

Penerapan Teori Graf dalam Pembacaan Genom Meliputi *Mapping, Assembling, dan Sekuensing DNA*

Pipin Kurniawati - 13513089

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganessa 10 Bandung 40132, Indonesia

pipinkurniawati@students.itb.ac.id

Abstrak—Genetika merupakan salah satu cabang ilmu biologi yang berkembang pesat. Perkembangan tersebut tidak terlepas dari aplikasi ilmu sains lainnya, salah satunya adalah aplikasi ilmu matematika. Langkah fundamental dalam memahami genetika makhluk hidup adalah dengan pembacaan genom. Pemetaan genom, DNA assembling, dan sekuensing DNA—yang termasuk langkah-langkah pembacaan genom—merupakan proses yang melibatkan banyak data sehingga rentan terjadi kesalahan. Oleh karena itu, memiliki pemahaman dasar akan teori graf merupakan hal yang penting dalam menganalisis ketiga tahap proses tersebut. Makalah ini membahas penerapan teori graf dengan pemodelan graf Hamilton, graf Euler, graf de Bruijn, dan graf pohon dalam biologi molekuler, lebih tepatnya pada proses pemetaan, assembling, dan sekuensing rantai DNA

Kata Kunci—pemetaan DNA, sekuensing DNA, graf Hamilton, pohon

DNA tentunya semakin banyak pula kemungkinannya. Sehingga, proses pembacaan genom untuk mendapatkan informasi dari DNA tersebut hampir mustahil dilakukan. Salah satu solusi untuk menghindari permasalahan tersebut adalah dengan menggunakan teori graf. Dengan menggunakan teori graf, proses pemetaan, assembling, dan sekuensing DNA dapat dimodelkan dalam bentuk yang lebih mudah diamati untuk tujuan-tujuan tertentu.

Pada makalah ini penulis ingin membahas penerapan teori graf dalam proses pembacaan genom suatu organisme yang meliputi tahap pemetaan, sekuensing, dan assembling DNA. Struktur utama makalah ini terdiri dari bagian II yang berisi dasar teori, bagian III berisi penerapan teori graf dalam proses *mapping* DNA, selanjutnya penerapan graf dalam sekuensing dan assembling DNA secara berturut terdapat di bagian IV dan V.

I. PENDAHULUAN

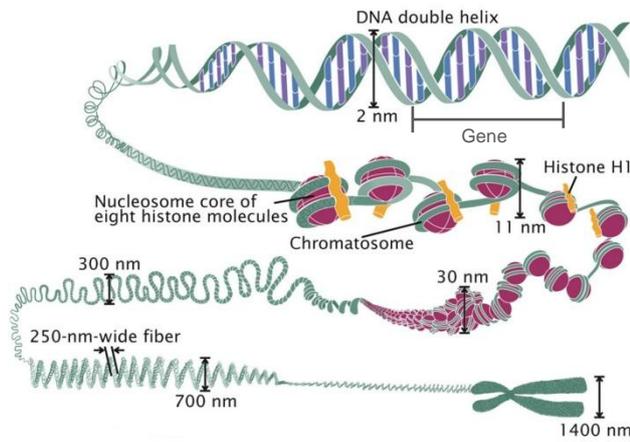
Seiring dengan perkembangan zaman, ilmu pengetahuan pun semakin berkembang. Salah satunya yaitu ilmu biologi. Dalam beberapa dekade terakhir, ilmu biologi, khususnya genetika berkembang sangat pesat. Penelitian mengenai genom banyak dilakukan. Salah satunya pembacaan genom manusia. Hal tersebut memberi dampak yang cukup signifikan terhadap berbagai aspek di kehidupan. Ilmu genetika telah menunjukkan kemajuan besar dalam memahami substansi dasar kehidupan sejak ditemukannya model rantai DNA *double helix* oleh Englisman Francis Crick dan James Watson. Namun pencapaian ini tidak akan terwujud tanpa bantuan dari disiplin ilmu lain, seperti kimia, fisika, matematika, dan bioinformatika.

Pembacaan genom suatu makhluk hidup merupakan hal fundamental dalam riset dasar biologi molekuler maupun berbagai bidang terapan seperti antropologi dan forensik. Seperti yang kita ketahui, kromosom makhluk hidup terdiri dari sekuens DNA yang sangat panjang. Sangat tidak mungkin melakukan pembacaan genom secara utuh. Oleh karena itu dalam prosesnya DNA harus dipotong menjadi fragmen-fragmen yang berukuran lebih kecil. Namun permasalahan yang dihadapi adalah terlalu banyak kemungkinan hasil sekuens dalam proses rekonstruksinya. Padahal hanya terdapat sedikit kemungkinan yang benar. Semakin panjang suatu sekuens

II. DASAR TEORI

A. DNA dan Genom

Salah satu objek dasar dalam genetika adalah rantai DNA, yaitu berupa informasi kode genetik makhluk hidup. DNA merupakan *string* huruf dalam alfabet {A, C, G, T} yang merepresentasikan Adenin (A), Sitosin (C), Guanin (G), dan Timin (T). Keempat alfabet tersebut adalah basa nitrogen (basa-N) yang terikat pada cincin karbon 1' (C1) nukleotida—monomer DNA—dengan jenis deoksiribosa. Nukleotida bergabung dengan nukleotida lain melalui ikatan fosfodiester antara cincin karbon 3' (C3) nukleotida pertama dan cincin karbon 5' (C5) nukleotida kedua. DNA rantai pendek yang berisi 6-12 nukleotida disebut oligonukleotida. Sejumlah DNA yang berisi kode spesifik terhadap suatu protein yang kemudian bertanggungjawab terhadap ciri genetis tertentu disebut gen. Rantai DNA yang berisi informasi gen dengan jumlah tertentu berikatan dengan protein membentuk struktur padat yang dapat menebal, memendek, dan menyerap zat warna yang disebut kromosom. Sedangkan genom merupakan keseluruhan materi genetik pada suatu organisme. Gambar 2.1 dibawah ini menggambarkan detail struktur kromosom. Struktur tersebut mengilustrasikan keterkaitan nukleotida, DNA, gen, dan kromosom.



Gambar 2.1 Ilustrasi keterkaitan DNA, gen, dan kromosom
sumber: faculty.fmcc.suny.edu/mcdarby

Hingga saat ini struktur DNA direpresentasikan dengan pemodelan rantai ganda atau yang lebih dikenal dengan *double helix*. Struktur ini terbentuk akibat adanya ikatan hidrogen antara pasangan basa-N dengan komplementernya. Adapun suatu pasangan basa dikatakan komplement apabila basa adenin (A) berpasangan dengan timin (T) dan sebaliknya atau basa guanin (G) berpasangan dengan sitosin (S) dan sebaliknya.

B. Pemetaan, Assembling, dan Sekuensing DNA

Pembacaan sekuens DNA merupakan langkah dasar dalam penelitian biologi molekuler. Namun proses tersebut bukan hal yang mudah dilakukan mengingat bahwa proses pembacaan tersebut tidak dapat dilakukan secara langsung misalnya dengan menggunakan mikroskop. Pada dasarnya proses pembacaan sekuens genom meliputi tiga tahap: *mapping*, *assembling*, dan *sequencing*.

Pemetaan Genom adalah proses untuk mengetahui lokus—letak suatu gen pada kromosom—dan jarak suatu gen secara relatif terhadap gen-gen lainnya. Proses ini menghasilkan urutan posisi sejumlah lokus pada suatu kelompok pautan (*linkage group*). *Linkage group* adalah bagian dari suatu kromosom.

Sekuensing DNA adalah proses penentuan urutan nukleotida yang terdapat pada fragmen DNA. Hasil sekuensing DNA inilah yang nantinya akan merepresentasikan sekuens nukleotida pada rantai DNA dalam proses pembacaan genom.

Sequence assembling adalah proses rekonstruksi rantai DNA utuh dari fragmen-fragmen pendek DNA. Proses ini berjalan sesaat setelah proses sekuensing DNA.

Proses pembacaan genom diawali dengan pemotongan rantai DNA utuh menjadi fragmen-fragmen yang berukuran lebih pendek terdiri dari 100000-1000000 nukleotida. Tahap pemotongan ini dilakukan dengan bantuan enzim restriksi endonuklease. Pada tahap ini, besar kemungkinan terjadinya kesalahan informasi urutan fragmen-fragmen DNA disebabkan banyaknya data. Bahkan sebesar 1% hingga 4% DNA yang dibaca menghasilkan beberapa kesalahan dan tidak merepresentasikan sebuah bagian DNA dari suatu DNA yang utuh. Oleh karena itu pemetaan DNA berguna untuk

menyimpan informasi urutan rantai DNA tersebut sehingga tingkat kesalahan dapat diminimalisir. Tahap *mapping* dimulai dengan pemotongan fragmen DNA yang kedua kalinya menjadi fragmen yang berukuran 40000 nukleotida.

Tahap selanjutnya pada proses pembacaan genom yaitu sekuensing DNA. Penentuan sekuens DNA hanya dapat dilakukan terhadap fragmen DNA dengan ukuran maksimal 1000 nukleotida. Sehingga fragmen DNA hasil *mapping* yang berukuran 40000 nukleotida harus dipotong secara acak menjadi fragmen dengan ukuran yang sesuai untuk dilakukan tahap sekuensing. Dalam proses sekuensing DNA, mengingat struktur DNA berupa *double helix*, harus dilakukan verifikasi apakah basa yang berpasangan merupakan komplementernya atau bukan. Setelah tahap sekuensing dilakukan, fragmen-fragmen DNA memulai tahap rekonstruksi membentuk fragmen awal dengan ukuran 40000 nukleotida. Tahap ini disebut *sequence assembling*.

C. Graf

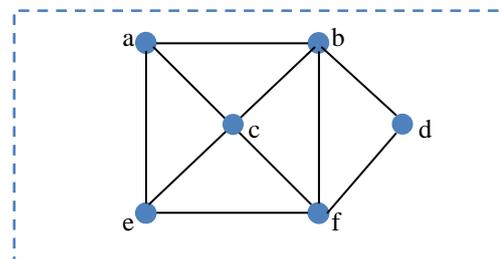
Graf digunakan untuk merpresentasikan objek – objek diskrit dan hubungan antara objek – objek diskrit. Graf $G = (V, E)$ terdiri dari :

- V = Himpunan tidak kosong dari simpul – simpul (vertices).
- E = Himpunan sisi (edges) yang menghubungkan sepasang simpul.

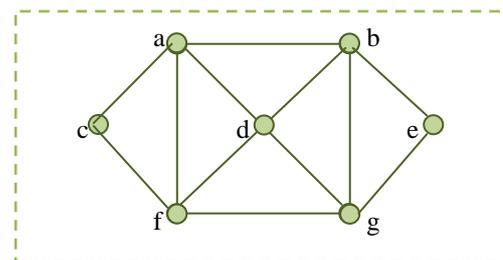
Berdasarkan orientasi arah pada sisi, secara umum graf dibedakan atas 2 jenis :

1. Graf Tak Berarah (*undirected graph*)
Graf yang sisinya tidak mempunyai orientasi arah
2. Graf Berarah (*directed graph* atau *digraph*)
Graf yang setiap sisinya diberikan orientasi arah.

Salah satu cabang teori graf yang banyak diterapkan di berbagai bidang adalah lintasan dan sirkuit Euler. Lintasan Euler adalah lintasan yang melalui masing-masing sisi didalam graf sebanyak tepat satu kali.



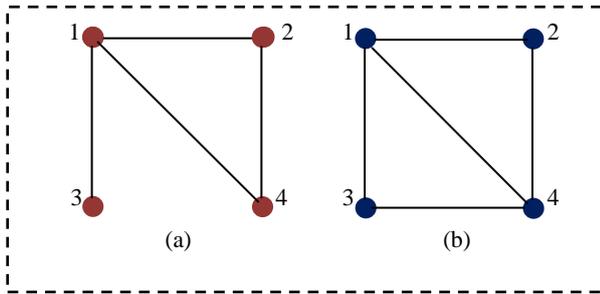
Gambar 2.2.1 Lintasan Euler a-b-d-f-b-c-f-a-c



Gambar 2.2.2 Sirkuit Euler c-a-b-e-g-b-d-g-f-d-a-f-c

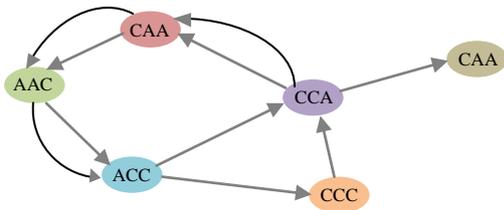
Sedangkan lintasan Euler yang kembali ke simpul asal sehingga terbentuk suatu lintasan tertutup (sirkuit) disebut sirkuit Euler. Gambar 2.2.1 dan 2.2.2 diatas adalah contoh graf dengan lintasan dan sirkuit Euler.

Selain lintasan dan sirkuit Euler, terdapat pula lintasan dan sirkuit Hamilton. Lintasan Hamilton adalah lintasan yang melalui tiap simpul didalam graf tepat satu kali. Sedangkan sirkuit Hamilton adalah lintasan yang melalui tiap simpul didalam graf tepat satu kali, kecuali simpul asal (sakaligus simpul akhir) yang dilalui sebanyak dua kali. Gambar 2.3(a) menunjukkan graf yang memiliki lintasan Hamilton 4-2-1-3 dan gambar 2.3(b) berupa graf dengan sirkuit Hamilton 1-2-4-3-1.



Gambar 2.3 Lintasan dan sirkuit Hamilton

Graf de Bruijn adalah salah satu jenis graf berarah yang simpulnya merepresentasikan suatu sekuens alfabet dan sisi-sisinya mengindikasikan letak pengulangan sekuens.



Gambar 2.4 Contoh graf de Bruijn

Gambar 2.4 diatas adalah contoh graf de Bruijn dengan sekuens CCAACCCAACCAA. Arah panah ganda antara dua simpul menunjukkan letak terjadinya *overlap*.

Selain jenis-jenis graf yang telah disebutkan diatas, ada pula graf terhubung tak-berarah yang tidak mengandung sirkuit. Jenis graf ini dikenal dengan pohon (*tree*). Graf pohon dengan n simpul memiliki $m = n - 1$ sisi.

III. PENERAPAN TEORI GRAF DALAM PROSES PEMETAAN DNA

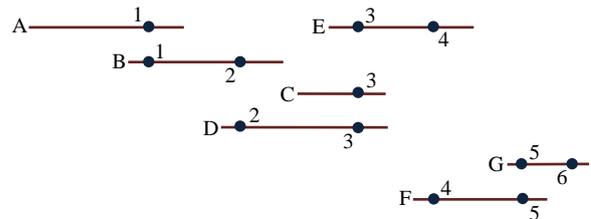
Seperti yang telah kita ketahui, pemetaan atau *mapping* adalah tahap awal dalam proses pembacaan genom suatu makhluk hidup. *Mapping* berfungsi sebagai memori untuk menyimpan informasi urutan fragmen-fragmen DNA saat proses pemotongan molekul DNA berlangsung. Hal ini dilakukan dengan cara menentukan lokasi penanda (*marker*) atau yang lebih dikenal dengan *probe*. *Probe* merupakan fragmen DNA atau RNA berisi informasi sekuens suatu gen yang diberi label radioaktif dengan

menggunakan molekul kimia yang dapat terdeteksi seperti biotin, *digoxigenin*, dan *fluorescein*. Terdapat dua metode yang digunakan dalam proses *mapping* yaitu teknik hibridisasi dan *restriction site analysis*. Namun pada makalah ini penulis ingin membahas penerapan graf dalam *genome mapping* dengan teknik hibridisasi. Hibridisasi adalah pembentukan ikatan dupleks stabil antara dua rangkaian nukleotida yang saling komplementer melalui perpasangan basa N.

Tahap awal dalam proses *mapping* adalah pemotongan rantai DNA menjadi beberapa fragmen dengan panjang yang berbeda. Kemudian seluruh fragmen DNA dikloning dan disimpan dalam suatu "*clone library*". Selanjutnya dilakukan tahap hibridisasi untuk mengetahui apakah suatu *probe* terhibridisasi dengan hasil kloning fragmen DNA tertentu yang terdapat di *library*. Informasi hasil proses hibridisasi berisi sekuens *probe* dan kloning fragmen DNA yang kemudian diterjemahkan kedalam bentuk "*DNA molecule map*" atau peta DNA.

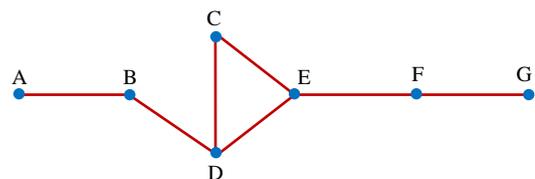
Penerapan graf dalam pemetaan DNA dilakukan dengan cara merepresentasikan setiap *clone* fragmen DNA sebagai simpul graf. Dua buah simpul terhubung dengan tanda panah tak-berarah apabila terdapat suatu *probe* yang terhibridisasi dengan kedua *clone* tersebut.

Gambar 3.1 dibawah ini merupakan contoh *mapping* dengan teknik hibridisasi. A, B, C, D, E, F, dan G adalah hasil kloning fragmen DNA yang disimpan dalam "*clone library*". Sedangkan 1, 2, 3, 4, 5, dan 6 adalah *probe* atau penanda.



Gambar 3.1 Fragmen DNA dengan *probe*

Gambar 3.2 dibawah adalah contoh representasi *mapping* DNA dengan teknik hibridisasi dengan pemodelan graf tak-berarah. Simpul A dan B terhubung karena *clone* A dan B terhibridisasi akibat adanya *probe* 1 di masing-masing fragmen. Begitu pula dengan pasangan simpul lainnya.



Gambar 3.2 Representasi *mapping* DNA dengan graf

IV. PENERAPAN TEORI GRAF DALAM PROSES SEKUENSING DNA

Hingga saat ini teknik biokimia paling mutakhir yang

dapat digunakan dalam proses sekuensing fragmen DNA adalah hibridisasi (*hybridization*). Teknik hibridisasi bertujuan untuk mendeteksi semua fragmen DNA berupa oligonukleotida dengan panjang l (8-12 basa-N). Seluruh oligonukleotida tersebut kemudian disimpan dalam suatu "library" yang berisi 4^l kemungkinan fragmen DNA dengan panjang l . Seluruh 4^l oligonukleotida tersebut disusun dalam bentuk *bio-chip*. Selanjutnya dilakukan perbandingan antara "library" dengan fragmen DNA lainnya. Proses hibridisasi menghasilkan himpunan oligonukleotida pembentuk rantai DNA utuh yang disebut spektrum. Spektrum merupakan substring dari huruf-huruf dalam alfabet {A, C, G, T} yang melambangkan basa-N. Panjang n sekuens rantai DNA utuh dapat dideteksi dengan elektroforesis gel.

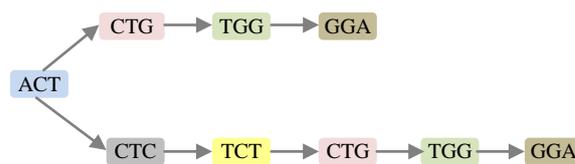
Algoritma dasar yang digunakan dalam proses sekuensing mengasumsikan spektrum bersifat ideal, yaitu spektrum diasumsikan tidak memiliki tingkat kesalahan. Spektrum terdiri dari $n - l + 1$ string berbeda yang merepresentasikan oligonukleotida dengan n adalah panjang rantai DNA utuh dan l adalah panjang oligonukleotida. Sebanyak $l - 1$ substring dari string didalam spektrum disusun secara *overlap* terhadap string yang bertetangga. Kemudian dilakukan pencarian sekuens fragmen oligonukleotida menggunakan pohon pencarian (*search tree*). Tiap oligonukleotida direpresentasikan sebagai simpul (*node*). Dua simpul (u, v) dalam spektrum dihubungkan dengan tanda panah berarah jika terjadi *overlap* antara $l - 1$ substring terakhir simpul u dan $l - 1$ substring pertama simpul v . Selanjutnya dari *search tree* tersebut dapat disederhanakan menjadi graf dengan lintasan Hamilton.

Adapun proses penentuan panjang fragmen oligonukleotida hingga menentukan pohon pencarian dari spektrumnya adalah sebagai berikut:

- Misalnya suatu rantai DNA utuh dengan panjang $n = 8$ memiliki sekuens basa-N ACTCTGGA. Lalu dilakukan proses hibridisasi menghasilkan oligonukleotida dengan panjang $l=3$
- $|\text{Spektrum}| = n - l + 1 = 8 - 3 + 1 = 6$
Spektrum ideal yang terbentuk {ACT, CTC, TCT, CTG, TGG, GGA}
- Selanjutnya adalah membuat pohon pencarian dengan jumlah minimum simpul sama dengan kardinalitas spektrum. Akar pohon pencarian dapat ditentukan dengan cara mencari simpul yang $l - 1$ substring pertamanya tidak terjadi *overlap* dengan $l - 1$ substring terakhir simpul lainnya
- Lakukan perbandingan antara 2 simpul dalam spektrum. Jika terjadi *overlap* antara $l - 1$ substring terakhir simpul u dan $l - 1$ substring pertama simpul v maka hubungkan kedua simpul dengan panah berarah dari u ke v . Ulangi langkah ini terhadap seluruh simpul pada spektrum

Pohon pencarian untuk kasus diatas terdapat pada gambar 4.1. Pohon pencarian untuk persoalan diatas menghasilkan 2 kemungkinan sekuens DNA, yaitu ACTGGA dan ACTCTGGA. Jika dihasilkan lebih dari 1

hasil sekuens, maka pilih sekuens dengan panjang yang sama dengan n .

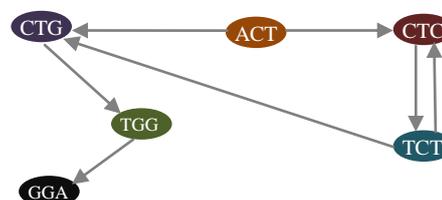


Gambar 4.1 Pohon pencarian pada proses sekuensing

Dengan menggunakan prinsip yang sama, sekuens DNA dapat ditentukan dengan menerapkan lintasan Hamilton. Contoh penerapan tahapannya menggunakan rantai DNA dengan sekuens ACTCTGGA. Adapun langkah-langkahnya sebagai berikut:

- Misalnya suatu rantai DNA utuh dengan panjang $n = 8$ memiliki sekuens basa-N ACTCTGGA. Lalu dilakukan proses hibridisasi menghasilkan oligonukleotida dengan panjang $l=3$
- $|\text{Spektrum}| = n - l + 1 = 8 - 3 + 1 = 6$
Spektrum ideal yang terbentuk {ACT, CTC, TCT, CTG, TGG, GGA}
- Selanjutnya adalah membuat lintasan Hamilton dengan jumlah simpul sama dengan kardinalitas spektrum
- Lakukan perbandingan antara 2 simpul dalam spektrum. Jika terjadi *overlap* antara $l - 1$ substring terakhir simpul u dan $l - 1$ substring pertama simpul v maka hubungkan kedua simpul dengan panah berarah dari u ke v . Ulangi langkah ini terhadap seluruh simpul pada spektrum

Representasi dengan lintasan Hamilton lebih sederhana karena jumlah simpul yang digunakan lebih sedikit sehingga dapat meminimalisir kesalahan jika diterapkan pada rantai DNA dengan ukuran lebih panjang.



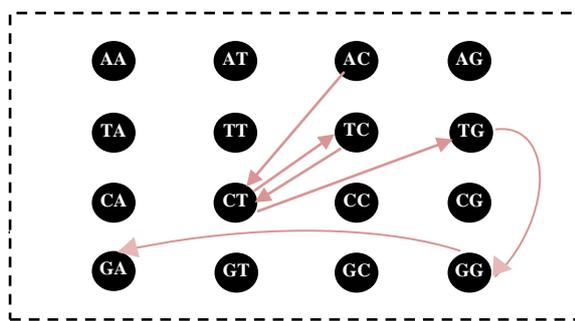
Gambar 4.2 Lintasan Hamilton DNA dengan sekuens ACTCTGGA

Metode lain yang dapat diterapkan sebagai representasi proses sekuensing DNA adalah dengan menggunakan lintasan Euler seperti yang dikemukakan oleh P.A Pevzner. Adapun langkah-langkah penentuan lintasan Euler seperti yang dijabarkan oleh Pevzner yaitu sebagai berikut:

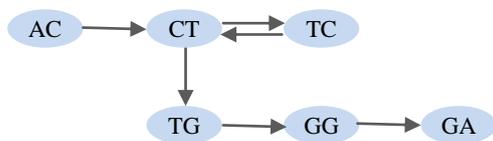
- Misalnya suatu rantai DNA utuh dengan panjang $n = 8$ memiliki sekuens basa-N ACTCTGGA. Lalu dilakukan proses hibridisasi menghasilkan oligonukleotida dengan panjang $l = 3$
- $|\text{Spektrum}| = n - l + 1 = 8 - 3 + 1 = 6$

Spektrum ideal yang terbentuk {ACT, CTC, TCT, CTG, TGG, GGA}

- Tentukan panjang anggota himpunan S dengan cara mengurangi panjang fragmen dalam spektrum dengan $l - 1$
 $s = l - 1 = 3 - 1 = 2$
- Selanjutnya menentukan semua anggota himpunan S.
 $|S| = 4^s = 4^2 = 16$
 $S = \{AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, GG\}$
- Kemudian periksa antardua simpul. Apabila dua simpul tersebut membentuk potongan fragmen yang memiliki urutan basa-N sesuai dengan salah satu fragmen dalam spektrum, maka hubungkan kedua simpul tersebut dengan graf berarah. Misal, simpul AC dan simpul CT, yang apabila dihubungkan membentuk ACT yang sesuai dengan fragmen oglinukleotida ACT pada spektrum. Maka hubungkan kedua simpul tersebut dengan tanda panah berarah dari AC menuju CT. Ulangi langkah ini terhadap seluruh pasangan simpul pada himpunan S. Jika telah selesai, ambil simpul-simpul yang terhubung. Atur posisi simpul-simpul tersebut sedemikian rupa sehingga membentuk graf de Bruijn.



Gambar 4.3 Pemeriksaan antardua simpul



Gambar 4.4 Graf de Bruijn sekuens DNA ACTCTGGA

Gambar 4.3 menunjukkan tahap pemeriksaan antardua simpul. Seluruh simpul yang terhubung dengan tanda panah berarah diambil untuk kemudian disusun menjadi graf de Bruijn. Kemudian dari graf de Bruijn tersebut dapat ditentukan lintasan Euler untuk mempermudah proses penentuan sekuens DNA. Pada gambar 4.4 diatas, lintasan Euler yang terbentuk adalah AC-CT-TC-CT-TG-GG-GA sehingga sekuens yang dihasilkan adalah ACTCTGGA.

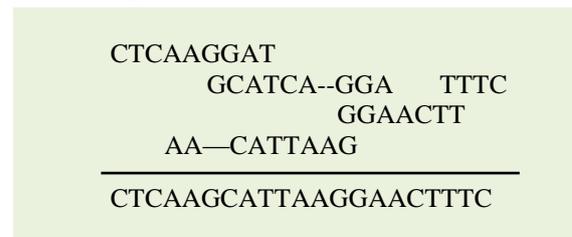
V. PENERAPAN TEORI GRAF DALAM PROSES PENYATUAN FRAGMEN DNA

Prosedur sekuensing DNA hanya dapat membaca urutan DNA dari suatu fragmen dengan panjang 1000

nukleotida. Oleh karena itu, tahap penyatuan DNA (*assembling*) dilakukan untuk menyatukan seluruh fragmen-fragmen DNA hasil sekuensing menjadi rantai DNA utuh.

Secara umum, pendekatan yang digunakan dalam proses assembling terdiri dari tiga tahap, yaitu deteksi overlap, *layout* (penyusunan) fragmen, dan penentuan konsensus.

Pada tahap awal, yaitu deteksi *overlap*, ditentukan tingkat kesesuaian antardua fragmen. Pemotongan acak rantai DNA menjadi fragmen-fragmen yang lebih pendek pada proses sekuensing menyebabkan timbulnya kemungkinan *overlapping* fragmen. Sehingga tujuan dari tahap ini adalah untuk mencari fragmen-fragmen yang saling *overlap*. Lebih jelasnya, untuk setiap pasang sekuens fragmen (s, t), ditentukan satu atau lebih substring terakhir dari s yang *overlap* dengan satu atau lebih substring awal dari t .



Gambar 5.1 Contoh assembling DNA

s	t	Shift
CTCAAGGAT	GCATCAGGA	5
CTCAAGGAT	AACATTAAG	3
GCATCAGGA	GGAACTT	6
GGAACTT	TTTC	5
AACATTAAG	GCATCAGGA	2
AACATTAAG	GGAACTT	8

Tabel 5.1 Spesifikasi overlap fragmen pada gambar 5.1

Gambar 5.1 menunjukkan contoh deteksi overlap pada *assembling* DNA. *Overlap* yang terjadi antar kedua fragmen tidak memiliki karakteristik tertentu. *Overlap* mungkin saja mengandung beberapa kesalahan seperti munculnya sekuens huruf pada salah satu fragmen sedangkan pada fragmen lainnya berupa *gap* atau sekuens huruf yang berbeda. Misalnya seperti yang terjadi pada fragmen DNA CTCAAGGAT dan AA-CATTAAG. Tabel 5.1 menunjukkan pasangan fragmen yang saling *overlap* dan jumlah pergeseran diantara keduanya.

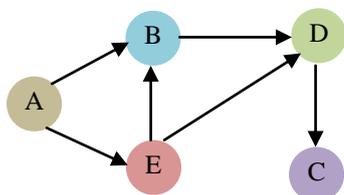
Tahap kedua dan ketiga dari proses *assembling* DNA memiliki prosedur yang hampir sama dengan tahap awal, hanya saja prosesnya dilakukan dengan melibatkan beberapa pendekatan biologis seperti membandingkan frekuensi kemunculan suatu basa-N pada rantai DNA hasil rekonstruksi dengan rantai DNA asli dan sebagainya.

Assembling DNA dapat direpresentasikan dalam graf dengan pemodelan multigraf. Simpul (*vertice*) dalam multigraf merepresentasikan fragmen DNA. Sedangkan sisi (*edge*) merepresentasikan lokasi *overlap*. Dalam prosesnya, fragmen DNA diberi suatu kode simbol untuk menyederhanakan penulisan. Adapun korespondensi

simbol dan fragmen ditunjukkan pada tabel 5.2 dibawah ini.

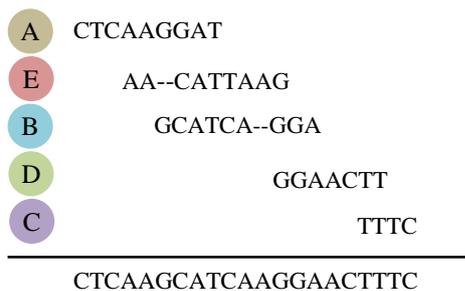
Kode	Fragmen DNA
A	CTCAAGGAT
B	GCATCA--GGA
C	TTTC
D	GGAACTT
E	AA--CATTAAG

Tabel 5.2 Kode simbol untuk tiap fragmen



Gambar 5.2 Multigraf assembling DNA

Gambar 5.2 menunjukkan representasi *assembling* DNA menggunakan multigraf. Rekonstruksi sekuens basa-N dapat dilihat melalui tanda panah berarah. Jika suatu simpul menunjuk ke lebih dari satu simpul lainnya maka urutan simpul selanjutnya ditentukan dari arah panah simpul-simpul tersebut. Contohnya pada kasus diatas, simpul A menunjuk ke simpul B dan E. Kemudian untuk menentukan urutan simpul selanjutnya, lihat keterhubungan simpul B dan E. Simpul B dan E terhubung dengan tanda panah berarah dari E ke B. Sehingga urutan simpul yang tepat adalah A-E-B. Adapun sekuens simpul pada gambar 5.2 yaitu A-E-B-D-C.



Gambar 5.3 Sekuens DNA hasil representasi multigraf

Gambar 5.3 menunjukkan sekuens basa-N pada rantai DNA hasil rekonstruksi melalui proses penyatuan DNA (*assembling*). Urutan simpul yang diperoleh dari representasi graf kemudian diubah menjadi fragmen DNA sesuai dengan kode simpul yang terdapat pada tabel 5.2. Penentuan lokasi antardua fragmen disesuaikan dengan menyatukan titik awal terjadinya *overlap* antara 2 fragmen tersebut hingga diperoleh sekuens DNA yang sesuai. Pada contoh diatas, hasil rekonstruksi DNA dari fragmen-fragmennya menghasilkan sekuens basa-N CTCAAGCATCAAGGAACTTTC.

VI. KESIMPULAN

Berbagai konsep mengenai teori graf telah diaplikasikan dalam berbagai bidang. Penerapan teori graf dengan pemodelan lintasan Hamilton, Euler, graf pohon, dan graf de Bruijn dalam pembacaan genom seperti yang telah dibahas dalam makalah ini adalah suatu kemajuan dalam bioinformatika. Penggunaan teori graf di sangat terasa manfaatnya apabila digunakan dalam proses pembacaan DNA dengan skala yang besar. Tentunya pembacaan genom atau DNA dengan menerapkan teori graf ini memiliki banyak keunggulan. Keunggulan-keunggulan tersebut antara lain:

1. Lebih mudah dibaca oleh manusia. Penggunaan graf tentunya membuat keseluruhan proses pembacaan DNA lebih tergambar dengan adanya grafik. Sehingga dalam proses translasi ke bahasa komputer, misalnya, lebih mudah dilakukan.
2. Algoritma menjadi lebih efektif. Karena penggunaan metode ini menghilangkan proses pengulangan pada tahapan proses tersebut misalnya saat sekuensing dan *assembling* sehingga membuat proses tersebut memakan waktu lebih lama.

VII. UCAPAN TERIMA KASIH

Saya mengucapkan terima kasih kepada Bu Harlili dan Pak Rinaldi Munir atas bimbingannya selama ini dalam kuliah IF2120 Matematika Diskrit. Saya juga berterima kasih kepada teman-teman seperjuangan satu studi atas semangat, bantuan, dan dukungannya.

REFERENSI

- Jane B. Reece, Michal L. Cain, Lisa A. Urry. *Campbell Biology*, 9th Edition. Pearson. 2011
- Rosen, Kenneth H. *Discrete Mathematics and Its Applications*, 7th Edition. The McGraw-Hill Companies. 2012
- Angeleska, A., Jonoska, N., & Saito, M. (2009). DNA Recombination Through Assembly Graphs. *Discrete Applied Math* 157, 3020-3037.
- J. Blazewick, P. Formanowicz. *Selected Combinatorial Problems of Computational Biology*. Science Direct. 2003
- W. Bains, G.C. Smith, A novel method for nucleic acid sequence determination, *Journal of Theoretical Biology* 135 (1988) 303-307.
- http://www.yourgenome.org/landing_dgg.shtml
Waktu akses : 08 Desember 2014, 10:20 WIB

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, 10 November 2014

Pipin Kurniawati - 13513089