

Bioinformatics: Challenges in Integrating Biological Information

Arif Sardi^{1*}

¹Program Studi Biologi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Ar-Raniry, Banda Aceh, Indonesia;

Article History

Received : October 21th, 2022

Revised : November 20th, 2022

Accepted : December 01th, 2022

*Corresponding Author:

Arif Sardi,

Program Studi Biologi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Ar-Raniry, Banda Aceh, Indonesia;

Email: arif.sardi@ar-raniry.ac.id

Abstract: Bioinformatics has become a buzzword in today's world of science. About twenty or thirty years ago, people saw biology and computer science as two completely different fields. However, at present, it seems that a new bridge has been built that connects these two fields, and this new field is called bioinformatics. Bioinformatics is a challenging and exciting new multidisciplinary field. This field combines methods of statistics, mathematics, computer science and information technology to solve problems based on biological data. Advances in experimental technology have generated large amounts of data. Therefore, managing and extracting useful information from these sources is essential if we are to discover new knowledge to advance our understanding of life. In this article, we try to discuss what is bioinformatics?, The basic knowledge needed to understand bioinformatics and the reasons that are very important for biologists to learn and apply bioinformatics. In addition, we also discuss an introduction to biological terminology that will be widely used in the field of bioinformatics. We are confident that this review will provide valuable insights and serve as a starting point for researchers to become acquainted with and study bioinformatics.

Keywords: bioinformatics; biology; computer science.

Pendahuluan

Bioinformatika telah menjadi kata kunci di dunia sains saat ini. Sekitar dua atau tiga dekade yang lalu, orang melihat biologi dan ilmu komputer sebagai dua bidang yang sama sekali berbeda (Sofi dkk, 2022). Satu bidang akan belajar tentang makhluk hidup dan fungsinya sedangkan yang lain akan belajar tentang komputer dan teori-teori yang mendasarinya (Tiwary, 2022). Namun, saat ini, sepertinya sudah terbangun jembatan baru yang menghubungkan kedua bidang ini, dan bidang baru ini dinamakan bioinformatika. Ilmu ini muncul sebagai kombinasi dari Ilmu Komputer dan Biologi (Gauthier, 2018).

Eksperimen biologi hanya dilakukan dengan dua acara sebelum munculnya bioinformatika terdiri dari *in vivo* dan *in vitro*. Bidang bioinformatika dianggap sebagai *in silico* (artinya silikon dalam bahasa Latin), di mana

eksperimen biologis dilakukan dalam *chip* silikon, atau lebih tepatnya dalam mikroprosesor (Warsinske, 2016). Salah satu keuntungan utama dari metode *in silico* adalah kita dapat menjalankan simulasi dan melakukan eksperimen tanpa menggunakan organisme atau reagen apa pun.

Bioinformatika adalah bidang multidisiplin baru yang menantang dan menarik. Bidang ini menggabungkan metode statistika, matematika, teknologi informasi dan ilmu komputer untuk memecahkan masalah berbasis data biologi. Kemajuan teknologi eksperimental telah menghasilkan data dalam jumlah besar. Oleh sebab itu, mengelola dan mengekstraksi informasi yang berguna dari sumber-sumber ini adalah hal yang penting apabila kita ingin menemukan pengetahuan baru untuk meningkatkan pemahaman kita tentang dunia kehidupan.

Hasil dan Pembahasan

Definisi bioinformatika

Bioinformatika adalah area penelitian interdisipliner yang mempertemukan antara ilmu komputer dan ilmu biologi. Bioinformatika sebagai penyatuan biologi dan informatika: bioinformatika melibatkan teknologi yang menggunakan komputer untuk penyimpanan, pengambilan, manipulasi, dan distribusi informasi yang berkaitan dengan makromolekul biologis seperti DNA, RNA, dan protein (Luscombe *et al.*, 2001). Penekanan di sini adalah pada penggunaan komputer karena sebagian besar tugas dalam analisis data genom sangat berulang atau kompleks secara matematis. Penggunaan komputer mutlak diperlukan dalam mengolah genom untuk pengumpulan informasi dan pengembangan pengetahuan.

Bioinformatika berbeda dari bidang terkait yang dikenal sebagai biologi komputasi. Bioinformatika terbatas pada sekuens, struktural, dan analisis fungsional gen dan genom serta produk-produk terkaitnya dan sering dianggap sebagai biologi molekuler komputasi. Namun, biologi komputasi mencakup semua area biologis yang melibatkan perhitungan. Misalnya, Metode data-analitis dan teoritis, pemodelan matematika dan teknik simulasi komputasi untuk mempelajari sistem biologis, perilaku, dan sosial (Lewis University, 2022).

Analisis biologi menghasilkan jumlah data biologis yang sangat banyak dan tentunya akan sangat sulit untuk menganalisisnya menggunakan cara manual. Di sinilah Ilmu Komputer datang untuk membantunya. Berbagai teknik komputasi digunakan untuk menganalisis kumpulan data biologis secara lebih akurat dan efisien melalui proses otomatis. Oleh karena itu, bioinformatika dapat dianggap sebagai bidang ilmu data untuk memecahkan masalah dalam biologi dan kedokteran.

Pentingnya belajar dan menerapkan bioinformatika

Bioinformatika telah menjadi suatu ilmu interdisipliner, seorang ahli biologi yang memiliki pengetahuan dalam bioinformatika akan mempunyai keuntungan lebih baik dalam memahami data penelitian yang diperoleh. Bioinformatika tidak hanya menjadi penting untuk penelitian dasar genom dan biologi

molekular, tetapi memiliki dampak besar pada banyak bidang bioteknologi dan ilmu biomedis (Khan, 2018; Shamer dkk, 2016).

Aplikasi bioinformatika dalam bidang biologi dan kedokteran adalah dalam desain obat berbasis pengetahuan genetika, analisis DNA forensik, dan bioteknologi pertanian. Studi komputasi interaksi protein-ligan memberikan dasar rasional untuk identifikasi cepat penemuan obat sintetik baru. Pengetahuan tentang struktur tiga dimensi dari protein memungkinkan molekul yang dirancang mampu mengikat ke lokasi reseptor protein target dengan afinitas dan spesifisitas yang tinggi. Pendekatan berbasis informatika ini secara signifikan mengurangi waktu dan biaya yang diperlukan untuk mengembangkan obat-obatan dengan potensi yang lebih tinggi, efek samping yang lebih sedikit, dan toksisitas yang lebih rendah daripada menggunakan pendekatan *trial and error* (Zloh dan Kirton, 2018)

Perlu kita sadari bersama bahwa pemahaman terkait genomik dan bioinformatik, saat ini siap untuk merevolusi sistem perawatan kesehatan kita dengan mengembangkan obat-obatan yang disesuaikan dengan kebutuhan. Berbagai penelitian sedang dan telah dilakukan untuk mengidentifikasi perubahan genetik pada pasien. Hal ini memungkinkan para tenaga medis untuk melakukan perawatan yang lebih baik dan bahkan langkah-langkah pencegahan yang terukur. Dengan mengidentifikasi perubahan genetik pada pasien maka dapat dilakukan perawatan yang lebih baik dan bahkan langkah-langkah pencegahan yang mungkin. Jenis kanker ataupun penyakit tertentu, yang disebabkan oleh perubahan genetik dapat diidentifikasi sebelumnya dan dapat diobati sebelum kondisinya memburuk (Shamer dkk., 2016).

Hasil dari analisis filogenetik molekuler telah diterima sebagai bukti di pengadilan pidana. Beberapa statistik Bayesian yang canggih dan metode berbasis kemungkinan (*likelihood-based*) untuk analisis DNA telah diterapkan dalam analisis identitas forensik (Liu & Harbison, 2018). *Tools* bioinformatika juga digunakan dalam pertanian. *Database* genom tanaman dan analisis profil ekspresi gen telah memainkan peran penting dalam pengembangan varietas tanaman baru yang memiliki produktivitas lebih tinggi dan lebih tahan terhadap penyakit (Aslam *et al.*, 2017).

Pengetahuan dasar yang diperlukan dalam memahami bioinformatika

Langkah awal sebelum mempelajari bioinformatika, kita harus belajar sedikit tentang biologi; genetika dan genomik secara spesifik. Ini akan termasuk mempelajari tentang gen, DNA, RNA, struktur protein serta berbagai proses sintesis lainnya. Setelah itu, kita harus mempelajari tentang sekuens biologis (misalnya, sekuens yang ditemukan dalam DNA, RNA dan protein) dan teknik untuk menemukan serta menganalisis berbagai pola dan situs informatif di dalamnya.

Pemahaman yang baik dalam statistic penting karena nantinya kita akan berurusan dengan sejumlah besar data. Hal ini diperlukan dalam hal menganalisis data sesuai dengan persyaratan spesifik. Tentu saja kita juga membutuhkan keterampilan pemrograman yang baik. Python, R, dan Bash adalah bahasa pemrograman yang paling umum digunakan dalam analisis data biologis. Mengenai hal mana yang akan digunakan terlebih dahulu tergantung pada tujuan dan data yang dianalisis. Kita juga dapat menggunakan bahasa lain seperti C/C++ dan Java. Setelah memiliki pemahaman dasar tentang konsep-konsep dasar, kita dapat menjelajahi area lain seperti bioinformatika struktural, sistem biologi dan jaringan biologis (Bonnal *et al.*, 2019; omgenomics, 2017).

Manusia adalah makhluk yang menarik dan tentu saja genomnya lebih menarik. Seluruh genom yang disimpan dalam molekul DNA adalah *blueprint* tentang bagaimana kemungkinan untuk menyandikan data yang sangat besar dalam satu entitas kecil (DNA) akan tetapi mengkode mereka secara tepat untuk menghasilkan manusia yang unik dengan karakter unik yang dimiliki mereka sendiri. Namun, perubahan tertentu dalam ekspresi gen dapat menyebabkan penyakit genetik yang fatal. Sistem layanan kesehatan memerlukan langkah-langkah untuk mengidentifikasi penyakit-penyakit semacam itu dan menyediakan perawatan serta tindakan pencegahan untuk membantu menyelamatkan nyawa manusia (Gaugner *et al.*, 2017).

Bioinformatika telah terbukti memiliki potensi besar untuk mengidentifikasi penyakit sebelum terjadi, menentukan pengobatan dan membantu membuat kehidupan manusia lebih

baik. Dengan inspirasi dan pengetahuan ilmu komputer, bidang-bidang seperti teknologi gen, ilmu pengobatan dan perawatan kesehatan dapat berevolusi dari menyembuhkan setiap pasien menjadi penyembuhan seluruh populasi.

Pengantar singkat untuk terminologi biologis

Data yang paling melimpah dalam Bioinformatika terdiri dari urutan DNA. DNA terdiri dari empat basa, A, G, C, dan T. Sekuens DNA dapat berupa sekuens pengkode ataupun non pengkode. Mungkin ada beberapa ribu basa dalam gen dan beberapa juta basa dalam genom bakteri tertentu. Ada sekitar 3 miliar basa dalam genom manusia (Cornell University, 2022). Sekuens protein adalah sekuens asam amino dan ada 20 jenis asam amino di alam yang membuat sekuens protein menjadi rangkaian yang disusun dari 20 macam huruf alfabet.

Rata-rata panjang protein adalah sekitar 300 asam amino. Ada jutaan urutan protein yang diketahui. Genotip adalah DNA dan sekuens protein bersama-sama menentukan fenotip (penampilan suatu organisme dan bagaimana ia menjalankan fungsinya). Struktur protein memainkan peran penting dalam menentukan fungsinya (National Human Genome Research Institute, 2022). Prediksi struktur protein dari urutan protein primer adalah salah satu tantangan terbesar yang belum terpecahkan dalam biofisika dan bioinformatika saat ini. Ada teknik *ab initio* yang didasarkan pada minimisasi energi, dan ada teknik berbasis pengetahuan seperti pemodelan homologi dan *threading*.

Ekspresi gen adalah proses menggunakan informasi dalam DNA untuk mensintesis *messenger*-RNA (transkripsi). mRNA yang ditranskripsi kemudian digunakan untuk mensintesis protein yang sesuai melalui proses translasi. Seluruh aliran kerja ini juga dikenal sebagai dogma sentral biologi molekular. Regulasi ekspresi gen, dengan kata lain, menentukan gen mana yang diekspresikan dalam jaringan atau kondisi tertentu dan gen mana yang tidak, hal ini juga merupakan tantangan dalam biologi molekular (Ritz, 2018).

Bidang transkriptomik menangani masalah ini dan kemajuan terbaru dalam bioteknologi memungkinkan para peneliti untuk menguji ekspresi gen dalam skala besar. *Microarrays* (*gene expression arrays* atau *gene chips*) memberikan sejumlah besar data tentang

aktivitas gen dalam kondisi tertentu atau dalam sampel biologis tertentu. Tes interaksi juga memberikan informasi tentang protein mana yang berinteraksi satu sama lain dan memungkinkan para peneliti untuk mendapatkan pandangan sampai tingkat sistem sel.

Analisis data biologi molekuler memiliki banyak tantangan seperti redundansi dan multiplisitas (keberagaman), *noise* (gangguan) dan ketidaklengkapan. Ada ribuan sumber data untuk data biologi molekuler. Koleksi database online Nucleic Acids Research (NAR) pada November 2019 mencatat 1.613 database yang dipilih dengan cermat yang mencakup aspek biologi molekuler dan sel. Basis data yang paling populer adalah NCBI Entrez, EBI Ensemble dan basis data KEGG.

Keterbatasan bioinformatika

Menyadari keterbatasan, menghindari ketergantungan, ekspetasi berlebihan terhadap output bioinformatika penting dilakukan setelah mengenali kekuatan bioinformatika. Bahkan, bioinformatika memiliki sejumlah keterbatasan yang melekat. Bioinformatika dan biologi eksperimental adalah kegiatan yang independen, tetapi saling melengkapi. Bioinformatika bergantung pada sains eksperimental untuk menghasilkan data mentah yang akan dianalisis. Interpretasi yang berguna dari data eksperimen dan petunjuk penting untuk penelitian eksperimental lebih lanjut.

Prediksi bioinformatika bukanlah bukti formal dari setiap konsep. Mereka tidak menggantikan metode penelitian eksperimental tradisional yang menguji hipotesis sebenarnya. Selain itu, kualitas prediksi bioinformatika tergantung pada kualitas data dan kecanggihan algoritma yang digunakan. Urutan data dari analisis berkapasitas tinggi sering mengandung kesalahan. Jika urutannya salah atau anotasi salah, hasil dari analisis hilir juga menyesatkan. Sangat penting untuk mempertahankan perspektif yang realistis tentang peran bioinformatika.

Bioinformatika sama sekali bukan bidang yang sempurna. Kebanyakan algoritma tidak memiliki kemampuan dan kecanggihan untuk benar-benar mencerminkan kenyataan. Mereka sering membuat prediksi yang salah yang tidak masuk akal ketika ditempatkan dalam konteks biologis. Kesalahan dalam penyetaraan sekuens,

misalnya, dapat mempengaruhi hasil analisis struktural atau filogenetik. Hasil perhitungan juga tergantung pada daya komputasi yang tersedia.

Banyak algoritma yang akurat tetapi menyeluruh tidak dapat digunakan karena memperlambat komputasi. Algoritma yang kurang akurat tetapi lebih cepat terkadang harus digunakan. Pertaruhan antara akurasi dan kelayakan komputasi akan terjadi (Xiong, 2012). Oleh karena itu, penting untuk mengingat potensi kesalahan yang dihasilkan oleh program bioinformatika. Perhatian harus selalu dilakukan ketika menafsirkan hasil prediksi. Ini adalah praktik yang baik untuk menggunakan banyak program, jika tersedia, dan melakukan beberapa evaluasi. Prediksi yang lebih akurat sering dapat diperoleh jika seseorang menarik konsensus dengan membandingkan hasil dari algoritma yang berbeda

Kesimpulan

Bioinformatika adalah bidang yang menyimpan potensi besar untuk merevolusi penelitian biologi dalam beberapa dekade mendatang. Saat ini, bidang tersebut sedang mengalami ekspansi besar-besaran. Selain menyediakan alat komputasi yang lebih andal dan lebih ketat untuk analisis sekuens, struktural, dan fungsional, tantangan utama untuk pengembangan bioinformatika di masa depan adalah mengembangkan alat untuk menjelaskan fungsi dan interaksi semua produk gen dalam sel. Hal ini menyajikan tantangan yang luar biasa karena membutuhkan integrasi bidang pengetahuan biologis yang berbeda dan berbagai alat matematika dan statistik yang kompleks. Pemahaman yang lebih dalam tentang fungsi selular, model matematika diperlukan untuk mensimulasikan berbagai reaksi intraselular dan interaksi di seluruh tingkat sel. Simulasi molekuler dari semua proses selular ini disebut biologi sistem. Mencapai tujuan ini akan mewakili lompatan besar menuju pemahaman penuh terkait sistem kehidupan. Itulah mengapa simulasi tingkat sistem dan integrasi dianggap sebagai masa depan bioinformatika. Membuat model jaringan yang sedemikian kompleks dan membuat prediksi tentang perilaku mereka menghadirkan tantangan dan peluang luar biasa bagi para ahli bioinformatika. Tujuan akhir dari

upaya ini adalah untuk mengubah biologi dari ilmu kualitatif menjadi ilmu kuantitatif dan prediksi.

Ucapan Terima kasih

Terima kasih kepada pihak yang telah membantu baik secara moral maupun material sehingga penulisan artikel ini dapat diselesaikan dengan baik.

Referensi

- Aslam, Z., Khattak, J.Z.K., Ahmed, M., & Asif, M. (2017). A Role of Bioinformatics in Agriculture. In: Ahmed M, Stockle C (eds) *Quantification of Climate Variability, Adaptation and Mitigation for Agricultural Sustainability*. Springer, Cham; pp 413-434
- Bonnal, R.J.P., Yates, A., Goto, N., Gautier, L., Willis, S., Fields, C., Katayama, T., & Prins, P. (2019). Sharing Programming Resources *Between Bio* Projects*. In: Anisimova M. (eds) *Evolutionary Genomics. Methods in Molecular Biology*, 1910. Humana, New York, NY
- Cornell University (2022). *Bacterial Genomes*. <https://micro.cornell.edu/>
- Gauthier, J., Vincent, A.T., Charetee, S.J., & Derome, N. (2018). A brief history of bioinformatics. *Briefings in Bioinformatics*, Volume 20, Issue 6.
- Khan, N.T. (2018). The Emerging Role of Bioinformatics in Biotechnology. *Journal of Biotechnology and Biomedical Science*. 1(3).
- Lewis University (2022). *Computational Biology and Bioinformatics: Two Fields Changing the World*. <https://online.lewisu.edu/>
- Liu, Y.Y., & Harbison, S.A. (2018). A review of bioinformatic methods for forensic DNA analyses. *Forensic Science International: Genetics* 33 (2018) 117–128
- Luscombe, N.M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med*. 40(4):346-358.
- National Human Genome Research Institute. (2022). *Genotype*. <https://www.genome.gov/genetics-glossary>
- Omgenomics (2017). *For bioinformatics, which language should I learn first*. omgenomics.com
- Ritz, A. (2018). Programming the Central Dogma: An Integrated Unit on Computer Science and Molecular Biology Concepts. *SIGCSE '18: Proceedings of the 49th ACM Technical Symposium on Computer Science Education*
- Shamer, K., Badgeley, M.A., Miotto, R., Glicksberg, B.S., Morgan, J.W., & Dudley, J.T. (2016). Translational bioinformatics in the era of real-time biomedical, health care and wellness data streams. *Briefings in Bioinformatics*, Volume 18, Issue 1
- Sofi, M.Y., Shafi, A., & Masoodi, K.Z. (2022). *Bioinformatics for Everyone*. Academic Press. London. ISBN: 978-0-323-91128-3, pp: 1
- Tiwary, B.K. (2022). *Bioinformatics and Computational Biology*. Springer Nature Singapore Pte Ltd. Singapore. ISBN 978-981-16-4240-1, pp: 2
- Warsinske, H. (2016). *In silico biology: How Math and Computer Science Teach Us About Life*. Misciwriter.com
- Xiong, J. (2012). *Essential Bioinformatics*. Cambridge University Press. Texas
- Zloh, M., & Kirton, S.B. (2018). The Benefits of in Silico Modeling to Identify Possible *Small-Molecule Drugs and Their Off-Target Interactions*. *Future Med Chem.*; 10(4):423-432