Penggunaan Lintasan Euler dalam Penyederhanaan Sekuensing DNA

Asep Saepudin / 13511093[1]
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
[1]asepsaepudin@students.itb.ac.id

***Abstrak*— Dalam makalah ini, akan dibahas bagaimana aplikasi dari teori graf, yaitu lintasan Euler dan graf de Bruijn, digunakan dalam dunia biologi molekular, tepatnya pada proses sekuensing DNA. Sekuensing DNA merupakan proses yang melibatkan banyak data dan kemungkinan hasil maupun kemungkinan kesalahan. Penggunaan lintasan Euler dan graf de Bruijn di sini merupakan salah satu langkah mempermudah dan mengefektifkan proses sekuensing DNA serta meminimalkan kemungkinan kesalahan yang terjadi. Lintasan Euler dan graf de Bruijn adalah konsep dasar informatika yang merupakan dua dari banyak konsep-konsep di dunia informatika yang digunakan dalam bidang bioinformatika.**

***Kata Kunci*—Sekuensing DNA, lintasan Euler, graf de Bruijn.**

I. PENDAHULUAN

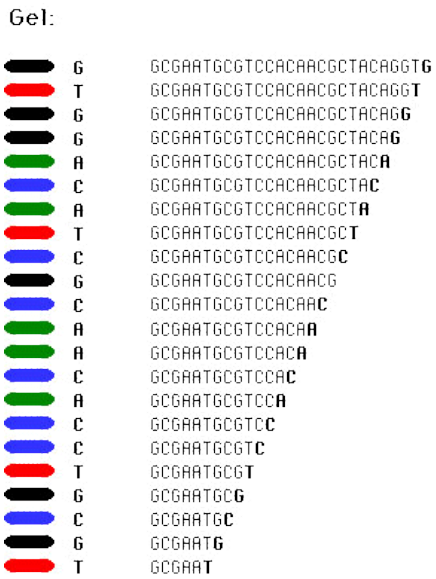
Sekuensing DNA adalah proses penentuan urutan nukleotida pada suatu fragmen DNA. Tujuan dari proses ini adalah untuk menentukan identitas maupun fungsi fragmen suatu DNA dengan membandingkannya dengan fragmen DNA lain yang telah diketahui. Teknik ini digunakan dalam riset dasar biologi maupun berbagai bidang terapan seperti kedokteran, bioteknologi, forensik, dan antropologi. Dalam 20 tahun terakhir, metode yang digunakan dalam sekuensing DNA adalah berlandaskan pada algoritma *overlap-layout-consensus*. Permasalahan yang dihadapi dengan penggunaan algoritma ini adalah terlalu banyak kemungkinan hasil dalam proses rekonstruksinya. Padahal, hanya ada sedikit kemungkinan yang benar. Semakin panjang suatu sekuens DNA tentunya semakin banyak pula kemungkinannya. Sehingga, proses sekuensing DNA untuk mendapatkan informasi dari DNA tersebut hampir mustahil dilakukan. Salah satu solusi untuk menghindari permasalahan tadi adalah dengan menggunakan teori graf, yaitu lintasan Euler dan graf de Bruijn. Metode dengan lintasan Euler dan graf de Bruijn ini akan mempermudah proses sekuensing DNA serta membuat algoritma menjadi lebih efektif. Selain itu, dengan metode ini pula kita bisa mengerucutkan berbagai kemungkinan sehingga kemungkinan yang akan didapat menjadi lebih sedikit.

II. DASAR TEORI

1. Sekuensing DNA

Sekuensing DNA adalah proses penentuan urutan nukleotida pada suatu fragmen DNA. Sekuensing DNA dapat kita analogikan sebagai *puzzle* acak gambar. Perbedaannya hanya terletak pada fakta bahwa kita "bermain" dengan ribuan bagian dengan ukuran yang bermacam-macam dan di antaranya terdapat bagian yang identik. Bagian-bagian dari *puzzle* tersebut merupakan potongan-potongan DNA dan *puzzle* yang telah diselesaikan dapat disebut genom. Secara khusus, penyusunan potongan DNA merupakan langkah untuk menentukan suatu genom hanya dengan mempelajari bagiannya. Karena sebuah DNA yang lengkap mengandung jutaan nukleotida, maka tidak ada jalan untuk mempelejari struktur genom tersebut secara utuh. Dengan teknologi saat ini, para ilmuan dan peneliti mampu mempelajari potongan DNA yang panjang nukleotidanya berkisar antara 500 hingga 1200. Suatu DNA dibangun oleh empat basa: adenin (A), timin (T), guanin (G), dan sitosin (C). Berbagai macam cara bisa dilakukan untuk merekonstruksi suatu DNA dari fragmennya. Namun, ada banyak masalah yang timbul dalam penyusunan potongan DNA ini yang membuat proses penyelesainnya menjadi lebih susah. Bahkan untuk 1% hingga 3% DNA yang dibaca pada proses sekuensing DNA menghasilkan beberapa kesalahan dan tidak mempresentasikan sebuah bagian DNA dari suatu DNA yang utuh. Jika kita tilik lebih jauh, masalah lain muncul kembali mengingat struktur DNA yang berupa *double-helix* di mana *helix* yang satu merupakan komplemen *helix* lainnya. Sehingga dalam proses sekuensing DNA yang dilakukan, kita juga harus memverifikasi apakah basa yang berpasangan merupkan komplemennya atau bukan. Adapun suatu pasangan basa dikatakan komplemen apabila basa adenin (A) berpasangan dengan timin (T) dan sebaliknya atau basa guanin (G) berpasangan dengan sitosin (S) dan sebaliknya.

Dalam 20 tahun terakhir, algoritma yang digunakan dalam proses sekuensing DNA adalah algoritma *overlap-layout-consensus*. Algoritma ini bisa digambarkan melalui gambar berikut:



Gambar 2.1 Algortima *overlap-layout-consensus*

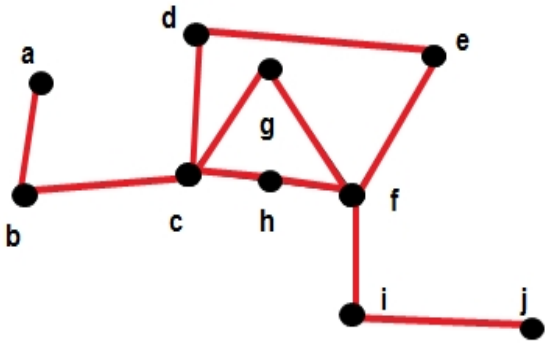
Tahapan dari algoritma ini dimulai dari proses *overlap*, yaitu proses pencocokan hasil pembacaan potongan DNA (ditunjukkan oleh blok berwarna pada Gambar 2.1) disertai pencarian kemungkinan adanya *overlapping* (pengulangan DNA dengan panjang tertentu yang memiliki urutan basa yang sama). Setelah itu, proses dilanjutkan ke tahap *layout.* Di tahap ini, terjadi pembacaan basa di sepanjang potongan DNA. Basa yang telah dibaca lalu "dicatat" di ujung tempat hasil pembacaan (ditunjukkan oleh serangkaian huruf di bagian kanan pada Gambar 2.1). Setelah tahap ini selesau, langkah terakhir yang harus dilakukan adalah *consensus.* Tahap ini adalah tahap di mana hasil pembacaan diklarifikasi dan diteliti sesuai dengan langkah pada tahap *layout.*

Kelemahan dari algoritma ini adalah tingkat keefektifannya yang rendah, bahkan ntuk sekuens DNA yang pendek di mana proses sekuensing menjadi tidak sederhana karena proses iterasi yang banyak sehingga algoritma ini tidak cocok digunakan untuk sekuensing DNA yang memiliki sekuens DNA yang panjang seperti manusia. Semakin panjang suatu sekuens maka algoritma ini semakin tidak efektif karena semakin banyak iterasi yang dilakukan. Oleh karena itu, proses sekuensing DNA untuk sekuens DNA yang panjang hampir mustahil dilakukan.

2. Lintasan dan Sirkuit Euler

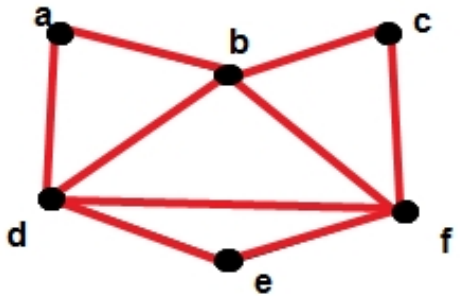
Lintasan dan sirkuit Euler merupakan salah satu cabang dari teori graf yang memiliki banyak aplikasi. Salah satunya aplikasi dari lintasan Euler di bidang bioinformatika adalah untuk proses rekonstruksi DNA dari fragmennya.

Lintasan Euler adalah lintasan yang melalui masing-masing sisi di dalam graf tepat satu kali. Lintasan Eluer dari gambar di bawah adalah **a-b-c-d-e-f-g-c-h-f-i-j**.



Gambar 2.2 Lintasan Euler a-b-c-d-e-f-g-c-h-f-i-j

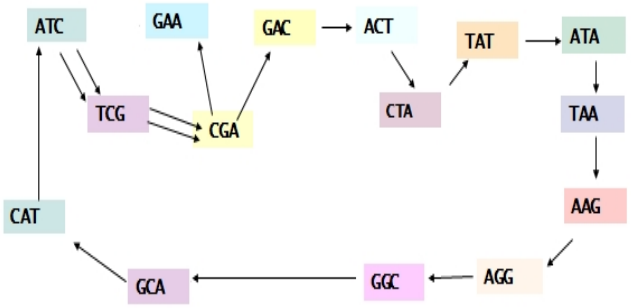
Sedangkan lintasan Euler yang kembali ke simpul asal sehingga terbentuk suatu lintasan tertutup (sirkuit) disebut sirkuit Euler. Sirkuit Euler dari gambar di bawah adalah **a-d-b-f-e-d-f-c-b-a.**



Gambar 2.3 Sirkuit Euler a-d-b-f-e-d-f-c-b-a

3. Graf de Bruijn

Graf de Bruijn adalah graf berarah yang simpulnya merepresentasikan suatu sekuens dengan menggunakan alfabet dan sisi-sisinya mengindikasikan di mana kemungkinan sekuens-sekuens tersebut mengalami pengulangan. Contoh graf de Bruijn dengan sekuens potongan DNA **ATCGACTATAAGGCATCGAA** ditunjukkan ole gambar di bawah:



Gambar 2.4 Graf de Bruijn dengan sekuens potongan DNA

Gambar 2.4 adalah graf de Bruijn dengan panjang potongan tiap simpul 3 dan sisi berarah di antara 2 simpul merepresentasikan potongan DNA dengan panjang 4. Contoh: simpul ATC dan simpul TCG merepresentasikan potongan DNA ATCG. Sedangkan arah panah ganda antara dua simpul menunjukkan terjadinya *overlapping*.

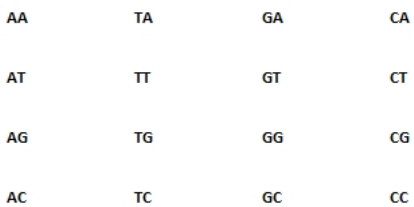
III. *SEQUENCING BY HYBRIDIZATION*

Hibridisasi adalah pembentukan ikatan dupleks stabil antara dua rangkaian nukleotida yang saling komplementer melalui perpasangan basa N. Hibridisasi dapat menunjukkan suatu keseragaman sekuens. Pasangan DNA-DNA, DNA-RNA, RNA-RNA dapat dibentuk dengan proses ini. Sedangkan *sequencing by hybridization* merupakan suatu metode sekuensing yang memanfaatkan lintasan Euler dan graf de Bruijn. *Sequencing by hybridization* adalah masalah dasar dalam suatu proses rekonstruksi DNA. Tujuan *sequencing by hybridization* adalah menentukan sekuens suatu nukleotida dari fragmen DNA yang tidak diketahui yang masukan datanya merupakan hasil eksperimen hibridisasi secara biokimia, yang disebut spektrum. Spektrum ini merupakan substring dari huruf-huruf yang melambangkan basa nukleotida.

Setelah ditentukan spektrum-spektrum yang panjangnya konstan tersebut, maka bisa dibuat graf de Bruijn yang tiap simpulnya merupakan representasi dari spektrum. Setelah graf de Bruijn berhasil dibuat, lalu tentukan lintasan Eulernya.

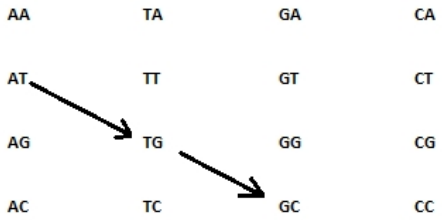
Adapun proses penentuan panjang potongan fragmen DNA hingga menentukan graf de Bruijn dari spektrumnya adalah sebagai berikut:

1. Diketahui suatu fragmen DNA yang akan diteliti, misal **ATGCGTGGCA**.
2. Lalu tentukan panjang potongan fragmen yang diinginkan (n), misal n = 3. Maka, potongan fragmennya jika dikumpulkan dalam suatu himpunan bernama S adalah S = {**ATG**, **TGC**, **GCG**, **CGT**, **GTG**, **TGG**, **GGC**, **GCA**}. Kardinalitas |S| = 8.
3. Lalu tentukan panjang spektrumnya. Rumus dari panjang spektrum adalah n-1. Jumlah spektrum yang digunakan adalah kardinalitas dari S dikurang 1 (|S| - 1). Dari sini, didapat panjang spektrum (m) = 2 dan jumlah spektrum yang digunakan = 7.
4. Langkah selanjutnya adalah membuat graf de Bruijn dengan simpul berjumlah sama dengan jumlah spektrum yang digunakan dan jumlah sisi berjumlah sama dengan kardinalitas himpunan S. Adapun langkah untuk membuat graf de Bruijn-nya adalah sebagai berikut:
   a) Tuliskan semua kemungkinan spektrum yang digunakan. Semua kemungkinan yang ada dari suatu panjang spektrum (m) adalah 4^m. Karena pada contoh kasus ini panjang spektrum yang digunakan adalah 2, maka jumlah simpulnya ada $4^2 = 16$, yaitu P = {**AA, AT, AG, AC, ... , CC**}



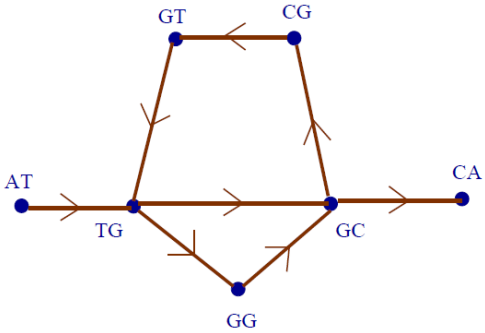
Gambar 3.1 Kemungkinan Simpul

   b) Lalu, periksa antardua simpul. Apabila dua simpul tersebut membentuk potongan fragmen yang urutan basanya ekivalen dengan anggota himpunan S, maka hubungkan dua simpul tersebut dengan graf berarah. Misal, simpul **AT** dan simpul **TG**, yang apabila dihubungkan membentuk **ATG** yang merupakan anggota dari himpunan S. Maka, hubungkan simpul **AT** dengan simpul **TG**. Begitu juga dengan simpul **TG** dan simpul **GC**.



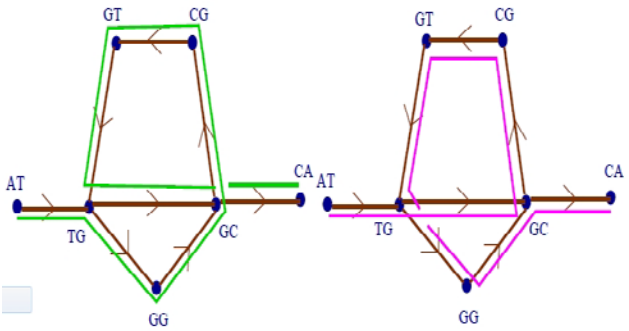
Gambar 3.2 Simpul dihubungkan dengan graf berarah

   c) Lakukan langkah serupa hingga seluruh potongan fragmen yang ada di himpunan S saling terhubung. Buang simpul yang tidak digunakan. Hasilnya ditunjukkan oleh gambar berikut



Gambar 3.3 Graf de Bruijn yang Dihasilkan

5. Setelah dihasilkan graf de Bruijn, langkah selanjutnya yang harus dilakukan adalah menentukan lintasan Euler-nya. Sekuens yang dihasilkan dari lintasan Euler merupakan sekuens DNA target yang memiliki kemungkinan cocok dengan DNA yang sedang diteliti. Adapun lintasan Euler dari graf berarah di atas ditunjukkan pada gambar berikut



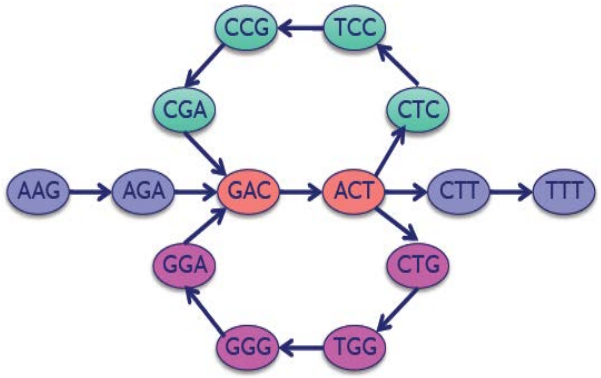
Gambar 3.4 Lintasan Euler yang Dihasilkan

Dari lintasa Euler di atas, maka sekuens DNA target yang mungkin adalah **ATGGCGTGCA** dan **ATGCGTGGCA**.

IV. VELVET ASSEMBLER

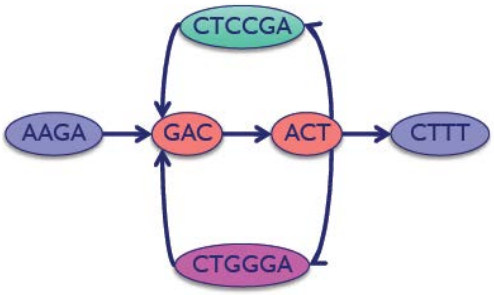
Velvet assembler adalah suatu metode yang menyederhanakan graf de Bruijn yang dihasilkan oleh metode *sequencing by hybridization*. Selain itu, metode ini juga mampu mencegah kesalahan-kesalahakn data yang terdapat pada sekuens yang dihasilkan (DNA target). Berikut adalah contoh penyederhanaan menggunakan *velvet assembler*.

Diketahui himpunan potongan fragmen dari suatu sekuens DNA S = {**AAGA, ACTC, ACTG, ACTT, AGAC, CCGA, CGAC, CTCC, CTGG, CTTT, GACT, GGAC, GGGA, TCCG, TGGG**}. Dengan menggunakan langkah yang sama seperti pada penjelasan sebelumnya, didapat graf de Bruijn sebagai berikut



Gambar 4.1 Graf de Bruijn yang Dihasilkan

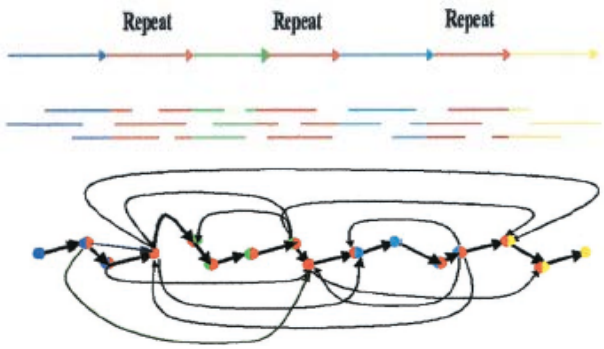
Dari graf di atas, didapat 2 sekuens DNA target, yaitu **AAGACTCCGACTGGGACTTT** dan **AAGACT-GGGACTCCGACTTT**. Dengan menyederhanakan subgraf linear pada graf di atas, akan didapat sebuah graf yang lebih sederhana. Graf tersebut ditunjukkan pada gambar di bawah



Gambar 4.2 Graf de Bruijn yang Lebih Sederhana

Dapat dilihat dari gambar bahwa graf yang lebih sederhana ini tetap tak menghilangkan kemungkinan sekuens DNA yang dihasilkan, yaitu **AAGACTCCGACTGGGACTTT** dan **AAGACTGG-GACTCCGACTTT.**

Jika kita perhatikan Gambar 4.1, di sana dapat dilihat terjadi pengulangan pada sisi **GAC-ACT**, di mana sisi tersebut dilewati sebanyak tiga kali. Apabila kita linearkan jalur yang seharusnya ditempuh dari awal hingga akhir, maka grafiknya akan menjadi seperti gambar di bawah



Gambar 4.3 Jalur de Broijn pada Gambar 4.1

Dengan menggambarkan tiga garis pada sisi **GAC-ACT** menjadi satu baris, maka kita bisa mengubah grafik pada Gambar 4.3 menjadi seperti grafik pada Gambar 4.1.

Gambar 4.4 Proses Penyederhanaan dengan Prinsip G*lue*

Inilah salah satu keuntungan penggunaan graf de Broijn dan lintasan Euler pada proses sekuensing DNA di mana proses iterasi atau pengulangan dapat disederhanakan. Penyederhanaan proses pengulangan ini mampu membuat graf yang dibentuk dapat diubah menjadi lebih sederhana. Ditambah lagi, penyederhanaan pengulangan ini juga menghapus kemungkinan-kemungkinan sekuens DNA lain sehingga mengerucutkan jumlah kemungkinan sekuens yang ada jauh lebih sedikit. Itu artinya, proses sekuensing DNA bisa berlangsung lebih cepat. Selain itu, proses sekuensing DNA menggunakan lintasan Euler pada graf dengan jutaan simpul juga lebih mudah dilakukan karena terdapat algoritma lintasan Euler. Inilah yang membedakan sekuensing DNA dengan cara menggunakan metode lintasan Euler dibanding dengan metode konvensional.

V. KESIMPULAN

Berbagai konsep dari teori graf telah diaplikasikan dalam berbagai bidang, di makalah ini telah didemonstrasikan penggunaan lintasan Euler dan graf de Broijn dalam proses sekuensing DNA merupakan salah satu langkah maju dalam dunia bioinformatika. Penggunaan dua teori graf di atas sangat terasa manfaatnya apabila digunakan dalam proses sekuensing DNA dengan skala yang besar. Dibanding dengan metode algoritma *overlap-layout-consensus*, tentunya sekuensing DNA dengan cara ini memiliki banyak keunggulan. Keunggulan-keunggulan tersebut antara lain:

1. Algoritma yang lebih efektif. Karena penggunaan metode ini menghilangkan proses pengulangan pada proses sekuensing yang membuat proses tersebut memakan waktu lebih lama.

2. Lebih mudah dibaca oleh manusia. Penggunaan graf tentunya membuat proses sekuensing DNA lebih terbayang dengan adanya grafik. Sehingga dalam proses translasi ke bahasa komputer, misalnya, lebih mudah dilakukan.

3. Metode *velvet assembler* adalah metode yang menyederhanakan graf de Bruijn hasil *sequencing by hybridization* tanpa mengubah lintasan Euler yang terbentuk pada graf tersebut.

REFERENCES

[1] R. Munir, Diktat Kuliah IF2091 Struktur Diskrit. Bandung: Program Studi Teknik Informatika Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung, 2008.

[2] Jonathan Kaptcianos, Graph Theory Aiding DNA Fragment Assembly.

[3] Graham Ellis, Computational Molecular Biology.

[4] Jacek Bla˙zewicz, Piotr Formanowicza, Marta Kasprzaka, Wojciech T. Markiewiczb, Sequencing by Hybridization with Isothermic Oligonucleotide Libraries.

[5] Richard Arratia, Bela Bollobas, Don Coppersmith, Gregory B. Sorkinc, Euler Circuits and DNA Sequencing by Hybridization.

[6] Jacek Blazewicza, Marta Kasprzaka, Computational Complexity of Isothermic DNA Sequencing by Hybridization.

[7] Pavel A. Pevzner, Haixu Tang, and Michael S. Waterman, An Eulerian Path Approach to DNA Fragment Assembly.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 18 Desember 2012

Asep Saepudin / 13511093