

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kehidupan manusia merupakan rangkaian dari berbagai macam penerimaan, pengolahan dan penyampaian informasi, baik itu cara bicara, pola hidup, sejarah, dan semua yang berhubungan dengan kegiatan manusia. Penyampaian informasi tidak hanya disampaikan di tahap komunikasi manusia pada umumnya, misalnya berbicara atau menyampaikan sesuatu secara visual, penyampaian informasi juga terjadi di tahap molekular sel, yaitu DNA. DNA adalah struktur informasi unik terkecil yang dimiliki oleh setiap organisme dan DNA berperan penting dalam seluruh kehidupan yang terjadi di muka bumi ini.

Ilmu yang mempelajari tentang DNA secara spesifik, salah satunya, adalah mikrobiologi. Mikrobiologi sudah berkembang pesat dari abad ke abad dengan meneliti bagaimana suatu informasi dari DNA tiap organisme dapat diwariskan atau bahkan dimodifikasi. Seiring dengan perkembangan ilmu mikrobiologi, kehidupan di dunia juga mengalami perkembangan dan perubahan yang kita sebut dengan evolusi. Evolusi yang kita kenal berhubungan erat dengan perubahan penyampaian informasi DNA oleh tiap organisme. Informasi DNA atau *sequence* dari DNA yang ada mengalami perubahan yang disebut mutasi [2].

Menurut Deepa Agashe dalam penelitiannya di jurnal biologi molekular dan evolusi, mutasi atau perubahan informasi sel terjadi berulang kali pada banyak varian gen [3]. Dengan semakin banyaknya perubahan informasi sel dari tiap organisme, wujud atau cara bertahan hidup sebuah organisme semakin beragam. Dalam hal ini, ilmu mikrobiologi mendapat tantangan dalam membandingkan organisme satu dengan yang lainnya; apakah suatu organisme tersebut memiliki kerabat atau hubungan erat dengan organisme yang memiliki wujud ataupun cara bertahan hidup yang mirip.

Dengan berkembangnya data tentang informasi DNA yang dimiliki, ilmu mikrobiologi saja tidak cukup dalam mengolah data dan melakukan perbandingan antar *sequence* DNA dari organisme-organisme yang ingin diteliti. Untuk itu, diperlukan ilmu komputasi informatika yang dapat membantu menganalisa

sequence DNA yang ada, membandingkan, dan dapat mengambil kesimpulan dari hasil analisa tersebut.

Bioinformatika adalah aplikasi dalam bidang sains biologi yang berfokus pada analisis data sekuens biologi. Bioinformatika cenderung digunakan dalam komputasi mikrobiologi dan dengan adanya bioinformatika, perkembangan pesat dari data *sequence* DNA yang muncul hampir tiap hari dapat diolah dan ditangani secara cepat. Bioinformatika mempunyai peran penting dalam pensejajaran atau perbandingan *sequence* DNA. Dalam *sequence* DNA, bioinformatika menggunakan algoritma untuk mensejajarkan suatu DNA dan mencari kecocokan dari DNA tersebut sehingga kemiripan dan korelasi antar DNA dari organisme dapat dianalisis dengan baik.

Untuk melakukan pensejajaran sekuens DNA, program dari ilmu bioinformatika menggunakan berbagai macam algoritma. Pensejajaran DNA memiliki dua macam teknik, yaitu pensejajaran global ataupun lokal. Dalam pensejajaran global, seluruh karakter DNA diproses sementara dalam pensejajaran lokal hanya sebagian dari karakter DNA yang diproses [2]. Beberapa algoritma yang digunakan dalam pensejajaran lokal antara lain adalah Smith-Waterman, FASTA, BLAST, dan masih banyak algoritma yang sedang dikembangkan. Sementara untuk pensejajaran global, algoritma Needleman-Wunsch masih kerap digunakan dan juga dikembangkan agar lebih efisien [4]. Di sisi lain, pensejajaran global mengalami banyak perkembangan dan banyak algoritma-algoritma baru diciptakan dengan mengadaptasi dari beberapa algoritma dasar, salah satunya adalah metode Lempel-Ziv, yaitu metode teknik *data compression* yang kemudian diadaptasi untuk pensejajaran DNA [5].

Kecepatan proses dan ketepatan pensejajaran, menjadi tolak ukur penting terhadap pensejajaran yang baik oleh suatu algoritma. Untuk itu, mencari keseimbangan yang optimal terhadap kecepatan dan ketepatan pensejajaran menjadi dasar pembuatan aplikasi bioinformatika ini dengan mengimplementasikan dan menganalisa kinerja dua algoritma pensejajaran global yaitu Needleman-Wunsch dan Lempel-Ziv.

Dari studi literatur yang telah dilakukan kinerja kedua algoritma ini belum pernah dibandingkan. Oleh karena itu, tugas akhir ini mengangkat hal baru yaitu

meneliti lebih jauh mengenai fungsi, analisa, dan perbandingan dua algoritma ini, khususnya mengenai kecepatan dan ketepatan serta implementasinya di dalam sebuah program. Harapannya adalah tugas akhir ini dapat berkontribusi kepada bidang bioinformatika, khususnya, pemahaman tentang kinerja dua algoritma tersebut dan, umumnya, bidang ilmu informatika serta penerapannya dalam biologi seluler.

1.2 Rumusan Masalah

Dari latar belakang yang sudah disampaikan, maka dapat dirumuskan beberapa masalah yang menjadi dasar pembuatan program bioinformatika, yaitu:

1. Bagaimana membuat implementasi Needleman-Wunsch dan Lempel-Ziv ke dalam suatu program?
2. Bagaimana menentukan algoritma yang tepat untuk tiap kasus pensejajaran DNA yang berbeda?
3. Bagaimana proses dan analisa dari kedua algoritma dapat membantu penelitian biologi molekuler ?

1.3 Tujuan Pembahasan

Tujuan pembahasan dari analisa algoritma bioinformatika, yaitu:

1. Mengembangkan aplikasi bioinformatika yang memiliki proses pengolahan data sekuens dengan algoritma Needleman-Wunsch dan Lempel-Ziv.
2. Memberikan analisis dari proses kinerja tiap algoritma untuk menentukan algoritma yang tepat di tiap kasus berbeda.
3. Memberikan analisis dari kedua algoritma dalam proses pensejajaran DNA untuk membantu memberikan hasil yang cepat dan tepat dalam penelitian biologi molekuler.

1.4 Ruang Lingkup

Ruang lingkup untuk program ini meliputi 3 hal yaitu aplikasi, perangkat lunak, dan perangkat keras.

1. Ruang lingkup dari pembuatan aplikasi bioinformatika ini adalah:

- a. Aplikasi ini akan memiliki dua metode pengolahan sekuens DNA berdasarkan dua algoritma yaitu Needleman-Wunsch dan Lempel-Ziv.
 - b. Tiap metode pengolahan sekuens DNA akan dites dari maksimal 1000 karakter hingga di atas 1000 karakter.
2. Perangkat lunak yang digunakan:
 - a. Microsoft Office 2013
 - b. Netbeans
 - c. JDK 8.0.2
 - d. Windows 10 Pro
 3. Perangkat keras yang digunakan untuk mendukung pembuatan aplikasi ini adalah:
 - a. Prosesor Intel Celeron @ 1.50GHz, 64bit.
 - b. RAM 4GB
 - c. HDD 500GB

1.5 Sumber Data

Sumber data untuk mendukung analisis dari program bioinformatika adalah *string* barisan DNA yang didapat dari website National Center for Biotechnology Information (NCBI). String yang digunakan akan diambil secara acak sebanyak 100 buah sekuens DNA dengan panjang kurang dari 1000 karakter untuk 50 *dataset* pertama dan panjang lebih dari 1000 karakter untuk 50 *dataset* kedua. *Dataset* ini digunakan untuk pengujian kecepatan dan skor dari hasil pensejajaran DNA menggunakan algoritma Needleman-Wunsch dan Lempel-Ziv.

1.6 Sistematika Penyajian

Sistematika penyajian untuk laporan analisa program bioinformatika dengan algoritma pensejajaran DNA ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini menjelaskan dasar-dasar pembahasan mengenai DNA, proses penyaluran informasi organisme, barisan DNA, dan peran bioinformatika dalam pensejajaran barisan DNA.

BAB 2 KAJIAN TEORI

Berisi teori-teori yang menjadi dasar penyusunan laporan.

BAB 3 ANALISIS DAN PERANCANGAN

Berisi analisis dari tiap algoritma, serta perancangan program bioinformatika yang memiliki empat algoritma pensejajaran barisan DNA.

BAB 4 IMPLEMENTASI

Berisi penerapan dari program terhadap barisan data DNA.

BAB 5 PENGUJIAN

Berisi pengujian dari tiap algoritma yang ada, serta pengujian keseluruhan program bioinformatika.

BAB 6 SIMPULAN DAN SARAN

Berisi kesimpulan dari laporan, serta saran dari penulis untuk program yang dirancang.

