

Phân tích Line x Tester và GGebipplot

(Line tester analysis and ggebipplot)

Nguyễn Đình Hiền

Bộ môn Công nghệ phần mềm

A- Phân tích Line x Tester

Lai đỉnh toàn phần là ph- ơng pháp lai thử chủ yếu để xác định khả năng kết hợp chung (GCA) do Devis đề xuất năm 1927, Jenkins và Bruce phát triển năm 1932.

Các dòng hoặc giống, cần xác định khả năng kết hợp, đ- ọc lai với một số cây thử (Tester). Ph- ơng pháp này rất có ý nghĩa ở giai đoạn đầu của quá trình chọn lọc, khi số l- ợng dòng còn quá lớn. Yếu tố thành công của lai đỉnh là chọn đúng cây thử.

Chọn xong dòng và cây thử, các cặp lai đ- ọc so sánh bằng ph- ơng pháp bố trí thí nghiệm kiểu khối ngẫu nhiên đầy đủ (Randomized Block Design).

Nếu chỉ tập trung vào cây lai thì có phân tích Topcross, nếu so sánh thêm kết quả của bố mẹ (các dòng và cây thử) thì có phân tích Line x Tester. Tuy ch- ơng trình phân tích LineTester phức tạp hơn Topcross nh- ng có thể viết chung vào một ch- ơng trình, ng- ời sử dụng tùy tình hình mà lựa chọn.

Các b- ớc phân tích gồm:

I/ Phân tích ph- ơng sai

Việc phân tích ph- ơng sai chia bao gồm các phần sau:

1- Phân tích ph- ơng sai 1

Phần này nhằm tách biến động chung ra thành 3 nguồn:

- Biến động do khối (Replication).
- Biến động do công thức (Treatment).
- Biến động do sai số ngẫu nhiên (Error).

Phần phân tích này có thể làm riêng, nhờ ch- ơng trình phân tích ph- ơng sai một nhân tố, bố trí kiểu khối ngẫu nhiên. (Dùng ch- ơng trình viết riêng cho bố trí thí nghiệm, hoặc các bộ ch- ơng trình thống kê nh- Irristat, Statgraf, Mstatc ...).

2- Phân tích ph- ơng sai 2

Phần này dùng khi phân tích Line*Tester, nhằm tách biến động do công thức thành các nguồn biến động sau:

- Biến động do bố mẹ (Parents).
- Biến động do các cặp lai (Crosses).
- Biến động còn lại (Bố mẹ đối lại cặp lai - Parents vs Crosses)

Nếu chỉ phân tích Topcross thì không cần phần phân tích ph- ơng sai này.

3- Phân tích ph- ơng sai 3

Dùng ph- ơng pháp phân tích ph- ơng sai 2 nhân tố (Coi dòng là nhân tố A, cây thử là nhân tố B), có t- ơng tác, để phân tích các giá trị trung bình của các cặp lai qua các khối, có thể tách biến động chung thành 3 nguồn :

- Biến động của dòng (Lines).
- Biến động của cây thử (Testers).
- T- ơng tác (Lines * Testers).

Sai số dùng ở đây là sai số thu đ- ọc ở phần phân tích ph- ơng sai 1

III/ Ước khả năng kết hợp chung và khả năng kết hợp riêng

Phần này sử dụng các kết quả của ph- ơng pháp phân tích thí nghiệm 2 nhân tố:

- Khả năng kết hợp chung của dòng i là hiệu quả của mức i của nhân tố A.
- Khả năng kết hợp chung của cây thử j là hiệu quả của mức j của nhân tố B.
- Khả năng kết hợp riêng giữa dòng i và cây thử j là ảnh h- ưởng của t- ơng tác

A_iB_j

Phần cuối cùng là tìm các thành phần ph- ơng sai di truyền, dựa trên các công thức về các thành phần di truyền trong các gia đình Half-sib (HS) và Full-sib(FS).

III/ Thí dụ cụ thể (tài liệu (4))

Có D = 5 dòng (gọi là 1,2...5) và C = 3 cây thử (gọi là 6..8).

Các cặp lai (gọi là 1 × 6, 2 × 6, . . ., 5 × 8).

Tất cả có $M = D \times C + D + C = 5 \times 3 + 5 + 3 = 23$ công thức.

Thí nghiệm lập lại K = 4 lần (R1,R2,R3,R4)

Số liệu của các cặp lai và của bố (Dòng - Lines), mẹ (Cây thử - Tester) đ- ọc ghi lại trong bảng

Công thức	R1	R2	R3	R4	Tổng $X_{1.}$
1 × 6	74.40	70.86	60.94	68.00	274.20
1 × 7	91.82	99.18	118.88	120.68	430.56
1 × 8	48.08	62.10	58.54	41.84	210.56
2 × 6	59.06	65.62	81.62	86.76	293.06
2 × 7	84.16	109.74	102.14	94.52	390.56
2 × 8	96.92	91.44	79.86	74.38	342.60
3 × 6	109.86	98.16	93.26	102.26	403.54
3 × 7	117.20	100.28	116.16	112.52	446.16
3 × 8	109.68	116.48	123.92	120.86	470.94
4 × 6	103.14	109.66	90.98	119.40	423.18
4 × 7	53.40	60.86	74.46	69.08	257.80
4 × 8	53.86	48.30	40.64	44.62	187.42
5 × 6	98.46	73.10	89.18	75.86	336.60
5 × 7	81.36	72.82	89.82	83.74	327.74
5 × 8	86.62	94.18	90.32	108.16	379.28
1	104.86	84.32	76.92	76.48	342.58
2	88.02	106.54	89.82	108.68	393.04
3	77.91	71.34	77.52	69.48	296.28
4	80.82	106.52	83.28	95.92	366.56
5	59.96	52.48	52.98	50.98	216.40
6	96.44	98.82	99.14	107.16	401.56
7	91.44	99.66	83.28	89.46	363.84
8	91.78	84.82	69.92	81.48	328.00

Phân tích ph- ơng sai 1

Nguồn biến động	Btd	Tổng BP	Trung bình	F
Khối(Replication)	3	83.000	27.667	0.304

	Ik	SSk	Smk	Smk/Sme
Công thức (Treatements)	22 Im	32552.940 SSm	1479.679 Smm	16.249 Smm/Sme
Sai số (Error)	66 Ie	6010.295 SSe	91.065 Sme	
Toàn bộ (Total)	91 Itb	38646.235		

Phân tích ph- ong sai 2

Nguồn biến động	Btd	Tổng BP	Trung bình	F
Công thức (Treatements)	22 Im	32552.940 Ssm		
Bố mẹ (Parents)	7 Ibm	6299.620 SSbm	899.946 Smbm	9.882 Smbm/Sme
Cặp lai (Crosses)	14 Icl	26199.654 SScl	1871.404 Smcl	20.500 Smcl/Sme
Bố mẹ vs cặp lai (Parents vs Crosses)	1 Ir	53.666 SSr	53.666 Smr	0.589 Smr/Sme
Sai số (Error)	66 Ie	6010.295 Sse	91.065 Sme	

Phân tích ph- ong sai 3

Nguồn biến động	Btd	Tổng BP	Trung bình	F
Dòng (Lines)	4 Id	10018.361 SSd	2579.590 Smd	1.457 Smd/Smdc
Cây thử (TesterS)	2 Ic	1718.925 SSc	859.493 Smc	0.485 Smc/Smdc
Dòng× Cây thử (Lines ×Testers)	8 Idc	14162.368 SSdc	1770.296 Smdc	19.440 Smdc/Sme
Sai số (Error)	66 Ie	6010.295 SSe	91.065 Sme	

Bảng các giá trị trung bình

	Cây1	Cây2	Cây 3
Dòng 1	68.550	107.640	52.640
Dòng 2	73.265	97.640	85.650
Dòng 3	100.885	111.540	117.735
Dòng 4	105.795	64.450	46.855
Dòng 5	84.150	81.935	94.820

4 - Ước l- ợng khả năng kết hợp

Khả năng kết hợp chung

Gọi Y_{ij} là số liệu trong bảng 2 (Tổng các X_{ijk} trên tất cả các khối k)

Tổng theo dòng $Y_{i.}$, tổng theo cây $Y_{.j}$, Tổng toàn bộ $Y_{..}$

Ta có - ớc l- ợng khả năng kết hợp chung (KNKHC) của dòng

$$GD_i = Y_{i.}/(C \times K) - Y_{..}/(D \times C \times K)$$

$$\begin{aligned}
GD_1 &= 915.32 / (3 \times 4) - 5174.20 / (5 \times 3 \times 4) = -9.960 \\
GD_2 &= 1026.22 / (3 \times 4) - 5174.20 / (5 \times 3 \times 4) = -0.718 \\
GD_3 &= 1320.64 / (3 \times 4) - 5174.20 / (5 \times 3 \times 4) = 23.817 \\
GD_4 &= 868.40 / (3 \times 4) - 5174.20 / (5 \times 3 \times 4) = -13.870 \\
GD_5 &= 1043.62 / (3 \times 4) - 5174.20 / (5 \times 3 \times 4) = 0.732
\end{aligned}$$

Có thể kiểm tra để thấy $\sum GD_i = 0$ ($i = 1, \dots, 5$)

Đối với cây thử

$$\begin{aligned}
GC_j &= Y_j / (D \times K) - Y_{.j} / (D \times C \times K) \\
GC_6 &= 1730.58 / (5 \times 4) - 5174.20 / (5 \times 3 \times 4) = 0.292 \\
GC_7 &= 1852.82 / (5 \times 4) - 5174.20 / (5 \times 3 \times 4) = 6.404 \\
GC_8 &= 1590.80 / (5 \times 4) - 5174.20 / (5 \times 3 \times 4) = -6.697
\end{aligned}$$

Có thể kiểm tra để thấy $\sum GC_j = 0$ ($j = 6, \dots, 8$)

Khả năng kết hợp riêng

Khả năng kết hợp riêng (KNKHR) của dòng i với cây j S_{ij} tính theo công thức:

$$S_{ij} = Y_{ij} / K - Y_i / (C \times K) - Y_{.j} / (D \times K) + Y_{.j} / (D \times C \times K)$$

$$\begin{aligned}
\text{Thí dụ } S_{16} &= 274.20 / 4 - 915.32 / (3 \times 4) - 1730.58 / (5 \times 4) + 5174.20 / (5 \times 3 \times 4) \\
&= -8.019
\end{aligned}$$

T-ong tự có các khả năng kết hợp riêng khác : $S_{17}, S_{18}, \dots, S_{58}$

Bảng các khả năng kết hợp riêng

	Cay 1	Cay 2	Cay3	Bien dong
dong 1	-8.019	24.959	-16.940	475.730
dong 2	-12.546	5.717	6.828	106.974
dong 3	-9.461	-4.918	14.378	148.832
dong 4	33.136	-14.321	-18.815	817.165
dong 5	-3.111	-11.438	14.548	164.695

Sai số của các - ớc l- ợng

Độ lệch chuẩn của các - ớc l- ợng của khả năng kết hợp chung, khả năng kết hợp riêng, của hiệu số các - ớc l- ợng đ- ợc cho bởi các công thức sau:

$$\begin{aligned}
SE(G_i \text{ của dòng}) &= (Sme / (C \times K))^{1/2} = (91.065 / 12)^{1/2} = 2.755 \\
SE(G_j \text{ của cây}) &= (Sme / (D \times K))^{1/2} = (91.065 / 20)^{1/2} = 2.134 \\
SE(S_{ij} \text{ của dòng} \times \text{cây}) &= (Sme / K)^{1/2} = (91.065 / 4)^{1/2} = 4.771 \\
SE(G_i - G_j \text{ của dòng}) &= (2 \times Sme / (C \times K))^{1/2} = (2 \times 91.065 / 12)^{1/2} = 3.896 \\
SE(G_i - G_j \text{ của cây}) &= (2 \times Sme / (D \times K))^{1/2} = (2 \times 91.065 / 20)^{1/2} = 3.018 \\
SE(S_{ij} - S_{kl} \text{ của dòng} \times \text{cây}) &= (2 \times Sme / K)^{1/2} = (2 \times 91.065 / 4)^{1/2} = 6.748
\end{aligned}$$

5 - Tính các thành phần ph- ơng sai di truyền

Phần đóng góp của dòng, cây thử, dòng $\times$ cây thử

$$\begin{aligned}
&\text{Phần do dòng : } \frac{SSd \times 100}{SSc \times 100} = \frac{10018.361 \times 100}{26199.54} = 39.384 \\
&\text{Phần do cây : } \frac{SSc \times 100}{SSc \times 100} = \frac{26199.54 \times 100}{1718.925 \times 100} = 15.242
\end{aligned}$$

$$\begin{aligned}
\text{Phần do cây thử : } & \frac{\text{SScl}}{\text{SSdc} \times 100} = \frac{26199.54}{14162.368 \times 100} = 6.561 \\
\text{Phần do dòng} \times \text{cây : } & \frac{\text{SScl}}{\text{Smd} - \text{Smdc}} = \frac{26199.54}{2579.590 - 1770.296} = 54.056 \\
\text{Cov(HS - dòng)} &= \frac{\text{C} \times \text{K}}{\text{Smc} - \text{Smdc}} = \frac{3 \times 4}{859.463 - 1770.296} = 67.441 \\
\text{Cov(HS - cây)} &= \frac{\text{D} \times \text{K}}{\text{Smc} - \text{Smdc}} = \frac{5 \times 4}{859.463 - 1770.296} = -45.542 \\
\text{Cov(HS - trung bình) mà chúng ta ký hiệu là Cov(HS)} &= \frac{1}{\text{K} \times (2 \times \text{D} \times \text{C} - \text{D} - \text{C})} \times \frac{(\text{D}-1) \times \text{Smd} + (\text{C}-1) \times \text{Smc}}{(\text{D} + \text{C} - 2)} - \text{Smdc} \\
&= \frac{1}{4 \times (30 - 5 - 3)} \times \frac{4 \times 2579.590 + 2 \times 859.463}{5 + 3 - 2} - 1770.296 \\
&= 2.681 \\
\text{Cov(FS)} &= \frac{(\text{Smd} - \text{Sme}) + (\text{Smc} - \text{Sme}) + (\text{Smdc} - \text{Sme}) + \text{Cov(HS)} \times \text{K} \times (6 - \text{D} - \text{C})}{3 \times \text{K}} \\
&= \frac{2579.590 + 859.463 + 1770.296 - 3 \times 91.065 + 2.681 \times 4 \times (6 - 5 - 3)}{3 \times 4} \\
&= 409.559
\end{aligned}$$

Ph- ơng sai của tổ hợp chung $\sigma^2_{\text{GCA}} = \text{Cov(HS)} = (1 + F)/4 \times \sigma^2_{\text{A}} = 2.681$

Với $F = 0$ có $\sigma^2_{\text{A}} = 10.724$

Với $F = 1$ có $\sigma^2_{\text{A}} = 5.362$

Ph- ơng sai của tổ hợp riêng $\sigma^2_{\text{SCA}} = (\text{Smdc} - \text{Sme})/K = (1770.296 - 91.065)/4 = 419.808 = ((1 + F)/2)^2 \times \sigma^2_{\text{D}}$

Với $F = 0$ có $\sigma^2_{\text{D}} = 1679.232$

Với $F = 1$ có $\sigma^2_{\text{D}} = 419.808$

6- Ch- ơng trình máy tính Linetest

Để phân tích lại đỉnh chúng tôi viết ch- ơng trình Linetest bằng ngôn ngữ Turbo Pascal, ch- ơng trình chạy nhanh và đơn giản cho ng- ời sử dụng.

Số liệu chạy để trong một tệp và gồm:

- 3 dòng đầu viết tiêu đề (Nội dung chính của đề tài, địa điểm, thời gian thí nghiệm, ng- ời chủ trì, ng- ời thực hiện ...)

- Dòng thứ t- ghi các tham số : Số dòng (D), số cây thử (C), số khối (K)

- Từ dòng thứ năm trở đi là số liệu:

Mỗi dòng là các số liệu (Chỉ tiêu theo dõi) trong các khối của một công thức

Các công thức là :

$D \times C$ tổ hợp lai của Dòng và Cây thử , bắt đầu là Dòng 1 $\times$ Cây1, sau đó là

Dòng 1 $\times$ Cây 2, . . . , Dòng 1 $\times$ cây C. Tiếp theo là Dòng 2 $\times$ Cây1, ...

Sau các tổ hợp lai là D dòng (Bố)

Cuối cùng là C cây thử (Mẹ)

Nếu chỉ nghiên cứu- cây lai thì không có các phần dòng và cây thử. Trong quá trình

chạy máy sẽ hỏi và ng- ời dùng sẽ trả lời bằng cách gõ một chữ trong số các chữ máy đề sẵn, thí dụ " In ra màn hình hay máy in (A / I /T) ", nếu muốn cho kết quả ra màn hình thì gõ A, nếu muốn in ra máy in thì gõ I, muốn ghi vào tệp thì gõ T.

B- Ggebiplot và Line x Tester

Chúng tôi đã giới thiệu ch- ơng trình Ggebiplot của Weikai Yan và ứng dụng để tính độ ổn định của các giống qua các môi tr- ờng (Gene x Environment).

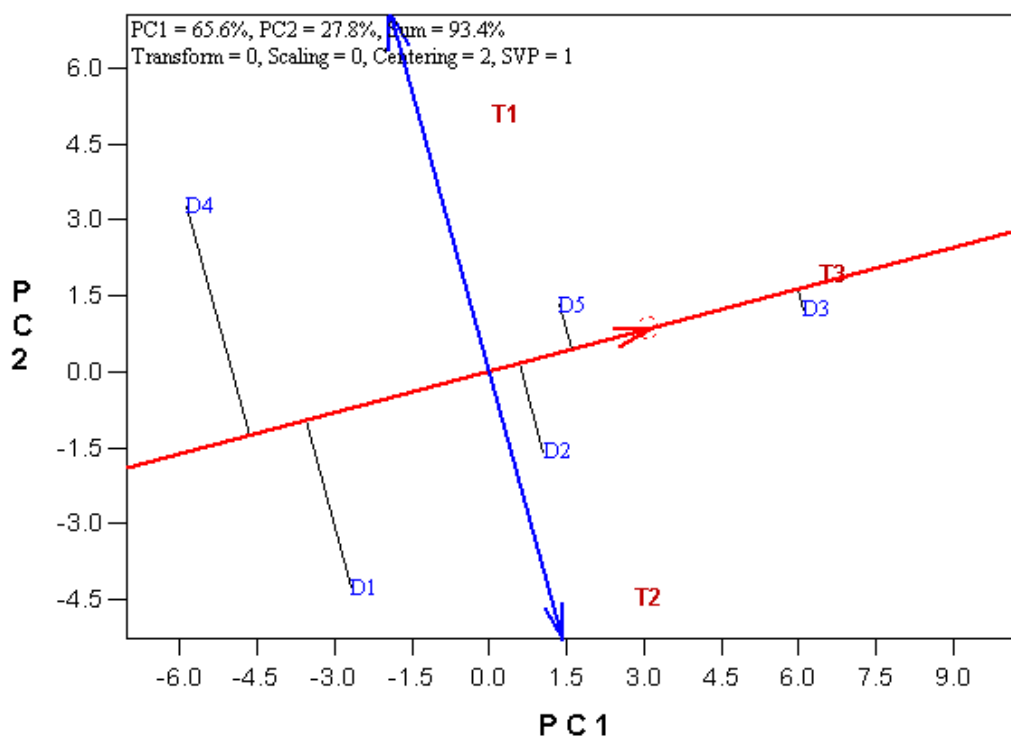
Hiện nay chúng tôi ch- a có các tài liệu nói về ứng dụng của Ggebiplot vào bài toán Line tester nh- ng so sánh lập luận dùng trong bài toán ổn định thấy có thể hình dung tester nh- môi tr- ờng, do đó có thể dùng mô hình Dòng x Tester thay cho Gene x Environment.

Lấy bảng các kết quả trung bình của các cặp lai Dòng x Tester thí dụ trên làm bảng số liệu để phân tích.

Có rất nhiều kết quả do Ggebiplot đ- a ra và tất cả chỉ mang tính gần đúng vì chỉ dùng đến 2 thành phần chính, nh- trong thí dụ này các kết quả chỉ gần đúng ở mức độ 90%. Kết quả trên hình vẽ nên dễ nhìn thấy và mang tính gợi ý để xem kỹ hơn các bảng kết quả mà ch- ơng trình Linetest cung cấp.

Đối với thí dụ trên chúng tôi đ- a ra 2 hình vẽ: hình vẽ đầu cho ta hình chiếu đồng thời cả Dòng và Tester lên một mặt phẳng chung. Căn cứ vào hình chiếu trên trục thứ nhất có thể thấy các khả năng kết hợp chung của các dòng. Dòng D₃ có KNKHC cao nhất, D₁ và D₄ có KNKHC thấp nhất, điều này hoàn toàn phù hợp với bảng các khả năng kết hợp chung.

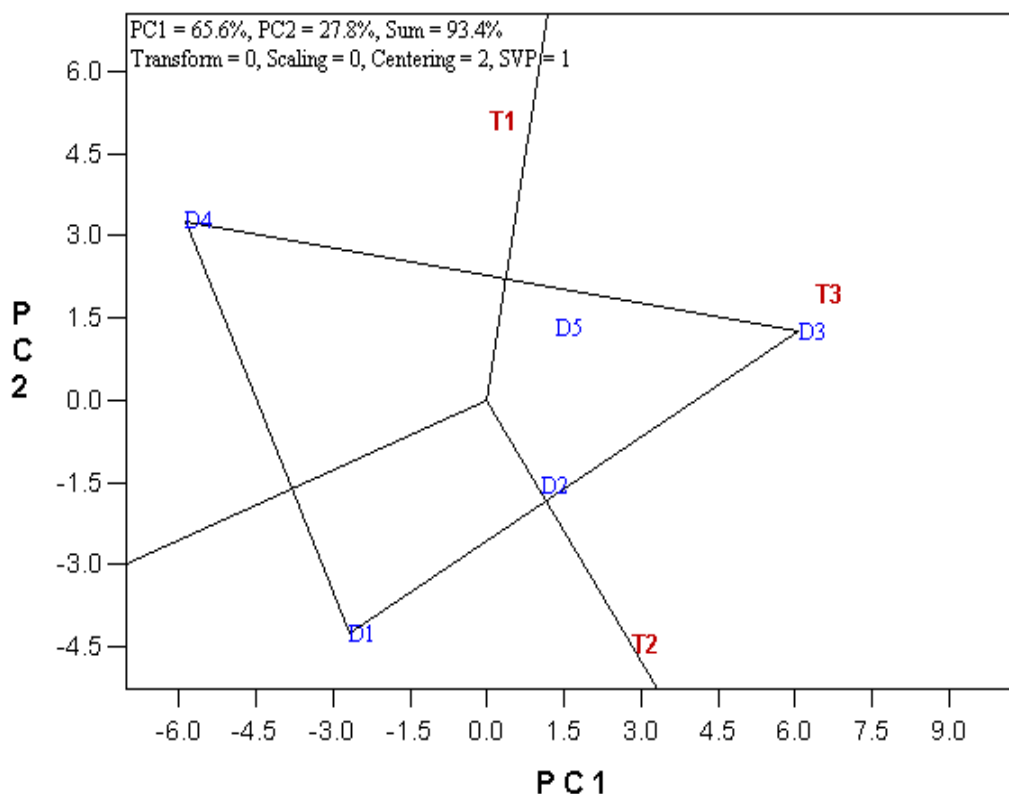
Các Dòng có khoảng cách theo h- ớng trục thứ hai cách xa trục thứ nhất có biến động lớn (các số liệu có ph- ơng sai lớn) thí dụ Dòng D1 và D4 có biến động lớn, điều này phù hợp với bảng các giá trị trung bình Dòng x Cây thử ở trên.



The Average Tester Coordination view

Hình vẽ thứ hai cho thấy mối quan hệ giữa Dòng và Tester.

Xét các đỉnh của đa giác bao trùm các dòng (tam giác D₄, D₃, D₁) sau đó vẽ 3 đường trục giao với 3 cạnh, 3 đường đó chia toàn bộ mặt phẳng thành 3 khu vực (sector). Khu vực có đỉnh là dòng D₃ chứa cây thử T₃ như vậy dòng D₃ và cây thử T₃ có khả năng kết hợp riêng cao, khu vực có đỉnh là dòng D₄ có cây thử T₁ như vậy dòng D₄ và cây thử T₁ có KNKHR cao, khu vực có đỉnh là dòng D₁ tuy không có cây thử nào như vậy dòng D₁ và cây thử T₂ có KNKHR cao. Các kết luận này phù hợp với các giá trị trong bảng các khả năng kết hợp riêng ở trên.



Which wins where or which is best for what

Trên đây là một số nhận xét sơ bộ về khả năng dùng Ggebiplot trong phân tích Line x Tester. Khi nào có tài liệu mới hoặc có những ứng dụng mới chúng tôi xin tiếp tục giới thiệu với các bạn đọc.

Tài liệu tham khảo

- 1- Nguyễn đình Hiền. *Chương trình Linetest. Bộ môn công nghệ phần mềm ĐHNN 1.*
- 2- Weikai Yan. *Software Ggebiplot. (Bản Beta Ggebiplot trên mạng tại địa chỉ www.ggebiplot.com)*
- 3- Dabholkar A R *Elements of Biometrical Genetics New Delhi 1992. Trang 326-365*
- 4- Singh R K and Chaudhary B D *Biometrical methods in quantitative genetic analysis Kalyani publishers New Delhi 1985. Trang 253-278.*
- 5- Weikan Yan *Ggebiplot - a Windows application for Graphical Analysis of multienviroment trial dats and another types of two way data Agronomy Journal 93 1111-1118 (2001)*

Hà nội 25/07/2006

3. Ngô Hữu Tình và Nguyễn Đình Hiền. Chương trình phân tích lai đỉnh từng phần. Tuyển tập kết quả nghiên cứu chọn lọc và lai tạo giống ngô. Nhà xuất bản Nông nghiệp 1995, trang 46 -52.
4. Ngô Hữu Tình và Nguyễn Đình Hiền. Chương trình phân tích lai đỉnh (Top cross). Tạp chí Nông nghiệp - Công nghiệp thực phẩm. Số 1/1996, trang 10 -12.