

# Penggunaan *Workflows* Dalam Aplikasi Bioinformatika Geneious Untuk Menganalisis Data Genomik

Benny Maisa

## Abstrak

Perkembangan penelitian biomolekuler yang meningkat pesat dalam beberapa dekade terakhir memberikan dampak peningkatan kebutuhan teknologi komputasi untuk menganalisis data genomik. Hasil penelitian berupa data digital yang berukuran sangat besar membutuhkan bantuan teknologi komputer untuk menyediakan berbagai solusi penyelesaian. Cabang ilmu bioinformatika, yang merupakan kombinasi cabang ilmu biologi, ilmu komputer, sistem informasi, matematika, kedokteran, farmakologi, fisika, kimia, dan statistik; memberikan solusi dengan beragam aplikasi untuk berbagai kebutuhan analisis genomik. Tinjauan pustaka ini bertujuan menunjukkan penggunaan *workflows* dalam aplikasi bioinformatika Geneious untuk menganalisis data genomik. Geneious adalah sebuah aplikasi bioinformatika *user-friendly* yang mengintegrasikan berbagai aplikasi analisis data genomik. Geneious menyediakan *workflows* standar siap pakai dan memberikan fitur pembuatan *workflows* sesuai dengan kebutuhan metode penelitian bioinformatika. Penggunaan *workflows* memberikan efisiensi bagi peneliti terutama dalam menganalisis data genomik dalam ukuran besar dan memerlukan integrasi beberapa aplikasi bioinformatika untuk proses analisis dan interpretasi.

**Kata kunci:** bioinformatika, *workflows*, data genomik, Geneious

## Abstract

*The development of biomolecular research that has increased rapidly in the past few decades has had an impact on the increasing need for computational technology to analyze genomic data. The results of the research in a very large digital data require the help of computer technology to provide a variety of solutions. Bioinformatics, which is a combination of branches of biology, computer science, information systems, mathematics, medicine, pharmacology, physics, chemistry, and statistics; provide solutions with various applications for various genomic analysis needs. This literature review aims to demonstrate the use of workflows in Geneious bioinformatics applications to analyze genomic data. Geneious is a user-friendly bioinformatics application that integrates various applications of genomic data analysis. Geneious provides standard ready-to-use workflows and provides features for making workflows according to the needs of the bioinformatics research method. The use of workflows provides efficiency for researchers, especially in analyzing genomic data in large sizes and requires the integration of several bioinformatics applications for the process of analysis and interpretation.*

**Keywords:** bioinformatics, workflows, genomic data, Geneious

**Affiliasi penulis:** Bagian Mikrobiologi FK Unand  
**Korespondensi:** Benny Maisa, e-mail: bennymaisa@med.unand.ac.id

## PENDAHULUAN

Selama hampir 4 dekade terakhir, penelitian biomolekuler meningkat dengan signifikan. Hal ini menyebabkan terjadinya ledakan dalam jumlah data genomik.<sup>1</sup> Perkembangan yang sangat pesat tersebut meningkatkan kebutuhan pada teknologi komputasi untuk membantu analisis data genomik.<sup>2</sup> Sehingga lahirlah cabang ilmu baru, bioinformatika, yang merupakan kombinasi cabang ilmu biologi, ilmu

komputer, sistem informasi, matematika, kedokteran, farmakologi, fisika, kimia, dan statistik.<sup>3,4</sup>

Bioinformatika secara umum berarti penggunaan metode komputasi untuk meneliti data biologis. Secara khusus berarti pengembangan metode komputasi untuk meneliti struktur, fungsi, dan evolusi gen, protein, dan keseluruhan genom suatu organisme.<sup>1</sup> Bioinformatika menggabungkan berbagai teknologi informasi yang akan mempercepat kalkulasi dan klasifikasi data gen dalam jumlah sangat besar untuk dapat dianalisis dan diinterpretasi oleh para ahli biologi.<sup>5</sup>

Beragamnya jenis data genomik membutuhkan beragam metode dan aplikasi berbeda untuk menganalisis data tersebut. Berbagai aplikasi telah dikembangkan sesuai dengan jenis data genomik, dan sebagian besar di antaranya bersifat *open-source*. Namun kekurangannya, beragam aplikasi *open-source* ini tidak terintegrasi dengan baik satu sama lain.

Geneious adalah salah satu aplikasi bioinformatika dengan kelebihan *user-friendly* dan mengintegrasikan berbagai aplikasi-aplikasi yang umumnya tersedia *open-source* dalam suatu fungsi bernama *workflows*.<sup>2</sup> *Workflows* adalah sebuah sistem untuk mengeksekusi beberapa langkah komputasi atau data manipulasi dengan tujuan efisiensi dan otomatisasi analisis data genomik.<sup>6,7</sup>

Artikel ini akan memaparkan beragam penggunaan *workflows* dalam aplikasi Geneious yang dapat digunakan untuk analisis data genomik.

## METODE

Penelitian ini merupakan studi literatur untuk mengetahui penggunaan *workflows* dalam aplikasi Geneious.

## PEMBAHASAN

Dalam aplikasi Geneious versi R11, terdapat 24 *workflows* standar siap pakai. *Workflows* yang tersedia adalah *workflows* yang umum digunakan dalam analisis genomik. Selain itu, pengguna juga dapat mendesain sendiri *workflows* jika tidak tersedia di dalam daftar *workflows* yang ada.

### Fungsi-fungsi dalam workflows

Fungsi-fungsi utama dalam *workflows* adalah berupa pembuatan *workflows* baru, edit, *copy*, dan *delete*. Selain itu, *workflows* dapat di-*export* untuk digunakan pada pengguna Geneious atau pada komputer lainnya. *Workflows* dari sumber ketiga juga dapat digunakan dengan menggunakan fungsi *import*.<sup>8</sup>

Tersedia lebih dari 120 pilihan langkah (*steps*) yang dapat dipilih jika menambahkan atau mengubah sebuah *workflows*. Masing-masing langkah dapat diubah konfigurasinya untuk menyesuaikan dengan kebutuhan data yang akan dianalisis.<sup>8</sup> Tidak ada batasan jumlah langkah dalam *workflows*, namun perlu pertimbangan *resources* komputer. Setiap aplikasi yang akan digunakan dalam *workflows*

membutuhkan *resources* yang tidak sedikit, sehingga integrasi beberapa aplikasi tersebut dalam satu eksekusi akan memberikan beban komputasi bervariasi pada komputer.

### *Workflows menyederhanakan berbagai fungsi dalam aplikasi Geneious*

Geneious terdiri atas 9 fitur utama, yaitu: (1) *Assembly & Mapping*; (2) *NGS Visualization & Analysis*; (3) *Alignment*; (4) *Analysis & Anotation*; (5) *Phylogenetic Trees*; (6) *Molecular Cloning*; (7) *Primer Design*; (8) *Database Searching*; dan (9) *Workflows*.<sup>9</sup> Fitur terakhir Geneious dapat menyederhanakan berbagai fitur lainnya untuk bekerja secara terintegrasi sehingga dapat mengefisienkan proses eksekusi berbagai fungsi di dalamnya.

### Penggunaan Workflows

Berikut adalah penggunaan beberapa *workflows* yang sering digunakan dalam penelitian bioinformatika:

1. *Align DNA then build tree*<sup>8</sup>  
*Workflow* ini diawali dengan eksekusi fungsi *alignment* pada beberapa *sequences* DNA menggunakan algoritma MUSCLE, lalu hasil *alignment* dilakukan eksekusi fungsi *tree builder* menggunakan algoritma *Geneious Tree Builder*. Hasil akhir *workflow* ini digunakan untuk menilai kedekatan filogeni antara beberapa *sequences* DNA (misal, gen).
2. *Align DNA via Muscle, ClustalW, and Geneious*<sup>8</sup>  
*Workflow* ini mengeksekusi tiga algoritma *DNA alignment* sekaligus dalam satu perintah. Hasil *workflow* ini untuk membandingkan hasil *alignment* satu algoritma dengan lainnya.
3. *Identify organism*<sup>8</sup>  
*Workflow* ini digunakan untuk mengidentifikasi organisme dari *DNA sequences*. Identifikasi dilakukan menggunakan aplikasi BLAST yang tersedia dalam bentuk *plugin* dalam aplikasi Geneious. *Workflow* ini akan menunjukkan organisme yang memiliki kesesuaian *sequences* paling dekat dengan *sequences* yang diidentifikasi. Hanya hasil dengan kesesuaian tertinggi (*Top Hit*) yang akan ditampilkan dalam hasil *workflow* ini.
4. *Map reads then find variations/SNPs*<sup>8</sup>

*Workflow* ini membutuhkan sebuah genom referensi sebagai peta. Beberapa buah *reads/sequences* dari suatu genom sampel akan dipetakan pada genom referensi yang berasal dari organisme yang sama. Hasil pemetaan ini akan dibandingkan kembali dengan genom referensi untuk melihat variasi genetik pada organisme tersebut. Genom referensi biasanya diperoleh dari *database* NCBI.

5. *Map reads to each reference sequence*<sup>8</sup>

*Workflow* ini membutuhkan beberapa buah genom referensi. Satu atau lebih *reads/sequences* akan dipetakan pada lebih dari satu genom referensi. Hasil pemetaan ini digunakan untuk membedakan hasil pemetaan pada masing-masing genom referensi.

6. *Sequence Search, Align, and Build Tree*<sup>8</sup>

*Workflow* ini menggabungkan dua *workflows* sebelumnya, *Align DNA then build tree* dan *Identify organism*. Diawali dengan mengeksekusi pencarian kemiripan *sequences* beberapa sampel dengan *database* menggunakan BLAST, lalu sampel-sampel yang memiliki kemiripan *sequences* akan digabungkan dan dilakukan *alignment* menggunakan algoritma MUSCLE. Setiap hasil *alignment* akan diproses untuk membentuk pohon filogeni menggunakan algoritma *Geneious Tree Builder*. Dari beberapa buah *sequences* sampel akan diperoleh hasil akhir berupa beberapa pohon filogeni sampel yang memiliki kemiripan *sequences*.

## SIMPULAN

Penggunaan *workflows* dalam aplikasi bioinformatika berkaitan dengan metode penelitian terhadap data genomik. *Workflows* memberikan keuntungan efisiensi bagi peneliti untuk menganalisis data genomik, terutama dalam ukuran besar dan membutuhkan integrasi beberapa algoritma.

## DAFTAR PUSTAKA

- Higgs PG, Attwood TK. *Bioinformatics and Molecular Evolution*. Malden, MA: Blackwell Publishing; 2005. 1-365 hal.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, dkk. *Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data*. *Bioinformatics* [Internet]. 15 Juni 2012 [dikutip 20 Oktober 2018];28(12):1647–9. Tersedia pada: <http://www.ncbi.nlm.nih.gov/pubmed/22543367>
- Can T. *Introduction to Bioinformatics*. In: *Methods in molecular biology* (Clifton, NJ) [Internet]. 2014 [dikutip 30 November 2018]. hal. 51–71. Tersedia pada: <http://www.ncbi.nlm.nih.gov/pubmed/24272431>
- Li C-Y, Chen R-M, Chien B-C, Hu R-M, Tsai JJP. *Unsupervised clustering of time series gene expression data based on spectrum processing and autoregressive modeling*. In: Tsai JJP, Ng K-L, editor. *Computational Methods With Applications In Bioinformatics Analysis*. Singapore: World Scientific Publishing Co. Pte. Ltd.; 2017. hal. 1–21.
- Chen P-L, Chen R-M, Chien B-C, Hu R-M, Tsai JJP. *Gene ontology-based analysis of time series gene expression data using support vector machines*. In: Tsai JJP, Ng K-L, editor. *Computational Methods With Applications In Bioinformatics Analysis*. Singapore: World Scientific Publishing Co. Pte. Ltd.; 2017. hal. 22–52.
- Ko G, Kim P-G, Yoon J, Han G, Park S-J, Song W, dkk. *Closha: bioinformatics workflow system for the analysis of massive sequencing data*. *BMC Bioinformatics* [Internet]. 19 Februari 2018 [dikutip 21 November 2018];19(S1):43. Tersedia pada: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2019-3>
- Liu B, Madduri RK, Sotomayor B, Chard K, Lacinski L, Dave UJ, dkk. *Cloud-based bioinformatics workflow platform for large-scale next-generation sequencing analyses*. *J Biomed Inform* [Internet]. Juni 2014 [dikutip 22 November 2018];49:119–33. Tersedia pada: <http://www.ncbi.nlm.nih.gov/pubmed/24462600>
- Biomatters Limited. *Geneious* [Internet]. Biomatters Limited; 2018. Tersedia pada:

[https://assets.geneious.com/installers/geneious/release/Geneious\\_win64\\_11\\_1\\_5\\_with\\_jre.exe](https://assets.geneious.com/installers/geneious/release/Geneious_win64_11_1_5_with_jre.exe)

9. Biomatters Limited. Geneious - Molecular Biology and NGS Analysis Tools [Internet]. Biomatters Limited. 2018 [dikutip 20 Oktober 2018]. Tersedia pada:  
<https://www.geneious.com/geneious/>