

**Panduan Pengoperasian
Program Numerical Taxonomy System (NTSYS-pc)
Versi 1.8 dan WinBoot untuk Analisis Klaster**

Penyusun
Masdiar Bustamam
Mahrup

Penyunting
Endang M. Septiningsih



**Balai Penelitian Bioteknologi dan Sumberdaya Genetik Pertanian
2003**

ISBN 979-956-27-7-5

Balai Penelitian Bioteknologi dan Sumberdaya Genetik Pertanian

Jl. Tentara Pelajar 3A, Bogor 16111

Tel. (0251) 337975, 339793

Faks. (0251) 338820

E-mail: borif@indo.net.id

KATA PENGANTAR

Untuk mempelajari struktur populasi suatu kelompok individu termasuk analisis kekerabatan (*similarity*) atau jarak genetik (*distance*) dan penyebaran kelompok tertentu dalam populasi, perlu dilakukan analisis kluster. Langkah awal dari klusterisasi kelompok individu perlu ditentukan "parameter" atau dasar pengelompokan yang digunakan, kemudian diikuti oleh analisis statistik dengan menggunakan metode multivariat atau *principal component analysis* (PCA). Parameter untuk pengelompokan dapat berupa sifat morfologi, fisiologi, ketahanan terhadap tekanan abiotik/biotik tertentu, dan lain-lain.

Dengan kemajuan dalam bidang biologi molekuler, terbuka kesempatan untuk klusterisasi kelompok individu berdasarkan pola pita atau profil DNA pada gel-elektroforesis. Profil DNA dapat dihasilkan melalui penggunaan teknologi RFLP atau PCR, tergantung pada tujuan penelitian, dana, dan waktu yang tersedia untuk suatu analisis. Berdasarkan profil DNA, individu tertentu dapat dikelompokkan secara objektif dengan membandingkan "laju migrasi pita-pita DNA" dari dua individu berbeda pada gel-elektroforesis. Berdasarkan skoring pita DNA (band) pada gel-elektroforesis dapat disusun matrik dari kekerabatan atau jarak genetik, dan selanjutnya dilakukan klusterisasi berdasarkan data matrik yang dihasilkan.

Pada populasi yang relatif kecil dan menggunakan parameter analisis yang sederhana, klusterisasi dapat dilakukan secara manual. Akan tetapi pada populasi dengan jumlah individu yang relatif besar atau dengan menggunakan berbagai parameter (beberapa primer PCR, atau beberapa metode analisis DNA), maka pengelompokan atau klusterisasi secara manual akan menjadi sangat kompleks. Untuk itu, klusterisasi dapat dilakukan melalui program komputer yang tersedia.

Dalam tulisan ini akan disampaikan uraian singkat cara pengoperasian analisis kluster dengan menggunakan koefisien DICE untuk menyusun matrik persamaan/jarak genetik dan *Unpair Group with Arithmetic Mean* (UPGMA) untuk analisis kluster yang terdapat dalam program NTSYS dan WinBoot.

Penyusun

DAFTAR ISI

	Halaman
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR GAMBAR	vii
PENDAHULUAN	1
MENYUSUN MATRIK PERSAMAAN/JARAK	2
CARA MENGHITUNG PERSAMAAN/JARAK GENETIK	2
Koefisien Sederhana	3
Koefisien Dice	3
Menghitung Data Matrik secara Manual	4
KLASTERISASI DENGAN UPGMA	5
Klasterisasi secara Manual	5
MENJALANKAN PROGRAM NTSYS-pc	8
Memasukkan Data Biner	8
Memasukkan Data Biner pada MS Excel untuk Input Analisis Matrik pada NTSYS	10
Memasukkan Data Biner pada MS Excel untuk Input Program WinDist	11
Menjalankan Program untuk Analisis Data Matrik Persamaan	13
Melalui Program NTSYS-pc	13
Melalui Program WinDist	14
Menjalankan Program untuk Analisis Klaster	16
Menjalankan Program NTSYS-pc untuk Menampilkan atau Mencetak Phenogram/Dendrogram	16
Menjalankan Program WinBoot untuk Analisis Bootstrap	18
<i>CONSENSUS TREE</i>	23
DAFTAR PUSTAKA	24

DAFTAR GAMBAR

	Halaman
Gambar 1. Jendela NTSYS waktu pertama kali dibuka	8
Gambar 2. NTEDITOR, tempat memasukkan data biner	9
Gambar 3. Memasukkan data biner pada MS Excel untuk data input bagi NTSYS-pc	10
Gambar 4. Perubahan data *.prn dari MS Excel menjadi *.dat pada Notepad	11
Gambar 5. Memasukkan data biner pada MS Excel untuk data input bagi WinDist	12
Gambar 6. Dialog editor dari program MS Excel yang muncul ketika keluar dari program	12
Gambar 7. Jendela program SIMQUAL pada menu QUALITATIVE NTSYS-pc	14
Gambar 8. Penampilan jendela WinDist setelah data dianalisis dengan menekan Compute	15
Gambar 9. Penampilan data matrik dari WinDist yang dibuka di Notepad	15
Gambar 10. Jendela program SAHN Clustering	16
Gambar 11. Tampilan jendela <i>Tree display</i>	18
Gambar 12. Tampilan gambar Phenogram setelah ditekan F3	18
Gambar 13. Tampilan jendela program WinBoot yang sedang melakukan proses penghitungan	20

ISBN 979-956-27-7-5

Balai Penelitian Bioteknologi dan Sumberdaya Genetik Pertanian

Jl. Tentara Pelajar 3A, Bogor 16111

Tel. (0251) 337975, 339793

Faks. (0251) 338820

E-mail: borif@indo.net.id

Panduan Pengoperasian Program Numerical Taxonomy System (NTSYS-pc) versi.1.8. dan WinBoot untuk Analisa Klaster

PENDAHULUAN

Untuk mempelajari struktur populasi suatu kelompok individu termasuk analisis kekerabatan (*similarity*) atau jarak genetik (*distance*) dan penyebaran kelompok tertentu dalam populasi, perlu dilakukan analisis klaster. Langkah awal dari klasterisasi kelompok individu perlu ditentukan "parameter" atau dasar pengelompokan yang digunakan, kemudian diikuti oleh analisis statistik dengan menggunakan metode multivariat atau *principle component analysis (PCA)*. Parameter untuk pengelompokan dapat berupa sifat morfologi, fisiologi, ketahanan terhadap tekanan abiotik/biotik tertentu dan lain-lain.

Dengan kemajuan dibidang biologi molekuler, terbuka kesempatan untuk klasterisasi kelompok individu berdasarkan pola pita atau profil DNA pada gel elektroforesis. Profil DNA dapat dihasilkan melalui penggunaan teknologi RFLP atau PCR, tergantung pada tujuan penelitian, dana, dan waktu yang tersedia untuk suatu analisis (Bustamam dan Moeljopawiro, 1998). Berdasarkan profil DNA, individu tertentu dapat dikelompokkan secara obyektif dengan membandingkan "laju migrasi pita-pita DNA" dari dua individu berbeda pada gel elektroforesis. Berdasarkan skoring pita-pita DNA (band) pada gel elektroforesis dapat disusun matrik dari kekerabatan atau jarak genetik, dan selanjutnya dilakukan klasterisasi berdasarkan data matrik yang dihasilkan (Yap dan Nelson, 1996).

Pada populasi yang relatif kecil dan menggunakan parameter sederhana, klasterisasi dapat dilakukan secara manual. Akan tetapi pada populasi yang relatif besar atau dengan menggunakan berbagai parameter (beberapa primer PCR atau beberapa metode analisis DNA), maka pengelompokan atau klasterisasi secara manual akan menjadi sangat kompleks. Untuk itu, klasterisasi dapat dilakukan melalui program komputer yang tersedia.

Dalam tulisan ini akan disampaikan uraian singkat cara pengoperasian analisis klaster dengan menggunakan koefisien DICE untuk menyusun matrik persamaan/jarak genetik dan "*Unpair Group with Aritmatic Mearl*" (*UPGMA*) untuk analisis klaster yang terdapat dalam program NTSYS (Rohlf, 1993) dan Winboot (Yap dan Nelson, 1996).

Perlu diingat bahwa untuk program NTSYS-pc ada beberapa versi baru yaitu : versi 2.0; perubahan utama adalah Program yang berjalan pada basis DOS dikonversi ke windows. Hal ini menyebabkan adanya perubahan besar pada "user interface" hanya saja bagaimana program itu sendiri dijalankan adalah sama (kalaupun ada perubahan pada modul-modulnya, hanyalah bersifat "minor"). Pada versi 2.1, ada beberapa perubahan pada "user interface", dapat dibaca data entry dari MS.Excel langsung oleh berbagai modul NTSYS-pc, merupakan salah satunya. Terdapat pula beberapa modul baru dan modul-modul yang disempurnakan.

MENYUSUN MATRIK PERSAMAAN/JARAK

Persamaan (similarity) dan jarak genetik (distance) merupakan data "estimasi kuantitatif" yang menerangkan asosiasi atau perbedaan antara sepasang individu (unit). Nilai persamaan biasanya berkisar antara 0 (nol) dan 1 (satu); dengan bertambahnya nilai persamaan berarti bertambah pula nilai kemiripan (likeness value) pasangan individu yang dibandingkan. Sebaliknya, dengan bertambahnya nilai kemiripan (likeness value) maka nilai jarak (genetic distance) akan berkurang.

CARA MENGHITUNG PERSAMAAN/JARAK GENETIK

Sokal dan Sneath (1963) menyusun matrik persamaan/jarak genetik berdasarkan data biner "*ada - tidak ada pita DNA*" pada lokus tertentu. Data ini kemudian dimasukkan kedalam tabel 2x2, dimana nilai satu (1) menunjukkan "*ada*" dan nilai nol (0) berarti "*tidak ada*" pita DNA (band).

		OTU _j	
	a	B	N _{jI}
OTU _j	c	D	N _{j0}
	n _{iI}	n _{i0}	n

Di mana:

OTU = Operational Taxonomic Unit (individu/specimen)

- a = jumlah band (pita DNA) yang "*ada*" pada OTU_i dan OTU_j (positive matches)
- b = jumlah band yang "*ada*" pada OTU_i dan "*tidak ada*" pada OTU_j
- c = jumlah band yang "*ada*" pada OTU_j tapi "*tidak ada*" pada OTU_i

d = jumlah posisi yang "tidak ada" band pada OTU $_i$ dan OTU $_j$ (negative matches)
 n_{i1} = $a + c$ yaitu jumlah posisi yang "ada" band pada OTU $_i$
 n_{i0} = $b + d$ yaitu jumlah posisi "tidak ada" band pada OTU $_i$
 n_{j1} = $a + b$ yaitu jumlah posisi yang "ada" band pada OTU $_j$
 n_{j0} = $c + d$ yaitu jumlah posisi band yang "tidak ada" pada OTU $_j$
 n = $a + b + c + d$ yaitu jumlah pasangan yang "tidak ada" band (negative matches)

ditambah dengan dua variabel lainnya yaitu:

m = $a + d$ yaitu jumlah pasangan yang sama posisi bandnya
 u = $b + c$ yaitu jumlah pasangan yang tidak sama posisi bandnya

Koefisien sederhana

Cara sederhana untuk menghitung matrik persamaan/jarak genetik adalah dengan menggunakan rumus:

$$S_{SM} = \frac{M}{N}$$

Akan tetapi pada kedua koefisien ini "negative matches" ikut terhitung, sehingga data yang dihasilkan kadang-kadang sulit untuk diinterpretasikan Yap dan Nelson (1996). Karena itu mereka menyarankan untuk menghitung matrik persamaan/jarak dengan menggunakan koefisien Dice yang ada pada program NTSYS dan WinBoot.

Koefisien Dice

Koefisien Dice tidak mengikutsertakan nilai-nilai nol (negative matches). Tetapi, "*pasangan dengan band sama*" dihitung dua kali "*pasangan dengan band tidak sama*" sehingga rumusnya menjadi:

$$S_b = \frac{2a}{2a + u} \quad \text{atau} \quad S_b = \frac{2a}{n_i + n_j}$$

dimana a adalah jumlah band yang sama posisinya pada OTU $_i$ dan OTU $_j$, n_i adalah jumlah band pada OTU $_i$ dan n_j adalah jumlah band pada OTU $_j$. Berdasarkan rumus ini dapat pula ditentukan formula untuk menghitung jarak genetik menurut Dice yaitu:

$$D = 1 - S$$

Menghitung Data Matrik Secara Manual

Misalkan kita punya profil DNA sebagaimana terdapat pada Tabel 1 dan 2. Untuk menghitung data matrik persamaan secara manual dilakukan dengan:

1. Menghitung jumlah band pada OTU_i dan OTU_j, masing masing disebut sebagai n_i dan n_j
2. Menghitung jumlah band yang ada pada OTU_i dan OTU_j dan disebut sebagai n_{ij}
3. Menghitung data matrik dengan menggunakan rumus Dice

$$Sb = \frac{2a}{n_i + n_j}$$

4. Bila diperlukan untuk menghitung nilai jarak genetik, rumus ini diganti dengan:
d = 1 - S
5. Ulangi prosedur ini untuk pasangan-pasangan berikutnya.

Tabel 1. Contoh data biner dari profil DNA suatu populasi

No. band (lokus DNA)	OTU (individu)				
	A	b	c	d	e
1.	0	1	1	1	0
2.	1	0	0	1	1
3.	1	1	0	0	1
4.	1	1	1	1	0
5.	0	1	0	1	1
6.	0	1	1	0	1
7.	0	0	1	1	1
8.	1	0	1	1	0
9.	1	0	0	1	0
10.	0	1	1	0	1

Tabel 2. Data matrik persamaan setelah analisis dengan Dice

	a	b	c	d	e
a	1,0				
b	0,36	1,0			
c	0,36	0,67	1,0		
d	0,67	0,31	0,62	1,0	
e	0,36	0,67	0,50	0,46	1,0

KLASTERISASI DENGAN UPGMA

Pada prinsipnya analisa klaster bertujuan untuk klasifikasi yang didasarkan pada persamaan atau jarak genetik. Klaster analisis merupakan penyederhanaan dari data matrik. Didalam program NTSYS ada beberapa cara klasterisasi data matrik, akan tetapi cara yang dilakukan dengan UPGMA adalah yang paling sederhana yaitu dengan membagi dua nilai matrik terkecil seperti tergambar pada ilustrasi berikut ini

Klasterisasi secara Manual

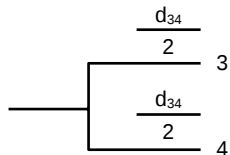
Untuk analisa klaster berdasarkan jarak genetik kemudian dapat dilakukan sebagai berikut:

1. Tentukan nilai jarak genetik sepasang OTU yang terkecil, artinya sepasang OTU yang sangat mirip untuk "similarity matrix" atau sangat berbeda untuk "distance matrix"
2. Kelompokkan kedua individu/OTU itu menjadi satu klaster. Kemudian bagi dua untuk mendapatkan titik cabang jarak antara individu yang dibandingkan
3. Perbaharui data matrik dengan menggabungkan nilai jarak pasangan OTU terkecil menjadi satu klaster yang dibagi dua (lihat contoh berikut)
4. Ulangi prosedur ini untuk nilai jarak terkecil berikutnya

Misalkan kita punya data matrik jarak genetik (d) sebagai berikut:

OTU	1	2	3	4
2	d 12			
3	d 13	d 23		
4	d 14	d 24	d 34	
5	d 15	d 25	d 35	d 45

Jarak antara d_{34} adalah yang terkecil, maka individu 3 dan 4 terkelompok dan bercabang pada titik tengah ($d_{34}/2$)



Maka data matrik berikutnya akan menjadi:

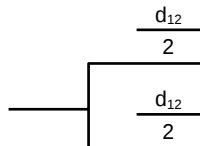
$$d_{1(34)} = \frac{d_{13} + d_{34}}{2}$$

$$d_{2(34)} = \frac{d_{23} + d_{34}}{2}$$

$$d_{5(34)} = \frac{d_{35} + d_{45}}{2}$$

OTU	1	2	(3,4)
2	d_{12}		
(3,4)	$d_{1(34)}$	$d_{2(34)}$	
5	d_{15}	d_{25}	$d_{5(34)}$

Kemudian ditentukan nilai jarak yang terkecil berikutnya dari matrik baru ini, dan ulangi langkah nomor dua untuk nilai terkecil itu. Misalkan kombinasi 1 dan 2 adalah yang terkecil maka 1 dan 2 terkelompok menjadi 1,2:



Dan dibuat lagi data matrik baru sebagai berikut:

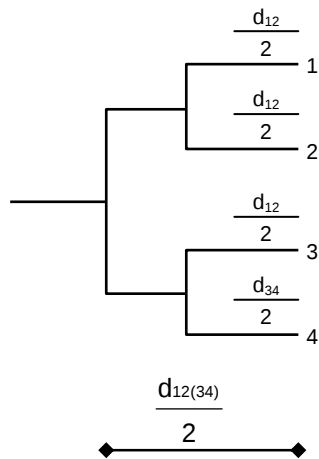
OTU	1,2	(3,4)
(3,4)	$d_{(1,2)(3,4)}$	
5	$d_{15 (1,2)}$	$d_{5(34)}$

Pada data matrik baru ini $d_{5(12)}$ dihitung seperti perlakuan diatas, sedangkan $d_{(12)(34)}$ dihitung dengan mengambil nilai tengah antara $d_{1(34)}$ dan $d_{2(34)}$ sebagai berikut:

$$d_{(12)(34)} = \frac{d_{1(34)} + d_{2(34)}}{2}$$

Waktu membaca matrik baru diperkirakan $d_{(12)(34)}$ adalah yang terkecil. Artinya (1,2) dan (3,4) terkelompok menjadi [(1,2)(3,4)] pada jarak:

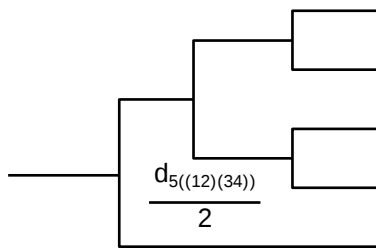
$$\frac{d_{12(34)}}{2}$$



Perhatikan nilai jarak $d_{12(34)}/2$ adalah jarak dari titik percabangan dari kelompok $[(12)(34)]$ ke ujung cabang, bukan ke titik cabang berikutnya. Matrik jarak yang baru dihitung dengan mengambil nilai tengah antara $d_{5(12)}$ dan $d_{5(34)}$, sehingga kita hanya punya satu entri untuk matrik berikutnya.

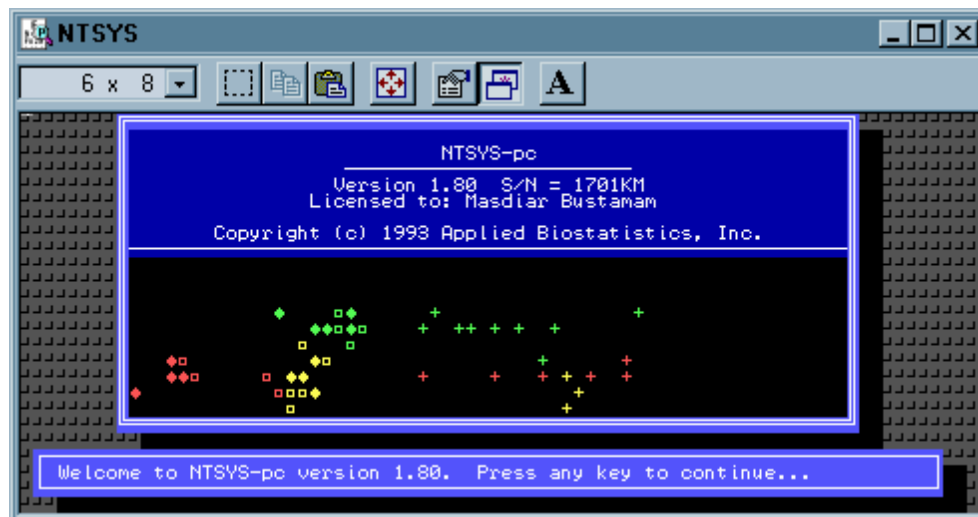
Individu	$[(12)(34)]$
5	$d_{5[(12)(34)]}$

Sehingga jarak $d_{5[(12)(34)]}$ dapat dihitung dengan mengambil nilai tengah antara $d_{5(12)}$ dan $d_{5(34)}$.



MENJALANKAN PROGRAM NTSYSpc

Mulailah menjalankan program NTSYSpc, setelah jendela NTSYS tampak dilayar, tekan sembarang tombol untuk masuk ke menu utama.



Gambar 1. Jendela NTSYS waktu pertama kali dibuka

Memasukkan Data Biner

Ada dua cara memasukkan data biner, yaitu dengan mengetikkan langsung pada program NTSYSpC (NTEDITOR) atau dengan mengtikkannya terlebih dahulu pada program MS Excel.

Memasukkan Data Biner Pada NTEDITOR

Pada NTEDITOR data biner diketikkan dalam bentuk kode ASCII yang dilakukan dengan cara sebagai berikut:

1. Tekan F10 pada menu utama untuk masuk ke menu file.
2. Pilih directory kemudian tekan enter untuk menentukan directory mana yang akan dijadikan tempat untuk menyimpan file lalu tekan enter.
3. Masih di menu file, pilih edit untuk memasukkan data, tentukan nama file (misal coba2.dat) lalu tekan enter.
4. Ketikkan data biner pada jendela NTEDITOR seperti pada gambar 2 di bawah.



Gambar 2. NTEDITOR, tempat memasukkan data biner

Memasukkan Data Biner melalui Program MS Excel

Ada dua tujuan memasukkan data biner pada MS. Excel; yaitu untuk (1) memasukkan data biner dalam bentuk ASCII pada excel untuk kemudian dilanjutkan dengan pengolahan data matrik pada program NTSYS (Rohlf, 1993), ini akan sangat berguna bila data perlu diolah dengan multivariate dan principal component analysis (PCA), dan (2) memasukkan data biner untuk kemudian digunakan sebagai input data dalam pengolahan data matrik pada program WinDist (Yap dan Nelson, 1996). Dengan kata lain, WinDist berguna untuk menterjemahkan data biner dalam file *.txt dari MS Excel menjadi data matrik dalam file *.mat yang akan menjadi data input dalam pengolahan *Shan Clustering* pada NTSYSpc.

Memasukkan Data Biner Pada MS Excel untuk Input Analisa Matrik Pada NTSYS

1. Masukkan data biner pada program MS Excel dengan urutan yang sama seperti cara pengetikan ASCII pada NTEDITOR.
2. Baris pertama, terdiri dari 4 kolom yaitu:
 - Angka 1 pada kolom 1, merupakan parameter untuk analisa matrik
 - Angka nL pada kolom ke 2, menunjukkan *jumlah line* atau *jumlah OTU* yang diuji, misalnya 6L
 - Pada kolom 3, adalah angka yang menunjukkan jumlah seluruh band (lokus DNA) dari skoring data, misalnya 10

- Pada kolom ke 4, diketikkan nilai 0 (nol) adalah untuk data yang hilang (missing data), sebaiknya hindari nilai missing data > 0
3. Pada baris ke 2, dimasukkan nama OTU yang diuji, satu kolom per OTU.
 4. Dikosongkan satu baris, kemudian dibawahnya dimasukkan data biner 0 dan 1 berurutan kebawah dan kesamping seperti termaktub pada point 2 di atas (Gambar 3).

	A	B	C	D	E	F	G	H	I	J	K
1	1	6L	10	0							
2	OTU1	OTU2	OTU3	OTU4	OTU5	OTU6					
3											
4	0	1	1	0	0	1	1	0	1	1	
5	1	1	0	0	1	1	0	1	0	1	
6	1	1	1	1	0	0	1	0	1	1	
7	0	0	1	1	1	0	0	1	1	0	
8	0	0	1	0	1	0	0	1	1	0	
9	1	0	1	1	0	0	0	1	0	0	
10											
11											
12											

Gambar 3. Memasukkan data biner pada MS Excel untuk data input bagi NTSYS-pc

5. Agar data dapat dibuka dan diolah di NTSYSpc, Kemudian data ini disimpan dengan ekstensi *.prn. (space delimited). Buka data ini pada program **Notepad** guna merubah format *.prn menjadi *.dat dengan jarak satu spasi untuk masing-masing nilai (Gambar 4).

```

1 6L 10 0
OTU1 OTU2 OTU3 OTU4 OTU5 OTU6

0 1 1 0 0 1 1 0 1 1
1 1 0 0 1 1 0 1 0 1
1 1 1 1 0 0 1 0 1 1
0 0 1 1 1 0 0 1 1 0
0 0 1 0 1 0 0 1 1 0
1 0 1 1 0 0 0 1 0 1

```

Gambar 4. Perubahan data *.prn dari MS Excel menjadi *.dat pada Notepad

Memasukkan Data Biner Pada MS Excel untuk Input Program WinDist

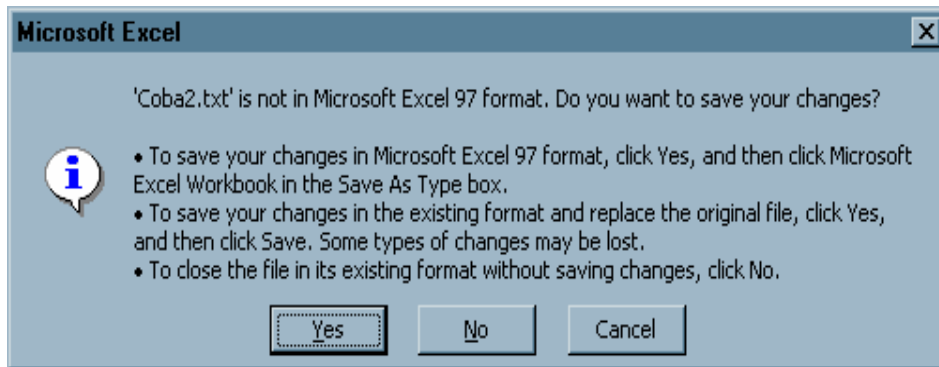
Berbeda dengan pemasukan data untuk NTSYSpc, input data untuk WinDist dimasukkan dengan cara sebagai berikut:

1. Pada jendela MS Excel masukkan data biner (skoring band atau lokus DNA)
2. Pada cell pertama baris pertama diketikkan "jumlah OTU atau Individu" yang diuji, dan cell ke dua baris pertama adalah "jumlah band/lokus DNA seluruhnya".
3. Pada kolom pertama dari baris berikutnya dimasukkan "nama OTU/Individu yang diuji". Perlu diingat bahwa nama OTU tidak boleh lebih dari delapan karakter, atau berupa angka berurutan (1, 2, 3 dst.).
4. Pada kolom berikutnya kesamping dan kebawah dimasukkan data "skoring band/lokusDNA" yang akan diuji. Setelah itu, data disimpan dalam format file *.txt (Tab delimited) seperti terlihat pada gambar 5.

	A	B	C	D	E	F	G	H	I
1	10	6							
2	otu-1	0	1	1	0	0	1		
3	otu-2	1	1	1	0	0	0		
4	otu-3	1	0	1	1	1	1		
5	otu-4	0	0	1	1	0	1		
6	otu-5	0	1	0	1	1	0		
7	otu-6	1	1	0	0	0	0		
8	otu-7	1	0	1	0	0	0		
9	otu-8	0	1	0	1	1	1		
10	otu-9	1	0	1	1	1	0		
11	otu-10	0	1	1	0	0	1		
12									
13									
14				band / lokus DNA					
15									

Gambar 5. Memasukkan data biner pada MS Excel untuk data input bagi WinDist

5. Keluarlah dari program Excel, setelah ditentukan nama file dan direktory dimana data itu disimpan. Bisanya, waktu akan keluar ada peringatan dari excel seperti terlihat pada gambar berikut, pilihlah tombol Yes jika ingin menyimpan file dalam format *.xls (tidak selalu direkomendasikan). Bila ini pilihannya berarti ada dua file serupa dalam format berbeda, hal ini kadang-kadang memudahkan, bila ada kekeliruan dalam memasukkan data atau data perlu disimpan dalam bentuk format lain untuk keperluan analisa lainnya (seperti analisa PCA).



Gambar 6. Dialog editor dari program MS Excel yang muncul ketika keluar dari program

Menjalankan Program untuk Analisa Data Matrik Persamaan

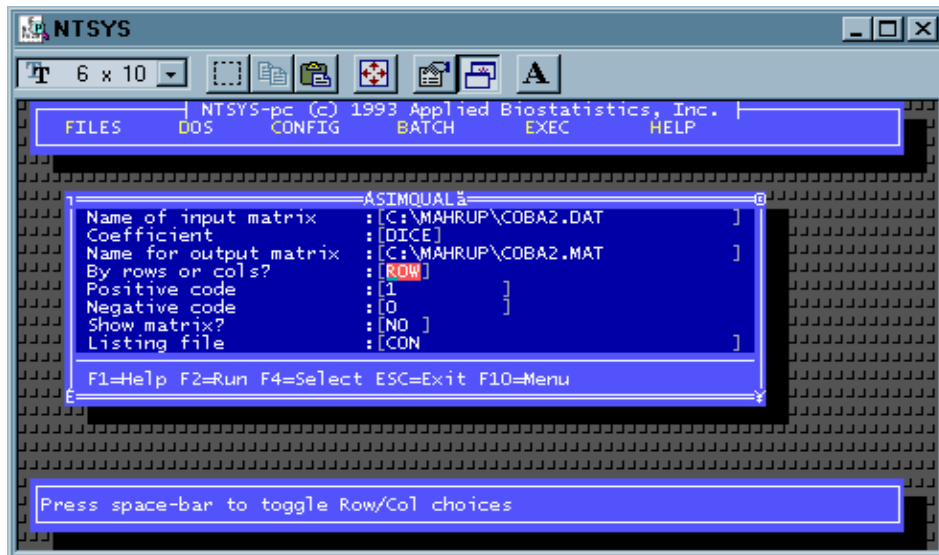
Sebagaimana dikemukakan pada pendahuluan bahwa data matrik dapat disusun langsung pada program NTSYSpc atau melalui program pembantu "WinDist". Pada WinDist, data matrik dapat dibuka melalui program notepad dan dicetak atau dikopikan ke Word sesuai keperluan untuk pembuatan laopran penelitian.

Melalui Program NTSYSPC

Pada program NTSYS data matrik diolah menggunakan koefisien DICE melalui program SIMQUAL.

Menjalankan Program SIMQUAL

1. Pada menu utama pilih QUALITATIVE untuk mengaktifkan menu SIMQUAL
2. Ketikkan nama file (drive\directory\file.*dat) pada baris *Name of input file*, atau tekan F4 untuk mencari file yang dimaksud.
3. Pilih DICE pada baris Coeficient
4. Pada baris *Name for output matrix*, ketikkan nama file (drive\directory\file.*mat). File dengan ekstensi *.mat ini merupakan input data bagi program Sahn Clustering.
5. Baris *By rows or cols?* pilihlah row, dengan menekan space bar.
6. Baris-baris berikutnya tidak perlu dirubah, selanjutnya tekanlah F2 untuk menjalankan program ini.
7. Tekanlah Esc. untuk kembali ke menu utama.

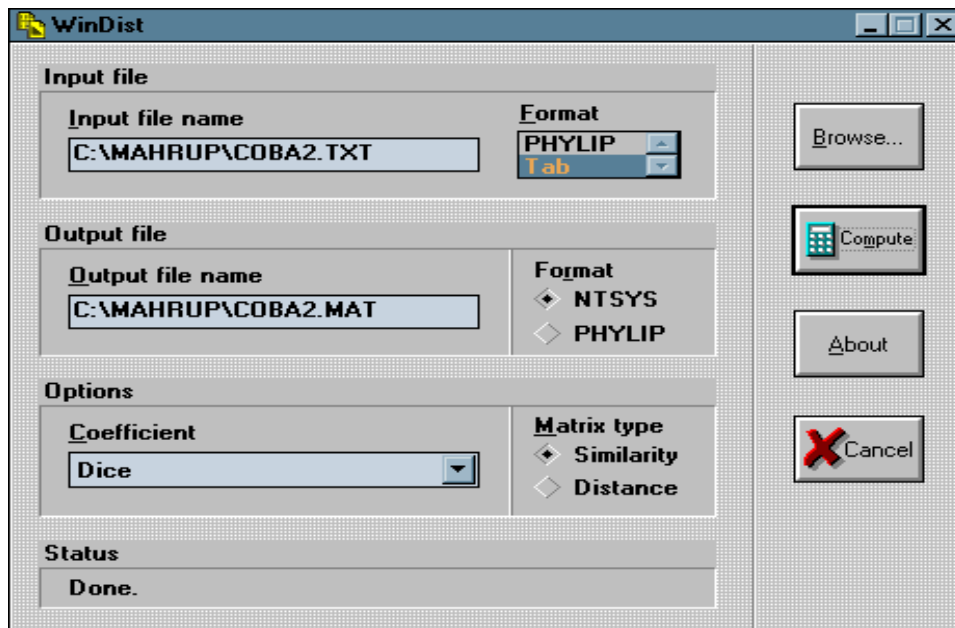


Gambar 7. Jendela program SIMQUAL pada menu QUALITATIVE NTSYS-pc

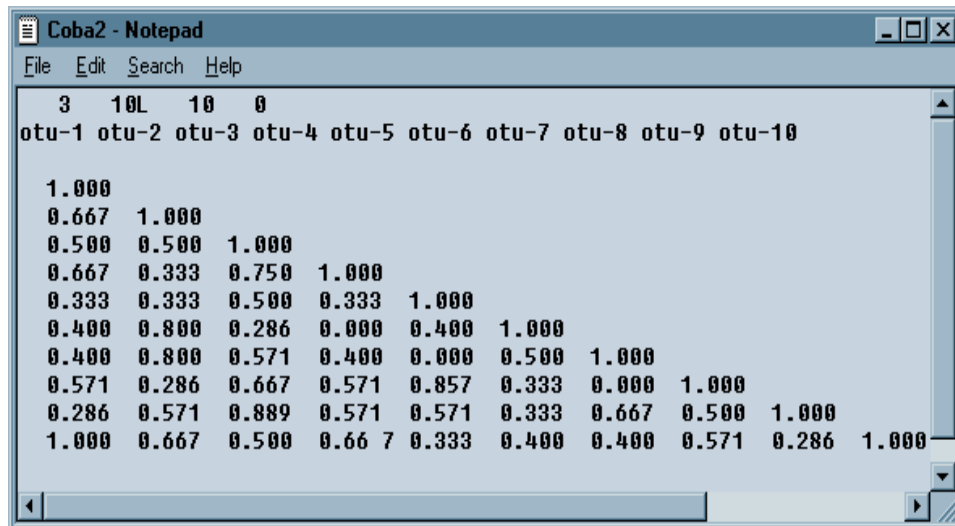
Melalui Program WinDist

WinDist adalah program pendukung untuk menganalisa data matrik dengan menggunakan koefisien dice. Data *.mat yang dihasilkan dapat digunakan untuk analisa kluster pada program NTSYSpc. Dengan kata lain, analisa matrik pada WinDist sama dengan menjalankan program SIMQUAL pada NTSYSpc.

1. Pertama- tama, buka program WinDist, dan ketikan input file yaitu file dengan *.txt yang dibuat di Excel. Browse file *.txt dan isikan pada file input, setelah file input terisi maka secara otomatis file output dalam format *.mat akan terisi pula
2. Pilihlah Tab bila data Excel dalam format *.txt (tab delimited) atau Phylip bila data Excel dalam format *.prn (space delimited). Pilih tipe matrik yang diinginkan "similarity/distance"
3. Kemudian tekan tombol Compute untuk prosesing analisa data. Setelah analisa selesai status ready akan berubah menjadi done (lihat gambar berikut).
4. Kemudian keluarlah dari jendela Windist dengan menekan cancel atau tanda silang pada sudut kanan atas jendela WinDist.
5. Bila data matrik ingin ditampilkan dalam pembahasan publikasi, data dapat dibuka melalui Notepad (Gambar 9), atau dikopi dari Noteped ke program MS.Word.



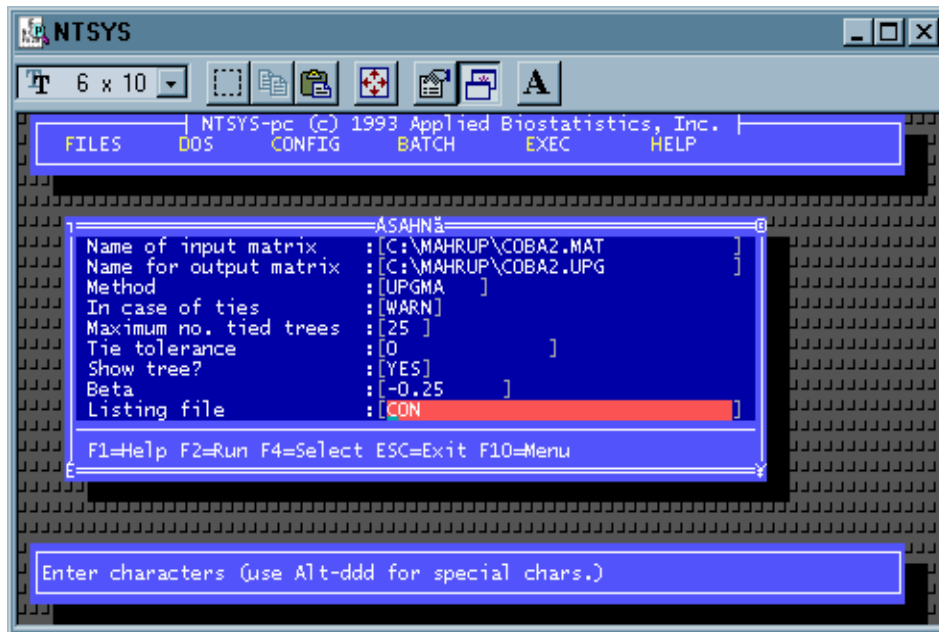
Gambar 8. Penampilan jendela WinDist setelah data dianalisis dengan menekan Compute



Gambar 9. Penampilan data matrik dari WinDist yang dibuka di Notepad

Menjalankan Program untuk Analisa Klaster

1. Pilih SAHN Clustering pada menu utama, kemudian jendela Sahn akan tampak di layar monitor (Gambar 10).
2. Pada baris *Name of input matrix* isilah dengan file (drive\directory\file.*mat) yang telah dibuat sebelumnya, baik dengan program SIMQUAL maupun dengan WinDist.
3. kemudian pada baris *Name for output matrix* ketiklah nama file (drive\directory\file.*upg), file ini nantinya akan digunakan dalam proses penampilan Phenogram.
4. Adapun baris-baris selanjutnya tidak perlu dirubah, kemudian tekanlah F2 untuk menjalankan proses klasterisasi.
5. Tekan Esc. untuk kembali ke menu utama.

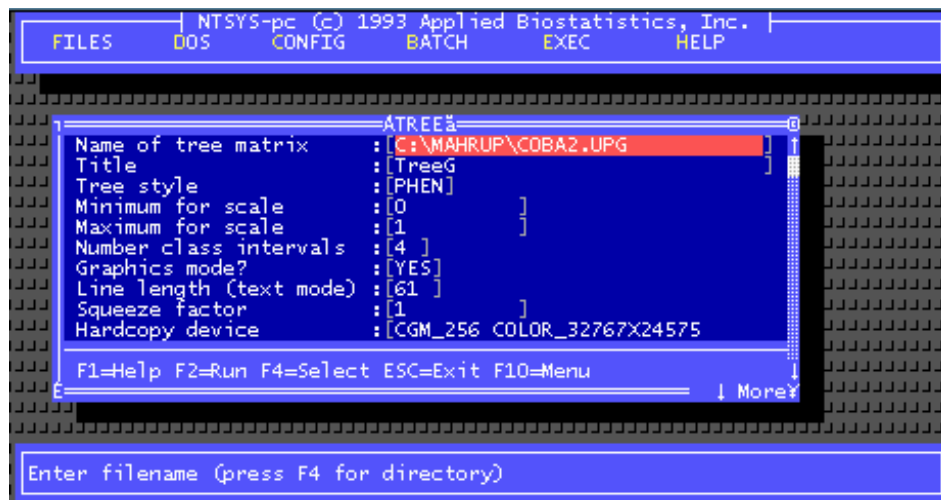


Gambar 10. Jendela program SAHN Clustering

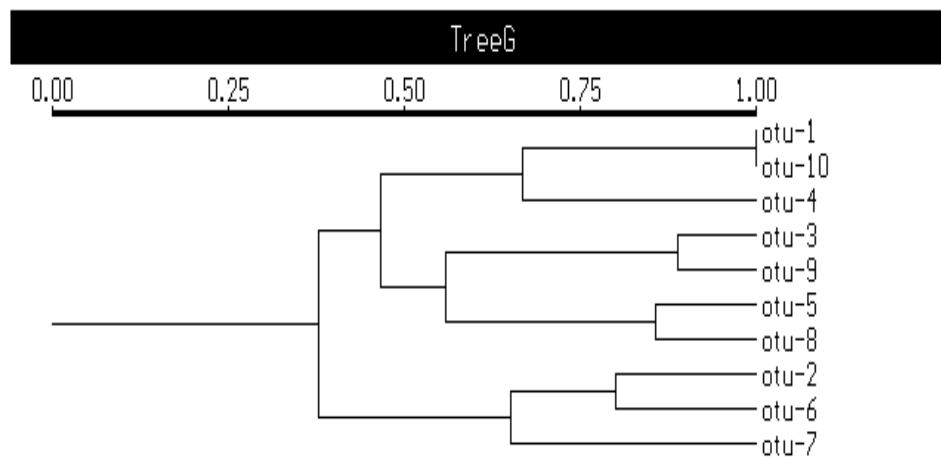
Menjalankan Program NTSYSpc untuk Menampilkan atau Mencetak Phenogram / Dendrogram

1. Pada menu utama pilih *Tree display*, kemudian jendela **Tree** akan tampak seperti pada Gambar 11. di bawah.
2. Kemudian pada baris *Name of tree matrix* isilah dengan file (drive\directory\file.*upg) yang telah dibuat pada tahanan Sahn Clustering.
3. Baris *Title* sifatnya opsional, artinya boleh dibiarkan tidak diganti.
4. Kemudian pilih PHEN (=Phenogram) pada baris *Tree Style*.
5. Ketiklah angka nol (0) pada baris *Minimum for scale*, angka satu (1) pada baris *Maximum for scale*.
6. Pada baris *Number class intervals* biasanya diisi dengan angka empat (4), tetapi angka ini bisa saja diisi dengan angka lebih kecil atau lebih besar.
7. Pastikan **Yes** pada baris *Graphic mode?*, jika tidak gantilah dengan jalan menekan space bar sehingga berubah menjadi Yes.
8. Dua baris berikutnya tidak perlu diadakan perubahan.
9. Kemudian pada baris *Hardcopy device*, pilihlah printer sesuai dengan nama printer yang terpasang pada komputer (bila akan dicetak), tetapi jika ingin mengkopinya ke file sehingga nantinya dapat disunting pada program lain seperti Ms.Word atau Power Point, isilah baris ini dengan CGM (=Computer Graphic Metafile) atau sejenisnya sesuai dengan software yang ada pada komputer. Dengan begitu file ini nantinya akan merupakan file gambar, sehingga jika ingin menyuntingnya harus di-Insert-kan sebagai file gambar (picture).
10. Baris port or file diisi dengan alamat printer pada komputer (biasanya lpt1) jika baris di atasnya dipilih printer. Tetapi jika yang dipilih CGM, maka baris ini harus diisi dengan nama file (drive\directory\file.*cgm) supaya dapat tersimpan menjadi file gambar.
11. Biarkan baris Listing file seperti apa adanya (Con).
12. Kemudian tekan F2 untuk menjalankan program dan F3 untuk menampilkan gambar. Pada sebagian komputer kadang-kadang F3 tidak perlu lagi ditekan, karena gambarnya langsung keluar setelah ditekan F2.
13. Tekan + atau – pada keyboard untuk membesarkan atau mengecilkan gambar, mengecilkan gambar yang tidak tampak seluruhnya di layar monitor merupakan sesuatu yang sangat penting, karena jika ingin mengkopinya ke file maka hanya gambar yang tampak di layar sajalah yang akan terkopi. Tekan Alt-P pada keyboard atau gunakan mouse untuk menekan Alt-P=Print pada bagian bawah layar.

14. Tekan Esc. untuk keluar dari tampilan gambar, kemudian tekan kembali Esc. sampai keluar pertanyaan peringatan untuk keluar atau tidak dari program. Tekan Y untuk keluar.



Gambar 11. Tampilan jendela *Tree display*



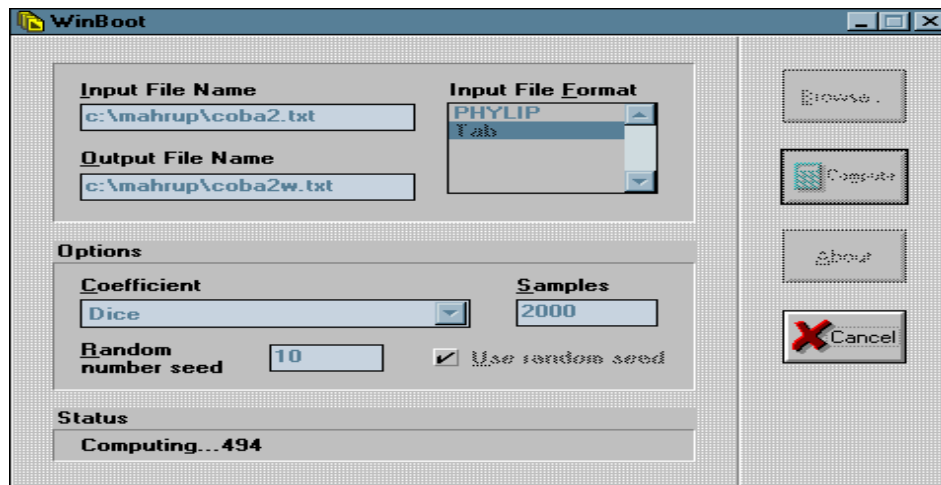
Gambar 12. Tampilan gambar Phenogram setelah ditekan F3

Menjalankan Program WinBoot untuk Analisa Bootstap

Analisa Sahn Clustering pada NTSYSpc kadang-kadang menunjukkan bahwa dari data yang diolah mungkin dihasilkan lebih dari satu bentuk phenogram, akan tetapi NTSYS tidak dapat menunjukkan "Cofident limit" dari klasterisasi yang dilakukan, sementara untuk akurasi analisa statistik nilai "Cofident limit" sangatlah diperlukan. Hal ini tentu dapat dimaklumi agar anda dapat mengambil keputusan apakah phenogram yang dihasilkan sudah betul-betul robust atau hanya terjadi secara kebetulan saja. Bila nilai bootstrap lebih kecil dari 90 berarti phenogram yang dihasilkan belum dapat dijadikan acuan, dan parameter untuk klasterisasi perlu diperluas.

1. Mulailah dengan membuka program WinBoot untuk menjalankannya.
2. Dalam menjalankan program ini sebenarnya tidak jauh berbeda dengan menjalankan program Windist. karena itu file inputnyapun sama.
3. Isilah kolom *Input File Name* dengan nama file yang akan dianalisa, file ini merupakan file yang berekstensi *.txt yaitu file yang dibuat pertama kali (data biner) di Ms.Excel untuk input file WinDist.
4. Bila menggunakan Browse untuk mengisi kolom input file, maka kolom *Output file Name* akan langsung terisi setelah kolom ini terisi. Tetapi harus diingat bahwa ekstensi file output dari program ini sama-sama *.txt, karena itu bedakanlah file outputnya untuk menghindari tertimpnyaa file asal.
5. Pilih Tab pada kolom *Input File Format*, jika filenya berformat text (Tab delimited *.txt).
6. Dibagian *Options*, pada kolom *coefiscent* pilih **Dice**.
7. Pada kolom Samples bisa diisi dengan angka 100, 200, 400 atau 2000 sebagai angka maksimal. Angka ini menunjukkan banyaknya jumlah pengulangan dalam penghitungan analisa klaster atau phenogram yang harus dilakukan oleh WinBoot sesuai dengan apa yang dikemukakan oleh Hedges (dalam Yap dan Nelson, 1996). Untuk mendapatkan batas kepercayaan atau Cofident limit $\pm 1\%$ dapat dipilih angka 400 untuk perkiraan bootstrap 99% atau 2000 untuk mendapatkan nilai bootstrap 95%. Nilai 100 adalah nilai "default program", dan ini dapat digunakan untuk mendapat gambaran nilai Cofident limit analisa klaster secara cepat. Akan tetapi untuk mendapatkan Cofident limit yang lebih tinggi biasanya dipilih nilai 2000.
8. Klik kotak sebelah kiri dari *Use random seed* bila ingin melakukan random sampling pada analisa cluster yang dilakukan WinBoot. Dengan begitu *Random number seed* akan aktif, sehingga anda bisa mengisi kotak disebelahnya dengan angka.

9. Tekan *Compute* setelah semua pilihan ditentukan, komputer akan mulai menjalankan proses penghitungan sampai selesai. Proses ini akan terlihat pada kolom *Status* dibagian bawah jendela WinBoot yang kemudian akan berubah menjadi *done* setelah selesai.
10. Seperti halnya pada WinDist, file dari hasil WinBoot ini dapat disunting dengan program Notepad yang kemudian dapat dikopikan ke program lain seperti MS. Word.



Gambar 13. Tampilan jendela program Winboot yang sedang melakukan proses penghitungan

Berikut ini adalah contoh dari hasil analisa Bootstrap dengan pengulangan 400 kali.

Species in order:

otu-1
otu-2
otu-3
otu-4
otu-5
otu-6
otu-7
otu-8
otu-9
otu-10

Sets included in the consensus tree:

Set (species in order) How many times out of 400

.....	318
...*...*	271
..*.....*	225
.*.***...	178
.......*	129
.*.***...	127
..*...*	102
*.***.***	70

Sets NOT included in the consensus tree:

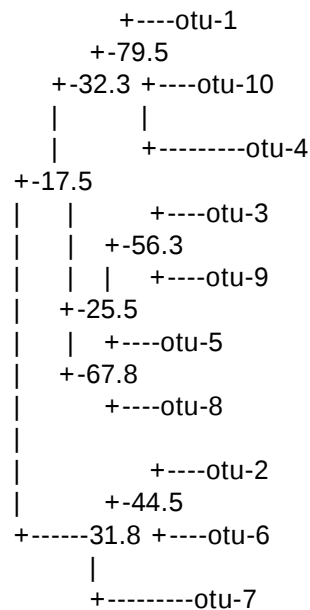
Set (species in order) How many times out of 400

.*.***...	109
**.....*	96
..**.....	90
..**.....*	85
..***.***	85
**...*...*	78
****.***	71
.***.***	68
...***...	57
*...***...	57
...*...	52
..***.***	51
..***.***	46
...*.***	41
*****.***	41
.....*	40
...	39
...	36
**...*	36
.***.***	34
*...***	34
.***.***	33
**.....	33
...	32
..*...*	31
**...*	30

....*	29
.**.*.*	26
..*.*	24
*****.*	23
.***	21
..**.*..	18
.*.***...	17
**.*.*.*	17
...*.*...	16
*****.*.*	16
...*.*..	15
*****.******	15
*****.*.*	15
..*.*.*	13
....**....	12
.*****.*.	11
..*.*..	10
...**.*..	9
.....	9
.*.***.*.	8
..***.*.*.	5
.**.******.	5
.***.*.*.	5
**.*.*....	5
.*.***.*..	4
..*.*.*	4
**.*.*..	4
**.*.*.*.*	4
....*.*.*.	3
.**.*.*.*.	3
.*.***....	2
***.*.*.*	2
*****.*	2
......	1
..*....	1
**.*.*.*.*	1

CONSENSUS TREE:

The numbers at the forks show the percentage of times the group consisting of the species which are to the right of that fork occurred



WinBoot computational run time: 0:00:03.671

DAFTAR PUSTAKA

- Bustamam, M. dan S. Moeljopawiro. 1998.** Pemamfaatan teknologi sidikjari DNA di bidang pertanian. Zuriat Vol. 9. No. 2. hlm. 77-90.
- Rohlf, F.J. 1993.** NTSYSpc: Numerical taxonomic and multivariate analysis system. Version 1.80 Setauket, Exeter software. Exeter Publishing Co. Ltd.
- Sokal, R.R. and P.H.A. Sneath. 1963.** Principles of numerical taxonomy. Freeman. San Fransisco. 393 p.
- Yap, I.V. and R.J. Nelson. 1996.** WinBoot: A program for performing bootstrap analysis of binary data to determine the Confident limits of UPGMA-based dendrogram. IRRI Discussion Paper Series No. 14, 20 pp. IRRI, Manila, Philippines.