تصنيف عدد من الكائنات الحية باستخدام تحليلي متعدد الأبعاد والتحليل العنقودي للسلسلة الببتيدية المتعددة لإنزيم Lactate dehydrogenase

علي عبد الحافظ إبراهيم
كلية العلوم ، جامعة النهرين

الخلاصة

يهدف البحث الحالي إلى تصنيف وتمييز 13 كائنا من الكائنات الحية المختلفة استنادا إلى السلسة الببتيدية المتعددة لإنزيم Lactate dehydrogenase باستخدام طريقتي تحليل المقياس المتعدد الأبعاد والتحليل العنقودي، كانت من أهم نتائج هذا البحث تصنيف وتمييز الكائنات إلى خمس مجموعات، تحتوي كل مجموعة منها على أعداد مختلفة من الكائنات الحية تتماثل فيما بينها، من جراء تماثل السلسة الببتيدية المتعددة لإنزيم Lactate dehydrogenase، وتختلف الكائنات الحية في المجموعة الواحدة عن المجموعات الأخرى، في حين تضمنت الاستنتاجات مايأتي:

أ- تطابق نتائج طريقتي المقياس المتعدد الأبعاد والتحليل العنقودي، في تصنيف الأجناس المختلفة من الكائنات الحية المختلفة (بيانات البحث)، إلى نفس العدد من المجموعات من جهة (بين المجموعات) ومن جهة أخرى فقد تماثلت في تصنيف وتحديد العدد ذاته من المشاهدات (الأجناس المختلفة من الكائنات الحية – بيانات البحث) في داخل المجموعة الواحدة (داخل المجموعات).

ب- تمكنت طريقة مقياس متعدد الأبعاد من إعطاء تفاصيل أكثر من طريقة التحليل العنقودي، وذلك من خلال تحديد وتصنيف مواقع المشاهدات (الأجناس المختلفة من الكائنات الحية – بيانات البحث) على خارطة ذات بعدين. في حين صنفت طريقة التحليل العنقودي المشاهدات المختلفة (الأجناس المختلفة من الكائنات الحية) كمجموعات مختلفة من دون أن تحدد مواقع هذه النماذج على الخارطة.

ج- امتازت طريقة التحليل العنقودي بدقة أكبر من طريقة المقياس المتعدد الأبعاد في تحديد درجة التقارب الوراثي بين الكائنات الحية المختلفة.

المقدمة

يعد إنزيم Lactate dehydrogenase من إنزيمات الأكسدة والاختزال، هذا الأنزيم يحفز و يحول حامض اللاكتيك (اللاكتيت) إلى باروفيت، ويستهلك في هذا التفاعل جزيئه واحدة من المرافق الأنزيمي NADH كمستقبل للهيدروجين لكل ثلاث ذرات كاربون، وبالتالي تحفيز أكسدة اللاكتيت إلى بايروفيت من خلال قابليته على نقل الهيدروجين بوجود المرافق الإنزيمي NAD^+ ، ويعد هذا الإنزيم من الإنزيمات المهمة في الجسم لكونه يساهم في مسار انحلال الكلوكوز لا هوائيا، [4]

ان إنزيم Lactate dehydrogenase يتمركز في سايتوبلازم الخلايا وبالتحديد في كل من السايتوسول والمايتوكوندريا والكلوروبلاست والبيروكسيسومات وللإنزيم خمسة متناظرات مشتقة من وحدتين فرعيتين للإنزيم يطلق عليها القلبية (H) والعضلية (M) بوزن جزيئي 35000 دالتون [1].

ويستخدم إنزيم Lactate dehydrogenase الذي يظهر خارج الخلية في التحري عن الضرر الذي يحدث للخلايا، إذ يتحرر الإنزيم، ويطلق إلى الدم المحيط بعد موت الخلايا لدى تعرضها إلى أي مسبب من مسببات الموت [3].

وتعد عملية تصنيف الأجناس المختلفة من الكائنات الحية، إلى أنواع ومجموعات متماثلة لمعرفة أصل ومراحل تطور ونشوء هذه الكائنات من ناحية، ومن ناحية أخرى إمكانية معرفة التغيرات الجينية والطفرات الوراثية للكائنات المتماثلة فيما بينها (التي تصنف ضمن المجموعة الوحدة) والمختلفة منها (التي تصنف ضمن مجموعات مختلفة)،

وعلى هذا الأساس فقد كان الهدف من هذا البحث هو تصنيف وتمييز عدد من الكائنات الحية من خلال استخدام السلسلة الببتيدية المتعددة لإنزيم Lactate dehydrogenase للكائنات الحية المختلفة باستخدام طريقتي تحليل المقياس المتعدد الأبعاد والتحليل العنقودي، ولتحقيق هذا الهدف فقد تم تقسيم البحث إلى ثلاثة محاور: أشتمل المحور الأول على إعطاء صورة مركزة عن الإطار النظري للتحليلين المقياس المتعدد الأبعاد والعنقودي، وتناول المحور الثاني الجانب التطبيقي ولكلا التحليلين، أما المحور الثالث والأخير فقد خصص لوضع الاستنتاجات والتوصيات انسجاما مع النتائج التي تم التوصل إليها في هذا البحث.

الجانب النظري

1- طريقة مقياس متعدد الأبعاد

تعتبر طريقة مقياس متعدد الأبعاد (Multidimensional Scaling)، إحدى طرائق متعدد المتغيرات، الذي يهتم بتحليل عينة م ا مكونة من n من العناصر (المشاهدات) مقاسه لـ p من المتغيرات. تعرف طريقة مقياس متعدد الأبعاد، بأنه ا عبارة عن سلسلة من الأساليب صممت لتكوين شكل بياني، يوضح هذا الشكل، العلاقة بين n من العناصر، بالاعتماد على جدول من المسافات بين هذه العناصر، ويمكن أن يكون هذا الشكل البياني ذا بعدا واحدا (إذا كانت جميع العناصر تقع على خط مستقيم)، أو ذا بعدين (إذا كانت جميع العناصر تقع في مستوي)، أو في ثلاثة أبعاد (إذا كانت جميع العناصر تمثل في الفضاء)، أو في عدد أعلى من الأبعاد (في الحالة التي التمثيل الهندسي لها غير ممكن). وتبدأ طريقة مقياس متعدد الأبعاد، بمصفوفة من المسافات (كمدخلات لهذه الطريقة)، بين n من العناصر (عدد هذه العناصر هو ½ n (n–1) من أزواج من النقاط) هي δij، حيث أنها تمثل المسافة الأقليدية (Euclidean distance) أو عدم التشابه (dissimilarity) بين العنصر i والعنصر j وتحسب مصفوفة المسافات باستخدام المعادلة الآتية [2] [9]:

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2} \quad \ldots\ldots\ldots\ldots\ldots (1)$$

حيث أن :

d_{ij} : المسافة بين العنصر i والعنصر j.

x_{ik} : قيمة المتغير x_k للعنصر i.

x_{jk} : قيمة المتغير x_k للعنصر j.

وقبل البدء بحسابات المسافات يتم تحويل المتغيرات الأصلية إلى متغيرات قياسية (Standardized Variables)، وبهذا فإن جميع المتغيرات p تكون متساوية من ناحية الأهمية، في تحديد هذه المسافات. ولأجل الحصول على مجال ذي عدد أبعاد أقل t ($t < p$) لمجموعة من النقاط n لحساب مصفوفة من المسافات d_{ij}، والتي لها نفس رتبة Rank مصفوفة المسافات الأصلية (مصفوفة المدخلات) δ_{ij}، لغرض تكوين شكل الرسم البياني لـ n من العناصر، يتم إتباع الأتي [5][6][7]:

أ- تهيئة شكل (صورة) أولية لـ n من العناصر، ولعدد من الأبعاد هو t، بعبارة أخرى، إن الإحداثيات $(x_1, x_2, x_3, ..., x_n)$ تفرض لكل عنصر في مجال ذي أبعاد عددها t.

ب- حساب المسافة الأقليدية d_{ij} (حيث أنها تمثل المسافة بين العنصر i والعنصر j) بين المشاهدات لهذا الشكل (الشكل ذي أبعاد عددها t)، وذلك بحساب انحدار المربعات الصغرى لـ d_{ij} على مسافات البيانات الأصلية δ_{ij}، كذلك يتم تقدير d_{ij} من خلال القيمة $\hat{d}_{ij}$ والتي تدعى بالتفاوت (disparities) وهو عبارة عن مسافة البيانات δ_{ij} المختارة ليماثل مسافة الشكل (المختار) d_{ij} أقرب قدر ممكن.

ج- قياس جودة التطابق بين مسافات الشكل (المختار) d_{ij} والتفاوت $\hat{d}_{ij}$، باستخدام إحصاءه ملائمة، تدعى بصيغة الإجهاد 1 والتي هي:

$$STRESS1 = \{\sum (d_{ij} - \hat{d}_{ij})^2 / \sum \hat{d}_{ij}^2\}^{1/2} \quad \ldots\ldots (2)$$

وتعرف هذه الإحصاءة بأنها عبارة عن مقياس المساحة، التي من خلالها يضغط على الشكل ألحيزي (الفضائي) للنقاط، لكي نتوصل قدر الإمكان إلى مسافة البيانات δ_{ij}.

د- تقارن قيمة مقياس الإجهاد مع قيمة صغيرة، فإذا كان مقياس الإجهاد أكبر من القيمة الصغيرة، يتم إيجاد شكل جديد، بحيث أن قيمة الإجهاد تقل أكثر من القيمة السابقة، ويستخدم مقياس متعدد الأبعاد طريقة الانحدار

الحـاد (steepest descent)، لإيجاد الشكل الجديد، الذي يتضمن بصورة أساسية حساب المشتقات الجزئية لدالة الإجهاد، لتحديد الاتجاهات، للحصول على زيادة قليلة (تحسين الصورة الحالية مقارنة مع الصورة الأصلية).

هـ- حالما يتم الحصول على قيمة الإجهاد المقنعة (أقل من القيمة الصغرى المحددة) نقلل عدد الأبعاد بواحد، وتعاد العملية (من الخطوة 2 إلى الخطوة 5) حتى يتم الوصول إلى أقل عدد من الأبعاد مع قيمة مقبولة للإجهاد.

2- طريقة التحليل العنقودي

تعد طريقة التحليل العنقودي أحد أساليب متعدد المتغيرات، وهي عبارة عن أسلوب يهتم بتصنيف البيانات إلى عناقيد أو مجموعات، وتتميز هذه العناقيد بكونها على درجة كبيرة من التشابه أو التجانس الداخلي (within cluster)، إلا إنها – العناقيد أو المجموعات – تكون غير متشابهة فيما بينها (between cluster). كما يعتبر التحليل العنقودي أداة مفيدة وفعالة لتحليل وتصنيف البيانات بطرق مختلفة ومتعددة. يستخدم التحليل العنقودي لتحليل مصفوفة من البيانات تحتوي على n من العناصر، ولكل منها عدد (p) من المتغيرات. وهناك العديد من الخوارزميات المستخدمة في التحليل العنقودي ومنها الخوارزمية الهرمية (Hierarchic) التي استخدمت في هذا البحث والتي تبدأ بحساب المسافات الأقليدية (Euclidean distance) لكل عنصر مع بقية العناصر. ومن ثم تتكون المجموعات من خلال عملية التكتل أو التعنقد. ومع بداية عملية التعنقد فان جميع العناصر تكون منفردة في المجموعات المختلفة. ثم تبدأ العناصر القريبة بالاندماج تدريجيا حتى تكون في النهاية جميع العناصر في مجموعة واحدة.

ويمكن توضيح عملية التعنقد بالخطوات الآتية [2] [9]:

أ-حساب المسافات الأقليدية التي تحدد درجة التقارب بين كل نوعين من العناصر المختلفة، وذلك استنادا إلى المعادلة (1).

ب- يتم البحث عن اقصر مسافة اقليدية في المصفوفة d_{ij} لغرض ربط العلاقة بين العنصر i والعنصر j.

ج-الاستمرار بعملية التعنقد أو التجمع بالاعتماد على اقصر مسافة اقليدية ممكنة إلى أن يتم ربط آخر مجموعتين (أو عنقود) لتكوين المجموعة النهائية، التي تكون على شكل هيكل شجرة من العناقيد أو المجموعات الأصلية، وتتميز هذه العناقيد بكونها على درجة كبيرة من التشابه أو التجانس الداخلي، إلا إنها – العناقيد أو المجموعات – تكون غير متشابهة فيما بينها.

3- اصطفاف السلاسل المتعدد

قبل البدء (عمليا) بتصنيف السلاسل الجزيئية، ينبثق سؤال مهم هو كيفية ترتيب (اصطفاف) السلسلة ؟. إن نوعية السلاسل ذات الاصطفاف المتعدد تحدد نوعية تصنيف هذه السلاسل. بعبارة أخرى، يعكس الاصطفاف للسلاسل التأثيرات الحيوية لها (المواقع المتجانسة) [8].

الجانب التطبيقي

1- الطريقة المستخدمة

1-1- طريقة تحليل مقياس متعدد الأبعاد

استخدمت طريقة مقياس متعدد الأبعاد (Multidimensional Scaling)، وتتضمن هذه الطريقة تكوين رسم بياني متعدد الأبعاد، باستخدام الطريقة غير المترية (Non metric method). هناك نوعان من طرائق مقياس متعدد الأبعاد، استنادا إلى مستويات القياسات ونسبة إلى المدخلات والمخرجات هما:

أ- **الطريقة لا مترية Non metric method**.

تفترض هذه الطريقة، مدخلات اعتيادية ومخرجات مترية.

ب- **الطريقة المترية Metric method**.

تفترض هذه الطريقة تماثل المدخلات والمخرجات في كونهما مترية.

1-2- طريقة التحليل العنقودي:

استخدمت طريقة التحليل العنقودي الهرمي Hierarchical Cluster Method حيث تتضمن هذه الطريقة بناء أو تشييد هرم أو شجرة على شكل هيكل. وهناك نوعان ضمن طريقة التحليل الهرمي هذه هي: الطريقة الأولى هي التعنقد (التكتل) (Agglomerate)، تبدأ المشاهدات أو العناصر منفردة في مجموعاتها الخاصة بها، المرحلة التالية تتضمن دمج المجموعتين الأكثر قربا لبعضهما البعض من باقي المجموعات المختلفة مكونة بذلك كتلة عنقودية جديدة،

بذلك فقد تم تقليص عدد المجموعات في هذه المرحلة والمراحل اللاحقة أيضا... وهكذا إلى أن نصل إلى أن جميع العناصر تتكتل في مجموعة واحدة كبيرة. أما الطريقة الثانية هي الانقسام (التشقق) (Divisive) فهي على عكس الطريقة الأولى، حيث أن هذه الطريقة تبدأ بمجموعة واحدة كبيرة تضم جميع المشاهدات (أو العناصر)، وفي المراحل اللاحقة فان المشاهدات الأكثر عدم تشابه (اختلاف) تنقسم عن المجموعة الكبيرة إلى مجموعات صغيرة. هذه العملية تستمر حتى تنفرد كل مشاهدة في مجموعة صغيرة.

2- بيانات البحث: [10]

تم الحصول على عدد من السلاسل الببتيدية المتعددة *Lactate dehydrogenase* لـ 13 نوع أو جنس، اشتملت على: Human و chicken و Dogfish و Lamprey و Barley و Maizey و Lacto casein و Bacillus stea و Lacto plant و Therma marine و Bifido و Thermua aqua و Mycoplasma والتي أدرجت في الجدول (1).

2-1 مصفوفة المسافات:

هي عبارة المسافة (عدم التماثل) بين جنس معين وجنس آخر، وفي حالة البيانات موضع البحث فأن مصفوفة المسافات تتألف من مصفوفة مربعة الشكل ذات بعد مكون من 13 صف وعمود، وقد أدرجت في الجدول (2).

3- نتائج البحث

3-1- نتائج طريقة مقياس متعدد الأبعاد

أدخلت مصفوفة البيانات (جدول(2)) في البرنامج الإحصائي (SPSS) وبعد التأكد من دقة البيانات تم تنفيذ البرنامج الخاص بطريقة مقياس متعدد الأبعاد، وكانت النتيجة الحصول على الآتي:

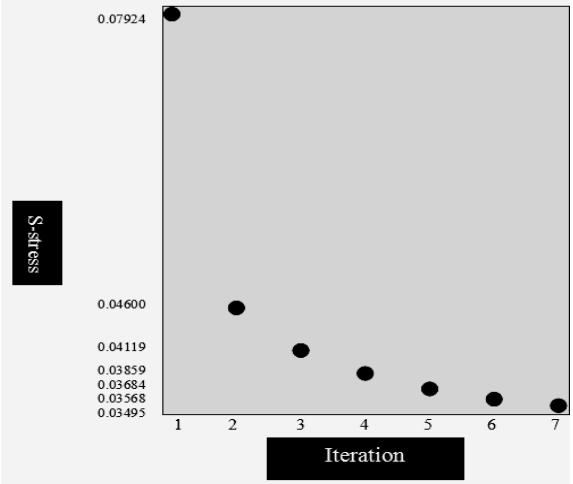
جدول (3)

نتائج التكرار (المسافات المثلى) باستخدام صيغة S-stress.

Iteration	S-stress	Improvement

1	.07924	
2	.04600	.03324
3	.04119	.00481
4	.03859	.00260
5	.03684	.00175
6	.03568	.00116
7	.03495	.00073

توقف التكرار في الخطوة (7)، ذلك أن مقدار التحسن في قيمة S-stress كان أقل من 0.001. ويعرف مقياس S-stress بأنه عبارة عن نسبة التباين (التفاوت) غير المحسوب من قبل مقياس متعدد الأبعاد، ويساعد هذا المقياس في تحديد العدد الملائم من الأبعاد في التحليل (بناء خارطة التحليل). ومن نتائج جدول (3)، رسم الشكل البياني الآتي، الذي يحدد أمثل عدد من الأبعاد لاستخدامه في التحليل، وكالآتي:



شكل (1): العلاقة بين التكرار وقوة الشد.

نلاحظ من الشكل أعلاه استقرار عملية التكرار للحصول على المسافات المثلى باستخدام صيغة S-stress، وبتحديد عدد الأبعاد مساوي إلى 2 ظهرت النتائج المدرجة في الجدول الآتي:

جدول (4)

إحداثيات المتغيرات لـ 13 جنس مختلف استنادا إلى البعدين الأول والثاني.

Stimulus Number	Stimulus Name	Stimulus Coordinate Dimension 1	2
1	VAR1	1.6747	-0.0748
2	VAR2	1.6557	-0.0098
3	VAR3	1.5692	-0.2615
4	VAR4	1.4327	-0.3466
5	VAR5	0.8736	0.3157
6	VAR6	1.0402	0.4090
7	VAR7	-1.2347	-0.0826
8	VAR8	-1.2229	0.4422
9	VAR9	-1.0966	0.1903
10	VAR10	-0.7517	-0.6574
11	VAR11	-0.5507	0.1743
12	VAR12	-1.6224	-1.0696
13	VAR13	-1.7669	0.9708

3-2- نتائج طريقة التحليل العنقودي

أدخلت مصفوفة البيانات جدول(2) في البرنامج الإحصائي (SPSS)، وبعد التأكد من دقة البيانات تم تنفيذ البرنامج الخاص بالتحليل العنقودي، كانت النتيجة الحصول على النتائج المبينة في الجدول (5).

جدول (5)

نتائج عملية التعنقد الناتجة من استخدام طريقة التحليل العنقودي لتحليل (13) جنس من الكائنات الحية المختلفة.

Stage	Cluster Combine	
	Cluster 1	Cluster 2
1	1	2
2	1	3
3	5	6
4	7	9
5	1	4
6	10	12
7	1	5
8	8	13
9	7	10
10	7	11
11	7	8
12	1	7

يتألف الجدول (5) من ثلاثة أعمدة، يمثل العمود الأول (من اليسار) المرحلة، وتأخذ الأعداد من 1 إلى 12 (التي تمثل عدد العناقيد المتشكلة)، ويمثل العمودين الثاني والثالث أرقام العناقيد المدمجة، ونتيجة لهذا الجدول يتكون الجدول يتكون الجدول (6).

جدول (6)

نتائج عملية التعنقد الناتجة من طريقة التحليل العنقودي لـ (13) جنس من الكائنات الحية المختلفة.

تابع جدول (6)

Number of clusters	Case	
1	X	X
2	X	X
3	X	X
4	X	X
5	X	X
6	X	X
7	X	X
8	X	X
9	X	X
10	X	X
11	X	X
12	X	X

يتألف الجدول (6) من (27) عمودا (13) منها تمثل الأجناس المختلفة، والأعمدة المتبقية تمثل الترابط بين كل جنسين متجاورين، في حين تمثل الصفوف أسبقية (تسلسل) ترابط (تعنقد) هذه الأجناس مع بعضها الآخر. ومن الجدول أعلاه يرسم الشكل (2).

شكل (2): مراحل التعنقد بين الأجناس المختلفة، يمثل العمود على اليمين المسافات بين هذه الأجناس وكل حسب أولوية التعنقد.

ومن النتائج المبينة في الجدول (6) والشكل (2) يمكن وصف عملية التعنقد بين الأجناس المختلفة من الكائنات الحية، والتي تتجمع ضمن المجاميع الآتية:

- **المجموعة الأولى:**

اشتملت هذه المجموعة المشاهدات (الأجناس) 1 و 2 و 3 و 4 و كالأتي:

* تعنقد المشاهدتين 1 و 2 في المرحلة الأولى وبمسافة مقدارها 0.02905.
* تعنقد المشاهدتين 1 و 3 في المرحلة الثانية وبمسافة مقدارها 0.04309.

- **المجموعة الثانية:**

ضمت هذه المجموعة المشاهدات (الأجناس) 5 و 6 وكالأتي:

* تعنقد المشاهدتين 5 و 6 في المرحلة الثالثة وبمسافة مقدارها 0.06908.

- **المجموعة الثالثة:**

اشتملت هذه المجموعة المشاهدات (الأجناس) 7 و 9 وكالأتي:

* تعنقد المشاهدتين 7 و 9 في المرحلة الرابعة وبمسافة مقدارها 0.09655.

- **المجموعة الرابعة:**

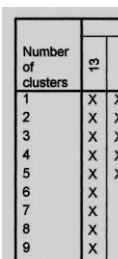
ضمت هذه المجموعة المشاهدات (الأجناس) 10 و 12 وكالأتي:

* تعنقد المشاهدتين 10 و 12 في المرحلة السادسة وبمسافة مقدارها 0.421.

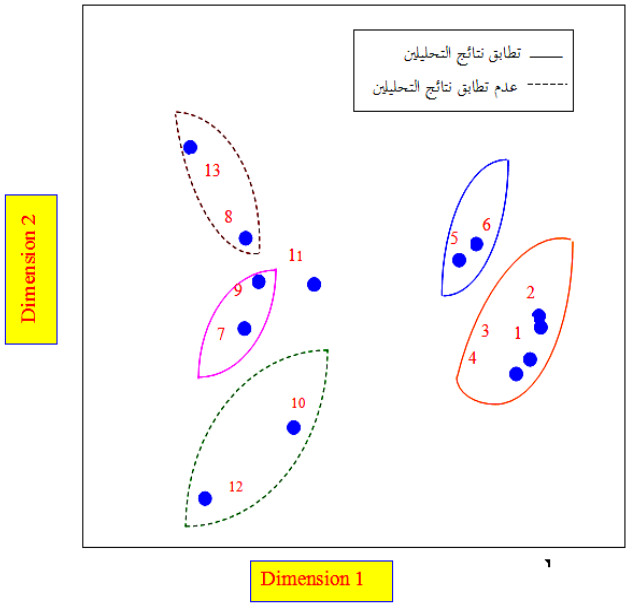
- **المجموعة الخامسة:**

ضمت هذه المجموعة المشاهدات (الأجناس) 8 و13 وكالأتي:

* تعنقد المشاهدتين 8 و 13 في المرحلة الثامنة وبمسافة مقدارها 0.531.



ومن خلال النتائج المبينة في جدول (4) (النتائج من طريقة مقياس متعدد الأبعاد) وشكل (2) وجدول (6) (الناتج من طريقة التحليل العنقودي) ينتج الأتي:



شكل (3): مواقع (13) جنس من الكائنات الحية المختلفة استنادا إلى البعدين الأول والثاني من خلال نتائج تحليل طريقتي مقياس متعدد الأبعاد والتحليل العنقودي .

4- الاستنتاجات والتوصيات

4-1- الاستنتاجات

أ- تطابق نتائج طريقتي مقياس متعدد الأبعاد والتحليل العنقودي، في تصنيف الأجناس المختلفة من الكائنات الحية المختلفة (بيانات البحث)، إلى نفس العدد من المجاميع من جهة (بين المجاميع) ومن جهة أخرى فقد تماثلت في تصنيف وتحديد العدد ذاته من المشاهدات (الأجناس المختلفة من الكائنات الحية – بيانات البحث) في داخل المجموعة الواحدة (داخل المجاميع).

ب-تمكنت طريقة مقياس متعدد الأبعاد من إعطاء تفاصيل أكثر من طريقة التحليل العنقودي، وذلك من خلال تحديد وتصنيف مواقع النماذج (الأجناس المختلفة من الكائنات الحية – بيانات الحث) على خارطة ذات بعدين. في حين صنفت طريقة التحليل العنقودي المشاهدات المختلفة (الأجناس المختلفة من الكائنات الحية) كمجاميع مختلفة دون أن تحدد مواقع هذه النماذج على الخارطة.

ج- امتازت طريقة التحليل العنقودي بدقة أكبر من طريقة مقياس متعدد الأبعاد في تحديد درجة التقارب الوراثي بين الكائنات الحية المختلفة.

4-2- التوصيات

أ- زيادة عدد السلاسل الببتيدية المتعددة ولإنزيمات مختلفة للوقوف على مدى التقارب الوراثي بين الكائنات الحية.

ب- استخدام طريقة التحليل العاملي لمقارنة نتائجها مع نتائج هذا البحث.

ج- بناء قاعدة بيانات تتضمن عدد من السلاسل الببتيدية المتعددة ولإنزيمات متنوعة للكائنات الحية المختلفة.

References

[1] Apple, F.S. and Rogers, M. A. (1986). Skeletal muscle lactate dehydrogenase isoenzyme alternation in men and women marathon runners. Journal of Applied Physiology 61,477-481.

[2] Bryan, F. J. Manly (1986), "Multivaraiate statistical method aprimer", University of Otago, Newzeland, Chapman and Hall.

[3] Drent, m. ;N.A.M. cobben, R. F. Henderson, E. F. M. wouters, M. Van Dieijen - visser (1996). Usefulness of lactate dehydrogenase and its isoenzymes as indicators of lung dumage or inflamution, Eur, Respired. J, 1996, 9, 1736-1742.

[4] Gietl, C. (1992). Biochem. Biophys. Acta 1100, 217-234.

[5] Maurice Kendall, Sc. D., F.B.A. (1975)., "MULTIVARAIATE ANALYSIS" Charles Griffin and company LTD, London and High Wycobe.

[6] Joseph F. Hair, Jr. (1987), "MULTIVARAIATE DATA ANALYSIS WITH READINGS", Macmillan Puplishing Company, New York. .

[7] Fieller,N. (2001) , Further Multivariate Analysis: Working Notes, NRJF, University of Sheffield.

[8] Sollenborn, M. (2003),Licentiate Thesis Proposal Clustering and Case-Based Reasoning for user Stereotypes, Department of Computer Science Engineering, Malardalen, University.

[9] Jansson, M. and Jimmy, J. (2003), Interactive Visualization of Statistical Data

using Multidimensional Scaling Techniques.
[10]http://www.google.com/search?hl=en&q=Bioinformatices+Master+course+Sequence+Analysis+%2213%2F12%2F04%22btnG=Google +Search

Abstract

This research paper aiming to classify and discriminate (13) different types of species, according to DNA sequences for lactate dehydrogenase enzyme, by using two analysis methods of multidimensional scaling and cluster.

The most important results are classified and discriminated species into five groups. Each group contains different numbers of these species, which are homogenous according to DNA sequences lactate dehydrogenase enzyme. These species are different from one group to another.

From our research paper we can drive the following conclusions :

A- The results of the two methods, multidimensional scaling and cluster analysis, are matching, for classifying the different types of species. It is, also, matching into the same number of the groups (between groups). The two methods, also, are corresponds in classifying and determined the same number of observation (different types of species), inside one group (inside groups).

B- Multidimensional scaling method gives more details, through classifying the observation (different types of species) on two dimensional map.

C- Cluster analysis method is more prices than multidimensional scaling method.