UGENE – A practical approach for complex computational analysis in molecular biology

Mikhail Fursov, Alexey Varlamov
mfursov@unipro.ru, varlax@unipro.ru
Center of Information Technologies "UniPro", Novosibirsk, Russia

UGENE (http://ugene.unipro.ru) is an open-source bioinformatics project. The ultimate goal of the project is to integrate popular bioinformatics tools and algorithms within a single visual interface that can be easily used by molecular biologists.

UGENE version v1.4 code base contains about 20 different plugins, each of them representing one of popular bioinformatics algorithms or methods. These include multiple alignment tools, Smith-Waterman algorithm implementation, HMMER2 tools, repeats and ORF analysis, restriction enzymes markup, search for transcription factor binding sites, integration with web databases like BLAST and CDD. For the complete list of plugins check the project web site: http://ugene.unipro.ru/plugins.html

In the next (1.5) version of UGENE we plan to support Primer3 package, add integrated BLAST tools and two secondary structure prediction algorithms: GOR and PSIPRED.

The key advantage of UGENE is a complete integration of the algorithms within a single visual interface. After installation UGENE does not require any additional software to be installed. From the developer's perspective it's important that all algorithms are refactored to use universal data model and common threading API.

All of the algorithms in UGENE are tuned to utilize multi-core environment. For example, MUSCLE multiple sequence alignment package embedded into UGENE is able to utilize multi-core environment, while the original is single threaded. HMMER2 package contains special optimizations for IBM Cell Broadband platform, platforms with Altivec and SSE2 instruction sets. With the next version UGENE will also add GPU based version of HMMER2 algorithm.

To ensure that integration of the 3rd party algorithms is correct the project maintains test base. The compatibility test base for UGENE v1.4 contains about 1200 tests. The v1.5 test base will have more than 2000 tests.

The main component of the visual interface of UGENE is W*orkflow Designer*. The Workflow Designer is used for construction of computational diagrams from the predefined set of visual algorithmic blocks, or processes. The key idea of Workflow Designer is to make the process of automation of routine tasks as simple as it's possible and make it available for non-programmers. The following page contains detailed description of Workflow Designer data model and interface: http://ugene.unipro.ru/plugin_workflow.html

UGENE is written with open-source QT4 C++ multi-platform library and QtScript scripting language. It is available for most of the popular platforms like Linux, Windows, MacOS X. UGENE is also included into Ubuntu and Fedora Linux distributions.

License: GNU General Public License v.2
Software and source code: http://ugene.unipro.ru/download.html