FILOGENETIKA MOLEKULER: METODE TAKSONOMI ORGANISME BERDASARKAN SEJARAH EVOLUSI

N.L.P. INDI DHARMAYANTI

Balai Besar Penelitian Veteriner, Jl. R.E. Martadinata No. 30, Bogor 16114

(Makalah diterima 4 Januari 2011 – 9 Maret 2011)

ABSTRAK

Filogenetika digambarkan sebagai klasifikasi secara taksonomi dari suatu organisme berdasarkan pada sejarah evolusi yaitu filogeninya mereka dan merupakan bagian integral dari ilmu pengetahuan yang sistematik yang mempunyai tujuan untuk menentukan filogeni dari organisme berdasarkan pada karakteristiknya. Analisis filogenetika sekuen asam amino dan protein biasanya akan menjadi wilayah yang penting dalam analisis sekuen. Analisis filogenetika juga digunakan untuk mengikuti perubahan yang terjadi secara cepat yang mampu mengubah suatu spesies, seperti virus. Pohon evolusi adalah sebuah grafik dua dimensi yang menunjukkan hubungan diantara organisme atau lebih spesifik lagi adalah sekuen gen dari organisme. Pemisahan sekuen disebut *taxa* (atau *taxon* jika tunggal) yang didefinisikan sebagai jarak filogenetika unit pada sebuah pohon. Pohon terdiri dari cabang-cabang luar (*outer branches*) atau daun-daun (*leaves*) yang merepresentasikan *taxa* dan titik-titik (*nodes*) dan cabang merepresentasikan hubungan diantara taxa, Ketika sekuen nukleotida atau protein dari dua organisme yang berbeda mempunyai kemiripan, maka mereka diduga diturunkan dari sekuen *common anchestor*. Terdapat tiga metode dalam filogenetika yang dibahas dalam makalah ini, yaitu: (1) *Maximum parsimony*, (2) *Distance* dan (3) *Maximum likehoood* yang secara umum digunakan untuk membentuk pohon evolusi atau pohon terbaik untuk mengamati variasi sekuen dalam kelompok. Masing-masing metode ini digunakan untuk tipe analisis yang berbeda dan penggunaanya disesuaikan dengan bentuk dan jenis data yang akan diolah.

Kata kunci: filogenetik, analisis, evolusi, sekuen nukleotida/protein

ABSTRACT

MOLECULAR PHYLOGENETIC: ORGANISM TAXONOMY METHOD BASED ON EVOLUTION HISTORY

Phylogenetic is described as taxonomy classification of an organism based on its evolution history namely its phylogeny and as a part of systematic science that has objective to determine phylogeny of organism according to its characteristic. Phylogenetic analysis from amino acid and protein usually became important area in sequence analysis. Phylogenetic analysis can be used to follow the rapid change of a species such as virus. The phylogenetic evolution tree is a two dimensional of a species graphic that shows relationship among organisms or particularly among their gene sequences. The sequence separation are referred as taxa (singular taxon) that is defined as phylogenetically distinct units on the tree. The tree consists of outer branches or leaves that represents taxa and nodes and branch represent correlation among taxa. When the nucleotide sequence from two different organism are similar, they were inferred to be descended from common ancestor. There were three methods which were used in phylogenetic, namely (1) *Maximum parsimony*, (2) *Distance*, and (3) *Maximum likehoood.* Those methods generally are applied to construct the evolutionary tree or the best tree for determine sequence variation in group. Every method is usually used for different analysis and data.

Key words: Phylogenetic, analysis, evolution, nucleotide/protein sequence

PENDAHULUAN

Filogenetika dikenal sebagai bidang yang berkaitan dengan ilmu biologi. Filogenetika menyediakan fasilitas dalam bidang epidemiologi manusia, ekologi, dan evolusi biologi. Ketertarikan peneliti menggunakan analisis filogenetika tidak jarang membuat sedikit membingungkan dikarenakan ketidakmengertian dalam menggunakan beberapa metode dalam analisis filogenetika. Pertanyaan yang sering muncul adalah metode analisis filogenetika mana yang akan digunakan? Pohon filogenetika mana yang bisa dipercaya? Pada makalah ini, penulis akan memaparkan bagaimana memilih metode filogenetika sesuai dengan data yang kita miliki untuk membuat pohon filogenetika yang dapat dipercaya.

Analisis filogenetika tidak terlepas dari evolusi biologis. Evolusi adalah proses gradual, suatu organisme yang memungkinkan spesies sederhana menjadi lebih komplek melalui akumulasi perubahan dari beberapa generasi. Keturunan akan mempunyai beberapa perbedaan dari nenek moyangnya sebab

sedang berubah dalam sebuah evolusi (ESTABROOK, 1984). Dalam mempelajari variasi dan diferensiasi genetik antar populasi, jarak genetik dapat dihitung dari jumlah perbedaan basa polimorfik suatu lokus gen masing-masing populasi berdasarkan urutan DNA (CAVALLI-SFORZA, 1997).

Analisis sistematika dilakukan melalui konstruksi sejarah evolusi dan hubungan evolusi antara keturunan dengan nenek moyangnya berdasarkan pada kemiripan karakter sebagai dasar dari perbandingan (LIPSCOMB, 1998). Jenis analisis yang diketahui dengan baik adalah analisis filogenetika atau kadang-kadang disebut *cladistics* yang berarti *clade* atau kelompok keturunan dari satu nenek moyang yang sama. Analisis filogenetik biasanya direpresentasikan sebagai sistem percabangan, seperti diagram pohon yang dikenal sebagai pohon filogenetika (BRINKMAN dan LEIPE, 2001).

Dalam sistem biologis, proses evolusi melibatkan mutasi genetik dan proses rekombinan dalam spesies untuk membentuk spesies yang baru. Sejarah evolusi organisme dapat diidentifikasi dari perubahan karakternya. Karakter yang sama adalah dasar untuk menganalisis hubungan satu spesies dengan spesies lainnya (SCHMIDT, 2003). Pohon filogenetik adalah pendekatan logis untuk menunjukkan hubungan evolusi antara organisme (SCHMIDT, 2003). Filogenetika diartikan sebagai model untuk merepresentasikan sekitar hubungan nenek moyang organisme, sekuen molekul atau keduanya (BRINKMAN and LEIPE., 2001). Salah satu tujuan dari penyusunan filogenetika adalah untuk mengkonstruksi dengan tepat hubungan antara organisme dan mengestimasi perbedaan yang terjadi dari satu nenek moyang kepada keturunannya (LI *et al*., 1999).

Analisis filogenetika molekuler

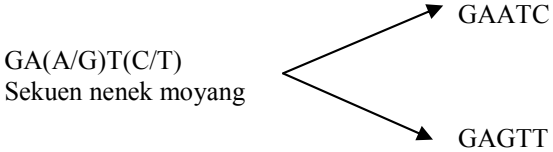
Konstruksi pohon filogenetika adalah hal yang terpenting dan menarik dalam studi evolusi. Terdapat beberapa metode untuk mengkonstruksi pohon filogenetika dari data molekuler (nukleotida atau asam amino) (SAITOU dan IMANISHI, 1989). Analisis filogenetika dari keluarga sekuen nukleotida atau asam amino adalah analisis untuk menentukan bagaimana keluarga tersebut diturunkan selama proses evolusi. Hubungan evolusi diantara sekuen digambarkan dengan menempatkan sekuen sebagai cabang luar dari sebuah pohon. Hubungan cabang pada bagian dalam pohon merefleksikan tingkat dimana sekuen yang berbeda saling berhubungan. Dua sekuen yang sangat mirip akan terletak sebagai *neighboring outside* dari cabang-cabang dan berhubungan dalam cabang umum (*Common branch*) (MOUNT, 2001).

Filogenetika digambarkan sebagai klasifikasi secara taksonomi dari organisme berdasarkan pada sejarah evolusi mereka, yaitu filogeni mereka dan merupakan bagian integral dari ilmu pengetahuan yang sistematik dan mempunyai tujuan untuk menentukan filogeni dari organisme berdasarkan pada karakteristik mereka. Lebih lanjut filogenetika adalah pusat dari evolusi biologi seperti penyingkatan keseluruhan paradigma dari bagaimana organisme hidup dan berkembang di alam (MOUNT, 2001).

Analisis filogenetika sekuen asam amino dan protein biasanya akan menjadi wilayah yang penting dalam analisis sekuen. Selain itu, dalam filogenetika dapat menganalisis perubahan yang terjadi dalam evolusi organisme yang berbeda. Berdasarkan analisis, sekuen yang mempunyai kedekatan dapat diidentifikasi dengan menempati cabang yang bertetangga pada pohon. Ketika keluarga gen ditemukan dalam organisme atau kelompok organisme, hubungan filogenetika diantara gen dapat memprediksikan kemungkinan yang satu mempunyai fungsi yang ekuivalen. Prediksi fungsi ini dapat diuji dengan eksperimen genetik. Analisis filogenetika juga digunakan untuk mengikuti perubahan yang terjadi secara cepat yang mampu mengubah suatu spesies, seperti virus (MCDONALD dan KREITMAN, 1991; NIELSEN dan YANG, 1998).

Hubungan analisis filogenetika dengan *alignment*/penjejeran sekuen

Ketika sekuen nukleotida atau protein dari dua organisme yang berbeda memiliki kemiripan, maka mereka diduga diturunkan dari sekuen *common ancestor*. Sekuen penjejeran akan menunjukkan dimana posisi sekuen adalah tidak berubah/*conserved* dan dimana merupakan *divergent*/atau berkembang menjadi berbeda dari *common ancestor* seperti diilustrasikan MOUNT (2001) pada ilustrasi di bawah ini. Sekuen 1 dan 2 diasumsikan berasal dari nenek moyang yang sama (*common ancestor*). Total terdapat dua sekuen yang berubah.



Studi sekuen biologi selalu tidak dapat dihindarkan dari penjejeran sekuen/*alignment*. Tujuan dari proses penjejeran adalah mencocokkan karakter-karakter yang homolog, yaitu karakter yang mempunyai nenek moyang yang sama (KEMENA dan NOTREDAME, 2009). Ketika menghomologikan sekuen, kolom dari penjejeran dapat digunakan untuk berbagai macam aplikasi seperti mengidentifikasi residu dengan struktur yang *analog* atau yang mempunyai fungsi

yang serupa atau untuk mengkonstruksi pohon filogenetika. Akurasi dari program penejejeran sekuen ynag lebih dari dua set/*multiple sequence alignment* telah dihasilkan oleh berbagai macam studi komperatif (BLACKSHIELDS *et al*., 2006; EDGAR dan BATZOGLOU 2006; NOTREDAME, 2007).

Metode paling umum dalam melakukan *multiple sequence alignment* adalah pertama melakukan penjejeran kelompok sekuen yang mempunyai hubungan dekat dan kemudian secara sekuensial ditambahkan sekuen yang berhubungan namun lebih berbeda. Penjejeran yang diperoleh diakibatkan karena sebagian besar sekuen yang mirip dalam kelompok sehingga tidak merepresentasikan sejarah yang sesungguhnya dari perubahan evolusi yang telah terjadi. Sebagian besar metode analisis filogenetika mengasumsikan bahwa masing-masing posisi sekuen protein atau asam nukleat yang berubah secara independen satu sama yang lain (kecuali evolusi sekuen RNA).

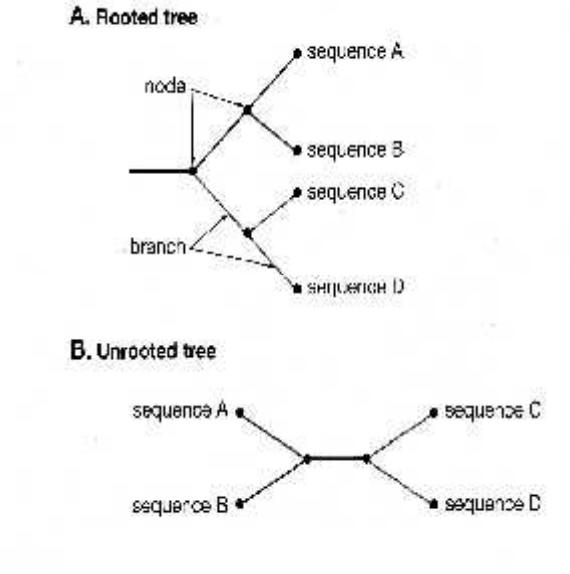
Seperti yang telah ditunjukkan sebelumnya, analisis sekuen yang sangat mirip dan mempunyai panjang yang sama adalah sangat jelas. Seringkali hasil penjejeran sekuen memperlihatkan adanya *gap* dalam penjejeran tersebut. *Gap* menunjukkan adanya insersi atau delesi dari satu atau lebih dari karakter sekuen selama evolusi. Protein yang dijejerkan semestinya mempunyai struktur tiga dimensi yang sama. Umumnya, sekuen dalan struktur *core* seperti protein tidak mengalami insersi atau delesi dikarenakan subtitusi asam amino harus cocok dengan lingkungan paket hidrofobik dari *core*. *Gap* sangat jarang ditemukan pada *multiple sequence alignment* yang menunjukkan sekuen *core*. Sebaliknya, beberapa variasi termasuk insersi, delesi sangat mungkin ditemukan di daerah *loop* pada bagian luar struktur tiga dimensi, sebab pada bagian ini tidak berpengaruh banyak terhadap struktur *core*. Daerah *loop* berinteraksi dengan molekul kecil, membran dan protein lain di lingkungan (MOUNT, 2001).

Gap dalam penjejeran merepresentasikan perubahan mutasi dalam sekuen termasuk insersi, delesi atau penyusunan ulang materi genetik. Ekspektasi bahwa panjang *gap* dapat terjadi sebagai akibat adanya introduksi tunggal yang memutuskan berapa banyak perubahan individu telah terjadi dan apa perintahnya. *Gap* diberi perlakuan (*treated*) dalam beberapa program filogenetik, tetapi tidak ada *clear-cut model* seperti bagaimana seharusnya mereka di perlakukan. Beberapa metode mengabaikan *gap* yang terjadi atau hanya memfokuskan dalam penjejeran yang tidak mempunyai *gap*. Meskipun *gap* dapat berguna sebagai petanda filogenetik di beberapa situasi.

Pendekatan lainnya untuk menangani *gap* adalah mencegah analisis situs individu dalam penjejeran sekuen, dan menggantikan dengan menggunakan skoring kemiripan/*similarity score* sebagai dasar dari analisis filogenetika.

Konsep pohon evolusi

Pohon evolusi adalah sebuah grafik dua dimensi yang menunjukkan hubungan diantara organisme atau lebih spesifik lagi adalah sekuen gen dari organisme. Pemisahan sekuen disebut *taxa* (atau *taxon* jika tunggal) yang didefinisikan sebagai jarak filogenetika unit pada sebuah pohon. Pohon terdiri dari cabang-cabang luar (*outer branches*) atau daun-daun (*leaves*) yang merepresentasikan *taxa* dan titik-titik (*nodes*) dan cabang merepresentasikan hubungan diantara *taxa*, yang diilustrasikan sebagai A-D pada Gambar 1.



Gambar 1. Struktur pohon evolusi

Sumber: MOUNT (2001)

Oleh karena itu, sekuen A dan B dipisahkan dari sekuen *common ancestor* yang direpresentasikan dengan titik-titik di bawahnya; C dan D adalah mempunyai kemiripan. Pada Gambar 1 menunjukkan bahwa sekuen A/B dan C/D memiliki *common ancestor* yang sama yang ditunjukkan dengan sebuah titik pada bagian paling rendah dari pohon. Hal ini sangat penting untuk mengenali bahwa masing-masing titik dalam pohon direpresentasikan sebuah pemisahan garis evolusi gen ke dalam dua spesies yang berbeda. Panjang masing-masing cabang pada titik berikutnya menunjukkan jumlah sekuen yang berubah yang terjadi sebelum level pemisahannya. Contohnya, panjang cabang antara titik A/B dan B menunjukkan spesies mempunyai rata-rata evolusi yang sama.

Total panjang semua cabang dalam pohon disebut sebagai panjang pohon. Pohon yang juga bercabang

dua atau *binary tree*, mempunyai dua cabang yang berasal dari masing-masing titik. Situasi ini adalah satu dari yang diperkirakan selama evolusi, dan hanya memisahkan spesies baru pada waktu itu. Pohon dapat mempunyai lebih dari satu cabang yang berasal dari sebuah titik jika pemisahan *taxa* juga sedemikian dekat sehingga mereka tidak dapat dipecahkan atau menjadi pohon yang sederhana.

Representasi alternatif dari hubungan sekuen diantara A-D pada Gambar 1A ditunjukkan pada Gambar 1B. Perbedaan diantara pohon A dan B yaitu pohon B adalah *unrooted tree*. *Unrooted tree* juga menunjukkan hubungan evolusi diantara sekuen A-D, tetapi tidak menyatakan lokasi dari moyang yang tertua/*oldest ancestry*. Sebagai contoh, B dapat diubah menjadi A dengan menempatkan titik yang lain dan menghubungkan *root* pada A dan B. *Root* dapat juga ditempatkan dimana saja dalam pohon. Jadi terdapat beberapa besar kemungkinan untuk *rooted* daripada *unrooted* untuk memberikan sejumlah *taxa* atau sekuen.

Dalam mengkonstruksi pohon filogenetika dapat diklasifikasikan menjadi 2 kategori yang digunakan sebagai strategi untuk menghasilkan pohon filogenetika terbaik. Kategori pertama adalah memeriksa semua atau sejumlah besar kemungkinan pohon filogenetika dan memilih satu yang terbaik dengan kriteria-kriteria tertentu. Biasanya disebut dengan metode *exhaustive-search*. Metode *maximum parsimony, Fitch Margoliash* dan *maximum likehood* termasuk dalam kategori ini. Kategori yang kedua adalah memeriksa hubungan topologi lokal dari pohon dan mengkonstruksi pohon terbaik dengan langkah demi langkah. Metode *Neighbor-joining* dan beberapa metode *Distance* lainnya adalah termasuk dalam kategori yang kedua ini (SAITOU dan IMANISHI, 1989). Dalam makalah ini akan dibahas tiga metode saja yaitu: (1) *Maximum parsimony*, (2) *Distance* dan (3) *Maximum likehoood* yang secara umum digunakan untuk membentuk pohon evolusi atau pohon terbaik untuk mengamati variasi sekuen dalam kelompok. Masing-masing metode ini digunakan untuk tipe analisis yang berbeda (MOUNT, 2001).

Metode *maximum parsimony*

Parsimony atau metode minimum *evolution* pertama kali digunakan dalam filogenetik oleh Camin and Sokal pada tahun 1965 (FELSENSTEIN, 1978). Metode ini memprediksikan pohon evolusi/ *evolutionary tree* yang meminimalkan jumlah langkah yang dibutuhkan untuk menghasilkan variasi yang diamati dalam sekuen. Untuk alasan ini, metode ini juga sering disebut sebagai metode evolusi minimum/*minimum evolution method*. Sebuah *multiple sequence alignment* dibutuhkan untuk memprediksi posisi sekuen yang sepertinya berhubungan. Posisi ini akan menampilkan kolom vertikal dalam *multiple sequence alignment*. Untuk masing-masing posisi yang disejajarkan, pohon filogenetika membutuhkan perubahan evolusi dalam jumlah terkecil untuk menghasilkan pengamatan perubahan sekuen yang diidentifikasi. Analisis ini terus menerus dilakukan terhadap masing-masing posisi dalam penjejeran sekuen. Akhirnya, pohon yang menghasilkan jumlah perubahan terkecil secara keseluruhan dihasilkan untuk semua posisi sekuen yang diidentifikasi. Metode ini berguna untuk sekuen yang mirip dan dalam jumlah yang sedikit. Alogaritma yang digunakan tidak rumit tetapi dijamin untuk dapat menemukan pohon yang terbaik, sebab semua kemungkinan pohon yang dibentuk berhubungan dengan kelompok sekuen yang diperiksa. Untuk alasan ini, metode ini cukup membutuhkan banyak waktu dan tidak berguna untuk data sekuen dalam jumlah besar dan asumsi lain harus dibuat untuk *root* pohon yang diprediksikan.

Metode jarak/*distance method*

Metode jarak bekerja pada jumlah perubahan diantara masing-masing pasangan dalam kelompok untuk mengkonstruksi pohon filogenetika dalam kelompok. Pasangan sekuen yang mempunyai jumlah perubahan terkecil diantara mereka disebut *neighbors*. Pada pohon, sekuen-sekuen ini menggunakan secara bersama-sama satu titik atau posisi *common ancestor* dan masing-masing dihubungkan titik oleh sebuah cabang. Tujuan dari metode jarak adalah metode untuk mengidentifikasi pohon pada posisi *neighbors* dengan benar, dan juga mempunyai cabang yang menghasilkan data orisinil sedekat mungkin. Penemuan *neighbors* terdekat diantara kelompok sekuen dengan metode jarak biasanya langkah pertama dalam memproduksi sebuah *multiple sequence alignment*.

Metode jarak pertama kali ditemukan oleh Feng dan Doolitle; pengelompokan program oleh penulis tersebut menghasilkan sebuah penjejeran dan pohon dari set sekuen protein (FENG dan DOOLITLE, 1996). Program CLUSTALW, digunakan untuk *neighbor-joining distance method* sebagai panduan untuk *multiple sequence alignment*. Program PAUP versi 4 merupakan pilihan untuk membentuk sebuah analisis filogenetika dengan *distance method*. Program PHYLIP *package* yang membentuk analisis *distance* termasuk program yang secara otomatis dibaca dalam sekuen dalam PHYLIP *infile format* dan secara otomatis menghasilkan *file* yang disebut dengan tabel *distance*.

Dalam pengukuran jarak genetik menggunakan model substitusi nukleotida, suatu sekuen DNA akan dibandingkan satu nukleotida dengan nukleotida lainnya. Jarak ini dapat mengukur suatu sekuen

nukleotida baik yang menyandi protein maupun tidak. Pada jarak matrik (*distance matrices*) yang dihasilkan, mereka mungkin digunakan sebagai *input* yang mengikuti program analisis jarak dalam PHYLIP. Program PHYLIP semua secara otomatis membaca *input file* yang disebut *infile* dan menghasilkan sebuah *outfile*. Jadi, nama *file* harus diedit ketika menggunakan program ini. Sebagai contoh, *distance outfile* harus diedit untuk memasukkan hanya tabel *distance* dan jumlah taxa, dan ketika *file* disimpan dengan nama sekuen *infile*. Analisis *distance* dalam program PHYLIP adalah sebagai berikut:

1. FITCH mengestimasi sebuah pohon filogenetika yang mengasumsikan penambahan panjang cabang menggunakan metode Fitch-Margoliash dan tidak mengasumsikan sebuah *molecular clock* (mengikuti rata-rata evolusi sepanjang cabang yang bervariasi).
2. KITCH mengestimasikan sebuah pohon filogenetika tetapi dengan mengasumsikan *molecular clock.*
3. NEIGHBOR mengestimasi pohon filogenetika menggunakan *neighbor joining* atau metode *unweighted pair group* dengan rata-rata aritmatika (UPGMA). Metode *neighbor joining* tidak mengasumsi *molecular clock* dan meghasilkan *unrooted tree.*

Metode UPGMA mengasumsikan sebuah *molecular clock* dan *rooted tree*. Metode ini secara normal menghitung skor similaritas yang didefinisikan sebagai jumlah total dari jumlah sekuen yang identik dan jumlah substitusi konservatif dalam penjejeran dua sekuen dengan gap yang diabaikan. Skor identitas antara sekuen menunjukkan hanya identitas yang mungkin ditemukan dalam penjejeran. Untuk analisis filogenetik digunakan skor jarak antara dua sekuen. Skor diantara dua sekuen adalah jumlah posisi yang tidak cocok/*mismatch* dalam penjejeran atau jumlah posisi sekuen yang harus diubah untuk menghasilkan sekuen yang lain. Gap mungkin diabaikan dalam kalkulasi atau diberi perlakuan seperti substitusi. Ketika sebuah skoring atau matrik substitusi digunakan, kalkulasi menjadi lebih komplek tetapi secara prinsip tetap sama.

Metode Fitch dan Margoliash

Metode FITCH dan MARGOLIASH (1987) menggunakan tabel yang diilustrasikan seperti pada Gambar 2. Sekuen-sekuen dikombinasi dalam tiga untuk mendefinisikan cabang-cabang pohon yang diprediksikan dan untuk menghitung panjang-panjang cabang dari pohon. Ini adalah metode *averanging distance* merupakan metode yang paling akurat untuk pohon dengan cabang yang pendek. Adanya cabang yang panjang bertendensi menurunkan tingkat kepercayaan dari prediksi (SWOFFORD *et al*., 1996).

A. *Sequence*

Sequence A ACGCGTTGGGCGATGGCAAC
Sequence B ACGCGTTGGGCGACGGTAAT
Sequence C ACGCATTGAATGATGATAAT
Sequence D ACACATTGAGTGATAATAAT

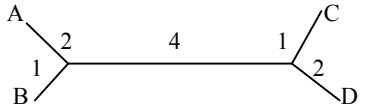
B. *Distance between sequences, the number of steps required to change one sequence into the other*

n_{AB} 3
n_{AC} 7
n_{AD} 8
n_{BC} 6
n_{BD} 7
n_{CD} 3

C. *Distance table*

	A	B	C	D
A	-	3	7	8
B	-	-	6	7
C	-	-	-	3
D	-	-	-	-

D. *The assumed phylogenetic tree for the sequences A-D showing branch lengths. The sum of the branch lengths between any two sequences on the trees has the same value as the distance between the sequences*



Gambar 2. Kelompok sekuen yang ideal dengan panjang cabang pohon yang diasumsikan

Sumber: MOUNT (2001)

Metode *neighbor - joining* (NJ)

Metode *neighbor-joining* sangat mirip dengan metode Fitch dan Margoliash kecuali tentang pemilihan sekuen untuk berpasangan ditentukan oleh perbedaan alogaritma. Metode *neighbor-joining* sangat cocok ketika rata-rata evolusi dari pemisahan *lineage* adalah di bawah pertimbangan yang berbeda-beda. Ketika panjang cabang dari pohon yang diketahui topologinya berubah dengan cara menstimulasi tingkat yang bervariasi dari perubahan evolusi, metode *neighbor-joining* adalah yang paling cocok untuk memprediksi pohon dengan benar (SAITOU dan MEI, 1987).

Neighbor-joining memilih sekuen yang jika digabungkan akan memberikan estimasi terbaik dari panjang cabang yang paling dekat merefleksikan jarak yang nyata diantara sekuen. Pada Gambar 4 menunjukkan pohon filogenetika yang dikonstruksi dengan metode *Neighbor-joining*.

Metode unweighted *pair group* dengan rata-rata aritmetika (UPGMA)

Metode jarak yang telah diuraikan di atas memberikan sebuah estimasi yang baik dari sebuah pohon evolusi dan tidak terpengaruh oleh variasi dalam rata-rata perubahan sepanjang cabang dari pohon. Metode UPGMA adalah metode sederhana untuk konstruksi pohon yang mengasumsikan rata-rata perubahan sepanjang pohon adalah konstan dan jaraknya kira-kira *ultrameric* (*ultrameric* biasanya diekspresikan sebagai *molecular clock tree*). Metode UPGMA dimulai dengan kalkulasi panjang cabang diantara sekuen paling dekat yang saling berhubungan, kemudian rata-rata jarak antara sekuen ini atau kelompok sekuen dan sekuen berikutnya atau kelompok sekuen dan berlanjut sampai semua sekuen yang termasuk dalam pohon. Akhirnya metode ini memprediksi posisi *root* dari pohon.

Pemilihan *Outgroup*

Jika kita ingin secara independen mendapatkan informasi yang meyakinkan dari sekuen lebih berhubungan, sebuah prosedur dapat diikuti dengan menambahkan sekuen pada pohon dan yang paling dekat dengan *root*. Modifikasi dapat meningkatkan prediksi dari pohon dengan metode di atas yaitu dengan menambahkan *outgroup* pada langkah akhir dari prosedur. Satu atau lebih sekuen jenis ini disebut sebagai *outgroup*. Sebagai contoh, sekuen A dan B berasal dari spesies yang telah diketahui terpisah satu dengan yang lain pada awal evolusi berdasarkan catatan fosil. A dan B kemudian diperlakukan sebagai *outgroup*. Pemilihan satu atau lebih outgroup dengan *distance method* dapat juga membantu dengan lokalisasi *root* dari pohon (SWOFFORD *et al.,* 1996). *Root* akan ditempatkan diantara *outgroup* dan titik yang menghubungkan sekuen. Sekuen dari *outgroup* semestinya berkorelasi dekat dengan sekuen-sekuen yang dianalisa, tetapi juga mempunyai perbedaan yang signifikan antara *outgroup* dengan sekuen yang lain daripada diantara sekuen itu sendiri (Gambar 3). Pemilihan sekuen *outgroup* yang terlalu jauh kemungkinan akan berperanan terhadap prediksi pohon menjadi salah akibat terdapat pebedaan yang secara random yang lebih banyak diantara sekuen *outgroup* dengan sekuen lainnya (LI dan GRAUR, 1991 dalam MOUNT, 2001). Perubahan *multiple sequence* pada masing-masing situs menjadi lebih mungkin dan akan lebih komplek untuk *genetic rearrangements* yang komplek. Untuk alasan yang sama, menggunakan sekuen yang terlalu berbeda dalam metode jarak dari prediksi filogenetik dapat berperanan terhadap kesalahan yang terjadi (SWOFFORD *et al*., 1996). Jumlah perbedaan yang meningkat, perubahan histori sekuen pada masing-masing situs menjadi lebih komplek dan menjadi sulit untuk memprediksi.

Gambar 3. Pohon filogenetika DNA *Maximum Likehood* sekuen hemagglutinin H7 virus Avian Influenza H7 low pathogenic

Sumber: MUNSTER *et al.* (2005)

Pendekatan *maximum likehood*

Metode ini menggunakan kalkulasi untuk menemukan pohon yang mempunyai hitungan variasi terbaik dalam set sekuen. Metode ini mirip dengan metode *maximum parsimony* dalam analisis yang dibentuk pada masing-masing kolom dalam *multiple sequence alignment*. Semua kemungkinan pohon yang terbentuk dipertimbangkan, sehingga metode ini hanya cocok untuk sekuen dalam jumlah kecil. Metode ini mempertimbangkan untuk masing-masing pohon, jumlah perubahan sekuen atau mutasi yang terjadi yang memberikan variasi sekuen. Metode *maximum likehood* menampilkan kesempatan penambahan untuk mengevaluasi pohon dengan variasi dalam rata-rata mutasi dalam *lineage* yang berbeda. Metode ini dapat digunakan untuk mengekplorasi hubungan antara

sekuen yang lebih beragam, dimana kondisi ini tidak dapat dilakukan dengan baik jika menggunakan metode *maximum persimony*. Kekurangan metode *maximum likehood* adalah membutuhkan pekerjaan komputer yang sangat intensif. Jika menggunakan komputer yang lebih cepat, metode *maximum likehood* dapat digunakan untuk model evolusi yang lebih komplek. Metode ini juga dapat digunakan untuk menganalisa mutasi pada *overlapping reading frame* pada virus (SCHADT *et al*., 1998). Pada Gambar 3 adalah sebagai contoh dari pohon filogenetika dengan menggunakan *maximum likehood.*

Prediksi filogenetik yang dipercaya

Analisis filogenetika set sekuen yang menjejerkan dengan baik adalah jelas sebab posisi yang bertanggung jawab dalam sekuen dapat diidentifikasi dalam *multiple sequence alignment* dari sekuen. Tipe perubahan dalam penjejeran posisi atau jumlah yang berubah dalam penjejeran antara pasangan sekuen menyediakan dasar untuk menentukan hubungan filogenetika diantara sekuen berdasarkan metode analisis filogenetika. Penentuan perubahan sekuen yang telah terjadi menjadi sulit sebab *multiple sequence alignment* mungkin tidak optimal dan sebab perubahan yang banyak terjadi pada penjejeran posisi sekuen. Pilihan metode *multiple sequence alignment* tergantung pada tingkat variasi diantara sekuen. Jika penjejeran yang cocok telah ditemukan, pertanyaannya adalah bagaimana prediksi filogenetika didukung oleh data dalam *multiple sequence alignment.*

Dalam metode *bootstrap*, data dilakukan *resampled,* dengan secara random memilih kolom vertikal dari sekuen yang dijejerkan untuk menghasilkan penjejeran, dan dalam pengaruh sebuah penjejeran baru dengan panjang yang sama. Masing-masing kolom digunakan lebih dari satu kali dan beberapa kolom mungkin tidak digunakan pada semua penjejeran yang baru. Pohon-pohon kemudian diprediksi dari beberapa penjejeran ini dari *resampled* sekuen (FELSENSTEIN, 1988). Untuk cabang-cabang dalam topologi filogenetika yang diprediksi menjadi signifikan jika set data *resampled* seharusnya berulangkali (sebagai contoh > 70%) memprediksi cabang-cabang yang sama.

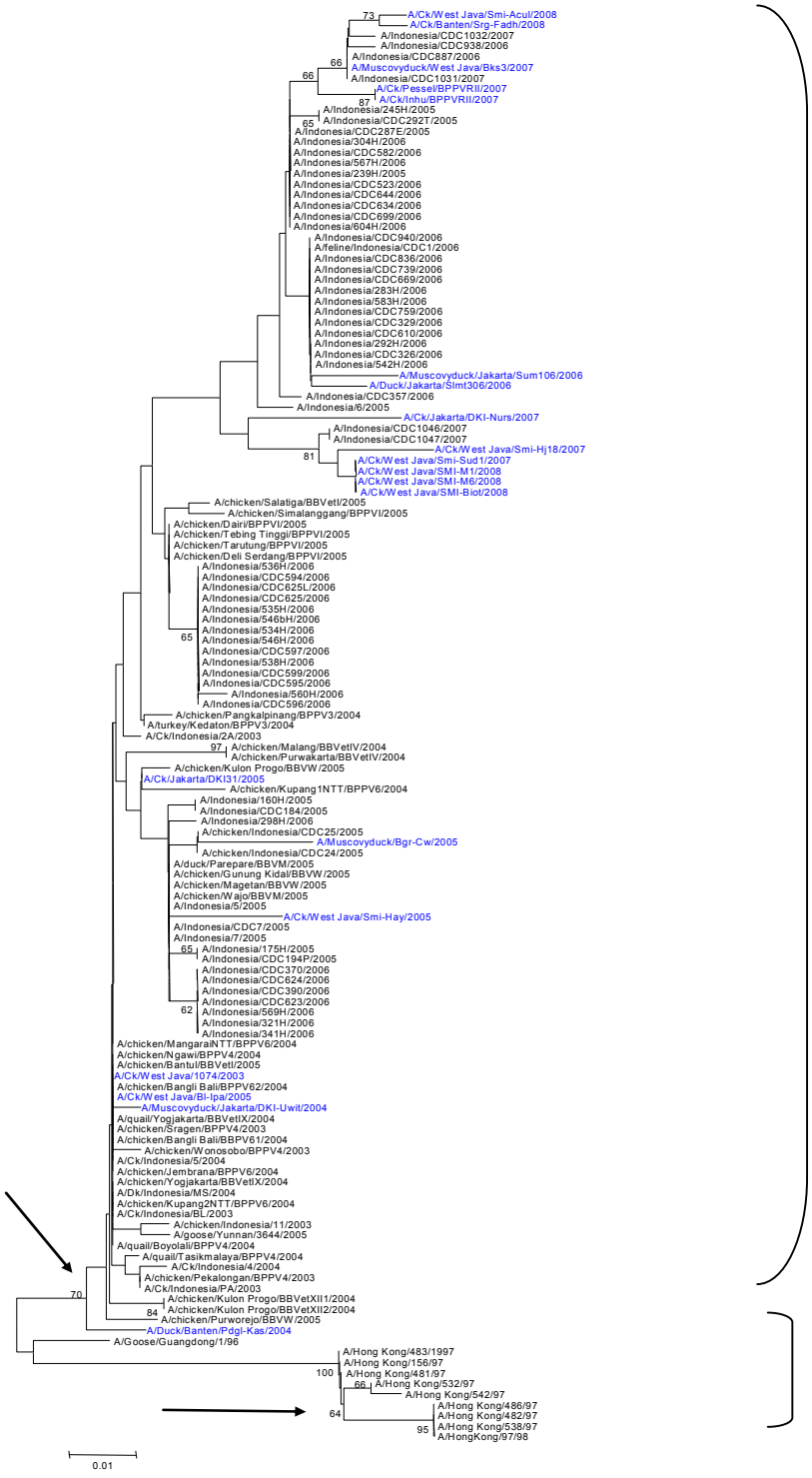
Analisis *bootstrap* adalah metode yang menguji seberapa baik set data model. Sebagai contoh validitas penyusunan cabang dalam prediksi pohon filogenetik dapat diuji dengan *resampled* dari kolom dalam *multiple sequence alignment* untuk membentuk beberapa penjejeran baru. Penampakan cabang dalam pohon dari sekuen *resampled* ini dapat diukur. Alternatifnya, sekuen kemungkinan harus dikeluarkan dari analisis untuk menentukan berapa banyak sekuen yang mempengaruhi hasil dari analisis. *Bootstrap analysis* didukung oleh sebagian besar paket *software* menguji cabang-cabang yang dapat dipercaya.

Sebagai contoh bagaimana analisis filogenetik dibuat, dapat dilihat pada Gambar 4.

Gambar 4 adalah sebuah pohon filogenetika molekuler yang disusun dari urutan sekuen nukleotida pada gen Matrix dari organisme virus Avian Influenza virus H5N1. Nilai *boostrap* ditunjukkan pada angka yang terletak pada cabang-cabang pohon filogenetika (tanda panah). Grup 2 yaitu kelompok gen Matrix dari virus influenza asal Hong Kong dan China yang digunakan sebagai *outgroup* dari keseluruhan analisis filogenetik pada Gambar 3, Grup 2 dan grup 1 merupakan virus influenza yang sama yang mempunyai kemiripan dengan virus influenza asal Indonesia sehingga sebagai *outgroup*, Grup 2, berkorelasi dekat dengan sekuen-sekuen yang dianalisis, tetapi juga mempunyai perbedaan yang signifikan antara *outgroup* dengan sekuen yang lain daripada diantara sekuen itu sendiri.

Jarak genetik berdasarkan metode algoritma pembentukan pohon akan menampilkan data berupa pohon filogenetika. Pohon filogenetika memberi informasi tentang pengklasifikasian populasi berdasarkan hubungan evolusionernya. Dalam rekontruksi pohon filogenetika, data molekul lebih banyak dipakai karena dianggap lebih stabil dalam proses evolusi dibandingkan dengan data morfologi. Pohon filogenetika dapat berakar (*rooted*) atau tidak berakar (*unrooted*), tergantung metode analisis yang dipergunakan. Akar pada pohon menggambarkan titik percabangan pertama atau asal masing-masing populasi dengan asumsi bahwa laju evolusi berjalan konstan (NEI, 1987). Pola percabangan pohon dibentuk berdasarkan jarak matrik antar pasangan populasi yang dapat menggambarkan fusi genetik yang terjadi pada kelompok tersebut (WEISS, 1995). Panjang cabang menggambarkan jumlah substitusi basa yang dapat berupa polimorfisme DNA atau haplotipe. Metode pengolahan data yang digunakan harus sesuai dengan set data yang ada, agar dapat menghasilkan pola percabangan (topologi) serta panjang cabang yang benar (CAVALLI-SFORZA, 1997). Topologi pohon yang salah akan mengakibatkan panjang cabang yang salah dan pohon secara keseluruhan tidak memberi informasi genetik apapun. Semua metode diatas mempunyai keunggulan dan kelemahan masing-masing, sehingga penggunaannya disesuaikan dengan bentuk dan jenis data yang akan diolah. Metode yang paling sering digunakan adalah metode *Neighbor-Joining* (NJ). Metode NJ merupakan metode yang disederhanakan dari metode *minimum evolution* (ME).

Untuk memperkecil kesalahan dalam mengkonstruksi pohon filogenetika dapat dilakukan sampling ulang dengan petanda genetik lain pada sampel yang sama dan kemudian membandingkan



Gambar 4. Pohon filogenetik gen M2 virus AI asal unggas di sekitar kasus H5N1 pada manusia. Grup 1 merupakan kelompok virus H5N1 asal Indonesia dan Grup 2 adalah virus asal China/Hong Kong sebagai *outgroup*. Kontruksi filogenetik menggunakan metode *neighbor-joining* dan analisis *bootstrap* (1.000 *replicates*) menggunakan model Kimura-Nei dalam software MEGA 4

Sumber: DHARMAYANTI *et al*. (2010)

kedua bentuk pohon tersebut. Akan tetapi tindakan tersebut membutuhkan biaya besar sehingga hampir tidak mungkin dilakukan. Sebagai gantinya EFRON (1979) memperkenalkan metode sampling ulang (*resampling*) dari data yang telah ada yang dikenal dengan analisis *bootstrap* untuk menguji validitas konstruksi pohon filogenetika.

KESIMPULAN

Pentingnya pemahaman bagaimana mengggunakan analisis filogenetika khususnya filogentika molekuler yang berasal dari data nukleotida atau asam amino sangat berperanan dalam pembuatan pohon filogenetik terbaik dan dapat dipercaya. Metode filogenetika yang telah dibahas dalam makalah ini digunakan untuk tipe analisis yang berbeda. Nilai *bootstrap* merupakan untuk menguji seberapa baik set data model yang kita gunakan. Jika nilai *bootstrap* rendah maka sekuen seharusnya dikeluarkan dari analisis untuk mendapatkan sebuah pohon filogenetika yang dapat dipercaya.

DAFTAR PUSTAKA

BLACKSHIELDS, G., I.M. WALLACE, M. LARKIN and D.G. HIGGINS. 2006. Analysis and comparison of benchmarks for multiple sequence alignment. Silico Biol. 6: 321 – 339.

BRINKMAN, F. and D. LEIPE. 2001. Phylogenetic Analysis. *In:* Bioinformatics: A Practical Guide to the Analisys of Gene and Protein. BAXEVANIS, A.D. and B.F.F. OUELLETTE (Eds.). John Willey & Sons. pp. 323 – 358.

CAVALLI-SFORZA, L.L. 1997. Genes, Peoples and Languages. Proc. Natl. Acad. Sci. USA. 94(15): 7719 – 7724.

DHARMAYANTI, N.L.P.I., F. IBRAHIM and A. SOEBANDRIO. 2010. Amantadine resistant of Indonesian influenza H5N1 subtype virus during 2003 – 2008. Microbiol Indones. 5(1): 11 – 16.

EDGAR, R.C. and S. BATZOGLOU. 2006. Multiple sequence alignment. Curr. Opin. Struct. Biol. 6: 368 – 373.

EFRON, B. 1979. Bootstrap Methods: Another Look at the Jackknife. Ann. Statist. 7(1): 1 – 26.

ESTABROOK, G. 1984. Phylogenetic trees and character-state trees. *In*: Perspectives on the Reconstruction Evolutionary History Cladistics. DUNCAN, T. and T. STUESSY (Eds.). Columbia University Press. pp. 135 – 151.

FELSENSTEIN, J. 1978. Cases in which Parsimony or Compatibility Methods will be positively misleading. Systematic Zoology 27(4): 401 – 410.

FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: Inferences and realibility. Annu. Rev. Genet. 22: 521 – 565.

FENG, D.F. and R.F. DOOLITTLE. 1996. Progressive alignment of amino acid sequences and construction of phylogenetic trees from them. Methods Enzymol. 266: 368 – 382.

FITCH, W.M. and E. MARGOLIASH. 1987. Construction of phylogenetic trees. Science. 155: 279 – 284.

KEMENA, C. and C. NOTREDAME. 2009. Upcoming challenges for multiple sequence alignment methods in the high-throughput era. Bioinformatics. 25: 2455 – 2465.

LI, S., D. PEARL and H. DOSS. 1999. Phylogenetic tree construction using Markov Chain Monte Carlo. Fred Hutchinson Cancer Research Center Washington. http://www.stat.ohio-state.edu/~doss/Research/mc-trees.pdf. (23 Januari 2011).

LIPSCOMB, D. 1998. Basics of Cladistic Analysis. Student guide paper. George WashingtonUniversity. http:// www.gwu.edu/~clade/faculty/lipscomb/Cladistics.pdf (23 Januari 2011).

MCDONALD, J.H and M. KREITMAN. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 351: 652 – 654.

MOUNT, D.W. 2001. Phylogenetic prediction. *In:* Bioinformatic, Sequence and Genome Analysis. Cold Spring Harbor laboratory. New York Press pp. 237 – 280.

MUNSTER, V.J., A. WALLENSTEN, C. BAAS, G. F. RIMMELZWAAN, M. SCHUTTEN, B OLSEN, A. D.M.E. OSTERHAUS and R.A.M. FOUCHIER. 2005. Mallards and Highly Pathogenic Avian Influenza Ancestral Viruses, Northern Europe. EID: 1545 – 1551.

NEI, M. 1987. Molecular Genetics. Columbia University New York Press.

NIELSEN, R. and Z. YANG. 1998. Likehood models for detecting positively selected amino acid sites and application to the HIV-1 envelope gene. Genetics. 148: 929 – 936.

NOTREDAME, C. 2007. Recent evolutions of multiple sequence alignment algorithms. PLOS Comput. Biol. 3: E123.

SAITOU, N. and M. MEI. 1987. The neighbor-joining method: A new method for constructing phylogenetic trees. Mol. Biol. Evol. 4: 406 – 425.

SAITOU, N. and T. IMANISHI. 1989. Relative efficiencies of the Fitch-Margoliash, Maximum-Parsimony, Maximum-Likehood, Minimum Evolution amd Neighbor-joining Methods of phylogenetic tree construction in obtaining the correct tree. Mol. Biol. Evol. 6(5): 514 – 525.

SCHADT, E.E., J.S. SINSHEIMER and K. LANGE. 1998. Computational advances in maximum likehood methods for molecular phylogeny. Genome Res. 8: 222 – 233.

SCHMIDT, H. 2003. Phylogenetic Trees from Large Datasets. Inaugural-Dissertation, Dusseldorf University. http://www.bi.uniduesseldorf.de/~hschmidt/publ/schmidt2003.phdthesis.pdf. (23 Januari 2011).

SWOFFORD, D.L., G.J. OLSEN., P.J. WADDELL and D.M. HILLS. 1996. Phylogenetic inference. *In*: Molecular Systematics, 2nd Edition. HILLS, D.M., C. MORITZ and B.K. MABLE (Eds.) Sinauer Associates, Sunderland, Massachusetts. Chap. 5 pp. 407 – 514.

WEISS, K.M. 1995. Genetic variation and human diseases: Principles and evolution approaches. Cambridge University Press, Cambridge. 354 p.