

Title: The Open HeliSphere™ project: True open source from the inventors of True Single Molecule Sequencing.

Author: Aaron Kitzmiller (akitzmiller@helicosbio.com)

License: GPL v2

Site: <http://open.helicosbio.com>

A recent publication describes the single-molecule sequencing-by-synthesis capabilities of Helicos' technology (Harris, et al. Single-Molecule DNA Sequencing of a Viral Genome Science 4 April 2008: 106-109), demonstrating for the first time the direct interrogation of unamplified strands of DNA to produce sequence information. To avoid the limitations on innovation and barriers to scientific discovery of closed instrument software, Helicos is opening its bioinformatics tool suite in a true open source project under a GPL v2 license. When fully launched, the project will feature:

- Tarball downloads of source code
- Mailing lists
- SVN checkout of trunk and release branches
- patch submission to the core code base
- bug tracking

Similar to the EMBASSY suite that complements the EMBOSS package, a second repository will be made available for tools and libraries wholly owned by users and potentially released under other licenses.

The project contains libraries and tools that support bioinformatic analysis beginning with the instrument-generated raw read container (SRF file) and extending through the creation of consensus alignment and summary reports. Components include:

- Core C/C++ and Perl libraries
- File manipulation tools for the Helicos SMS file format, including SRF and text conversion
- Alignment tools
- Tools for calculation of error rates and length distributions
- Complete Perl analysis pipelines
- Sample datasets