التحليل العنقودي لتراكيب وراثية مختلفة من فستق الحقل

رياض جبار منصور* ماجد شايع حمدالله**
*جامعة الكوت ـ كلية الزراعة
**جامعة بغداد ـ كلية الزراعة
E-mail: majidzoini@yahoo.com

الكلمات المفتاحية: فستق الحقل، تحليل عنقودي، تباعد وراثي

تاريخ الاستلام: 20 / 11 / 2012 تاريخ القبول: 16 / 1 / 2013

المستخلص:

نفذت تجربة حقلية في أربعة مواقع هي الأنبار وجلولاء والأسحاقي والكوت في الموسم 2002لأختباراداءاحد عشر تركيبا من فستق الحقل (ArachisHypogaea) وتجميعها في مجاميع متجانسة وتقدير درجة التباعد بينها بصيغة مسافات Euclidean باستخدام التحليل العنقودي Cluster Analysis . تم تحليل بيانات حاصل القرنات باستخدام طريقة التجميع الهرمي (hierarchical) بعد ان تم تقدير درجة التشابه عن طريق إنشاء مصفوفة Proximities. أشارت نتائج التحليل العنقودي الى تجمع التراكيب الوراثية في ست مجاميع تشكلت خلال عشر مراحل. تكونت كل مجموعة من تركيبين وراثيين باستثناء المجموعة السادسة التي تكونت من تركيب واحد فقط. دلت نتائج التحليل العنقودي ان التركيبين Medni و ICGV863كانا الأكثر تشابها لامتلاكهما اقل مسافة Euclideanوهذا يشير الى تشابه الجينات التي تسيطر على صفة حاصل القرنات، في حين حصلت اعلى مسافة Euclidean للتركيب C92 والتي تدل على البعد الوراثي لهذا التركيب ربما بسبب اختلاف اصله الوراثي . توزعت التراكيب المدروسة الى مجموعتين رئيسيتين ،ضمت المجموعة الأولى (A) تراكيب المجاميع I وIII و VI في حين ضمت المجموعة الثانية (B) بقية المجاميع والتي تميزت بمعدل حاصل اعلى من المجموعة A وهذا مؤشر عن وجود جينات مفضلة مسؤلة عن صفة الحاصل في تراكيب المجموعة B يمكن الاستفادة من انعزالاتها.يمكن استخدام التراكيب المتباعدة وراثيا لتحسين الأصناف المحلية وذلك بنقل جينات الحاصل المتفوقة بإحدى طرائق التربية لأن التباعد الوراثي سيضمن عدم تشابه الجينات بين التراكيب المختبرة.

CLUSTER ANALYSIS FOR DIFFERENT GENOTYPES OF PEANUT

Ryad J. Mansour* **Majid Sh. Hamdalla****
***University of Al-Kut - College of Agriculture**
****University of Baghdad - College of Agriculture**
E-mail: majidzoini@yahoo.com

Keywords: Peanuts, Cluster analysis, Hierarchical, Genetic diversity.

Recievd: 12 /11 / 2012 **Accept: 6 / 1 / 2013**

ABSTRACT:

A field experiment was carried out at four sites (Al-Anbar ,Al-AshaqiGaloulaa andAl-Kut) in 2002 to test the performance of 11 genotypes of peanuts (*ArachisHypogaea*) and accumulated into uniform groups, in addition to estimate the genetic diversity in term of euclidean distancethrough cluster analysis. Pod yield was analyzed based on hierarchical clustering technique after constitute resemblance matrix. The genotypes grouped into six clusters which they formed during teen stages. Each group consisted of two genotypes beyond groupVI which it consisted of one group. The cluster analysis indicated that genotypes Medni and ICGV863 were hadsimilar response because of having minimum Euclidean distance and this may be an indicator of similarity for the genes they carry on. The genotype C92 had the highest Euclidean distance and genetic diversity and this may be a result of difference of its pedigree. The tested genotypes were grouped into two major groups, group (A) included genotypes I, III , VI,while group B included the rest genotypes which had highest yield than A, this may be a result of existing favorable genes govern yield character. The local cultivars can be improved through transporting the yield genes from diverse, superior genotypes.

المقدمة:

يعد محصول فستق الحقل (*Arachis hypogaea*) من المحاصيل الزيتية المهمة ذات المردود الاقتصادي العالي (Hamidou et al,2011) وتتراوح نسبة الزيت في بذوره بين 50-55 %. ان اختلاف استجابة التراكيب الوراثية للعوامل البيئية يفرض على الباحثين ملاحظة اداءها خلال مواقع ذات تباين جغرافي(Roozeboom واخرون، 2008) من خلال تطبيق تجارب البيئات المتعددة (MEYT) لتحديد انماط التداخل الوراثي-البيئي الذي يعد القاعدة التي يُبنى عليها الأنتخاب للتكيف

(Phakamas واخرون، 2010). ان تقييم نسبة التباعد الوراثي من الأهداف المهمة التي يصبوا اليها مربي النبات لكل مشروع تربية ينفذه سواء كان انتخابا او تهجينا او إدخالا. ان استخدام مواد وراثية جديدة هو امر ضروري الأ انه يجب توفر عدد من الشروط فيها ومن اهمها انتكون ذات بعد وراثي عن التراكيب المحلية المطلوب تحسينها لأجل ضمان امتلاكها لجينات غير موجودة (لاتشابه) نظيراتها في التراكيب المحلية. يعتبر التحليل العنقودي من الأدوات الجيدة لمربي النبات لتقييم التباعد الوراثي وتحديد مواقع الصفات الكمية (QTL) والمحافظة على الاصول الوراثية، فضلا عن انه لايحتاج الى وضع افتراضات حول طبيعة توزيع البيانات(Yaو Ortiz, 1994). يعرف التحليل العنقودي (taxonomy Analysis segmentation or cluster) بأنه ذلك التحليل الذي يبحث في تحديد المجاميع المتشابهة لمجموعة من الحالات (التراكيب الوراثية) لصفات محددة بالاعتماد على تشابه استجاباتها للظروف البيئية المحيطة وتصنيف التراكيب الوراثية حسب تقاربها او تباعدها الوراثي (2011,Stuckburger) ،اذ يعتمد على حساب مسافات Euclidean والتي تعبر عن مقدار التباعد (dissimilarity or diversity) بين معدلات المتغيرات الكمية ومن ثم انتاج dendogram يبين توزيع التراكيب الوراثية المدروسة الى مجاميع بحسب ادائها واصولها الوراثية.

توجد طرائق متعددة لهذا التحليل منها طريقة Two-step التي تناسب اعداد البيانات الكبيرة الحجم ($n\leq1000$) وطريقة *K*-means التي تستخدم عند القدرة على تحديد عدد clusters مسبقا وطريقة التجميع الهرمي (hierarchical) وهي اهمها واكثرها شيوعا والتي تسمح بأختبار مسافة محددة واختبار طريقة الأرتباط. تناسب هذه الطريقة حجم البيانات الصغير ($n\leq250$) وتعرض النتائج بشكل dendogram. تتضمن طريقة التجميع الهرمي خطوتان، الأولى تكوين جدول لدرجة التشابه بين المشاهدات يدعى مصفوفة proximities يتم فيه تقدير مسافات euclidean بين المشاهدات المدروسة وهذا يتم بطرق مختلفة مثل Euclidean Distance و Squared Euclidean Distance و Cosine و Minkoeski و Chi-square و Size Difference و Simple Matching وغيرها من الطرق التي تستخرج درجة التشابه بطرق رياضية مختلفة. الخطوة الثانية يتم فيها تجميع المشاهدات الى مجاميع متعددة حسب درجة التشابه وهذا يتم بطرق مختلفة منها Between- Groups Linkage و Wiyhin-Groups Linkage و Nearest neighbor و Furthest Neaghbor و Centroid و Wards وغيرها ، حيث تعتبر كل مشاهدة مجموعة ثم تدمج كل مشاهدتين (مجموعتين) الأكثر تشابها (ذات مسافة euclidean الأقل من مصفوفة proximities) في مجموعة واحدة بعدها تدمج المجموعة الجديدة مع مجموعة اخرى متشابهة وهكذا تعاد العملية حتى يتم دمج جميع المشاهدات في مجموعة واحدة رئيسة. يواجه الباحثون مشكلتين عند أجراء التحليل العنقودي، الأولى هي تحديد الصفات المناسبة التي يجب الاعتماد عليها في هذا التحليل، اذ ان الصفات تختلف في مقدرتها على تمييز المتغيرات الى مجاميع فعلى سبيل المثال عند اختبار مجموعة اصناف تمتلك جميعها مساحة ورقية متقاربة، فهنا تعتبر صفة المساحة الورقية غير مفيدة في التحليل العنقودي لأنها ذات قدرة واطئة على تمييز تلك الأصناف الى مجاميع متمايزة لذلك فان خبرة الباحث هي المعيار في تحديد الصفة (الصفات) الملائمة. المشكلة الثانية هي كيفية تجميع القياسات المختلفة (الصفات ذات الوحدات المختلفة) الى رقم او قيمة مفردة تستخدم للتعبير عن درجة التشابه بين الأصناف. ان تحويل البيانات Transformation بطرق مختلفة قبل تكوين مصفوفة proximities سيكون الحل الأمثلوهذا يعتمد على خبرة الباحث ((2011, Stuckburger).

يهدف البحث الى (1) تقدير درجة التباعد الوراثي بين جينات التراكيب الوراثية لأحد عشر تركيبا من فستق الحقل تم اختبارها خلال اربع بيئات بدلالة حاصل القرنات و (2) تصنيف او تقسيم التراكيب المدروسة الى مجاميع للاستفادة منها كاباء في برامج التربية (3) التعريف بهذه التقانة الأخصائية التي استخدمت حديثا في العراق.

المواد وطرائق البحث:

طبقت تجربة بتصميم القطاعات الكاملة المعشاة بأربعة مكررات لمقارنة احد عشر تركيبا وراثيا من محصول فستق الحقل جميعها مدخلة من السودان باستثناء الصنف المحلي. ذات أصول وراثية مختلفة في الموسم 2002 في اربعة مواقع هي AL-Anbar و AL-Ashaqi و Galoulaa و AL-Kut . زرعت التراكيب الوراثية في جميع المواقع للمدة 25 نيسان – 1ايارالبذور على مروز بمسافة 90 سم وبين الجور 25سم بواقع بذرتين في الجورة. اضيف سماد السوبر فوسفات بمعدل 400 كغم . هكتار$^{-1}$ عند الحراثة واضيف سماد اليوريا بمعدل 80 كغم. هكتار$^{-1}$ بعد الانبات. اجريت عمليات العزق والتعشيب كلما دعت الحاجة لذلك واجريت عملية التحضين مرتين. تم اختيار عشر نباتات عشوائية و تم حصادها عند النضج. أجري تحليل التغايرالتجميعي (AMMI) وتحليل cluster بالاعتماد على طريقة agglomerative لتبسيط البيانات عن طريق تجميع التراكيب الوراثية في مجاميع للصفات المدروسة حسب تشابه نمط الاستجابة (1976,Williams وSneath وSoka, 1973). تم أجراء التحليل على مرحلتين، الأولى تضمنت تحليل البيانات بطريقة المكونات الرئيسة (1967,Jeffers)، ثم اجراء تحليل cluster في المرحلة الثانية. يتكون تحليل cluster من عدد من الخطوات تبدأ بتكوين مصفوفة لدرجة التشابه بين التراكيب المدروسة تدعى proximities matrix ثم يتم تكوين dendogram بألأعتماد على طريقة UPGMA (Sneath وSoka, 1973)، اذ تحسب

مسافات Euclidean التي تعبر عن درجة التشابه بين معدلات المجاميع من مصفوفة proximities matrixالتي تم تكوينها في المرحلة الأولى وذلك حسب المعادلة الأتية:

$$\text{Distance } (x,y) = \{\sum_i (x_i - y_i)^2\}^{½}$$

(Punitha et al ,2010). أستخدم برنامج SPSS v.17 لتحليل البيانات.

النتائج والمناقشة:

يشير (جدول-1) الى وجود اختلافات معنوية بين التراكيب الوراثية خلال البيئات المدروسة وهذه تعتبر اشارة اولية عن تباعد وراثي بينها يمكن الكشف عنه بالتحليل العنقودي. شارك التغاير الوراثي والتداخل الوراثي-البيئي بنسبة 24 % لكل منها. تفوق التركيب C97 في صفة حاصل القرنات على بقية التراكيب الوراثية وبنسبة 110% عن اوطأ التراكيب المدروسة. في حين اختلف تأثير البيئات نتيجة للتداخل الوراثي-البيئي وقد تفوقت البيئة Galoulaa على بقية البيئات (جدول-2) وهذا يتفق مع ما وجده Hamdalla واخرون (2013) عند تحليلهم لبعض تراكيب فستق الحقل تحت البيئات نفسها باستخدام تقانة GGE-Biplot الأخصائية.

يشير جدول-3 الى مراحل تشكل المخطط العنقودي. يدل عمود المعامل في جدول3 على طبيعة التباعد بين المجاميع، حيث تشير قيمة المعامل الصغير على ان المجاميع متشابهة بينما تدل قيمة المعامل الكبيرة على انخفاض التجانس. ان قيمة المعامل تناظرمسافات euclideanعلى المحور الأفقي للشكل1، الأ انها لاتتشابه معها من المجاميع اعتمادا على قيم euclidean الأقل حتى تندمج جميع التراكيب ضمن شكل اومجموعة حيث اعيدت صياغة قيم المعامل الى المدى المكون من 1- 25 حسب برنامج SPSS لتسهيل قرائتها. يلاحظ ان المرحلة الأولى بدأت بدمج اكثر تركيبين تشابها في الأداء وهما Medni (تسلسل 2) و ICGV863 (تسلسل 9) في مجموعة واحدة يطلق عليها المجموعة 2 اعتمادا على التسلسل الأصغر لهما، فقد امتلكا اقل قيمة لمسافة Euclidean والتي تم التعبير عنها بقيمة المعامل البالغة 578633. ان مسافة Euclidean الواطئة بين هذين التركيبين تدلل على تشابه جينات الحاصل بينهما. في المرحلة الثانية تم دمج التركيبين C95 (تسلسل3) و ICGV868 (تسلسل 5) ليكونا المجموعة 3 بقيمة معامل قدره 697871 وهكذا حتى دمجت جميع التراكيب في مجموعة واحدة في الخطوة العاشرة.

جدول-1: التحليل التجميعي لتباين حاصل القرنات لأحد عشر تركيبا وراثيا من فستق الحقل خلال أربع بيئات.

s.o.v.	d.f.	s.s.	m.s.	f-cal.	f-tab.	% s.s.	Model
Env.	3	139339342	46446447	87.31	**	41.8	عشوائي
Reps./env.	12	6383541	531962	16.oo	-	-	-
Gen.	10	81847870	8184787	39.79	**	24.6	ثابت
Gen.×Env.	30	80327708	2677590	13.02	**	24.1	عشوائي
Error	120	24684610	205705	-	-	-	-
Total	175	332583071	-	-	-	-	-

جدول-2: متوسط حاصل القرنات (طن . هكتار$^{-1}$) لأحد عشر تركيبا وراثيا من فستق الحقل خلال اربع بيئات.

	البيئات				
التركيب الوراثي	Al-Anbar	Al-Ashaqi	Galoulaa	Al-Kut	المعدل
ICGV87	5143	4329	4748	2348	4142
Medna	3201	3836	5610	1491	3535
C95	4036	3665	4047	3995	3761
C109	4292	5022	4960	3326	4400
ICGV868	4412	2998	4350	3433	3798
C92	2481	2101	4019	1134	2343
C97	4378	4569	6169	4593	4927
ICGV88	2050	3392	5114	1844	3100
ICGV863	3328	3884	6256	1869	3834
Local	2194	4443	4976	2060	3418
CGV862	3763	3381	6993	4634	4693
ا.ف.م.	682.3				317.5
المعدل	3571	3784	5204	2730	3822
أ.ف.م.	355.8				

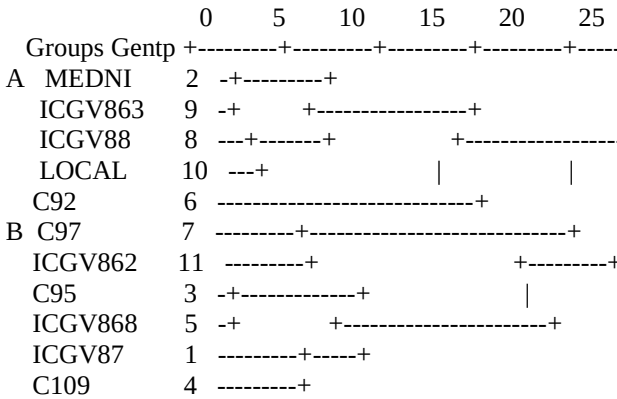
جدول-3 : المعامل المقدر بين المجاميع حسب التحليل العنقودي (cluster)

المرحلة	Cluster 1	Cluster 2	المعامل
1	2	9	578633
2	3	5	697871
3	8	10	1191314
4	1	4	2204177
5	7	11	2471457
6	2	8	2612739
7	1	3	3482799
8	2	6	6630302
9	1	7	8341212
10	1	2	1.071 E7

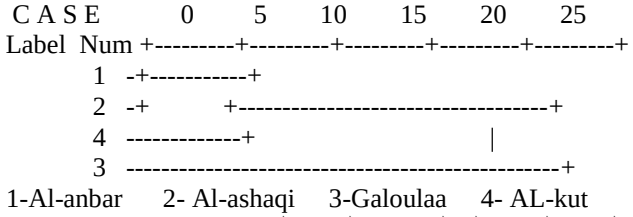
جدول- 4: عدد واسماء التراكيب لكل مجموعة حسب التحليل العنقودي(cluster).

المجاميع	اسم التركيب	عدد التراكيب الوراثية	المعدل
I	Medni , ICGV863	2	3684
II	C95 , ICGV868	2	3779
III	ICGV88 , Loca	2	3259
IV	ICGV871 , C109	2	4271
V	C97 , ICGV862	2	4810
VI	C92	1	2434

مسافات Euclidean



شكل 1 التجميع الهرمي (Hierarchical Clustering) لتوزيع تراكيب فستق الحقل على المجاميع (المحور العمودي) ومسافات Euclidean (المحور الأفقي)



شكل-2: التجميع الهرمي (Hierarchical Clustering) لتوزيع البيئات على المجاميع (المحور العمودي) ومسافات Euclidean (المحور الأفقي)

أنشيء مخطط التحليل العنقودي بالأعتماد على تقدير درجة التشابه بين التراكيب المدروسة (شكل-1) ، اذ توزعت التراكيب الوراثية الى ست مجاميع (جدول 4). ضمت كل مجموعة تركيبان وراثيان باستثناء المجموعة السادسة التي تكونت من تركيب وراثي واحد وهذا يشير الى البعد الوراثي لهذا التركيب عن بقية التراكيب الوراثية المدروسة الذي امتلك اقل معدل لصفة حاصل القرنات في حين كانت المجموعة الخامسة هي الأعلى لهذه الصفة لأنها ضمت التركيبين المتفوقين C97 و ICGV862.

يبين (الشكل- 1) طبيعة توزيع التراكيب الوراثية الى مجاميع مختلفة حسب درجة تشابه أداء هذه التراكيب الوراثية. يلاحظ من الشكل ان التراكيب توزعت الى مجموعتين قبل ان تندمج جميع التراكيب في مجموعة واحدة في المرحلة الأخيرة، ضمت المجموعة الأولى (A) تراكيب المجاميع I و III و VI في حين ضمت المجموعة الثانية (B) بقية المجاميع والتي تميزت بمعدل حاصل اعلى من المجموعة A وهذا مؤشر عن وجود جينات مفضلة مسؤلة عن صفة الحاصل في تراكيب المجموعة B يمكن الاستفادة من انعزالاتها. يشير شكل 2 الى نمط تجمع البيئات المدروسة الى مجاميع حسب اداء تراكيب فستق الحقل خلالها. يلاحظ ان البيئات توزعت الى مجموعتين، تكونت المجموعة الأولى من ثلاث بيئات هي AL-Anbar و AL-Ashaqi و AL-Kut في حين ضمت المجموعة الثانية بيئة Galoulaa فقط. ان تجمع البيئات بهذا النمط يشير الى تشابه اداء التراكيب خلال بيئات كل مجموعة خاصة البيئتان AL-Anbar و AL-Ashaqi اللذان امتلكا اقل مسافة euclidean ، بالتالي يمكن الأنتخاب في احدى هاتين البيئتين وتعميم النتائج على البيئة الأخرى. كانت البيئة Galoulaa الأبعد عن جميع البيئات الأ انها امتلكت اعلى معدل حاصل للتراكيب المدروسة.ان تحسين صفة الحاصل يعتمد على تجميع الجينات المفضلة لهذه الصفة من خلال انتخاب التراكيب التي تتميز بمعدل حاصل عال وذات تباعد وراثي عال عن التراكيب المحلية وهذا يمكن الكشف عنه بأستخدام تقانات الوراثة الجزيئية والتقانات الأحصائية الحديثة كتقانة التحليل العنقودي او تقانة GGE-Biplot وغيرها.

REFERENCE:

Hamdalla,M.Sh., R.J.Mansour, and H.Sarhan.2013. Using of Biplot technique for analysis of G.E. interaction for peanut pod yield. AL-Kufaa J. (under press).

Hamidou, F., P. Ratnakumar, O. Halilou, O. Mponda, T. Kapewa, E. Monyo, I. Faye, B. R. Ntare, S.N. Nigam, H.D. Upadhyaya, and V. Vadez. 2011. Selection of intermittent drought tolerant lines across years and locations in the reference of groundnut. http: //dx.doi.org /10. 1016/ i.fcr.2011.10.009..

Jeffers, J.N.R. 1967. Two case studies in the application of Principal component analysis. Appl. Stat. 16:225-236.

Phakamas,N., A.Patanothai, K.Panngpetch, S.Jogloy and G.Hoogenboom. 2010. Determination of adaptive responses of peanut genotypes and patterns of genotypes × location interaction using the CSM-CROPGRO- peanut model. Int.J. Plant Prod. 4(3): 223-232.

Punitha,B., P.Vindhiyavarman and N.Manivannan. 2010. Genetic divergence study in sunflower (Helianthus annuus L.). Electr. J. Plant Breed. 1(4): 426-430.

Roozeboom, K.L., Schapaugh, W.T., Tuinstra, M.R., Vanderlip, R.L., Milliken, G.A., 2008.Testing wheat in variable environments: genotype, environment, interaction effects, and grouping test locations. Crop Sci. 48, 317–330.

Sneath,P.H.A. and R.R.Sokai.1973. Numerical Taxonomy: The Principal and Practice of Numerical Classification. W.H.Freeman Co., San Francisco.

Stuckburger, D.W. 2011. Multivariate Statistics: concepts, models and applications. www.psychstat.mission.

Williams, W.T. 1976. Pattern Analysis in Agriculture Sciences. Elselver, Amsterdam.

Yau.S.K. and G.Ortiz.1994. Regression and analysis of elite wheat lines grown in rainfed environment. J. Genet. Breed. 48:183-190.