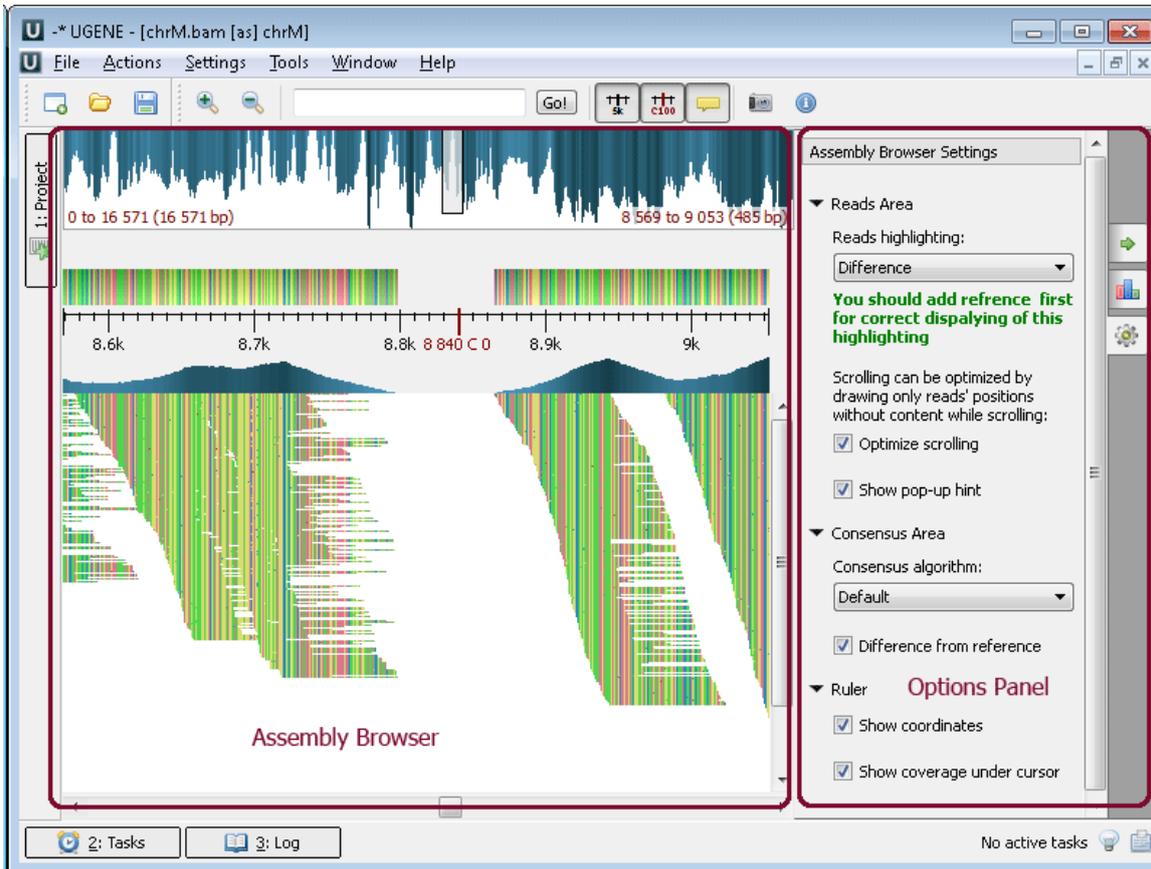


After calculation the tree appears in the MSA Editor in a separate window:



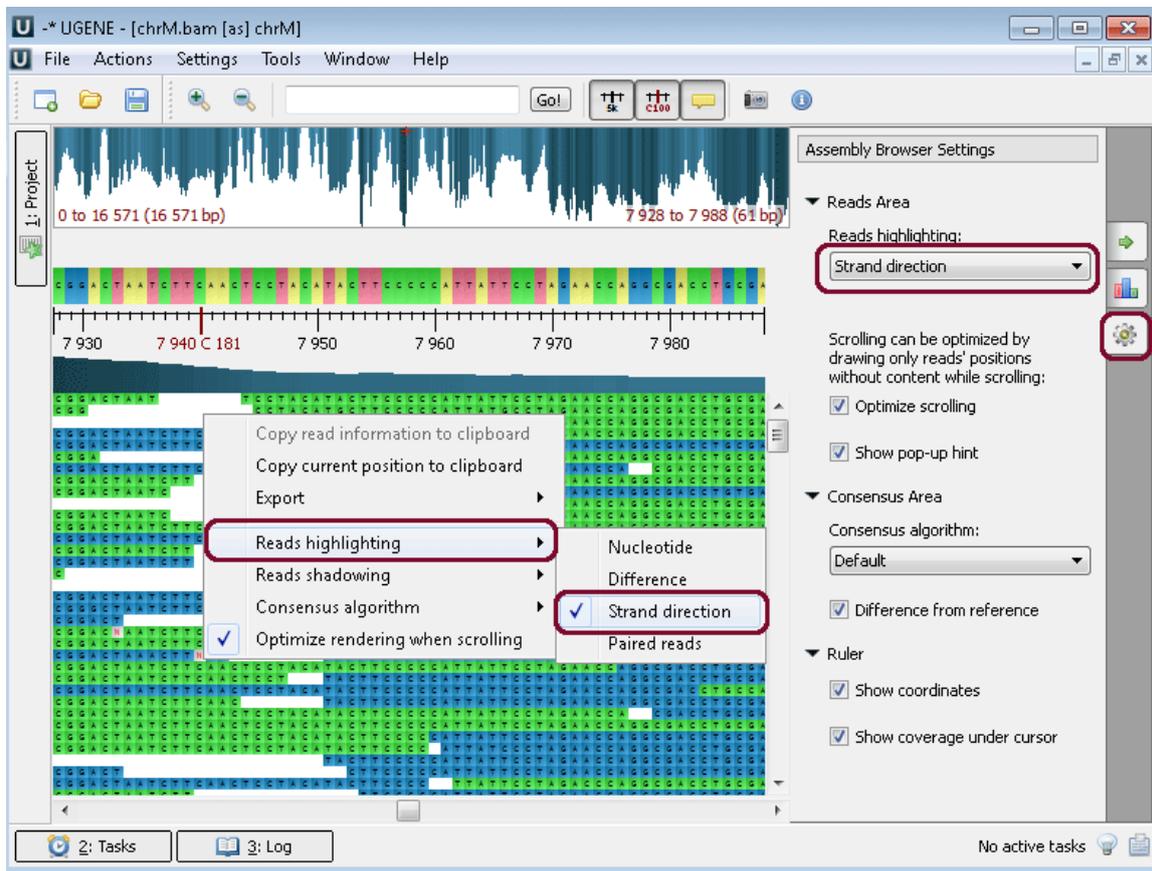
## View, align and assemble short reads

**SAM/BAM files (Assembly Browser).** *Assembly Browser* is used to visualize and efficiently browse large next generation sequence assemblies. Currently supported formats are SAM (Sequence Alignment/Map) and BAM, which is a binary version of the SAM format. Both formats are produced by SAMtools and described in the following specification: SAMtools. To activate the *Alignment Editor* open any assembly file. For example you can use the `$ugene/data/samples/Assembly/chrM.bam` file provided with UGENE. After opening the file in UGENE the *Assembly Browser* window appears:



Using the *Assembly Browser* you can: browsing and zooming assembly, getting information about reads, short reads visualization, associating reference sequence, consensus sequence, exporting.

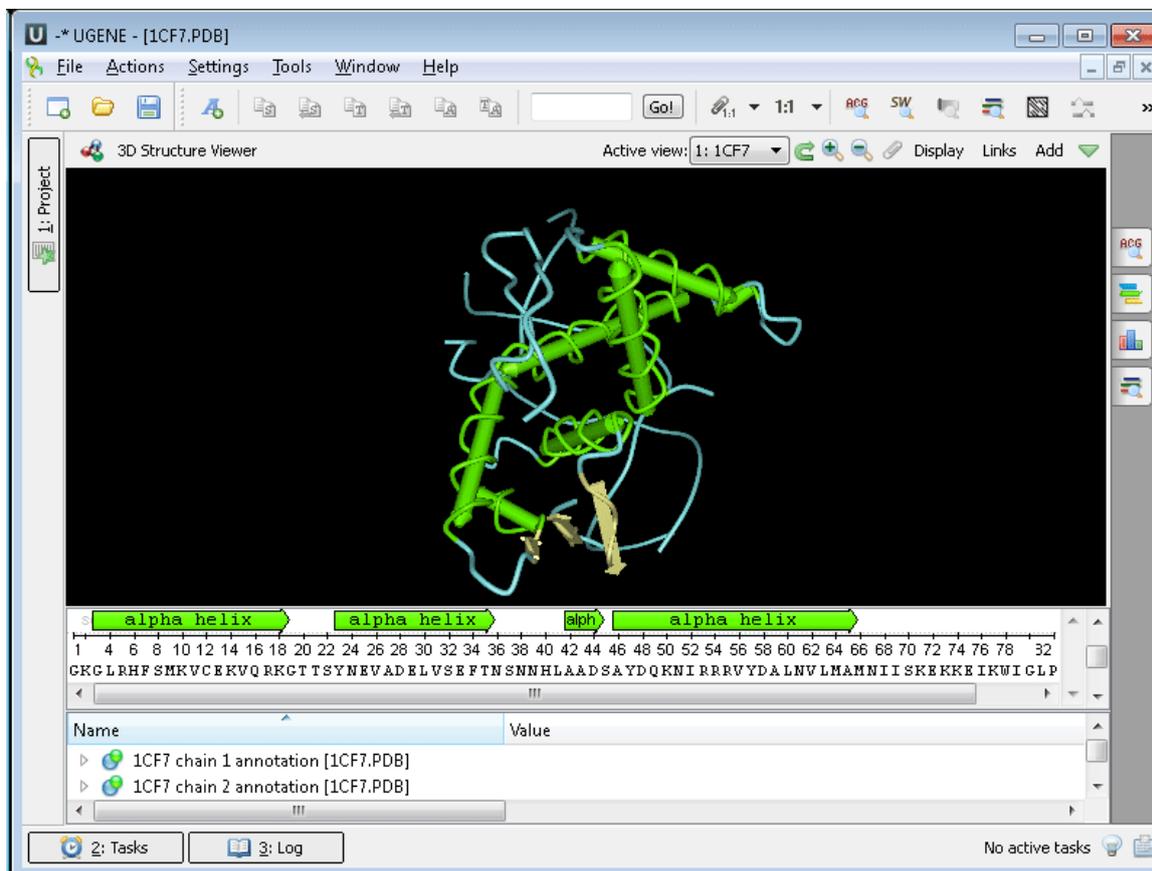
**Example 3:** Highlighting the strand of reads. You can do this using the context menu or the Options Panel.



## 3D structures and surface algorithms

**3D structures (3D Structure Viewer).** The *3D Structure Viewer* is intended for visualization of 3D structures of biological molecules. Using the *3D Structure Viewer* you can work with data from the Protein Data Bank (PDB) - a repository for the 3D structural data of large biological molecules, such as proteins and nucleic acids, maintained by the Worldwide Protein Data Bank (wwPDB). You can work as well with data from the NCBI Molecular Modeling DataBase (MMDB), also known as "Entrez Structure", a database of experimentally determined structures obtained from the RCSB Protein Data Bank.

The *3D Structure Viewer* is opened automatically when you open a PDB or MMDB file. For example, open `$ugene/data/samples/PDB/1CF7.PDB`. The *3D Structure Viewer* adds a view to the upper part of the *Sequence View*.



Using the *3D Structure Viewer* you can:

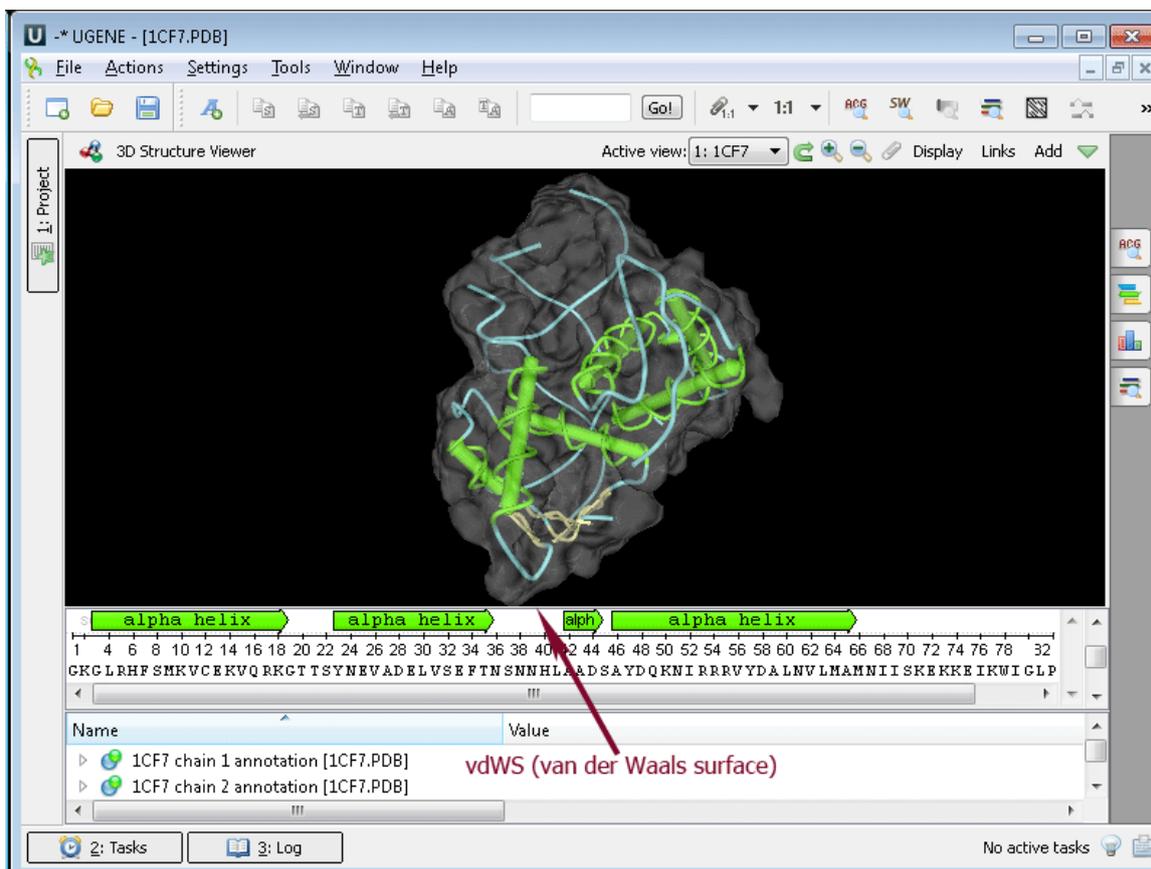
- Changing 3D Structure Appearance;
- Moving, zooming and spinning 3D structure;
- Highlighting sequence region;
- Selecting models to display;
- Exporting 3D structures image;
- Working with several 3D structure views.

**Example 5:** Calculating Molecular Surface. To calculate the molecular surface of a molecule select the *Molecular Surface* item in the *3D Structure Viewer* context menu or in the Display menu on the toolbar and check one of the following items:

- SAS (solvent-accessible surface)
- SES (solvent-excluded surface)
- vdWS (van der Waals surface)

To remove the molecular surface that has already been calculated select the O item. You can also select the *Molecular Surface Render Style* to modify the calculated molecular surface appearance:

- Convex Map
- Dots

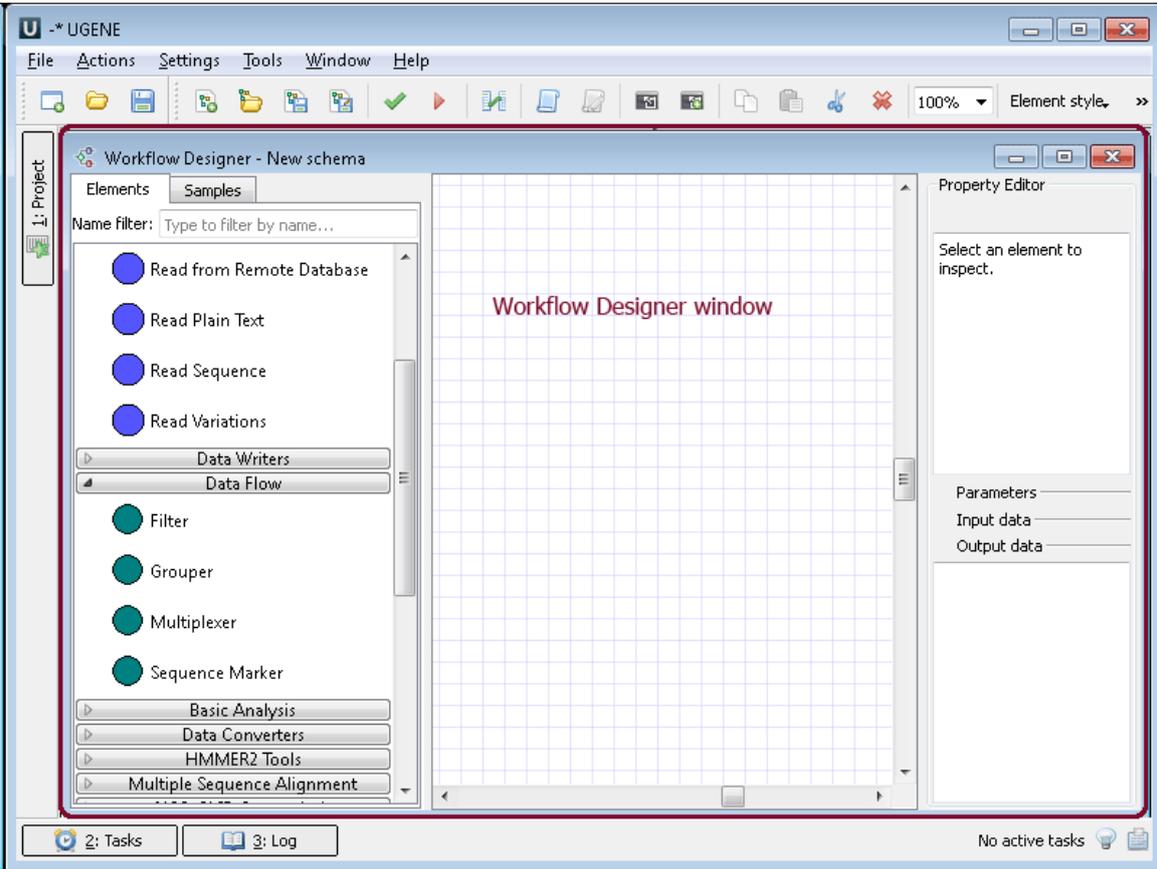


## Workflow Designer: pipelines and repeatable experiments

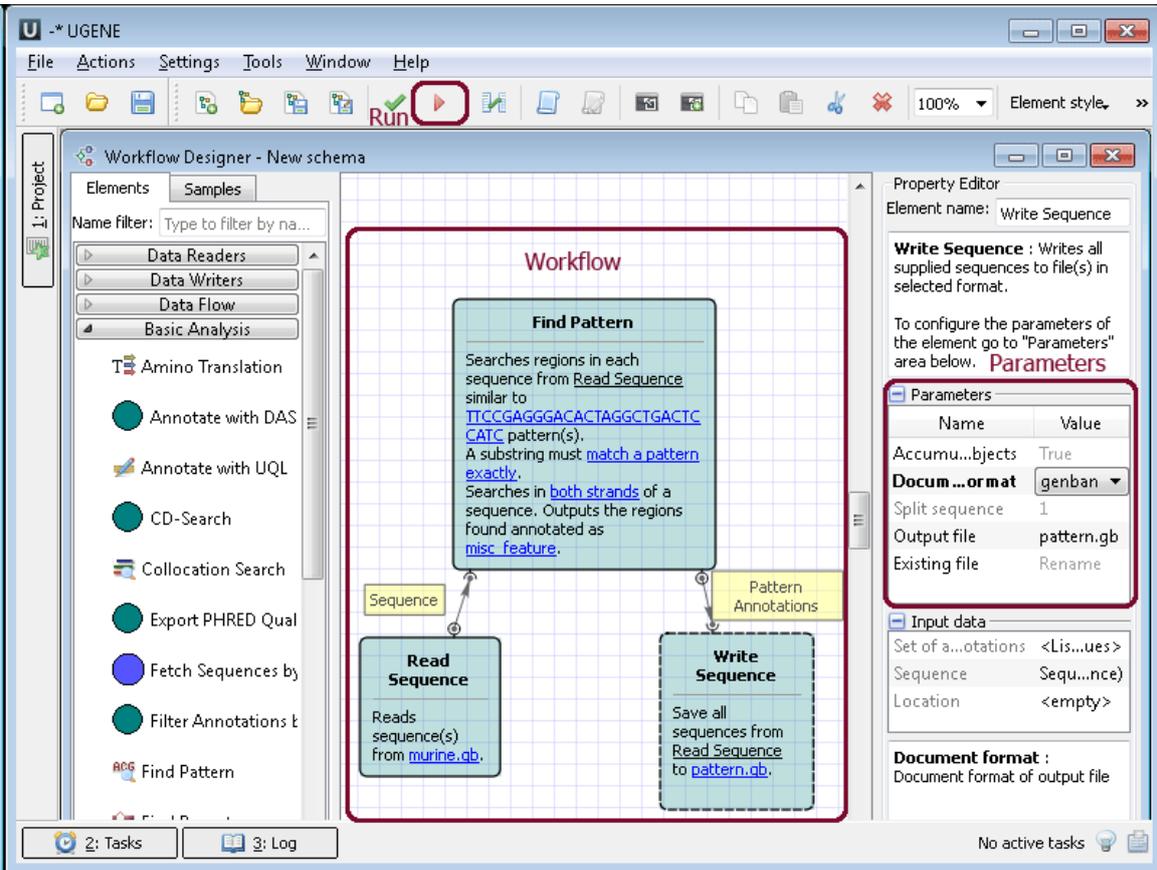
**Workflow Designer.** UGENE *Workflow Designer* is a central part of UGENE that allows a molecular biologist to create and run complex computational workflows even if he or she is not familiar with any programming language.

A workflow comprises reproducible, reusable and self-documented research routine, with a simple and unambiguous visual representation suitable for publications. A workflow can be run both locally and remotely, either using graphical interface or launched from the command line. Elements in workflow correspond algorithms integrated into UGENE. Additionally you can create custom workflow elements using integrated scripting language, or by connecting arbitrary external command line utility.

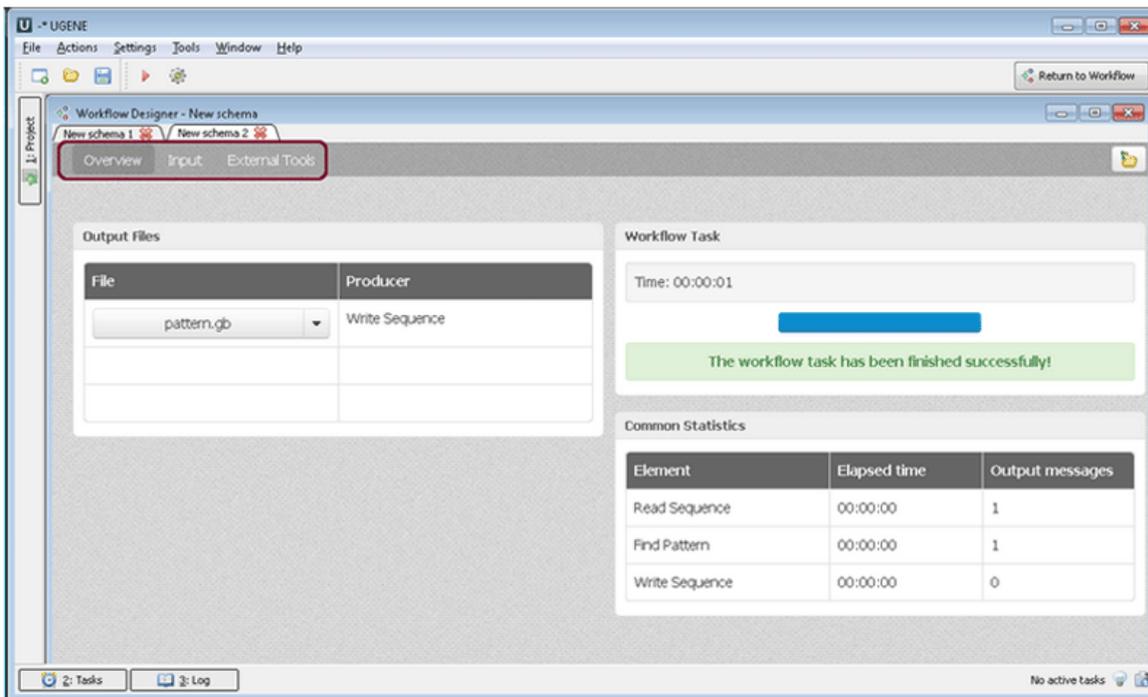
To launch the Workflow Designer select the *Tools Workflow Designer* item in the UGENE main menu. The following window appears:



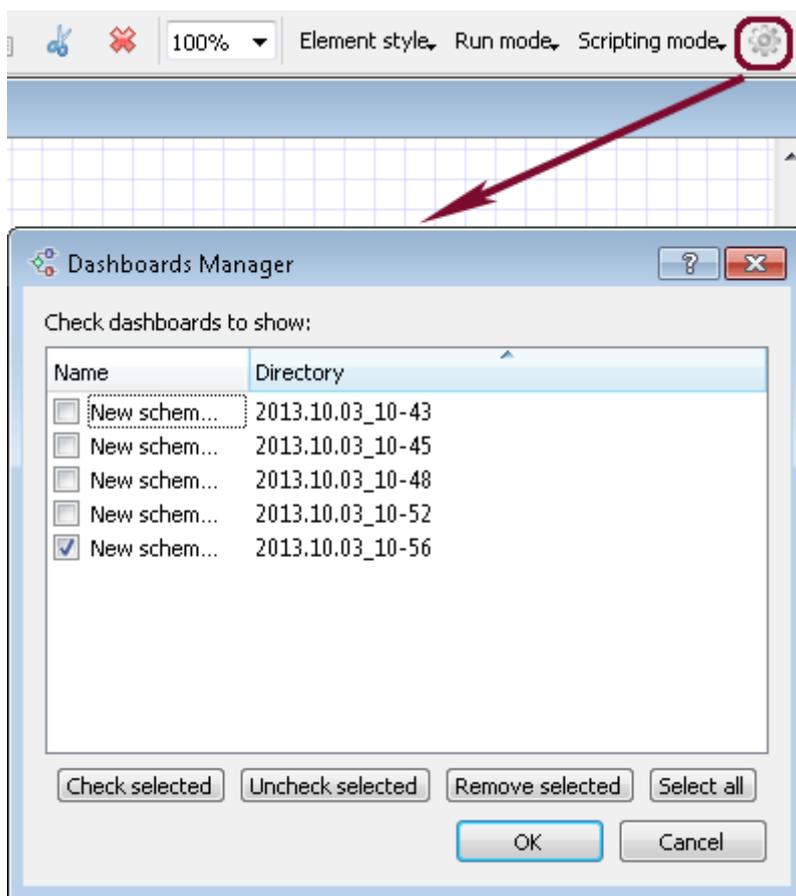
Example 4: You can find pattern in a sequence or in sequences and save it as annotations using the following workflow:



Create the workflow, choose parameters and click the Run button. If you want to search pattern in many sequences you can add these sequences into Read Sequence element. After the end of the running process a report appears. The report include all information about workflow.



All your workflows have been saved and you can navigation between it and use it with a help of the *Dashboards Manager*:



Note that workflows in UGENE are easy to read and share, can be reused multiple times and compiled into a separate standalone command line tools!

For more detailed information about Workflow Designer use the [Workflow Designer Documentation](#).