Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data

Peter Meintjes¹*, Chris Duran¹,²*, Matthew Kearse¹, Richard Moir¹, Amy Wilson¹, Steven Stones-Havas¹, Matthew Cheung¹, Shane Sturrock¹, Simon Buxton¹, Alex Cooper¹, Sidney Markowitz¹, Tobias Thierer¹, Bruce Ashton¹ and Joseph Heled¹

¹Biomatters Ltd., 2/76 Anzac Avenue, Auckland, 1010, New Zealand
²School of Agriculture and Food Science, University of Queensland, Brisbane, 4072, Australia

ABSTRACT
Summary: The two main functions of bioinformatics are the organization and analysis of biological data using computational resources. Geneious Basic has been designed to be an easy-to-use and flexible desktop software application framework for the organization and analysis of biological data, with a focus on molecular sequences and related data types. It integrates numerous industry-standard discovery analysis tools, with interactive visualizations to generate publication-ready images. One key contribution to researchers in the life sciences is the Geneious public application programming interface (API) that affords the ability to leverage the existing framework of the Geneious Basic software platform for virtually unlimited extension and customization. The result is an increase in the speed and quality of development of computation tools for the life sciences, due to the functionality and graphical user interface available to the developer through the public API. Geneious Basic represents an ideal platform for the bioinformatics community to leverage existing components and to integrate their own specific requirements for the discovery, analysis and visualization of biological data.

Availability and implementation: Binaries and public API freely available for download at http://www.geneious.com/basic, implemented in Java and supported on linux, Apple OSX and MS Windows. The software is also available from the BioLinux (Field et. al, 2006) package repository at http://nebc.nerc.ac.uk/downloads/geneious_packages/.

Contact: Peter Meintjes (peter@biomatters.com) and Chris Duran (chris@biomatters.com)

1 INTRODUCTION
Research in the biosciences increasingly depends upon bioinformatics for the effective organization and analysis of biological data and experimental results. Having access to the appropriate bioinformatics tools is crucial to the success of any research project. The field of bioinformatics itself has previously been segregated into the parallel realms of data organization (databases, access, integration, web services and search tools) and data analysis (efficient algorithms and statistical modeling). As a result, a large proportion of bioinformatics work involves the choreparing of data from a variety of sources through a pipeline of analysis tools. This approach typically involves a number of ad hoc labor-intensive steps to convert file formats and create transitional file types. Geneious Basic was created to provide a general framework for research-focused bioinformatics tasks to overcome these challenges and take advantage of modern computing trends. Highlights of the platform include automated database searching, data backup functionality and an extensible API for the integration of novel bioinformatics analysis tools.

2 METHODS
Geneious Basic is written in Java Swing to maximize interoperability among all commonly used operating systems. It is compiled under and requires Java 5 to run. The application provides core modules to enable the visualisation, manipulation and transfer of DNA sequences (linear, circular and short oligos such as primers and probes), amino acid sequences, pairwise and multiple alignments, phylogenetic trees, 3-D structures, sequence chromatograms, contig assemblies, microsatellite electropherograms and statistical graphs.

Fig. 1. Modular overview of the Geneious Basic software stack. Top-most modules have dependencies on lower modules. The unshaded modules represent the publicly accessible modules for plugin development.

The underlying software framework for Geneious Basic is modular and multi-tiered with a focus on handling bioinformatics data and tools (Fig. 1).
It integrates a comprehensive plugin system which is grounded in the extensible Geneious API. The public API component allows plugin developers to leverage the functionality and user interface of the Geneious platform whilst concentrating on the development of processes and algorithms. The API download from the Geneious website provides a number of skeleton plugin examples to use as a basis for new plugins, allowing developers with a basic level of java knowledge to develop fully-functional plugins that greatly extend the functionality of Geneious Basic.

The public API allows developers to leverage online sequence search web services such as NCBI BLAST. Also using the public API, developers can implement plugins that exploit external binaries and even online computational resources, a growing trend in the field of Bioinformatics (Schatz et. al, 2010).

3 RESULTS

The Geneious Basic graphical user interface (GUI) comprises of three main panels, a variety of pull-down menus and many right-click functions for common bioinformatics analyses. Geneious Basic displays selected file(s) in a variety of different ways including sequence view (linear and circular), dotplot view, query-centric alignment view, protein domain view, 3D structure view, text view and notes.

A number of analysis algorithms are implemented within Geneious Basic. For a given document, the user can select from among the appropriate analysis tools to process the data; for example, in the case of a sequence document, the user may choose to perform a BLAST (Altschul et al., 1990) search for a given query sequence against a specific online repository, align against other sequences or generate protein translations. The combinations of available tools and document types allow for the implementation of complex workflows in a single, consistent environment (Fig. 2).

Geneious Basic has an active community of plugin developers, using the public API to contribute third-party plugins and additional functionality within Geneious. The Species Delimitation Plugin, a plugin for the exploration of species boundaries. Mol Ecol Resour. 11:154-157. is one example of scientific research leveraging the Geneious public API for publishable academic research. At the time of the writing of this manuscript, there were nine available plugins for Geneious Basic (Table 1).

Table 1. Summary of available plugins for Geneious Basic

<table>
<thead>
<tr>
<th>Plugin name</th>
<th>Category</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>DualBrothers Recombination Detector</td>
<td>Phylogenetics</td>
<td>Marc Suchard</td>
</tr>
<tr>
<td>Green Button</td>
<td>Superecomputing</td>
<td>Biomatters Ltd.</td>
</tr>
<tr>
<td>Heterozygotes</td>
<td>Sequencing</td>
<td>Biomatters Ltd.</td>
</tr>
<tr>
<td>InterProScan</td>
<td>Protein</td>
<td>Michael Thon</td>
</tr>
<tr>
<td>MrBayes</td>
<td>Phylogenetics</td>
<td>Marc Suchard</td>
</tr>
<tr>
<td>Phobos Tandem Repeat Finder</td>
<td>Nucleotide</td>
<td>Christoph Mayer</td>
</tr>
<tr>
<td>PhyML</td>
<td>Phylogenetics</td>
<td>Vincent Lefort</td>
</tr>
<tr>
<td>Species Delimitation</td>
<td>Protein</td>
<td>Masters and Ross</td>
</tr>
<tr>
<td>Transmembrane Prediction</td>
<td></td>
<td>Marc Suchard</td>
</tr>
</tbody>
</table>

Fig. 2. The phylogenetic tree showing the relationship of a MT-CO1 sequence from an unknown source with similar public sequence. Sequences were identified from Genbank using BLAST. Sequences were downloaded then aligned using the Geneious Aligner and a phylogenetic tree built using PhyML. All of these steps were performed within Geneious Basic.

4 DISCUSSION

The functionality in Geneious Basic can be compared to a number of other desktop software packages, such as VectorNTI (Lu and Moriyaama, 2004), CLC Bio, Sequencer (GeneCodes), Lasergene (DNASTar) and MEGA4 (Tamura et al., 2007). In comparison to the above, the focus of Geneious Basic is to provide an extensible desktop platform in which existing components can be leveraged by third-party plugin developers. This is a unique and important contribution to bioinformatics, allowing users to customize and extend core to their needs without significant manual intervention, matching the success of extensible platforms in parallel spaces such as network analysis and workflow management (Smoort et al., 2010; Goecks et al., 2010). Also comparable is Geneious Pro (Drummond et. al, 2010), a commercial extension to Geneious Basic. It shares the base functionality of Geneious Basic and additionally provides tools focused on data sharing, collaboration and advanced next generation sequencing functionality.

Geneious Basic brings together a large and disparate number of complementary data sources, analysis methods and visualization tools. It is the intention of the authors that by encouraging development to the public API new analysis methods and visualization tools will continue to grow and diversify the range of bioinformatics tasks available to all researchers.

REFERENCES


