
PROSES IMPLEMENTASI BIOINFORMATIKA PADA DIGITALISASI DATA GENETIKA MANUSIA

Herman

Program Magister Informatika, Fakultas Teknologi Industri,
Universitas Ahmad Dahlan

Email: hermankaha@mti.uad.ac.id

Sunardi

Program Studi Teknik Elektro, Fakultas Teknologi Industri,
Universitas Ahmad Dahlan

Email: sunardi@mti.uad.ac.id

Tri Stiyo Famuji

Program Magister Informatika, Fakultas Teknologi Industri,
Universitas Ahmad Dahlan

Email: 2108048018@webmail.uad.ac.id

ABSTRAK

Perkembangan ilmu biologi menjadi pesat khususnya genetika yang mengakibatkan lonjakan berbagai eksperimen data genetika manusia sebagai pembawa informasi genetik untuk analisis pewarisan sifat dan duplikasi, serta untuk kegiatan forensik. Pada nukleus sel atau genetika terdiri dari karbohidrat, protein maupun lemak, dan terdiri dari zat yang mempunyai kandungan fosfor yang tinggi. Zat itu terdapat di dalam nukleus sel yang disebut *nuklein*. Nama ini kemudian diubah menjadi asam nukleat. Asam nukleat terdiri dari dua tipe, yaitu Asam Deoksiribonukleat (ADN) atau *Deoxyribo Nucleic Acid* (DNA) dan Asam Ribonukleat (ARN) atau *Ribonucleic Acid* (RNA). Digitalisasi data genetika dibutuhkan untuk mempermudah penelitian dan pengembangan. Bioinformatika merupakan data hasil eksperimen dari laboratorium para peneliti ahli biologi molekuler ataupun biomedis untuk mempermudah pengolahan data genetika manusia dengan teknologi komputasi. Data digital dari genetika dapat disimpan dalam *database* dengan format tertentu. Penelitian ini bertujuan untuk menjelaskan langkah-langkah untuk digitalisasi data dari genetika manusia dari sampel biologi menjadi data digital. Bentuk dari data digital genetika manusia dapat digunakan untuk penelitian oleh para ahli biologi menggunakan perangkat lunak yang dapat membaca file dengan format FASTA. FASTA merupakan salah satu jenis file untaian dari beberapa jenis file format untaian protein yang tersedia di GenBank (*database* penyedia untaian protein). Data digital dari genetika akan dapat digunakan pada penelitian selanjutnya oleh ahli biologi tanpa harus mengambil sampel biologi.

Kata kunci: bioinformatika, DNA, FASTA, genetika manusia, RNA.

ABSTRACT

The rapid development of biology, especially genetics, has resulted in a surge of various experiments on human genetic data as carriers of genetic information for the analysis of inheritance and duplication, as well as for forensic activities. The cell nucleus, or the genetic material, consists of carbohydrates, proteins, and fats, as well as substances that have a high phosphorus content. This substance is contained in the cell nucleus and is called nuclein, which was later changed to nucleic acid. Nucleic acids consist of two types, namely deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The digitization of genetic data is needed to facilitate research and development. Bioinformatics is the use of experimental data from the laboratories of molecular biologists or biomedical researchers to facilitate the processing of human genetic data

with computational technology. Digital data from genetics can be stored in a database in a specific format. This study aims to explain the steps for digitizing human genetic data from biological samples into digital data. The digital form of human genetic data can be used for research by biologists using software that can read files in the FASTA format. FASTA is one type of file format among several types of protein strand file formats available in GenBank (a protein strand provider database). The digital data from genetics will be used in future research by biologists without having to take biological samples.

Keywords: *bioinformatics, DNA, FASTA, human genetics, RNA.*

1. PENDAHULUAN

Perkembangan ilmu kedokteran tidak lepas dari perkembangan biologi. Secara umum, bioteknologi dan norma hak asasi manusia dapat berkembang bersama. Bioteknologi membawa risiko dan manfaat bagi realisasi hak asasi manusia. Di sisi lain, jalur perkembangan bioteknologi dapat dibentuk oleh norma-norma hak asasi manusia, terutama yang diakui secara universal sebagai kritis bagi manusia.

Perkembangan biologi sangat pesat, terutama di bidang informasi genetik manusia. Pesatnya perkembangan biologi telah menyebabkan peningkatan berbagai percobaan dengan informasi genetik manusia. Genetika adalah sesuatu yang telah diperhatikan orang selama berabad-abad. Bahkan sebelum Mendel mempresentasikan teorinya, orang tua sudah marah dengan adanya pernikahan saudara kandung [1]. Selain itu, konsep genetika juga diterapkan saat memilih calon jodoh bagi anaknya. Mereka justru memilih calon jodoh dari antara anak-anaknya, terlepas dari apakah mereka secara genetik baik atau buruk, bahkan sampai menggali informasi tentang keluarga mempelai wanita, terlepas dari apakah seseorang itu cacat atau tidak. Hal ini diwariskan secara turun-temurun untuk menghindari keturunan yang cacat fisik atau mental.

Data genetika manusia didapatkan dari hasil penelitian oleh para ahli dalam bidang ilmu biologi yang kedepannya dapat diolah menjadi informasi yang terdigitalisasi sehingga memudahkan untuk penelitian selanjutnya. Data tersebut perlu diolah menjadi informasi bermanfaat untuk masa depan. Data genetika manusia memiliki kegunaan diantaranya sebagai pembawa informasi genetik, untuk analisis pewarisan sifat dan duplikasi diri, serta fungsi DNA untuk kegiatan forensik. Teknologi DNA bergerak menuju digitalisasi. Penggunaan teknologi komputasi mutakhir dan program perangkat lunak mendorong para peneliti memecahkan kode genom manusia, merancang obat baru, dan menulis kode DNA yang dimodifikasi [2].

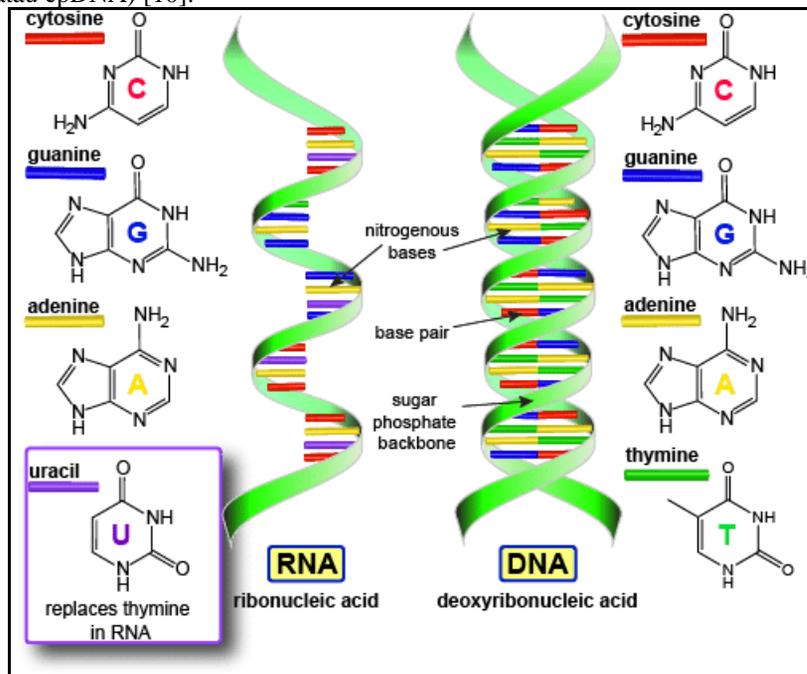
Pertama-tama, pembahasan genetika bukanlah molekuler. Genetika awalnya mempelajari bagaimana gen diturunkan dari orang tua ke keturunannya selama reproduksi dan bagaimana gen bekerja sama untuk mengontrol sifat-sifat seperti tinggi badan, warna rambut, dan struktur gigi. Seiring berkembangnya ilmu teknologi, informasi semakin dibutuhkan oleh masyarakat luas tak terkecuali informasi dalam dunia biologi yang menjadi topik yang sangat diminati. Pada tahun 1979 Paulen Hogeweg menciptakan ilmu bioinformatika [3]. Bioinformatika dapat dijabarkan sebagai sistem dari teknologi komputer untuk mengelola informasi biologi [4]. Bioinformatika mempunyai tujuan utama untuk menunjang proses biologi. Selain itu bioinformatika juga dapat digunakan untuk menyelesaikan masalah yang berkaitan dengan manusia.

Penelitian ini membahas lebih rinci tentang bioinformatika yang dapat digunakan untuk menyelesaikan pembuatan data genetika manusia dan penerapannya dengan pendekatan pada ruang lingkup ilmu komputer. Untuk memudahkan kegiatan-kegiatan pengkajian data genetika manusia dibutuhkan platform untuk penyimpanan data hasil kegiatan tersebut. Tujuan penelitian ini adalah menerangkan proses digitalisasi data genetika manusia.

1.1. Genetika Manusia

Genetika adalah cabang biologi yang menjelaskan persamaan dan perbedaan ciri-ciri yang diwarisi oleh makhluk hidup [5]. Genetika juga menjelaskan hubungan keturunan antara orang tua dan keturunan serta peran materi genetik [6]. Friedrich Miescher meneliti komposisi kimia inti sel atau *nucleus cell* pada tahun 1869 dan menemukan bahwa inti sel tidak terdiri dari karbohidrat, protein atau lemak, tetapi dari zat dengan kandungan fosfor yang sangat tinggi. Karena zat tersebut terkandung di dalam inti sel, maka disebut inti sel [1]. Nama itu kemudian diubah menjadi asam nukleat. Asam nukleat terdiri dari dua jenis yaitu asam *deoksiribonukleat* atau asam *deoksiribonukleat* (DNA) dan asam ribonukleat atau asam *ribonukleat* (RNA).

DNA dan RNA adalah polimer linier (*polinukleotida*) yang terdiri dari subunit protein atau monomer nukleotida. DNA adalah materi hereditas atau warisan di hampir semua organisme hidup. DNA seseorang, atau genom, bersifat unik dan dianggap sebagai informasi pribadi atau pribadi, yang mencakup informasi tentang keluarga dan leluhur seseorang [2]. Komponen penyusun nukleotida terdiri dari tiga jenis molekul protein, yaitu gula *pentosa* (*deoksiribosa* pada DNA atau *ribosa* pada RNA), basa nitrogen, dan gugus fosfat (Gambar 1). Basa yang ditemukan pada nukleotida adalah basa *purin* (*adenin* = A, *guanin* = G) dan basa *pirimidin* yaitu *cytosin* = C, *tymin* = T, *urasil* = U [7]. Untaian rantai DNA mengandung gen yang diperintahkan untuk menghasilkan protein. Kemudian protein melakukan pekerjaan di sel dan tubuh kita [8]. Protein adalah senyawa organik yang tersusun dari asam-asam amino yang tersusun secara linier dan melingkar membentuk lingkaran. Urutan asam amino suatu protein didefinisikan sebagai urutan gen yang tertulis dalam kode genetik [9]. DNA biasanya terkandung dalam inti sel (dikenal sebagai DNA inti), tetapi sejumlah kecil juga dapat ditemukan di *mitokondria* (disebut DNA *mitokondria* atau mtDNA) atau di *organel* khusus seperti plastida tanaman (disebut DNA *kloroplas* atau cpDNA) [10].



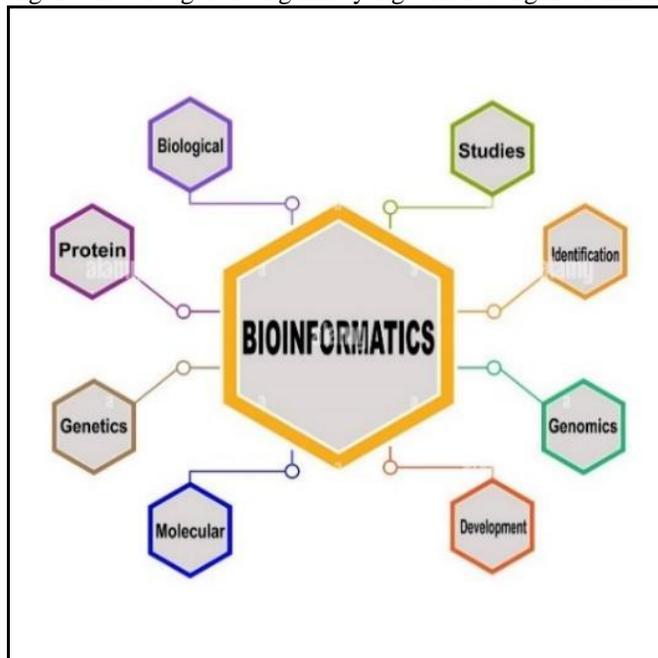
Gambar 1. DNA dan RNA [11]

1.2. Bioinformatika

Bioinformatika digambarkan sebagai kombinasi dari dua ilmu yaitu biologi molekuler dan teknologi informasi. Tujuan pengembangan ilmu baru ini adalah untuk memudahkan pengolahan

data sekuen biologi dengan menggunakan metode komputer terkini untuk mendapatkan informasi atau masukan yang berarti bagi perkembangan kedokteran atau bidang biologi lainnya. Ilmu bioinformatika telah mencapai tonggak yang sangat signifikan sejak Proyek Genom Manusia dimulai pada awal 1990-an. Proyek ambisius ini bertujuan untuk mengurutkan asam nukleat dan protein dari genom dan proteom manusia, membuka kemungkinan baru anotasi informatif untuk pengembangan ilmu kehidupan atau organisme [12].

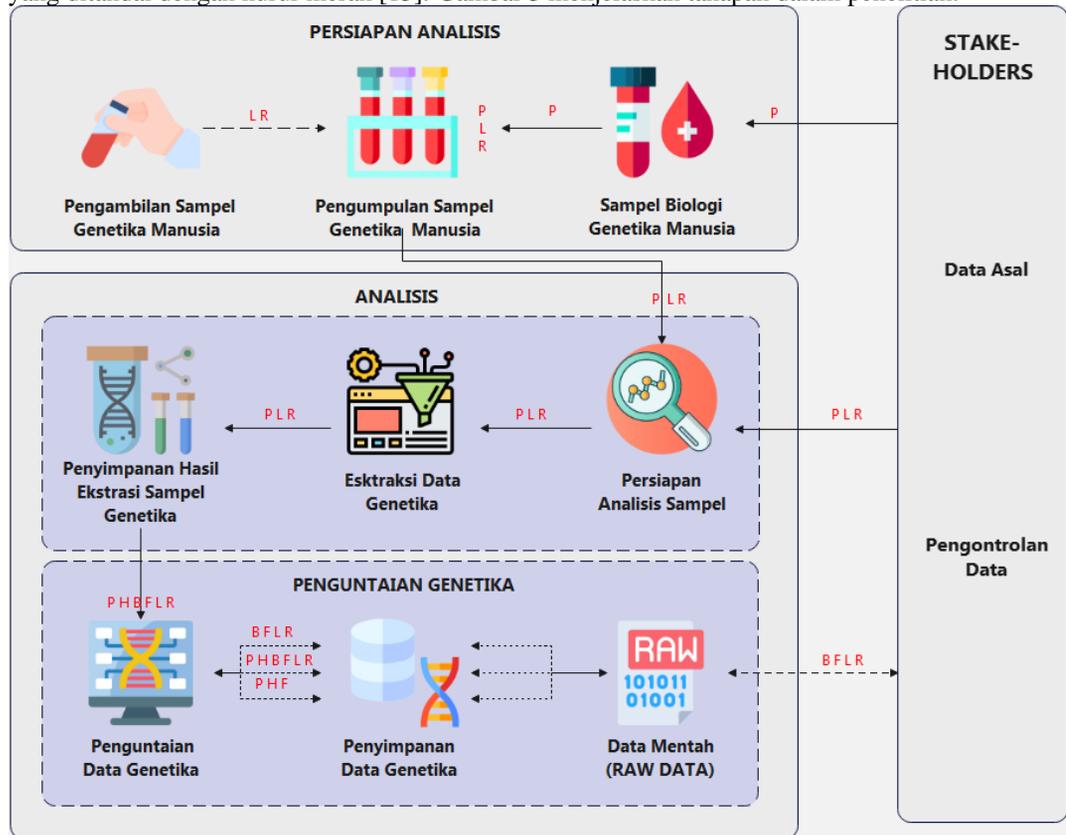
Bioinformatika adalah kombinasi dari biologi, ilmu komputer, matematika dan disiplin ilmu terkait lainnya yang menjadi disiplin ilmu tersendiri. Akses ke sumber bioinformatika canggih dengan cara yang mudah dan efisien telah menjadi sangat penting untuk penelitian dalam ilmu yang menuntut studi sistematis dan holistik organisme hidup [13]. Tujuan utamanya adalah untuk menawarkan perspektif baru untuk mencapai perspektif global untuk mendukung pengembangan bioteknologi di masa depan [6]. Bioinformatika diperlukan untuk penelitian pengeditan gen. Pertama, jelajahi alat komputasi canggih yang dikembangkan untuk studi pengeditan gen. Bioinformatika adalah ilmu yang menggabungkan biologi dan komputer untuk penyimpanan informasi, pengambilan informasi, pemrosesan data, dan distribusi informasi yang berkaitan dengan makromolekul biologis seperti DNA, RNA, dan protein. Bioinformatika memperoleh informasi dari analisis komputer data biologis. Ini bisa berupa informasi yang disimpan dalam kode genetik, tetapi juga hasil eksperimen dari berbagai sumber, statistik pasien, dan literatur ilmiah. Database biologis adalah kumpulan data yang terorganisir, atau lebih tepatnya informasi biologis yang terkait secara logis, yang bisa besar atau sangat kecil [14]. Gambar 2 menunjukkan gabungan dari berbagai bidang ilmu yang terkait dengan bioinformatika.



Gambar 2. Gabungan Ilmu yang Terkait dengan Bioinformatika [6]

2. METODOLOGI PENELITIAN

Sistem informasi genetika adalah sistem cyber-fisik yang dibagi menjadi tiga fase di mana orang berinteraksi dengan semua bagian sistem. Tahap persiapan analitik meliputi pengumpulan sampel biologis, pengumpulan sampel genetik manusia dan penyimpanan sampel genetik manusia. Fase analitik meliputi persiapan laboratorium basah, sekuensing DNA, dan subfase dari pipa bioinformatika. Pada tahap analisis, langkah persiapan untuk analisis sampel, ekstraksi dan penyimpanan informasi genetik menggunakan beberapa host atau perangkat. Tahap pengurutan genetik meliputi pengurutan informasi genetik, penyimpanan informasi genetik jangka panjang, dan informasi genetik mentah digital yang dapat digunakan untuk penelitian lebih lanjut. Setiap komponen sistem dan setiap pemangku kepentingan rentan terhadap eksploitasi vektor serangan yang ditandai dengan huruf merah [15]. Gambar 3 menjelaskan tahapan dalam penelitian.



Keterangan:

P – Physical (fisik)	B – Bioinformatika & perangkat	L – Local network atau jaringan
H – Hardware (perangkat keras)	lainnya	lokal
	F – Firmware/Sistem operasi	R – Remote network atau cloud
— Material, data, dan jalur kueri	- - - - - Bahan mentah, data, dan jalur kueri	

Gambar 3. Tahapan Penelitian [15]

3. HASIL DAN PEMBAHASAN

3.1. *Pengambilan Sampel Genetika Manusia*

Sampel genetik manusia digunakan sebagai sumber informasi untuk mengidentifikasi sesuatu. Untuk keperluan forensik, sampel genetik dapat memberikan informasi genetik tentang tubuh manusia mana pun. Hampir semua sampel biologis dapat digunakan untuk pengujian genetik. Darah, jaringan lunak, tulang dan gigi, air liur, akar rambut, sel kulit mati, dan sperma dapat berfungsi sebagai sampel biologis untuk pengujian genetik.

3.2. *Pengumpulan Sampel Genetika Manusia*

Data sampel manusia akan dikumpulkan terlebih dahulu sebelum dilakukan proses ekstraksi untuk dilakukan pengujian. Metode penyimpanan sampel menggunakan wadah tabung yang disimpan dalam freezer -20°C untuk menjaga kualitas DNA. Sampel yang digunakan untuk analisis biasanya diambil dari tempat kejadian cukup lama untuk mencapai laboratorium [16].

3.3. *Sampel Biologi Genetika Manusia*

Proses pengumpulan, penyimpanan, dan distribusi sampel telah menerima sedikit pengakuan sebagai poin yang sah untuk kompromi informasi genetik. Sampel biologis sebagai input ekosistem ini dapat dimodifikasi secara jahat untuk mengandung malware yang dikodekan, meskipun hingga saat ini hanya ditunjukkan dalam sistem dimana perangkat lunak penguntai direayasa secara artifisial untuk memasukkan kerentanan yang akan dipicu oleh malware yang dikodekan. Sampel biologis juga dapat didegradasi, dimodifikasi, atau dihancurkan untuk membahayakan material, dan data yang dihasilkan, integritas, dan ketersediaan. Gudang sampel dan peralatan penyimpanan sering dihubungkan ke jaringan untuk tujuan pemantauan, membuat rentan terhadap jaringan yang berdekatan, dan serangan jarak jauh. Biorepositori dan pengumpulan serta distribusi sampel dapat ditargetkan untuk mencuri banyak sampel biologis, seperti penipuan pengujian genetik yang diketahui dan eksfiltrasi sejumlah kecil sampel yang ditargetkan mungkin sulit dideteksi. Penyimpanan, transit, dan penghancuran bahan biologis sensitif harus dipertimbangkan oleh pemangku kepentingan sebagai aspek penting dari keamanan informasi genetik dan keamanan siber secara keseluruhan [15].

3.4. *Persiapan Analisis Sampel*

Sebelum memulai pemisahan DNA, sampel gen yang diambil dari lapangan harus dipersiapkan secara khusus terlebih dahulu. Sampel harus dibagi menjadi dua bagian, yaitu sampel yang akan diekstrak dan sampel yang akan disimpan sebagai koleksi yang ataupun diproses untuk tujuan yang lain. Catatan penting adalah bahwa ukuran sampel rata-rata yang harus disiapkan untuk analisis DNA hanya sekitar 50-100 mg (tergantung tujuan penelitian). Ukuran ini relatif kecil, yaitu seukuran 2 butir beras. Proses pemisahan ini dirancang untuk menjaga kesegaran dan kemurnian sampel serta meminimalkan risiko kontaminasi oleh organisme lain atau berbagai reagen kimia seperti formalin yang dapat mengganggu proses ekstraksi DNA. Sampel yang telah terpisahkan selanjutnya disimpan dalam larutan etanol (99,99%) atau minimal 96% untuk menjaga kualitas sampel dan kualitas DNA yang akan diekstrak. Proses ini disebut mengawetkan sampel untuk ekstraksi DNA [16].

3.5. *Ekstraksi Data Genetika*

Ekstraksi DNA adalah proses dan tahapan pertama yang dilakukan untuk mendapatkan total DNA dari manusia. Secara umum, ekstraksi DNA melibatkan beberapa tahapan proses yang

penting, mulai dari tahap preparasi hingga memperoleh ekstrak DNA yang terlarut dalam larutan *buffer* khusus. Larutan tersebut berguna untuk menjaga dan mempertahankan kondisi DNA secara kualitatif dan kuantitatif dalam jangka waktu yang lebih lama. Secara kualitatif, ini berarti larutan *buffer* harus mampu menjaga kualitas DNA yang terlarut dalam kondisi yang baik. Sekaligus secara kuantitatif, artinya larutan *buffer* harus mampu mempertahankan jumlah DNA yang terlarut sedemikian rupa sehingga jumlahnya tetap (tidak terdegradasi/rusak) dan cukup untuk digunakan pada tahap selanjutnya tanpa mempengaruhi kualitas atau kuantitas DNA. Tahapan proses ekstraksi DNA diawali dengan preparasi sampel dan pemilihan metode ekstraksi yang tepat untuk mendapatkan ekstrak DNA. Keberhasilan proses ekstraksi DNA dapat diukur dengan beberapa metode, antara lain pengecekan keberadaan pita DNA menggunakan metode *elektroforesis* atau mengukur konsentrasi DNA terlarut menggunakan metode *spektrofotometri* [16].

3.6. Persiapan Penyimpanan Hasil Ekstraksi Sampel Genetika

Pemangku kepentingan harus mempertimbangkan penyimpanan, pengangkutan, dan penghancuran bahan biologis sensitif sebagai bagian penting dari keseluruhan keamanan informasi genetik dan keamanan dunia maya [15]. Pembekuan sampel secara instan dianggap sebagai metode terbaik untuk pengawetan jangka panjang, karena aktivitas enzimatik dan kimia lainnya menurun pada suhu yang lebih rendah. Namun, metode penyimpanan lain yang memerlukan pengawet cair atau dukungan fisik (seperti baki berbasis kertas) dapat menawarkan keuntungan dibandingkan pembekuan, seperti: B. Lebih sedikit ruang dan konsumsi energi, tidak ada risiko kegagalan daya, penyimpanan jangka panjang lebih pendek. Biaya dan transfer serta pengiriman sampel lebih mudah. Misalnya, darah dapat disimpan pada suhu kamar dalam etanol 95-100%. Saat sampel dibekukan secara langsung, disarankan suhu -80°C atau serendah mungkin untuk memaksimalkan pengawetan DNA [17].

3.7. Penguntaian Data Genetika

Pengurutan data genetik merupakan langkah terakhir untuk menentukan urutan nukleotida dari fragmen yang diamplifikasi. Derivasi dilakukan dengan metode Sanger menggunakan sequencer DNA otomatis berdasarkan metode pelabelan terminator pewarna. Tahapan pembentukan untai DNA meliputi (1) preparasi DNA, (2) proses amplifikasi menggunakan primer universal, (3) penjernihan DNA, (4) electrophoresis (elektroforesis), dan (5) pembacaan analisis molekul kecil hasil stranding. Data yang diperoleh berupa elektroforogram berupa file ABI, dengan masing-masing nukleoid diidentifikasi dengan warna yang berbeda.

Instrumen penguntaian DNA merupakan titik fokus pembuatan data mentah dan peralatan laboratorium lain serta komunikasi data jaringan. Alat dan teknik keamanan standar diterapkan seperti pemindaian kerentanan, pemantauan paket, pemodelan ancaman, penilaian konfigurasi, forensik digital, dan penilaian tumpukan penuh, termasuk pembongkaran perangkat keras dan analisis dinamis dan statis dari berbagai komponen perangkat lunak. Tabel 1 merupakan perangkat DNA Sequencer berdasarkan dari laman website Oxford Nanopore Technologies [15]

Tabel 1. Daftar perangkat DNA sequencer [18]

	<i>Flongle</i>	<i>MinION & MinION Mk1C</i>	<i>GridION</i>	<i>PromethION 24</i>	<i>PromethION 48</i>
Read length	Fragment length = read length. Longest read now >4 Mb				
Run Time	1 min – 16 hrs	1 min – 72 hrs	1 min – 72 hrs	1 min – 72 hrs	1 min – 72 hrs
Number of flow cells per device	1	1	5	24	48
DNA sequencing yield per flow cell*	Up to 2.8 GB	Up to 50 GB	Up to 50 GB	Up to 290 GB	Up to 290 GB
DNA sequencing yield per device*	Up to 2.8 GB	Up to 50 GB	Up to 250 GB	Up to 7 GB	Up to 14 GB

3.8. Penyimpanan Data Genetika

Data penguntaian sinyal mentah disimpan dalam sistem penguntaian memori dan ditransmisikan ke satu atau lebih titik akhir. Pengiriman data dengan aman melalui jaringan lokal memerlukan konfigurasi teknologi informasi internal. Dokumentasi vendor biasanya menyebutkan penerapan *firewall* untuk mengamankan sistem penguntaian dan melakukannya dengan benar membutuhkan pengetahuan mendalam tentang jaringan aman dan kewaspadaan aktivitas jaringan. Dokumentasi juga secara umum menyebutkan penonaktifan dan pengaktifan protokol dan port jaringan tertentu dan tindakan lebih lanjut yang mungkin sulit bagi sebagian besar organisasi kecil hingga menengah, sementara juga menghilangkan kontrol dan mitigasi umum lainnya.

Lab dan sistem penguntaian DNA terhubung ke layanan pihak ketiga. Lab memiliki sedikit kendali atas postur keamanan koneksi ini. Platform cloud vendor penguntaian DNA dan independen diimplementasikan untuk pemrosesan bioinformatika, penyimpanan data, serta fungsi pemantauan dan pemeliharaan perangkat. Otentikasi multifaktor, akses berbasis peran dan tugas, dan banyak tindakan keamanan lainnya tidak umum di platform ini. Kesalahan konfigurasi pada layanan cloud dan komunikasi jarak jauh merupakan kerentanan utama terhadap informasi genetik yang ditunjukkan oleh pelanggaran sebelumnya. Tabel 2 merupakan daftar perangkat untuk penyimpanan data genetika.

Tabel 2. Tinjauan perangkat dan sistem penguntaian genetik populer [15]

<i>Vendor</i>	<i>Produk</i>	<i>Waktu (h)</i>	<i>Luaran (Gb)</i>	<i>Sistem Operasi</i>	<i>Komputasi</i>	<i>Jaringan</i>	<i>Cloud Service</i>
Illumina	iSeq	19	1	Windows 7 &	Standalone &/or external device	Wired or wireless	BaseSpace (AWS)
	MiniSeq	24	8	Windows 10			
	MiSeq	24	15				
	NextSeq	30	300				
	HiSeq	84	1.500				
	NovaSeq	44	6.000				
Oxford Nanopore Technologies	SmidgION ^M	-	~1	Android & IOS	External device	Wired or wireless	EPI2ME (AWS)
	Flonge ^M	16	2	Windows, Macintosh, Linux	External device	Wired	
	MinION Mk1B ^M	48	30		Standalone &/or external device		
	MinION Mk1C ^M	48	30	Linux (Ubuntu)			
		48	150				
		72	8.600				
	GridION Mk1						
	PromethION						
Pacific Biosciences	Sequel	20	50	Linux (Ubuntu & CentOS)	Standalone	Wired	SecureLink (AWS)
	Sequel II	30	4.000				
Applied Biosystems*	SeqStudio	2	~0.45	Windows 10	Standalone & external device	Wired or wireless	Thermo Fisher Cloud (AWS)
Ion Torrent*	3500/3500xL	2	-	Windows Vista			
	3730/3730xL	3	-	SP1			
	GeneStudio S5	8	50	Windows 2000 Pro	Standalone & external device	Wired	
	48	20	Linux (Ubuntu)				
	Genexus						

3.9. Data Mentah (Raw Data)

Data genetika yang sudah terdigitalisasi dan sudah siap digunakan untuk kegiatan penelitian menggunakan teknologi yang terkomputerisasi. Data tersebut dapat digunakan langsung ataupun disimpan menggunakan media penyimpanan yang sesuai dan keamanannya terjamin untuk memudahkan kegiatan-kegiatan penelitian seputar data tersebut untuk kedepannya. Format file dari raw data genetika menggunakan FASTA. FASTA adalah salah satu format yang paling banyak digunakan untuk menyimpan sekuens DNA dan protein [19]. Format tersebut merupakan format berbasis teks untuk mewakili urutan nukleotida atau urutan asam amino. Nukleotida atau asam amino diwakili menggunakan kode huruf tunggal. Gambar 4 merupakan contoh raw data dari

genetika data tersebut didapatkan dari laman website NCBI (National Center for Biotechnology Information).

Salah satu dari banyak format file sekuens protein yang dapat diakses di *GenBank* adalah FASTA (database penyedia sekuens protein). Karena cukup sederhana dan lebih mudah dibaca oleh banyak program komputer yang digunakan untuk analisis bioinformatika. Saat ini, keuntungan utama dari program FASTA adalah kemampuan untuk menggunakan matriks penilaian yang lebih luas, dan fasilitas untuk menggabungkan anotasi menjadi penyelarasan. Kekurangan mengimpor format FASTA, semua data harus dalam satu baris teks. Jika kesulitan mengimpor format FASTA, pastikan tidak ada teks berulang yang ditambahkan ke baris FASTA sebelum dijalankan oleh perangkat lunak pengurutan protein yang digunakan.

```
>ref|NC_000007.14|:54999755-55230888 Homo sapiens chromosome 7, GRCh38.p14 Primary Assembly
AATATATTTGACTACAAGTGATAGCCACTAGGAAATAAAGTCTCTGTTGTTTCATTCATGCAGGTTTAG
GCACGGCAAAGAGATGGCAGGGCACACACCTTCTGAGGCACACCACGGTTATCCAACCAAACTACTTTC
CAGATCTTACTAAAGGGAAGTCAAGGCCGAAACACTGAAGAAAAGCATCACTGTGCCAAAGTTT
TAGAGTCTGATGAGATAAAACACTATTTAGAGAAATGTGCACACCTCTTTTTAACTAGTGGTGAAT
GGATAACAAAAGGCAGTATAGCCACACAATGGAGTACTATGCAGCAGTAAAAAGAAACAACTACCAAT
AAATGCAGAAACATGGAGGGAATTTATTGAGCTGAGTGACAAAAGCCAAACACAAAATGCTTCATACTGA
ATGATCTTTTTGAAATCTAGATCTTTATGAACCTATACAAAATACAACTATTCTGTAGTGACCTCAG
ATAGATCACTAGTGTGGGAGGCCACCGGCAGAGGAGAGTTTTGCAAAGTAGCATGGAACCTCTCT
GGGTGATGGAAACATCCTGTGCCCTGATTGTAGCAGTGCTCATGGTGCACAGCAGTGAATGGACACT
CTAAATAATGCAGCCTTTTGCTTGAACATAACCTGATTAAAGTTGAAGATAAATGGCTAATTAGTACA
TGAACAAAGCTTGATGCCATTAGACAATAAAGAAATGCAAAAACAAAGATAAAACCTTTCATGCCCTGAAG
TTTATTGCATAGAAATTTATCTTAATTTAAAGGTGCAGATCTTTTAAAAAGAAATATAAATAAAAAATTT
TAAACTAGAAAACAAAGCACTCAAATCTACTTGCAAATGTGAAATTTGATTACAGGTTATTCAAGTTA
GCATTTGATGGAAATCTTTAAGTGCCTTCAGTAGGAAGCCAGGTAATAAATTAGCTTCAACTTAATG
GAGGTATAGTTACATAACAATAAGCCAGATCTATTTAGAGTGCCTACATAGACAGATAGATAGATAGATA
GATAGATAGATAGATAGATAGATAGATAGATGATGGATAGGTAGAGAGAGAAATGTGTGTGAGAGAGAC
AGAGAAAGAGAGAGAAAGAAAAGAGAAAGAAACAAAGAAAGAAAAGAGAGAAACAAAGACAGAAAGAA
AAATAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAAAGAAAGAAAAAAA
TACTGGGCAATAAATAAGTGAACAACCTAAGTATTGATAGGAAATGAGTCCAGGATATACTATTACAT
ATTTTTTAAAGTGCAGAAATAACAGTGTATGTATAGAGACTATCTCTGGAAAGACCCGCTCTATGGTTTGA
ATGTCCTCTCCAAAACCTCATGCTGAACTTAATTTCCAATGTGGCAGTATTAAGAGGTTAATATATTAGA
TTAATGAATCAATGTGTTATGATGAGAGAGAGGCTTGTGGCTTTATAAGAAAGAAAGGAAATCTGAGCTAG
CATGTTATGTAGCACGCTCAGCCCCCTTAACATGAGATGCCCTGCACCACCTGGGACTCTGCAGAGGGC
CCCACCAGCAAGAAGGCTCTCAGCAGATGCAAGTCCCCAACCTGTACCTCCCAGCCTCCATAAATGTAA
GAAATAAATTCATTTTCTTAAGTTACTCAGTTTCAGGTATGCTTTATAAGCAACAGAAATCAAAC
AAAACCCCAAGGAGCTGATGAGCGTGATCATCTCTAGTGATGGCTGGAGATGGGATGGGTTTTTCACT
GTACAGTTTTAAAACATAATTTCAAACCAAGTGAATTAAGTATTCAAATTTGAAATGAAATTAATTAATA
AAACATGTGGACATAAACAGAAAGCGTGACTTTTCTCAAGGGTCCGTCTCATTACCCTTTCTAAGGCA
AGCCTTGGTAGTGATGCCCAACCCACATGTCCTCTCTCTCTCTCTGBCACTCTCAGACACTCT
GCTTTGATGCCACCCCTCAGTACCCTCACTCCCAATTGCAGTGCCTGATCTGCAGAAAGCACTATCTCT
```

Gambar 4. Raw Data DNA dengan Format Penyimpanan FASTA File

4. KESIMPULAN

Berdasarkan hasil penelitian, data genetika manusia yang sudah digitalisasi dapat memudahkan para ilmuwan biomedik dalam pengembangan dan analisis untuk pengkajian gen manusia sehingga tidak perlu lagi untuk melakukan ekstraksi data genetika dari sampel biologi yang mentah. Pada penelitian selanjutnya pihak ilmuwan biomedik hanya tinggal menggunakan raw data atau data mentah dari hasil ekstraksi sampel biologi yang sudah dilakukan diekstraksi. Raw data atau data mentah digital dari genetika berformat sesuai dengan kebijakan yang berlaku yaitu menggunakan format file FASTA. File FASTA adalah salah satu dari beberapa format file sekuens

protein yang tersedia di GenBank (database penyedia sekuens protein). Jenis file FASTA lebih populer daripada format sekuens protein seperti .fa dan .fsa. karena file FASTA cukup sederhana dan formatnya lebih familiar untuk dibaca oleh banyak program komputer yang digunakan untuk kegiatan analisis bioinformatika.

DAFTAR PUSTAKA

- [1] Suryo, *Genetika Manusia Suryo*, 12th ed. Yogyakarta: Yogyakarta: Gadjah Mada University Press, 2016 © 2016 Gadjah Mada University Press, 2016.
- [2] S. Arshad, J. Arshad, M. M. Khan, and S. Parkinson, "Analysis of security and privacy challenges for DNA-genomics applications and databases," *J. Biomed. Inform.*, vol. 119, 2021, doi: 10.1016/j.jbi.2021.103815.
- [3] I. Yunita, K. Tjandradiredja, and S. Hansun, "Perkembangan Bioinformatics dalam Ruang Lingkup Ilmu Komputer," *J. Ultim.*, vol. 8, no. 1, pp. 65–69, 2017, doi: 10.31937/ti.v8i1.505.
- [4] Mahrus, Lalu Zulkifli, Saprizal Hadisaputra, and Ida Ayu Putu Armyani, "Penggunaan Bioinformatika dalam Pembelajaran Sains Untuk Menyelesaian Kesulitan Belajar Siswa pada Materi Genetika di SMPN 20 Mataram," *J. Pengabd. Magister Pendidik. IPA*, vol. 4, no. 4, pp. 290–295, 2021, doi: 10.29303/jpmipi.v4i4.1128.
- [5] I. B. M. Artadana and W. D. Savitri, *Dasar-Dasar Genetika Mendel dan Pengembangannya*, vol. xviii. 2018.
- [6] P. Sudrajad et al., "Pemanfaatan informasi genom untuk eksplorasi struktur genetik dan asosiasinya dengan performan ternak di Indonesia," *Livest. Anim. Res.*, vol. 19, no. 1, p. 1, 2021, doi: 10.20961/lar.v19i1.47658.
- [7] R. V. S. . Morihito, S. E. Chungdinata, T. A. Nazareth, M. I. Pulukadang, R. A. . Makalew, and B. Pinontoan, "Identifikasi Perubahan Struktur Dna Terhadap Pembentukan Sel Kanker Menggunakan Dekomposisi Graf," *J. Ilm. Sains*, vol. 17, no. 2, p. 153, 2017, doi: 10.35799/jis.17.2.2017.17368.
- [8] M. Mustami and C. Muthiadin, *Konsep Dasar Pewarisan Gen pada Manusia*. 2021.
- [9] I. C. Bu'ulolo, N. Simamora, S. Tampubolon, and A. Pinem, "Sequence Alignment Menggunakan Algoritma Smith Waterman," *J. Integr. Ed. Khusus*, vol. II, no. 2, pp. 2–7, 2010.
- [10] N. Z. Muhoirotin, Anita Widiastuti, Yulinda Aswan, Dian Puspita Yani, Hutabarat, *Genetika dan Biologi Reproduksi*, no. 1998. 2022.
- [11] Jagad.id, "DNA dan RNA : Definisi, Struktur, Fungsi, dan Perbedaan." <https://jagad.id/dna-dan-rna-definisi-struktur-fungsi-dan-perbedaan/> (accessed Jul. 09, 2022).
- [12] A. A. Parikesit, "Kontribusi Aplikasi Medis dari Ilmu Bioinformatika Berdasarkan Perkembangan Pembelajaran Mesin (Machine Learning) Terbaru," *Cermin Dunia Kedokt.*, vol. 45, no. 9, pp. 700–703, 2018, [Online]. Available: <http://www.kalbedmed.com/DesktopModules/EasyDNNNews/DocumentDownload.aspx?portalid=0&moduleid=471&articleid=225&documentid=65>.
- [13] L. Zimmermann et al., "A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core," *J. Mol. Biol.*, vol. 430, no. 15, pp. 2237–2243, 2018, doi:

10.1016/j.jmb.2017.12.007.

- [14] Y. Li, C. Huang, L. Ding, Z. Li, Y. Pan, and X. Gao, “Deep learning in bioinformatics: Introduction, application, and perspective in the big data era,” *Methods*, vol. 166, pp. 4–21, 2019, doi: 10.1016/j.ymeth.2019.04.008.
- [15] G. J. Schumacher, S. Sawaya, D. Nelson, and A. J. Hansen, “Genetic Information Insecurity as State of the Art,” *Front. Bioeng. Biotechnol.*, vol. 8, 2020, doi: 10.3389/fbioe.2020.591980.
- [16] S. Sjafaraenan, H. Lolodatu, E. Johannes, R. Agus, and A. Sabran, “Profil Dna Gen Follicle Stimulating Hormone Reseptor (Fshr) Pada Wanita Akne Dengan Teknik Pcr Dan Sekuensing Dna,” *Bioma J. Biol. Makassar*, vol. 3, no. 1, 2018, doi: 10.20956/bioma.v3i1.3909.
- [17] I. Di Lecce, J. Sudyka, D. F. Westneat, and M. Szulkin, “Preserving avian blood and DNA sampled in the wild: A survey of personal experiences,” *Ecol. Evol.*, vol. 12, no. 8, 2022, doi: 10.1002/ece3.9232.
- [18] T. O. NANOPORE, “Advancing human genomics with nanopore sequencing,” p. 28, 2019, [Online]. Available: <https://nanoporetech.com/sites/default/files/s3/white-papers/human-genetics-research-white-paper.pdf>.
- [19] H. Lopez-Fernandez *et al.*, “SEDA: A Desktop Tool Suite for FASTA Files Processing,” *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 19, no. 3, pp. 1850–1860, 2022, doi: 10.1109/TCBB.2020.3040383.