

To print, use Control+P (PCs), Command+P (Macs), or File-->Print.



SOFTWARE REVIEW

Artemis 4



Artemis

Reviewed by Matt Carson

Posted December 21, 2001 · Issue 117

Overall scores

<u>Installation</u>	Very good
<u>Learning curve</u> (beginner who can Web surf and word process)	Good, but use of more advanced features may take a little time
<u>Technical support</u>	Fair
<u>Features</u>	Very good
<u>Customizability</u>	Very good
<u>Utility to biologists</u>	Very good
Value for money	Excellent

Overview

Within the last several years, genome sequencing has become a major focus of both big business and academic science. Unfortunately, the latter often does not get a chance to take advantage of many of the more powerful tools used in these undertakings due to their large price tags. Luckily, there is [Artemis](#).

Artemis is a DNA sequence viewer and annotation tool with a wide variety of resources for biologists. Developed at the [Sanger Centre](#) in Cambridge, UK, this program is very useful for analyzing sequence, creating informative [European Molecular Biology Laboratory](#) (EMBL) files, and preparing projects for submission to [GenBank](#). Artemis displays contiguous DNA sequences. It translates the DNA into the six possible reading frames and displays the translations in a scrolling window along with the forward and reverse strands of the DNA sequence. Sequence features and stop codons are displayed in a synchronized scrolling window above the protein and DNA sequence. Each feature has an edit window where you can add information such as feature number, annotator name, and product. A large list of identifiers, for example, "poly A signal" and "gene," is available. The base range is also visible and can be edited manually. If a batch [BLAST](#) search is performed, the entire list of results for each open reading frame (ORF) can be viewed in its own window.

I have found myself using Artemis at least every other week to format data for input into other programs such as circular and linear map scripts and local BLAST database-creation tools. Following the Unix tradition of freely available top-notch software, Artemis is an application that proves itself useful beyond what you might expect from freeware.

Available platforms Unix, [GNU](#)/Linux, Macintosh, Windows

System requirements Artemis is designed to function on any machine with a recent version of [Java](#) installed. For Unix and GNU/Linux, any version of [Java](#) higher than 1.1.1 is sufficient, although Artemis was developed using 1.1.7. It is highly recommended that you use the Sun's Java and not the default GNU/Linux [Kaffe](#) distribution. For Windows, you need [Java](#) version 1.2 or better. Macintosh users need [Macintosh Runtime for Java](#) (MRJ) 2.2 or better.

Test platforms 30 MHz SuperSparc with 64 Mb Main memory; 250 MHz UltraSparc with 512Mb; just about every possible configuration of Linux box; G3 Macs with 256 to 640Mb memory, 400 and 500 MHz processors. Solaris 7 and 8, Red Hat Linux 7.1, Windows 2000, and Mac OS 9.

Price Free

How Long Did It Take to Learn to Use It Productively?

I began using Artemis a little over a year ago. At the time, my lab was about to begin the annotation of *Neisseria gonorrhoeae*, and five or six of us were going to be working most of each day for about two months to complete two passes through the genome. None of us had ever done this before. Setting up this operation required a few days and involved performing a batch BLAST on the genome and parsing the best hits with a [Perl](#) script. After playing with the software for a day or so, we were off and annotating.

The time it takes to produce results with Artemis depends on what the user wants to do. Viewing a DNA sequence is very easy and should take no more than five minutes to master, but setting up a large-scale annotation project may require some knowledge of shell scripts and Perl, as well as a few days to experiment with Artemis. Advanced administration may require some knowledge of Unix itself, as significant functionality can be tied to Artemis by the appropriate combination of external software and custom scripts.

Artemis is not something to be afraid of, however. The format is fairly intuitive, and for those who've spent a little time with computers, it should not be difficult to figure out how to make the software work.

Product Quality

Ease of installation Good. Precompiled binaries are available for all platforms but you must have Java installed.

User friendliness Fair to good

Interface Good. Java GUI. Advanced use of Unix/GNU Linux versions may require launching from the command line.

Intuitiveness of design Good

Customizability

Users can customize a large number of features by altering the "artemis/etc/options" file. Everything from font size to user-defined qualifiers is specified here in plain text, which allows the user to tailor the program to his or her specific needs. The Windows and Macintosh versions are not quite as easy to customize because the program comes in binary

form, and you must extract this file with a zip utility before you can alter it. Artemis will then read this options file if you place it in the directory where you installed the Artemis binary.

Ability to Program in Scripts, Add Extension Modules, etc.

Source code is freely available from the [Sanger Centre FTP site](#) for those with a knack for Java.

Plot-type data created in any external program can be loaded into Artemis and displayed against the DNA sequence view.

You can add shell scripts to the "artemis" directory to run custom analyses or perform any desired data manipulations. The availability of programming experience is essentially the only limiting factor. Perl, although it doesn't actually work within Artemis, can also be very useful for preparing files for import or export.

Ability to Import and Export in Different File Formats

Artemis takes raw and [FASTA](#) files for DNA sequence and GenBank (.GBK), [EMBL](#) (.TAB), or [GFF](#) (.GFF) files for sequence feature tables. When provided with an accession number, the Unix and Linux versions allow the user to import entries from the [European Bioinformatics Institute](#) (EBI). Artemis will export DNA or protein sequence in raw, FASTA, EMBL, or GenBank formats.

Useful or Unusual Features

One unique feature of Artemis as a sequence and feature viewer is that it does not require all features to be contained in a single GenBank or EMBL sequence-definition file. Once you've loaded a genome or other sequence, you can overlay as many different annotations, or combinations of annotations, on the sequence as you wish simply by loading additional EMBL format files that define the locations and feature identifiers for the additional features you wish to display. This turns out to be particularly useful for situations in which you want to have auxiliary information displayed, but not included in your main annotation. For example, the maintenance of a simple EMBL file containing sequence boundaries for spanning a large insert library allows quick indexing into that set whenever a sequence ambiguity shows up in a particular feature.

Another unusual feature, and one not particularly useful in my hands, is the "zoom" functionality. Artemis has two main windows: a lower DNA/protein sequence view and the feature view above. Each window implements a "zoom" function as a side scroll-bar, allowing the user to zoom into or out of the sequence displayed. This might sound like a neat feature, but if you use Artemis for a while, you will inevitably click in this scroll area and accidentally zoom out to an aerial view or into a base-by-base view of your genome that you can make no sense of whatsoever.

I have found many other features that are useful as well, although which are thought most useful depends greatly on the particular project. One of my favorites is the feature selector. This search utility allows the user to look for features by key (CDS, gene, etc.), qualifier (product, EC number, etc.), or any text within the feature. This is very handy for making a specialized .TAB file containing, for example, only ABC transporters. Artemis also includes a great navigator that makes searching through a genome by text, key, base pattern, or amino-acid string easy. The user can display GC and AG percent and deviation, a GC frame plot, or correlation scores for the entire sequence. There is also a lot of information about each ORF, such as a full statistics page containing everything from molecular weight to the percentage of each codon and amino acid, and hydrophobicity and hydrophilicity plots. On Unix and Linux systems, the run menu allows the user to execute any of various types of BLAST searches or a [sigcleave](#) analysis on any feature. And, thankfully, Artemis cooperates well with the [Concurrent Versions System](#) (CVS), which is very handy if you're lucky enough to have company when doing your annotation.

Version 4 provides many improvements over previous versions, including bug fixes as well as some new features, such as the ability to read the output of [MSPcrunch](#), a BLAST post-processing filter, and new support

for running the [Clustal W](#) and [Jalview](#) multiple alignment editors. It should be noted, however, that the use of these advanced programs may require shell scripting or Perl programming ability. See the Artemis [release 4](#) page for details.

Limitations

The Windows and Macintosh versions of this software are not as fully developed as the Unix and GNU/Linux versions, nor are they as well supported. I highly recommend running the Unix or Linux versions for large-scale annotation, not only because of the stability of the operating system, but also because of extra options such as immediate BLAST analysis of ORFs from within the program.

On the other hand, Artemis can run quite slowly at times, especially on Unix and Linux systems, when viewing very large genomes. This problem may be worse if the machine it is running on is equipped with a version of Java greater than 1.1.x. From my experience, the Macintosh and Windows versions stand out in this area, so for speed of display, one of these may be the better option.

Comparisons with Similar Software

I honestly can't say I've spent much time with comparable software. Most of the software used to accomplish the same tasks as Artemis are proprietary, and, being employed at a university, I usually use free software if it will do the job. The closest comparison I have seen is a free program called [AceDB](#), a genome database system used primarily for larger eukaryotic genomes. Another somewhat similar program is [Sequencher](#), although it is not an annotation tool and is really only similar in that it displays DNA sequence.

Recently I've had the opportunity to test [DNAMAN](#), as well as the Lasergene package from [DNASStar](#). Both of these packages provide considerably more functionality than Artemis, such as the ability to switch to alternate genetic codes, label restriction enzyme cut-sites, and perform multiple sequence alignments. They also have some disappointing limitations, such as the lack of Unix versions and limited interoperability, which could hamper their utility in high-throughput applications. After having tested these commercial applications for a while, I would recommend Artemis for the researcher looking for a cross-platform and collaborative sequence and annotation viewer, and the commercial applications for those who need to do a considerable amount of sequence analysis.

Technical Support and Documentation

Many users will be able to get started right away after viewing the informative [user manual](#). In order to keep up with current issues concerning Artemis, users may also want to join the [mailing list](#). For those seeking more, there is a discussion group. To sign up, send a message to majordomo@sanger.ac.uk with "subscribe artemis-users" in the body of the message. Remember, this is free software and comes without any warranty.

Target Users

Artemis is targeted toward academic scientists in the business of sequencing/annotating prokaryotic or lower eukaryotic genomes. It remains one of the few tools of its kind for the university environment. It's free and requires only intermediate computer skills to use, and it will work for anything from a junior-sized gene to a whole genome.

Publisher information	Wellcome Trust Sanger Institute Wellcome Trust Genome Campus Hinxton Cambridge, CB10 1SA, UK Tel: +44 (0)1223 834244 Fax: +44 (0)1223 494919 Email(Artemis developers): artemis@sanger.ac.uk Web site: www.sanger.ac.uk/Software/Artemis/v4/
Pricing structure	Free. Artemis version 4 is available under the terms of the GNU General Public License
Software class	Data analysis and visualization Molecular biology

***Matt Carson** is the director of bioinformatics/Unix systems administrator at the University of Oklahoma Health Sciences Center's Laboratory for Microbial Genomics.*



[E-mail this article to a friend](#)

Want to see a review of particular software? Send a suggestion.



[Feedback](#)

Previous Beagle Software Reviews

- ▶ [Storyspace 2 for Macintosh](#)
Reviewed by Charlie Schick (Posted December 7, 2001 · Issue 116)
- ▶ [PHP 4](#)
Reviewed by Marijan Adam (Posted November 23, 2001 · Issue 115)
- ▶ [JFile 5.0](#)
Reviewed by Charlie Schick (Posted November 9, 2001 · Issue 114)
- ▶ [Dreamweaver 4](#)
Reviewed by Robyn Ness (Posted October 26, 2001 · Issue 113)
- ▶ [REALbasic Professional Edition 3.5](#)
Reviewed by John Ray (Posted October 12, 2001 · Issue 112)
- ▶ [ACLUSTER 2.0](#)
Reviewed by Virginia Fitzpatrick (Posted September 28, 2001 · Issue 111)

[more](#)