Geneious R8: A powerful and comprehensive suite of molecular biology tools

Olsen, C.*[1], Qaadri, K.[1], Moir, R.[2], Kearse, M.[2], Buxton, S.[2],Cheung, M.[2], Hengjie, W.[2], Kuhn, J.[2], Stones-Havas, S.[2], Duran, C.[2]

[1] Biomatters, Inc. 60 Park Place, Suite 2100 Newark, NJ 07102

[2] Biomatters Ltd L2, 18 Shortland Street Auckland, 1010 New Zealand

Contact: christian@biomatters.com

geneious

Introduction

Biomatters' Geneious R8 is a bioinformatics software platform that allows researchers the command of industry-leading algorithms and tools for their genomic and protein sequence analyses. Researchers at all levels can easily manage, analyze, & share their sequence data via a single intuitive software application. Geneious R8 provides tools for next-generation sequence analysis, sequence alignment, molecular cloning, chromatogram assembly, and phylogenetics. **New features for this major version release include: CRISPR gene editing, a 16S Biodiversity Tool, a Species Identification plugin, Circular *de novo* assembler, Improved Workflow Builder, as well as many new plugins**. R8 affords real-time dynamic interaction with sequence data and empowers biologists to produce stunning publication-quality images to increase the impact of their research. By utilizing Geneious R8, biologists can easily improve their bioinformatic workflow efficiencies to free up more time for their research.

Background

One of the challenges with bioinformatics is the abundance of applications, websites, and algorithms available to the research community for sequence analysis. Two different approaches have been used in the design of bioinformatics software.

The first approach is the '*black-box*' approach, where the software and underlying algorithms are proprietary and not available for inspection, adjustment, or extension. In some cases, the file formats generated by these applications fall into the '*black-box*' category as they are proprietary and encrypted. The second approach is the '*glass-box*' approach, where the algorithms implemented in the software are made available and the software has the flexibility to be extended by the public. The files generated by '*glass-box*' applications are also open and require minimal effort for parsing and converting. The '*glass-box*' approach tends to be driven by the research community and is frequently found in academic and collaborative settings.

Geneious R8 was created using the '*glass-box*' approach in order to reduce the bioinformatics learning curve and allow more biologists to manage, analyze, and visualize their sequence data. Geneious R8 is a cross-platform Java application, which runs on any computer with a Java Virtual Machine installed on it (PC, MacOSX, Linux). Geneious R8 uses publicly available, peer-reviewed algorithms, which afford researchers a measure of confidence, with the ability to inspect how the algorithms were used in the peer-reviewed literature. Three domains and multiple web services have been incorporated into one user interface for genomic sequence analysis: **molecular cloning, NGS, and phylogenetics**. In addition to a robust feature-set, Geneious R8 empowers the researcher with intuitive data management features like a drag-and-drop file interface, local file directory organization, and the ability to connect to relational databases (Figure 1).
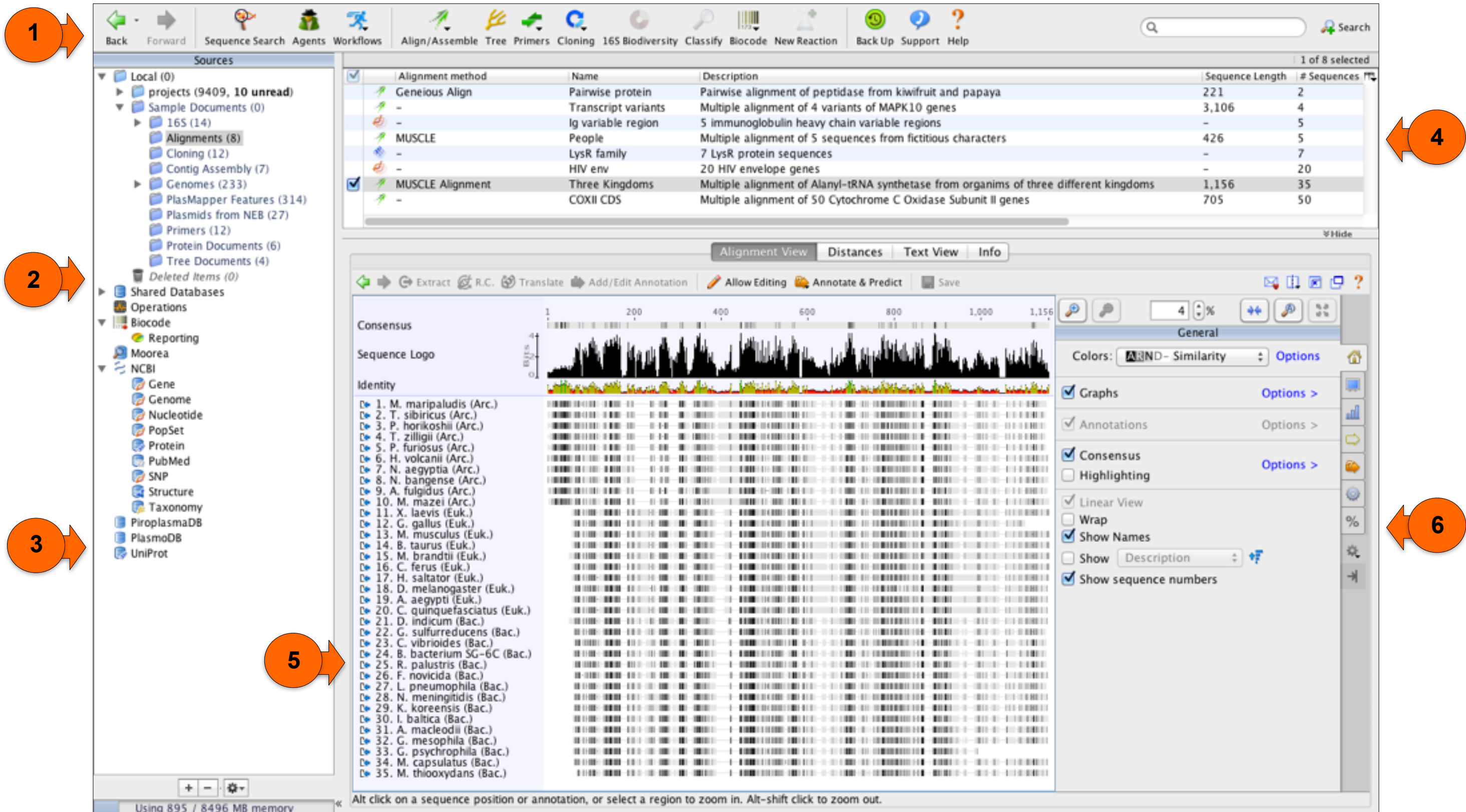

Figure 1: Geneious R8 interface. [1] – 'Menu Bar' – configurable for frequently used analyses; [2] – local directory; [3] - 'Live' database services; [4] – 'Document Table' includes file metadata; [5] - real-time interactive 'Document Viewer'; [6] – 'Options Panel' used to configure display in the document viewer. Everything displayed in the viewer may be exported in a variety of image file formats.

Features

Geneious R8 is a mature Java framework, which combines useful tools and functions into a single intuitive graphical user interface. By design, the user is faced with a single learning curve to help navigate the diversity of leading applications and algorithms, which seamlessly communicate behind the scenes. The framework is flexible and able to adapt to the changing demands of research via a public Java API. The extensible plug-in architecture allows any developer to leverage what has been developed for their own purpose (Figure 2). Much of the current functionality of Geneious stems from community driven feature requests and plugin creation (e.g. **Workflow builder** Figure 2), **Sequence Classifier**, **GenBank submission**, **Augustus gene prediction**, **Mira Assembler**). The selection of the feature set is fundamentally guided by the scientific community via collaborations and consultations with biologists and bioinformaticians.
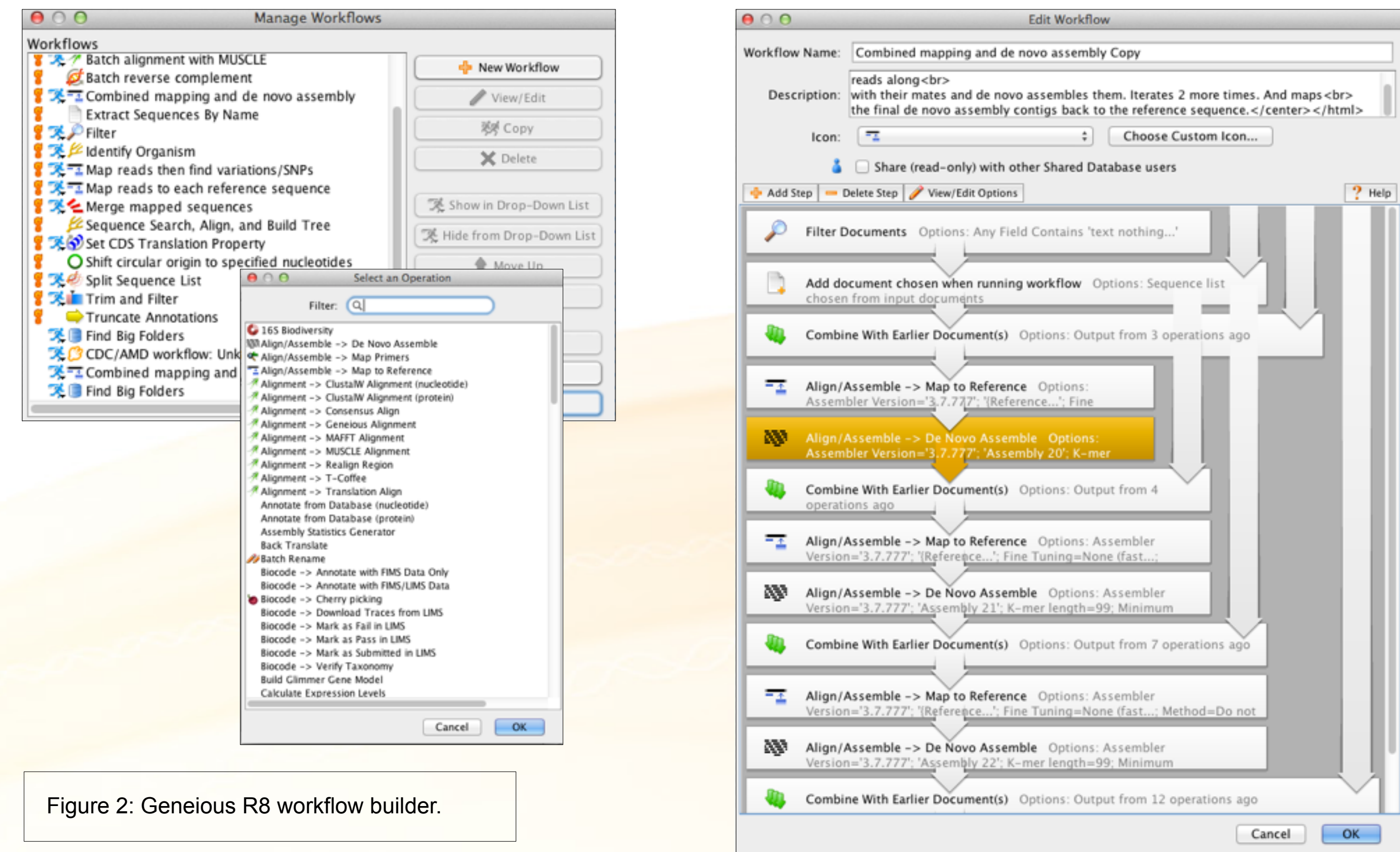

Figure 2: Geneious R8 workflow builder.

Molecular Cloning

Geneious R8 has added new cloning tools (Figure 3) for **CRISPR (clustered regularly interspaced short palindromic repeats)** gene editing as well as the CRT tool, which is used to rapidly find CRISPRs in bacteria and archaea. This new plugin runs CRT and saves the results as annotations on your genome of interest.

The interactive document viewer enables the user to quickly create/edit constructs, restriction enzyme lists, annotate restriction enzymes directly onto linear & circular DNA and then perform *in silico* digests into fragments. The ability to track the cloning process through parent-descendent relationships allows the user to fork and edit a cloning process for protocol development. Geneious R8 provides a customizable and searchable primer and restriction enzyme database. **Primer3** is included in Geneious R8[1,2] and allows the user to design degenerate primers, designate mismatches, and perform primer extensions.
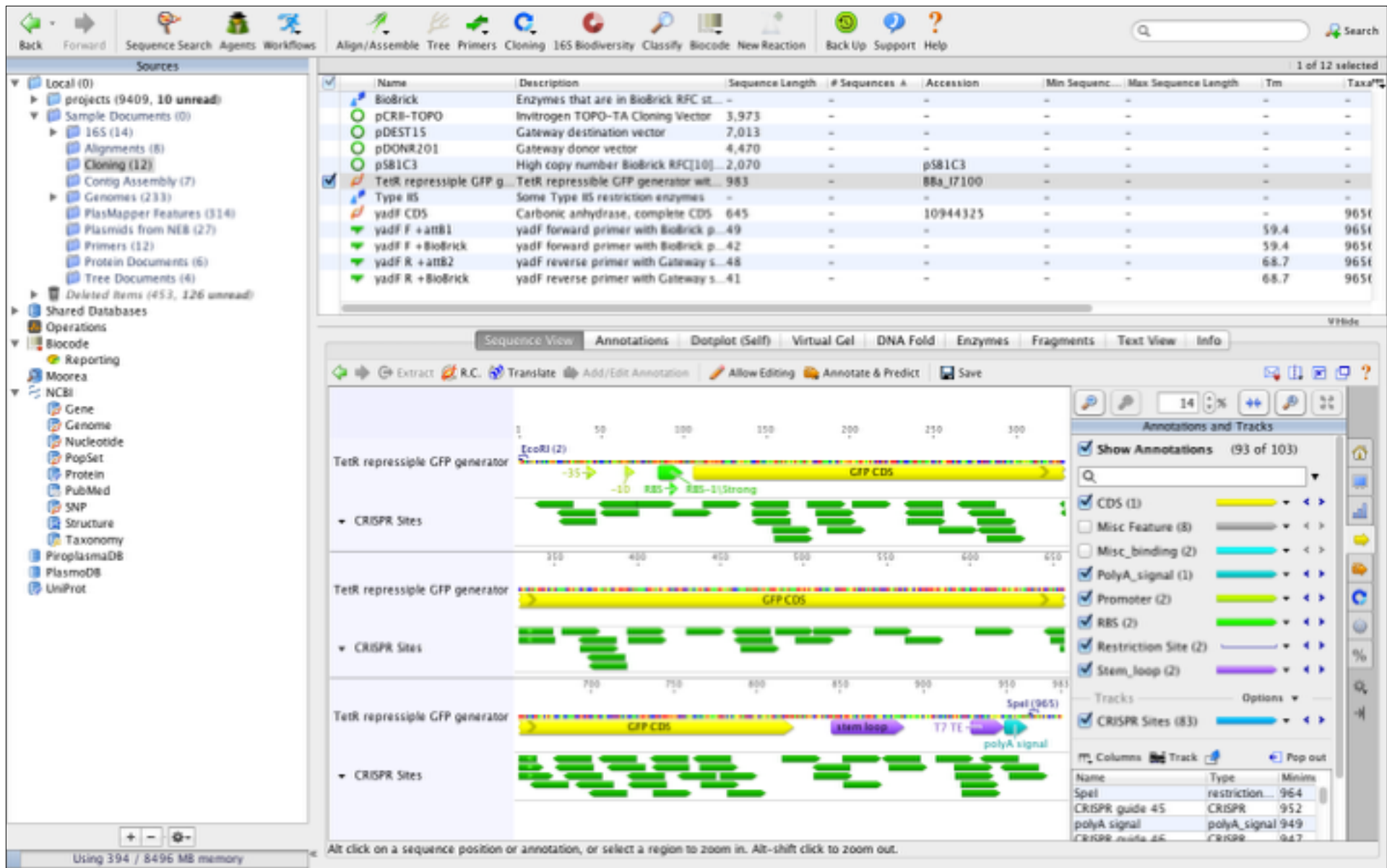

Figure 3: CRISPR gene editing

Next-Generation Sequence Analysis

Geneious R8 provides a robust suite of tools for processing sequence data from all sequencing platforms (eg. Illumina, Sanger, PacBio) and has been performance tuned to generate longer contigs with greater accuracy. Pre-processing tools like trimming, pairing, and barcode sorting allow biologists the ability to groom and visualize their sequence data without having to utilize command-line tools & scripts. Geneious R8 includes a number of assemblers including the Mira hybrid assembler, a circular *de novo* assembler, and tools for working with RNAseq transcriptome data. Users are now able to analyze gene expression in mapped mRNA data using the standard normalization methods RPKM & TPM.

16S rRNA Biodiversity Tool

The 16S rRNA Biodiversity Tool is a cloud-based tool for routine species classification and relative abundance measurement using high throughput 16S rRNA amplicon sequencing data from environmental samples. Preprocessed full length bacterial 16S rRNA sequences may be submitted, but any sub-region of 16S can be used. The user submits their next-generation sequence amplicon data through the Geneious R8 bioinformatics platform to a distributed cloud compute resource. The data are then analyzed using the Ribosomal Database Project Database (RDP) Classifier[3]. The RDP Classifier assigns sequences derived from bacterial and archaeal 16S genes and fungal 28S gene to the corresponding taxonomy model using a 'Naïve Bayesian Classifier' for rapid assignment of rRNA sequences. The Geneious 16S Biodiversity Tool accurately assigns a taxonomy (in the range of domain to genus) along with a confidence-estimate for each sequence by comparing them to the RDP database. A secure weblink is returned within Geneious' 'Document Viewer'. Upon clicking the link the output is then displayed in a web browser using Krona[4], which produces an interactive HTML5 hierarchical graph of the bacterial diversity in the sample.
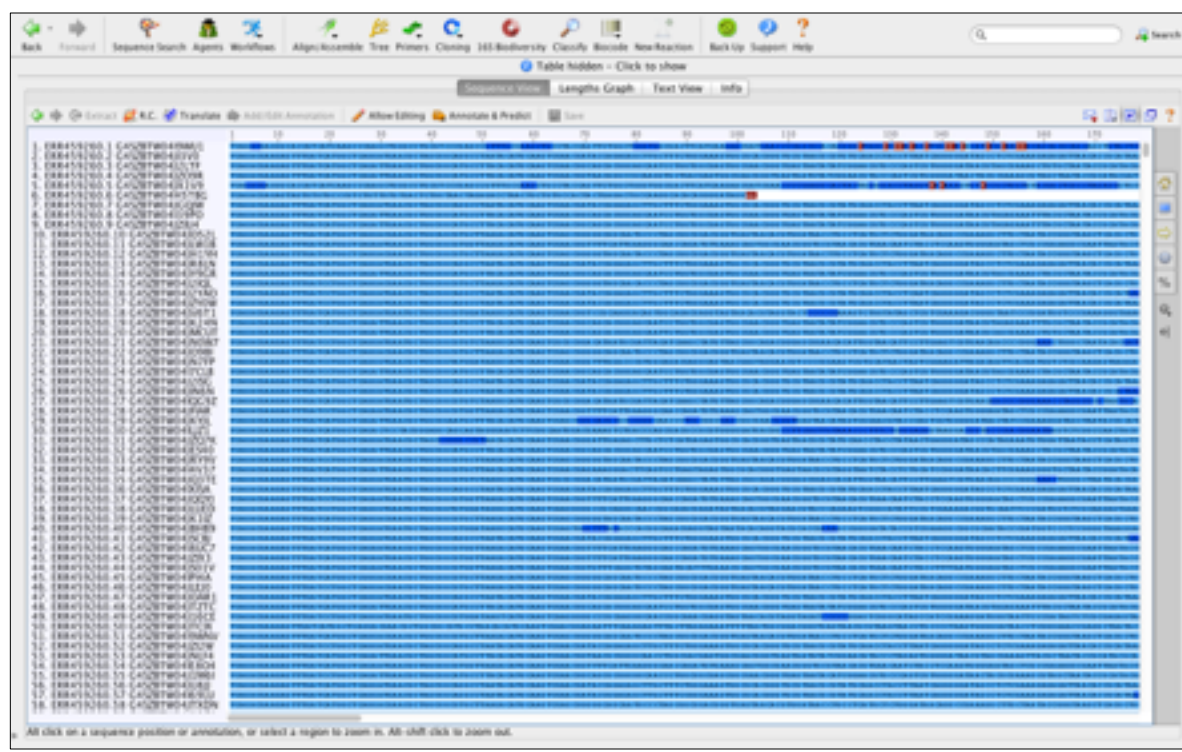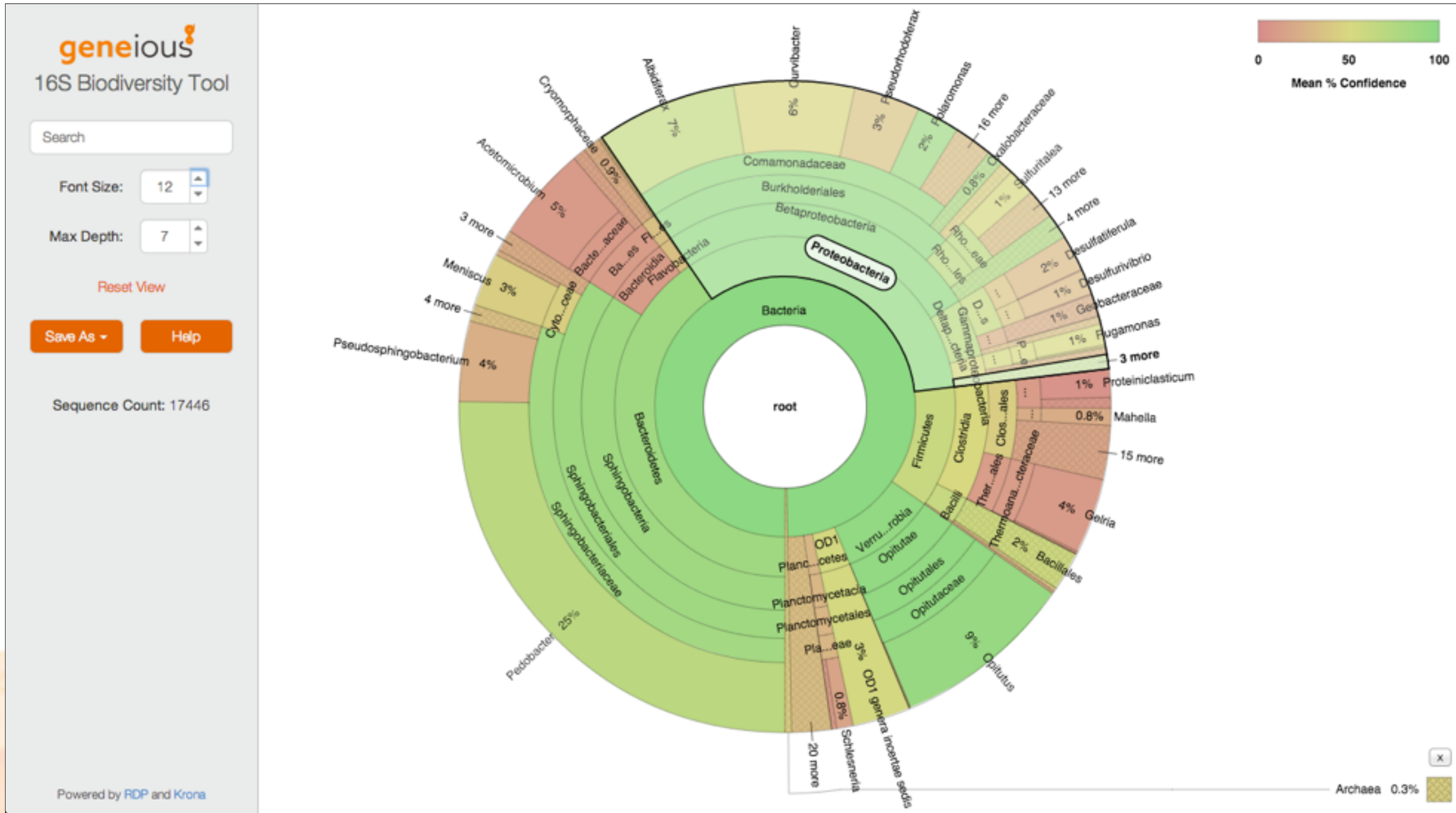

Figure 6: Preprocessed 16S rRNA amplicon data submitted to RDP and visualized using the Krona Interactive Hierarchical Browser

References

1. Untergrasser A, et. Al. 2012. Primer3 - new capabilities and interfaces. Nucleic Acids Research 40(15):e115.
2. Koressaar T, Remm M. 2007. Enhancements and modifications of primer design program Primer3 Bioinformatics 23(10): 1289-91.
3. Wang, Q, G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol. 73(16):5261-7.
4. Ondov BD, Bergman NH, and Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics. 2011 Sep 30; 12(1):385.