

**Penerapan Steiner Tree Pada Konstruksi Filogeni Minimal
Berdasar Rangkaian Asam Amino**
(*The Application Of Steiner Tree On Minimum Phylogeny Construction
Based On Amino Acid Sequence*)

Firdaus Ubaidillah
Staf Pengajar Jurusan Matematika FMIPA Universitas Jember

ABSTRACT

The sequence of amino acid in certain proteins on organism has given inspiration to construct phylogeny. Based on the difference, it can be determined mutation distance between species pairs. Using the mutation distance data, we can construct a weighted complete graph in which the nodes correspond to species and the edges weight correspond to the mutation distance. Once the complete graph is obtained, the next step is to determine the minimum spanning tree (MST). In order to construct the minimum phylogeny, we consider phylogenetic tree as Steiner tree problem in phylogeny through the reduction of its MST length so that it can form new tree with the least tree length. The weights on the connected edges with at least two degrees on the MST are reduced according to the dissimilarities of amino acid in the pairs at the suitable sites.

Keywords: minimum spanning tree, Steiner tree, phylogeny

PENDAHULUAN

Dalam teori evolusi dikemukakan bahwa semua spesies berasal dari satu leluhur yang sama. Leluhur dari semua spesies adalah spesies yang pertama kali ada. Setiap organisma atau spesies yang ada sekarang ini berasal dari satu induk yang sangat sederhana kemudian berangsur-angsur mengalami perubahan ke arah yang kompleks dalam proses waktu yang panjang.

Selama evolusi semua bagian organisma berubah. Anggota badan pada hewan berevolusi menjadi banyak bentuk dan ukuran, jaringan pembentuknya, sel-selnya dan molekulnya juga berubah. Evolusi molekul selalu dipandang sinonim dengan evolusi protein, karena evolusi di tingkat molekul sebagian besar dipelajari secara menyeluruh pada protein. Protein adalah klas molekul yang paling umum dan paling berdiversifikasi pada organisma. Semua protein dibentuk oleh rantai asam amino dan secara struktural dibedakan oleh rangkaian rantai asam amino penyusunnya.

Evolusi molekuler adalah evolusi di tingkat molekul. Sumber evolusi molekuler yang utama berasal dari mutasi yaitu perubahan dalam struktur DNA yang menyebabkan berbagai protein terbentuk. Ada beberapa macam mutasi antara lain mutasi kromosom berskala besar, terjadi bilamana seluruh daerah kromosom lenyap, duplikasi, diubah atau dipindah ke kromosom lain. Mutasi titik, yaitu apabila reproduksi basis yang keliru disalin

dari DNA tetua. Keadaan ini menyebabkan perubahan dalam suatu asam amino dari suatu protein (Ridley, 1991).

Salah satu cabang dalam ilmu biologi yang mempelajari tentang hubungan evolusioner antar sekumpulan spesies atau taksa adalah filogenetik. Pengertian tree filogenetik, singkatnya filogeni, adalah hubungan evolusioner antar sekumpulan taksa atau spesies yang dinyatakan dalam tree dimana daun-daunnya, yakni simpul-simpul berderajat satu, menyatakan spesies-spesies yang berbeda (Kannan *et al.*, 1998). Pengertian tentang tree filogenetik yang diambilkan dari Baca & Lagergren (1998; 2003) sebagai berikut. Misal C adalah himpunan karakter atau situs, dan untuk setiap $c \in C$ misal \mathcal{A}_c adalah himpunan state asam amino yang diperkenankan pada karakter atau situs c . Misal $m = |C|$ dan S adalah himpunan spesies yang diberikan, maka untuk spesies $s \in S$ adalah unsur dari $\mathcal{A}_1 \times \mathcal{A}_2 \times \dots \times \mathcal{A}_m$. Filogeni untuk himpunan S yang memuat n spesies adalah tree T yang mempunyai sifat-sifat :

(i). $S \subseteq V(T) \subseteq \mathcal{A}_1 \times \mathcal{A}_2 \times \dots \times \mathcal{A}_m$;

(ii). Setiap daun di T adalah di S .

Bobot atau panjang filogeni T didefinisikan sebagai

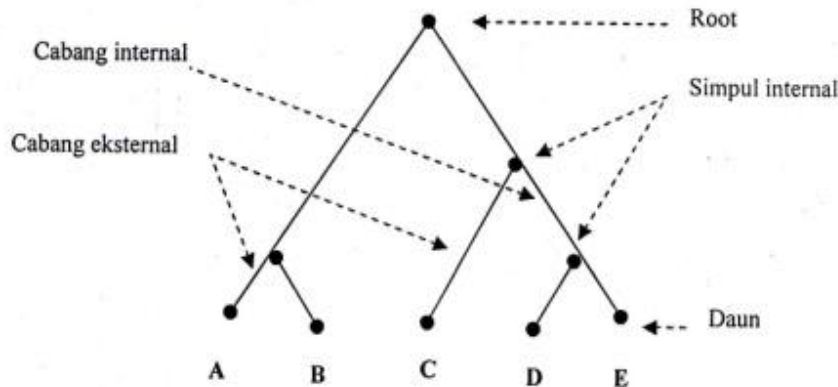
$$L(T) = \sum_{(u,v) \in E(T)} w(u,v)$$

dimana untuk dua spesies u dan v , $w(u,v)$ menyatakan banyaknya state karakter u dan v

berbeda. Dalam filogeni berbentuk rooted tree, simpul-simpul internal tidak diberi label.

Konstruksi filogeni banyak dipelajari dalam biologi molekuler maupun ilmu komputer dimana problem ini serupa dengan problem dalam optimasi jaringan. Salah satu alasan problem ini adalah karakteristik NP-complete (Gonnet, 1994), oleh karena itu masih menjadi problem terbuka.

Pada umumnya penyajian filogeni menggunakan *rooted tree*. Pada rooted tree memuat daun-daun, simpul-simpul cabang dan panjang cabang yang masing-masing menyatakan spesies, leluhur dan banyaknya perubahan atau mutasi selama evolusi. Gambar 1 menjelaskan pengertian sebuah tree filogenetik. Sebuah daun menyatakan serangkaian nukleotida atau asam amino.



Gambar 1. Filogeni dengan penjelasan.

Banyaknya topologi tree yang mungkin, tergantung dari banyaknya spesies. Jumlah topologi yang mungkin untuk rooted tree dengan n spesies, $n \geq 2$, adalah $(2n-3)(2n-5) \dots 3 \cdot 1 = (2n-3)!!$ (Felsenstein, 1978).

Sebagai contoh, untuk 20 spesies terdapat $8,2 \times 10^{21}$ topologi tree. Untuk beberapa kasus n , banyaknya topologi tree yang mungkin, diberikan dalam Tabel 1. Mengingat semakin besar nilai n maka jumlah topologi yang mungkin juga membesar secara eksponensial. Oleh karena itu, tidak mungkin menentukan semua topologi tree untuk n yang cukup besar.

Tabel 1. Banyaknya topologi rooted tree dengan n spesies

Ban/aknya spesies (n)	Banyaknya tree
1	1
2	1
3	3
4	15
5	105
6	945
7	10395
8	135135

9	2027025
10	34459425
11	654729075
12	13749310575
13	316234143225
14	7905853580625
15	213458046676875

METODE

Pada umumnya ada tiga metode dalam mengkonstruksi filogeni dari rangkaian asam amino. Ketiga metode itu adalah berdasarkan jarak, misalnya *least square* dan filogeni minimal; berdasarkan karakter, misalnya parsimoni; berdasarkan probabilitistik, misalkan *maximum likelihood*. Dalam konstruksi ini menggunakan metoda jarak (*distance method*) yang berdasar pada disimilaritas susunan asam amino. Tahap awal dalam mengkonstruksi filogeni adalah menyusun tabel yang berupa matriks. Setiap entri pada matriks menunjukkan *jarak mutasi* atau *jarak genetik* yaitu banyaknya perbedaan asam-asam amino pada situs yang bersesuaian.

Selanjutnya membangun graf lengkap atas n spesies dengan memandang spesies sebagai simpul dan jarak mutasi sebagai bobot sisi

yang menghubungkan kedua simpul. Banyaknya spanning tree berbeda yang dapat dibentuk dari graf lengkap ini adalah n^{n-2} (Tucker, 1995). Tiap-tiap spanning tree ini merupakan filogeni. Namun tidak mungkin menggunakan semua tree ini untuk konstruksi filogeni. Dengan beranggapan bahwa sesuatu yang terjadi di alam berjalan secara optimal, maka tree yang akan digunakan untuk konstruksi filogeni adalah spanning tree yang mempunyai bobot terkecil yaitu minimal spanning tree (MST). Salah satu MST atas graf lengkap itu digunakan sebagai tree referensi. Dari tree referensi ini dapat ditentukan semua MST atas graf lengkap tersebut (Narwen, 2002).

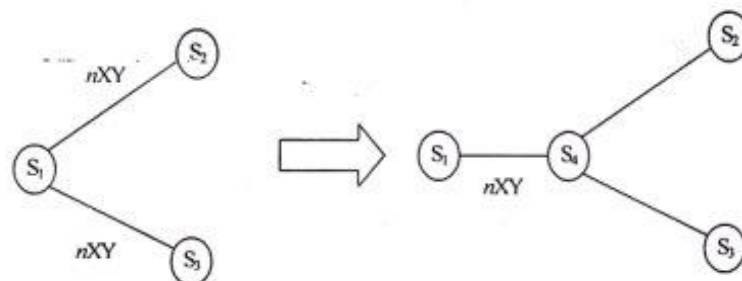
Menentukan filogeni minimal dari n spesies yang diberikan merupakan suatu problem Steiner tree dalam filogeni yaitu menentukan filogeni dengan bobot minimal yang membolehkan menambah spesies-spesies selain n spesies yang diberikan (Foulds, 1984; Baca & Lagergren, 2003). Pengertian Steiner tree diberikan oleh Lu, Tang & Lee (2003) sebagai berikut. Diberikan sebuah graf $G = (V, E)$ terhubung sederhana, subset simpul-simpul $S \subseteq V$, dan fungsi metrik $w : E \rightarrow R^+$. Steiner tree untuk S adalah sebarang tree di G yang merentang semua simpul-simpul di S . Steiner tree dapat memuat simpul-simpul tambahan selain yang diberikan. Bobot Steiner tree didefinisikan sebagai jumlah bobot semua sisi-sisinya. Problem Steiner tree adalah menentukan Steiner Minimum Tree yakni Steiner tree dengan panjang minimal di G .

Mengingat bobot Steiner minimal tree tidak lebih dari MST, maka salah satu cara untuk menentukan filogeni minimal dapat melalui MST-nya. Tidak semua MST itu dipilih untuk filogeni minimal. Untuk menentukan filogeni minimal melalui MST, tree dapat dipandang sebagai problem Steiner tree dalam ruang metrik (A, w) dimana $A = A_1 \times A_2 \times \dots \times A_n$

menyatakan himpunan spesies dan $w : A \times A \rightarrow R^+$ menyatakan jarak mutasi antar dua spesies. Filogeni minimal adalah MST terpilih dan yang menghasilkan pereduksian terbesar. Sebelum melakukan pereduksian, terlebih dahulu menentukan MST terpilih atas semua MST yang ada.

Pemilihan MST yang akan digunakan untuk konstruksi filogeni minimal adalah berdasarkan pemilihan sisi-sisi pada satu kelas ekuivalen yang dipilih lebih dari satu kali yang menghasilkan total nilai mutasi pasangan asam amino terkecil. Nilai mutasi pasangan asam amino adalah nilai minimal yang dihasilkan berdasarkan mutasi nukleotida kodon dari satu asam amino ke asam amino yang lain. Sankoff dalam (Felsenstein, 2004) memberikan nilai mutasi nukleotida pada satu kelompok basa purin atau pirimidin adalah 1, sedang nilai mutasi nukleotida dari kelompok basa purin ke pirimidin atau sebaliknya adalah 2,5. Hal itu berdasarkan bahwa mutasi antar satu kelompok basa ke kelompok basa yang lain lebih sulit terjadi dibandingkan dengan mutasi dalam satu kelompok basa.

Tahap selanjutnya adalah mengkonstruksi filogeni minimal dengan cara mereduksi bobot MST terpilih. Prosedurnya adalah mengambil salah satu MST lalu menggambar MST tersebut disertai bobot sisi dengan menampilkan perbedaan asam amino dengan menggunakan simbol $i\alpha\beta$ untuk menyatakan perubahan atau perbedaan situs ke- i dari asam amino α ke asam amino β . Sisi-sisi yang bersisihan dengan simpul yang sama dan memiliki kesamaan disimilaritas asam amino dapat digabungkan untuk menciptakan tree baru dengan menambah simpul baru pada MST itu. Proses kombinasi itu melalui suatu cara reduksi yang dijelaskan oleh Foulds (1984; 1986) seperti dalam Gambar 2.



Gambar 2. Proses penggabungan. S4 adalah simpul baru yang timbul akibat proses reduksi.

HASIL DAN PEMBAHASAN

Dalam konstruksi filogeni minimal ini menggunakan contoh kasus asam amino protein cytochrome c atas 23 spesies hewan yang diambil dari PIR (2004). Asam amino methionine (M) sebagai pengambil inisiatif dapat diabaikan sehingga panjang rangkaian

104. Data 23 spesies dan rangkaian asam amino untuk 30 situs pertama diberikan dalam Gambar 3. Jarak mutasi antar spesies dihitung berdasarkan jumlah perbedaan asam amino. Jarak mutasi selengkapnya diberikan pada Tabel 2.

	situs :	1	1111111112	222222223
		1234567890	1234567890	1234567890
S1_Arabian_camel		GDVEKGGKIF	VQKCAQCHTV	EKGGKHKTGP
S2_Spider_monkey		...F...R..	IM..S.....
S3_Chicken		..IF.....S.....
S4_Cattle	
S5_Dogfish	V.N.....
S6_Donkey	
S7_Domestic_duck	S.....
S8_Emu		..I.....S.....
S9_S.elephant_seal	
S10_E.gray_kangaroo	
S11_Domestic_horse	
S12_Rhesus_macaque		IM..S.....
S13_Man		IM..S.....
S14_Domestic_pigeon		.VIAGD....S.....
S15_mor.soon_river-p	R.....SA	QANL.....
S16_House_mouse	
S17_Pacific_lamprey	V.S.....	..A.....
S18_Domestic_rabbit	
S19_Ostrich		..I.....S.....
S20_King_penguin		..IKGK....S.....
S21_Snapping_turtle	
S22_Common_turkey		..I.....S.....
S23_W.rattlesnake		SM..GT....	.E.....

Gambar 3. Rangkaian asam amino untuk 23 spesies. Data yang ditampilkan merupakan 30 situs pertama dari 104 situs. Bagian kiri dari tiap-tiap baris menyatakan nomor dan nama spesies. Tanda titik menyatakan simbol asam amino yang sama dari asam amino spesies paling atas pada situs yang sama.

Dengan memandang spesies sebagai simpul dan jarak mutasi sebagai sisi yang menghubungkan kedua simpul memberikan hasil sebuah MST seperti Gambar 3 dengan panjang 132 yang digunakan sebagai tree referensi. Jumlah semua MST yang mungkin adalah 9216. Berdasarkan pemilihan MST, maka untuk konstruksi filogeni minimal hanya menghasilkan 8 MST. Kedelapan MST itu diberikan dalam Tabel 3.

Hasil reduksi dari kedelapan MST melalui proses penggabungan menghasilkan dua filogeni minimal berbeda dengan bobot 121.

Salah satu filogeni minimal itu diberikan dalam Gambar 5. Filogeni minimal yang lain memberikan hasil yang serupa kecuali berbeda di sisi (S8-S30).

Penyajian akhir filogeni minimal hasil tes bootstrap consensus tree dengan program MEGA 2.1 (Kumar dkk, 2004) diberikan dalam Gambar 6. Terlihat spesies-spesies berkelompok dalam beberapa klas. Spesies-spesies S24, S25, S26, S27, S28 diduga merupakan spesies dari klas mamalia dan S29 dari klas aves. Domestic pigeon dan king penguin berdasarkan klasifikasi hewan

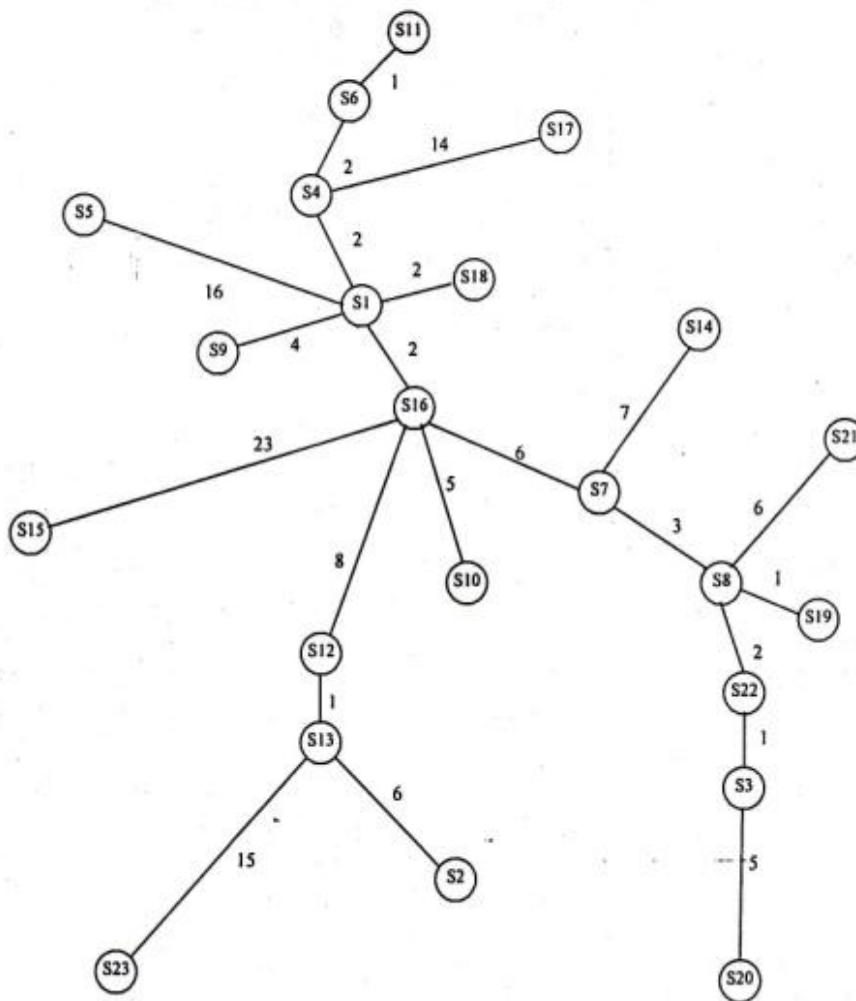
merupakan spesies dari klas aves tetapi dalam berkelompok dengan mamalia. konstruksi filogeni ini ternyata lebih dekat

Tabel 2. Jarak mutasi antar pasangan 23 spesies. Kolom terakhir adalah nama-nama spesies sesuai pada kolom pertama.

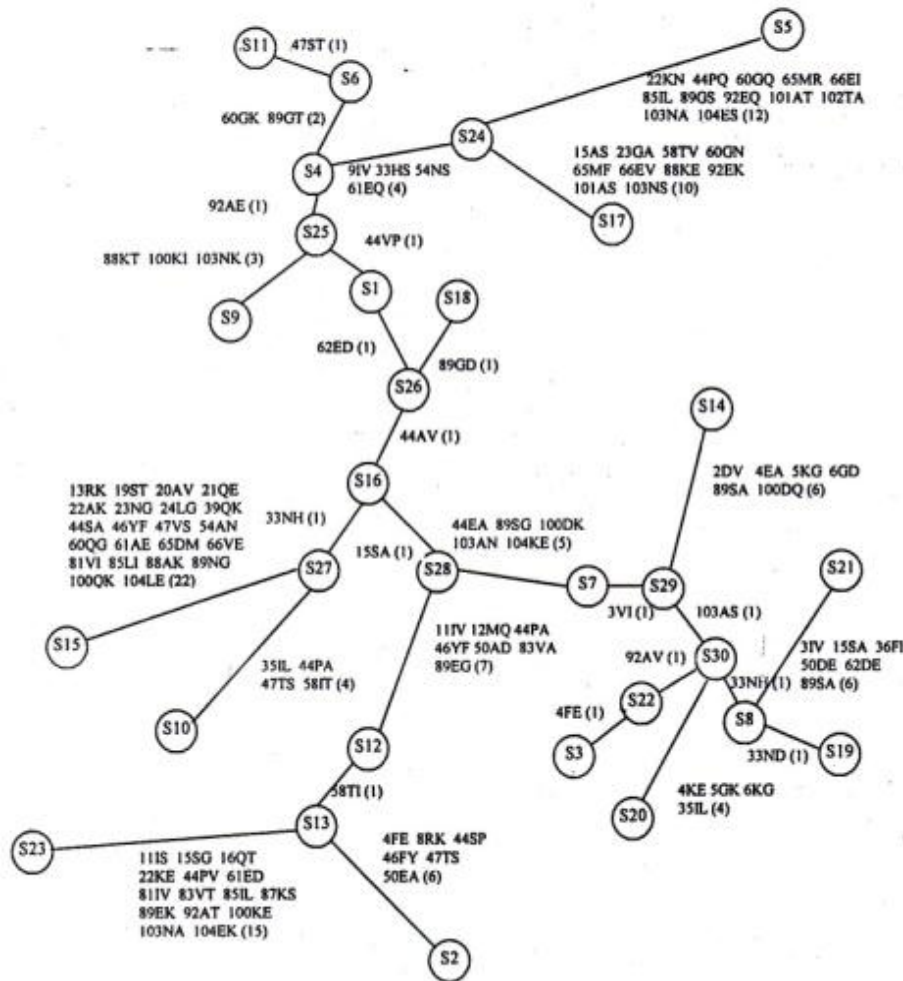
	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	Spesies
S1	12	10	2	16	4	7	9	4	6	5	9	10	12	24	2	15	2	9	12	8	9	19	Arabian camel
S2		14	13	26	14	13	15	15	11	13	7	6	17	30	11	23	11	15	17	16	15	20	Spider monkey
S3			10	20	11	4	3	11	13	12	13	14	8	28	9	19	9	3	5	9	1	20	Chicken
S4				16	2	8	10	4	6	3	9	10	13	25	3	14	4	10	13	9	9	20	Cattle
S5					16	17	19	18	20	17	23	24	23	27	17	16	17	19	23	19	19	26	Dogfish
S6						9	11	6	8	1	10	11	14	25	5	15	5	11	14	10	10	21	Donkey
S7							3	8	10	10	10	11	7	25	6	18	6	3	6	7	3	17	Domestic duck
S8								10	10	12	12	13	8	25	8	18	8	1	5	6	2	20	Emu
S9									8	7	11	12	13	25	5	15	6	10	13	9	10	21	Elephant seal
S10										7	11	10	15	24	5	17	6	11	13	11	12	21	Gray kangaroo
S11											11	12	15	25	6	16	6	12	15	11	11	22	Domestic horse
S12												1	15	27	8	20	8	12	15	14	12	16	Rhesus macaque
S13													16	28	9	20	9	13	16	15	13	15	Man
S14														29	11	23	11	8	7	12	8	22	Domestic pigeon
S15															23	29	23	26	30	26	27	30	Monsoon river-prawn
S16																16	2	8	11	9	8	19	House mouse
S17																	17	18	22	19	18	29	Pacific lamprey
S18																		8	11	9	8	18	Domestic rabbit
S19																			5	7	2	20	Ostrich
S20																				11	5	23	King penguin
S21																					8	21	Common turkey
S22																						19	Rattlesnake
S23																							Crotalus viridis

Tabel 3. MST-MST terpilih untuk konstruksi filogeni minimal

Bobot	Sisi
1	(S3-S22), (S6-S11), (S8-S19), (S12-S13)
2	(S1-S4), (S4-S6)
2	{{(S1-S16), (S16-S18)} atau {{(S1-S18), (S16-S18)}}
2	(S8-S22) atau (S19-S22)
3	(S7-S22)
4	(S4-S9)
5	(S10-S16)
5	(S20-S22)
6	(S2-S13), (S8-S21)
6	(S7-S16)
7	(S7-S14)
8	(S12-S16)
14	(S4-S17)
15	(S13-S23)
16	(S4-S5)
23	(S15-S16) atau (S15-S18)



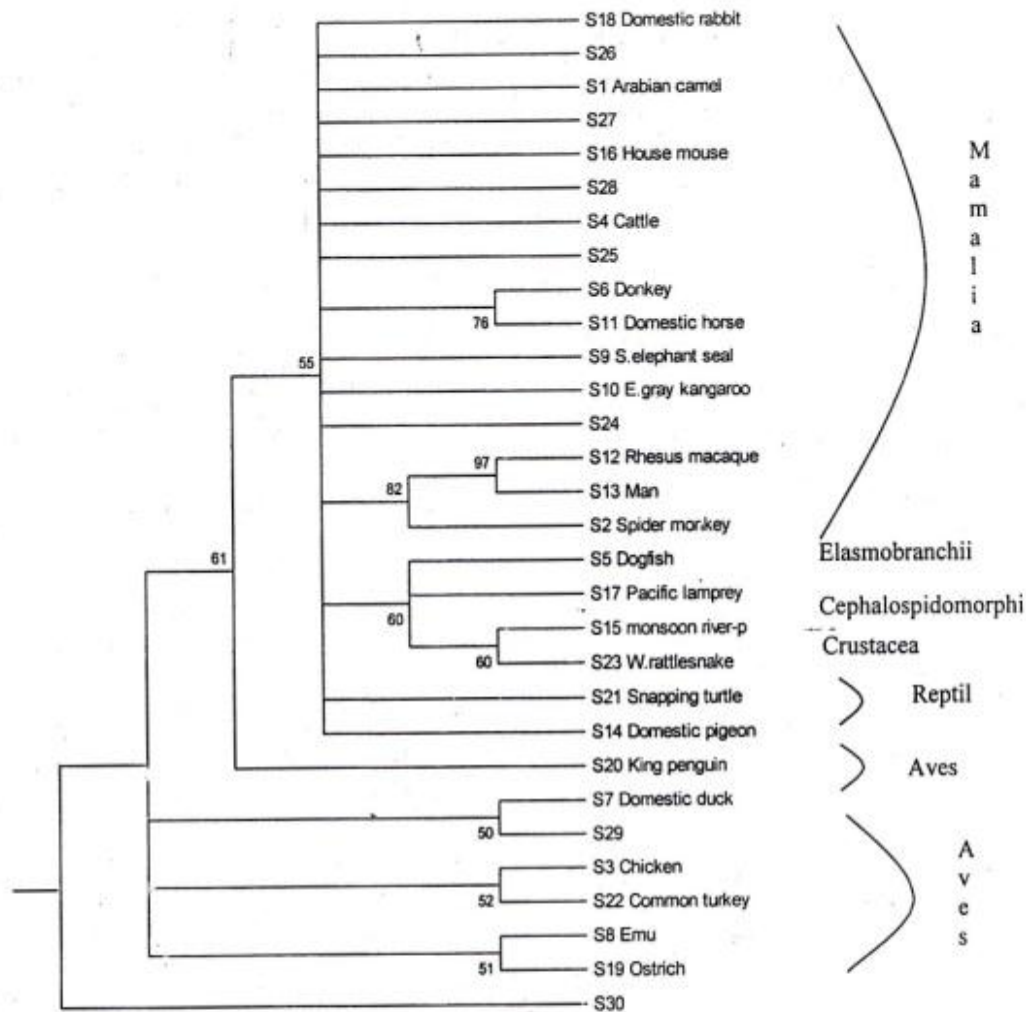
Gambar 4. Minimal spanning tree yang dikonstruksi dari 23 simpul spesies. Angka-angka pada sisi menyatakan bobot sisi.



Keterangan :

- | | | |
|--------------------|-----------------------------|--------------------------------------|
| S1 : Arabian camel | S9 : Southern elephant seal | S17 : Pacific lamprey |
| S2 : Spider monkey | S10 : Eastern gray kangaroo | S18 : Domestic rabbit |
| S3 : Chicken | S11 : Domestic horse | S19 : Ostrich |
| S4 : Cattle | S12 : Rhesus macaque | S20 : King penguin |
| S5 : Dogfish | S13 : Man | S21 : Snapping turtle |
| S6 : Donkey | S14 : Domestic pigeon | S22 : Common turkey |
| S7 : Domestic duck | S15 : Monsoon river-prawn | S23 : Western rattlesnake |
| S8 : Emu | S16 : House mouse | S24,...., S30 : hypothetical spesies |

Gambar 5. Salah satu dari dua filogeni minimal. Angka dengan penyertaan simbol-simbol alpabet dan angka dalam kurung masing-masing menyatakan perbedaan asam amino pada situs tersebut dan bobot sisi.



Gambar 6. Filogeni dari kombinasi metode UPGMA, *bootstrap* dengan 1000 replikasi dan tree *majority-rule consensus*. Angka-angka pada simpul internal adalah angka *bootstrap* dan sebelah kanan pengelompokan *klas* menurut klasifikasi hewan.

KESIMPULAN

Berdasarkan hasil dan pembahasan dapat disimpulkan beberapa hal dalam konstruksi filogeni minimal dengan menggunakan Steiner tree seperti berikut ini.

1. Dalam mengkonstruksi filogeni, spesies-spesies dapat dipandang sebagai simpul-simpul tree dan jarak mutasi yaitu banyaknya perbedaan asam amino pada situs yang bersesuaian antar pasangan spesies sebagai bobot sisi antar kedua simpul.
2. Menentukan filogeni minimal adalah suatu problem Steiner tree dalam filogeni yang dikonstruksi berdasar bobot tree yang

minimal dengan membolehkan menambah simpul-simpul baru melalui suatu minimal spanning tree yang direduksi demikian sehingga minimal spanning tree tersebut menghasilkan reduksi terbesar dengan kata lain menghasilkan tree dengan panjang minimal. Proses reduksi didasarkan pada dua sisi yang bersisian pada suatu simpul yang mempunyai kesamaan disimilaritas asam amino.

3. Hasil penyajian filogeni minimal dalam rooted tree melalui kombinasi metode analisis *bootstrap*, analisis cluster dan tree konsensus menunjukkan bahwa spesies-spesies yang secara klasifikasi biologi

- mempunyai hubungan kekerabatan lebih dekat telah menunjukkan kecenderungan berkelompok.

DAFTAR PUSTAKA

- Baca, D.F. and Lagergren, J., 1998. On the approximability of the Steiner tree problem in phylogeny, *Discrete Applied Mathematics*, **88** : 129 – 145.
- Database Protein, 2004. *Protein Information Resource* (PIR), <http://pir.georgetown.edu>.
- Felsenstein, J. 1978. The number of evolutionary trees, *Systematic Zoology*, **327** : 27 – 33.
- Foulds, L. R., 1984. *Combinatorial optimization for undergraduates*, Springer Verlag, New York
- _____, 1986. Testing the teory of evolution: a novel application of combinatorial optimization, *Discrete Applied Mathematics* **15** : 271-282.
- Gonnet, G.H, (1994). New algorithms for the computation of evolutionary phylogenetic trees. In Suhai, S. (ed), *Computational methods in genome Research*, 153-161.
- Kannan, S., Warnow, T and Yooseph, S., 1998. Computing the local consensus of trees, *SIAM J. Comput.*, **27** : 1695 – 1724.
- Kumar, S., Tamura, K., Jacobsen, I., dan Nei, M., 2004. Paket program *MEGA 2.1* (Molecular Evolutionary Genetics Analysis), <http://www.megasoftware.net>.
- Lu, C.L, Tang, C.Y. and Lee, R.C.T., 2003. The full Steiner tree problem, *Theoretical Computer Science*, **306** : 55 – 67.
- Narwen, 2002. *Suatu Metoda untuk Menentukan Semua Minimum Spanning Tree dalam suatu Graf*, Tesis Program Magister Matematika, Departemen Matematika ITB, Bandung.
- Ridley, M., 1991. *Masalah-masalah evolusi*, penerjemah Achmad Fedyani Saifuddin, UI Press, Jakarta.
- Tucker, A., 1995. *Applied combinatorics*, Ed.3, John Wiley & Sons Inc, New York.