المجلة العراقية للعلوم الاحصائية (19) 2011
ص ص [1–20]

استخدام ثلاث طرائق الإحصائية للتعرف على عدد من التغيرات في سلسلة المادة الوراثية

باسل يونس ذنون الخياط* **فاطمة محمود حسن**

الملخص

تعد مسألة إيجاد الطفرات الوراثية من المسائل ذوات الأهمية الخاصة بعلم الوراثة لعلاقتها الوثيقة بالتطور الوراثي للكائنات الحية . في هذه الورقة البحثية نقدم ثلاث طرائق إحصائية مقترحة لغرض التعرف على الطفرات الوراثية، الأولى تعتمد أساسا على النموذج الماركوفي، والثانية على نماذج الانحدار الذاتي، وأما الثالثة فتعتمد على خوارزمية مقترحة وموظفة على نموذج ماركوف المخفي. وتطبق الطرائق الثلاثة على متتابعة الحامض النووي منقوص الأوكسجين DNA للإنسان، ويتبين من التطبيق العملي ان هذه الطرائق تعطي نتائج جيدة في الكشف عن الطفرات الوراثية.وصلت في بعض الأحيان إلى 90%.

---

* استاذ/ قسم الاحصاء والمعلوماتية/كلية علوم الحاسوب والرياضيات

تاريخ التسلم :12 /10 /2010 ـــــــــــــــ تاريخ القبول : 7/12 /2010

Human deoxyribonucleic acid (DNA) sequence

Abstract

The problem of finding genetic mutations is an issue of special importance for its relationship with genetics evolution of living organisms. In this paper we present three statistical methods proposed for the creation of genetic mutations, the first depends mainly on the Markovian model، the second on the autoregressive models, and the third depends on the algorithm proposed based on the hidden Markov model. The three methods are applied to DNA for sequence with shortage oxygen of humans, and shows the practical application show that of these methods give good results in the detection of genetic mutations, reached in some cases to 90%.

مقدمة:

الطفرة mutation هي تغير في سلسلة قواعد النيوكليتيدات للجين بحيث يؤثر في الطراز المظهري ويمكن إن تحدث الطفرة على المستوى الجزيئي بحيث يؤدي إلى استبدال قاعدة نيوكليوتيدية واحدة بدلا من أخرى، ويمكن أن تحدث الطفرة على المستوى الجزيئي بحيث يؤدي إلى استبدال قاعدة نيتروجينية واحدة بدلا من أخرى، أو إضافة عدد من القواعد أو حذفها وقد تحدث الطفرة على المستوى الكروموسومي بحيث يمكن أن يتغير جزء من الكروموسوم فيحدث تغير مفاجئ في عدد الكروموسومات أو طريقة نظامها. وينشأ من هذا التغير ظهور صفة جديدة وقد تحدث الطفرة بنوعيها (الكروموسومي، والجيني) تلقائياً نتيجة خلل في انقسام الخلية، يترتب عليه خلل في عدد الكروموسومات أو نظام ترتيبها، كما يمكن أن تحدث الطفرة نتيجة التعرض للإشعاعات أو تناول مواد كيميائية من شأنها أن تؤثر في عمل الكروموسومات أو الجينات (Watson et al. ,2003). إن وجود الطفرة يؤدي إلى التباين الوراثي ويسمح للكائنات بالتكيف مع البيئات الجديدة، وفي الوقت نفسه قد يؤدي

ازدياد معدل الطفرات إلى عدم انتظام انتقال المعلومات الوراثية بدقة من جيـل إلـى آخر. وللقيام بتحليل الجينات فإن الأطباء يبحثون عن وجود طفرة، أي بمعنـى أنهـم يتأكدون من أن تتسلسل الحامض النووي سليم وليس فيه زيادة أو نقص أو تبـديل. إن تأثير الطفرات متنوع، اذ يمكن أن توقف الطفرات صناعة البروتين توقفا تامـا أو قـد تخفض كمية إنتاج البروتين أو تزيدها كما قد تضعف وظيفته .

يمكن تقسيم الطفرات التي تحدث على المستوى الجزيئي التي تطرأ في ال DNA بحسب نوع الضرر الذي تحدثه في الحامض النووي على نوعين:

أ. الطفرة الإبداليـة substitution mutation : تتم هذه الطفرات بتبديل نيوكليوتيد معين من الشريط مما يؤدي إلى تغيير نيوكليوتيد ال mRNA المنسوخ عنه (حيـث mمختصر لmessenger)، وهو عبارة عن شريط مفرد يتكون كشريط متمم لأحد شريطي الـ DNA في النواة، وهذا التغير بدوره قد يؤدي إلى تغيير حامض أمينـي واحد من الأحماض الامينية الناتجة عن عملية تجميعها قبل تشكيل البروتين مما يغيـر خصائص هذا البروتين (Drake ,1970).

ب. طفرات قراءة الإطار frame shift mutations: تـتم هـذه الطفـرة بحـذف نيوكليوتيد من DNA أو إضافته مما يؤدي إلـى تغييـر الأحمـاض الامينيـة مـن النيوكليوتيد المحذوف أو المضاف و حتى نهاية الشريط المنسوخ. وضرر هذا النـوع من الطفرات يكمن في أنها تحدث تغييرا مستمرا في شفرة الـ RNA المنـسوخ. إن شفرة الـ RNA يتم تقسيمها على شكل ثلاثيات، كل ثلاثية معنية بحامض أميني محدد. فعند إضافة أو حذف نيوكليوتيد ستنجرف جميعها كي تعطي بروتينا مختلفا كليا (قاري وجبر،2010)

يمكن تصنيف العوامل التي تساعد على حدوث الطفـرات الوراثيـة إلـى صـنفين رئيسيين، وهما:

1- عوامل داخلية: إذ يقوم إنـزيم التـضاعف أثنـاء عمليـة التـضاعف بوضـع النيوكليوتيدات في غير موضعها الصحيح، وتنتج الطفرة عند عدم قدرة الخلايـا علـى إصلاح كل الأخطاء الناتجة عن ذلك.

2- عوامل خارجية: كالإشعاعات المختلفة وعدد من المواد الكيميائية، أو قـسم مـن أنواع الفيروسات التي تؤدي إلى إحداث تغيير في تركيب القواعد النيتروجينية لجزيء DNA.

وتكمن خطورة هذه الطفرات عند حصولها قي الجينات الموجودة على الكروموسومات، مما يؤدي إلى التأثير في عملها أو إيقاف عملها ايقافا تاما، فيسبب ذلك حدوث الإخلال في الوظائف المرتبطة بهذه الجينات وظهور عديد مـن الأمـراض،(Alberts and other ,2002).

تتكون القواعد النيتروجينية مما يأتي:

أ. بيورينات: و تشمل قاعدتين هما: أدنين A و كوانين G ، و تتألف كل منهما من حلقتين.

ب. بيرميدينات: و تشتمل على ثلاث قواعد: ثايمين T و سايتوسين C و يوراسيل U،حيث يحتوي mRNA المراسل على U بدل الـ T الموجود في DNA و يتألف كل منها من حلقة واحدة

وفي هذا البحث سوف يدرس الحامض النووي DNA والذي تتكون قواعده من A,T,C,G.

يُقترح في هذه الورقة البحثية ثلاث طرائق لأجـل معرفـة نوعيـة القاعـدة النيتروجينية المحذوفة والطرائق هي :

الطريقة الأولى (الطريقة الماركوفية) :

لنفرض أن متتابعة الحامض النووي DNA هي من مرتبة معينة، ولـتكن p. إن الخطوات الرئيسة لإيجاد القيم المفقودة في متتابعة ألـ باستخدام هذه الطريقـة يمكـن إيجازها بالنقاط الآتية :

1. تقسم المشاهدات على مجموعتين، الأولى تستخدم لتقدير مصفوفة الانتقاليـة مـن المرتبة p، للتحقق من الشرعية
2. لإيجاد المُتنبأ عند الزمن $T+1$، $\hat{x}_{T+1}$، نلاحظ الحالات المُتحققة عنـد الأزمنـة T و $T-1$ و ... و $T-p+1$، $x_T, x_{T-1},..., x_{T-p+1}$، ولنفرض أن هذه الحالات هي $i_T, i_{T-1},..., i_{T-p+1}$، إذ إن هـذه الحـالات تنتمـي إلـى رمـوز فـضاء الحالـة $S=\{1,2,\ldots,N\}$. نُثبّت الـصف المقابـل للحـالات $i_T, i_{T-1},..., i_{T-p+1}$ مـن المصفوفة الانتقالية $\underline{P}$ ونلاحظ الاحتمالات المقابلة في الأعمدة التي تتقـاطع مـع هذا الصف. فلو كان العمود j يعطي أكبـر احتماليـة مـن احتمـالات الـصف $i_T, i_{T-1},..., i_{T-p+1}$، فإننا نختار المتنبأ لكي يكون $\hat{x}_{T+1} = j$. أي إن:

$$\hat{x}_{T+1} = \max_{j=1,2,...,N} p_{i_T, i_{T-1},..., i_{T-p+1}} \qquad (12)$$

ولتطبيق هذه الطريقة على مشاهدات الحامض النووي للإنـسان الطبيعـي، استخدمت أول 14071 مشاهدة لتقدير المصفوفة الانتقالية، وباقي المـشاهدات التـي عددها 2500 استخدمت لغرض اختبار شرعية النموذج. وباعتبـار أن المتتابعـة مـن الرتبة الثالثة فان كل قاعدة نيتروجينية عند الزمن n تعتمد على القواعـد النيتروجينيـة

الثلاث التي تسبقها. ومادام فضاء الحالة يكّون المجموعة S={A,T,C,G} ، فهناك 64 إمكانية ممكنة للحالات الثلاث السابقة ، مع ملاحظة ان مجموع عدد الحالات الثلاثية الناجمة يساوي عدد المشاهدات الكلية مطروحا منه 3 (الرتبة المفروضة). إن المصفوفة الانتقالية المدرجة في الملحق تم تكونها وذلك بقسمة كل عنصر من عناصر مصفوفة التكرارات على مجموع صفه.

الطريقة الثانية (طريقة الانحدار الذاتي):

إن نموذج الانحدار الذاتي من المرتبة p autoregressive model of (order p) هو عبارة عن معادلة فرقية difference equation من المرتبة p، يرمز له عادة AR(p). فإذا كانت $y_1, y_2, ..., y_T$ هي مشاهدات من متسلسلة زمنية، فإن الصيغة الرياضية لنموذج الانحدار الذاتي من المرتبة p تكون على النحو الآتي (الخياط، 2010):

$$y_n = a_1 y_{n-1} + a_2 y_{n-2} + ... + a_p y_{n-p} + e_n;\ n = p, p+1, ... \qquad (13)$$

إذ إن $a_1, a_2, ..., a_p$ هي كميات ثابتة مجهولة (معلمات النموذج) وان e_n هو متغير عشوائي يُعرف بالضوضاء Noise. ويمكن إعادة كتابة النموذج السابق بالصيغة البديلة الآتية (وهي الصيغة المعتمدة ب(MATLAB , 2009)):

$$A(q)\, y_n = e_n, \qquad (14)$$

إذ أن $A(q)$ هي متعددة حدود Polynomial من الدرجة p ولها الشكل العام الآتي:

$$A(q) = 1 - a_1 q^{-1} - a_2 q^{-2} - ... - a_p q^{-p} \qquad (15)$$

كما أن q^{-1} هو عامل الإزاحة الخلفي backward shift operator والذي يعرّف على النحو الآتي: $q^{-1} y_n = y_{n-1}$ ، $q^{-2} y_n = y_{n-2}$ ، $q^{-k} y_n = y_{n-k}$. ولو كانت $x_1, x_2, ..., x_T$ هي مشاهدات مُتحققة من سلسلة ماركوفية $\{X_n; n = 0,1,2,...\}$ ذات

فضاء الحالة $S = \{s_1, s_2, ..., s_N\}$. ولغرض ملاءمة نموذج انحـدار ذاتـي لهـذه المشاهدات فإن الخطوة الأولى التي نقوم بها هي ترميز الحالات إلى الأعداد الصحيحة 1,2,…,N ، ولتكن المشاهدات بعد الترميز هي $y_1, y_2, ..., y_T$. ثم نستدعي الدالـة الحاسوبية ar بالهيئة ar(y,p)، إذ أن p تمثل مرتبة النموذج المطلوبة.

لو مثلنا مواقع متتابعة الحامض النووي DNA بالخلايا الموضحة بالشكل الآتي:

.	.	.	$x(n-2)$	$x(n-1)$	$x(n)$	$x(n+1)$	$x(n+2)$	.	.	.

وكانت القيمة المفقودة هي $x(n)$، عندها يمكن تقدير القيمة المفقودة على النحو الآتي: نقوم ببناء نموذج انحدار ذاتي وذلك بعد ترميز الحالات الأربع (الحوامض الأمينيـة).

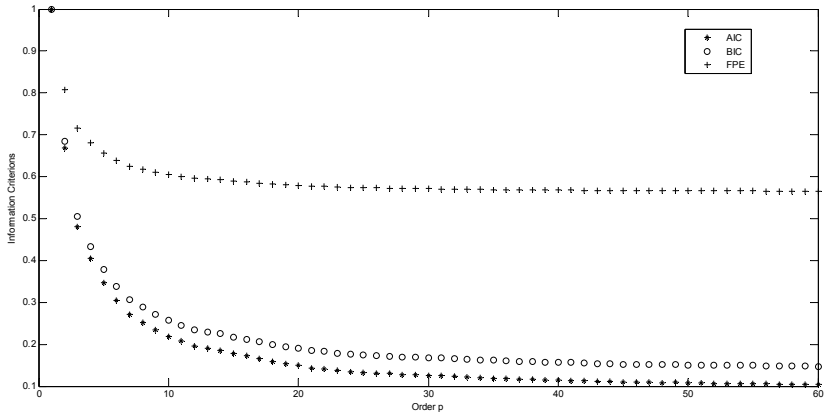
ثم نقوم بإيجاد القيمة المفقودة منه، ولتكن هذه القيمة $\hat{x}(n)$. ولغرض ملاءمة نمـوذج لإيجاد القيم المفقودة بوساطته فقد قمنا بالإجراء الآتي: أخـذنا أول 14071 مـشاهدة لغرض ملاءمة النموذج الرياضي، وتركت باقي المشاهدات التي عددها 2500 مشاهدة لغرض اختبار شرعية النموذج. وقد حاولنا استخدام معايير المعلومات للوصول إلـى المرتبة المناسبة لنموذج AR ، إلا أن المشكلة التي واجهتنا هي أن قيم هذه المعـايير تستمر بالهبوط دون الوصول إلى مرتبة معقولة. والشكل الآتي يبين قيم المعايير الثلاثة FPE و AIC و BIC ، بعد تحويلها إلى قيم معيارية (بقسمة قيم المعيار على أكبر قيمة فيه، تكون اكبر قيمة للمعيار هي 1).

حيث أن

AIC ≡ Akaike Information Criterion

FPE≡ Final Prediction Error Approach

BIC ≡ Bayesian Information Criterion



الشكل(1): القيم المعيارية لمعايير المعلومات الثلاثة FPE و AIC و BIC.

ومن خلال التفحص لقيم معايير المعلومات وقيم دالة الخسارة وجد أن المرتبة الثالثة مقبولة ، لذا فقد تمت ملاءمة نموذج AR(3)، وهذا النموذج يجهزنا بنموذج التنبؤ الآتي:

$$y(n)= 0.3195\ y(n-1) + 0.2805\ y(n-2) + 0.3383\ y(n-3) \quad (16)$$

إذ أن Loss function= 1.35167 و FPE= 1.35225.

مقارنة بين الطريقتين الأولى والثانية:

لغرض المقارنة بين نتائج الطريقتين المقترحتين لإيجاد القيم المفقودة في متتابعة DNA تم اختيار مواقع عشوائية في شريط الحامض النووي للإنسان الطبيعي، وعدّت قواعدها النيتروجينية مفقودة. ثم طُبقت الطريقتان المذكورتين انفاً لتقدير القواعد النيتروجينية المفقودة. والجدول الآتي يلخص نتائج التقدير.

الجدول (1): تقدير القواعد النيتروجينية المفقودة باستخدام الطريقتين المقترحتين.

الموقع	الرمز الحقيقي y(n)	الرمز المتنبأ به بالطريقة الماركوفية	الرمز المتنبأ به بطريقة الانحدار الذاتي
8	G	G	A
57	T	T	T
129	T	T	T
235	A	A	A
308	C	C	C
484	T	T	T
528	G	C	T
820	C	C	C
987	T	C	T
1001	C	C	A
1063	A	A	T
1184	C	C	C
1374	C	A	T
1490	C	C	C
1751	C	G	C
1991	C	G	C
2076	A	A	A
2121	T	T	T
2427	A	A	T
2467	T	T	T

ومن الجدول الأخير يتبين أن نسبة المطابقة بين القواعد النيتروجينية الحقيقية والمقدّرة بكل من الطريقتين المقترحتين هي على النحو الآتي:

نسبة المطابقة بين القواعد النيتروجينية الحقيقية والمتنبأ بها بالطريقة الماركوفية هي 75%.

نسبة المطابقة بين القواعد النيتروجينية الحقيقية والمتنبأ بها بطريقة الانحدار الذاتي هي 70%.

ومن النتائج الأخيرة يتبين أن كلتا الطريقتين تقدمان نتائج مقبولة من الناحية العملية.

الطريقة الثالثة (نموذج ماركوف المخفي):

نماذج ماركوف المخفية هي عبارة عن مجموعة منتهية من الحالات كل حالة تقترن بتوزيع احتمالي ، أما الانتقالات (transitions) ما بين الحالات(states) فتُحدد بوساطة مجموعة من الاحتمالات يُطلق عليها احتمالات الانتقال والتي يُمكن أن تنتج نتيجة أو مشاهدة بشكل عام ، تتولد الحالة الناتجة (أي المشاهدة) طبقاً لتوزيع الاحتمالية المقترنة ، حيث توجد احتمالية ناتجة فقط ولا توجد حالة ظاهرة يمكن أن تُشاهد ، ولهذا فإن الحالات تكون مخفية (Aazami ,2008). وفي الآونة الأخـيرة حازت نماذج ماركوف المخفية على اهتمام كبيـر فـي مجتمع المعلوماتية الحيوية، إذ إنها تُستخدم في حل المسائل الخاصة بتنبؤ التركيب الثانوي البروتيني (Protein Secondary Structure rediction)، وإيجاد الجينات، وتصنيف عائلة البروتين وتشكيل خرائط ارتباط الجينات لعلم الوراثة وخرائط فيزياوية (Thomsen ,2001). بشكل عام فان الفرق بين نماذج ماركوف الاعتيادية ونماذج ماركوف المخفية هو أن نماذج ماركوف الاعتيادية (MMs) تكون الحالات فيها ملاحظة بشكل مباشر، ولذا تكون احتمالات الانتقال من حالة إلى أخرى هي كل ما نحتاج إلى معرفته لتحديد النموذج. أما في حالة نماذج ماركوف المخفية فلا يمكننا الاطلاع على الحالات بشكل

مباشر لكن عوضاً عن ذلك تكون لدينا مجموعة من المشاهدات مع احتمال توليد كل مشاهدة من مختلف حالات النموذج ، عادة ما تتوفر لدينا احتمالات الانتقال والمشاهدة فضلاً عن متسلسلة المشاهدات لما تساعدنا في تقدير متسلسلة الحالات التي أدت إلى ذلك السلوك (Fonzo et al ،2007). ومن الجدير بالذكر أن (HMMs) يجب أن تحدد مدخلاته المتمثلة بالمعلمات الضمنية وهما (N, M) ومتسلسلة المشاهدات (O) وإعطاء قياسات معلماته $\lambda = (A, B, \pi)$.

أخذت متتابعة الحامض النووي DNA للمايتوكندريا للإنسان الطبيعي والتي تتكون من 16571 قاعدة نيتروجينية، واستخدمت خوارزمية Viterbi لحساب نوعية القاعدة النيتروجينية المحذوفة. وقد تم توظيف نموذج ماركوف المخفي لإيجاد الطفرات الوراثية من خلال الخوارزمية المقترحة الآتية:

الخوارزمية (1): (تقدير القاعدة النيتروجينية المحذوفة لمتتابعة الحامض النووي)

الخطوة (1): تهيئة بيانات الإدخال X والتي هي عبارة عن متتابعة الحامض النووي الرايبي منقوص الأوكسجين DNA .

الخطوة (2): ترميز القواعد النيتروجينية الأربع التي بتتابعاتها المختلفة تشكل متتابعة الحامض النووي كما يأتي حيث إن $A \equiv 1$ و $T \equiv 2$ و $C \equiv 3$ و $G \equiv 4$.

الخطوة (3): تتكون هذه الخطوة من خمس مراحل، وعلى النحو الأتي:

المرحلة الأولى: تحذف نسب مختلفة من كل نوع من أنواع القواعـد النيتروجينيـة الأربع والتي هي A,T,C,G ويوضع بدلا منها الحرف M للدلالة على الطفرة (M هو أول حرف من كلمة Mutation).

المرحلة الثانيـة: حذف جميع القاعدة النيتروجينية A أو بعضها ووضع الحـرف M بدلاً منها.

المرحلة الثالثة: حذف جميع القاعدة النيتروجينية T أو بعضها ووضع الحرف M بدلاً منها.

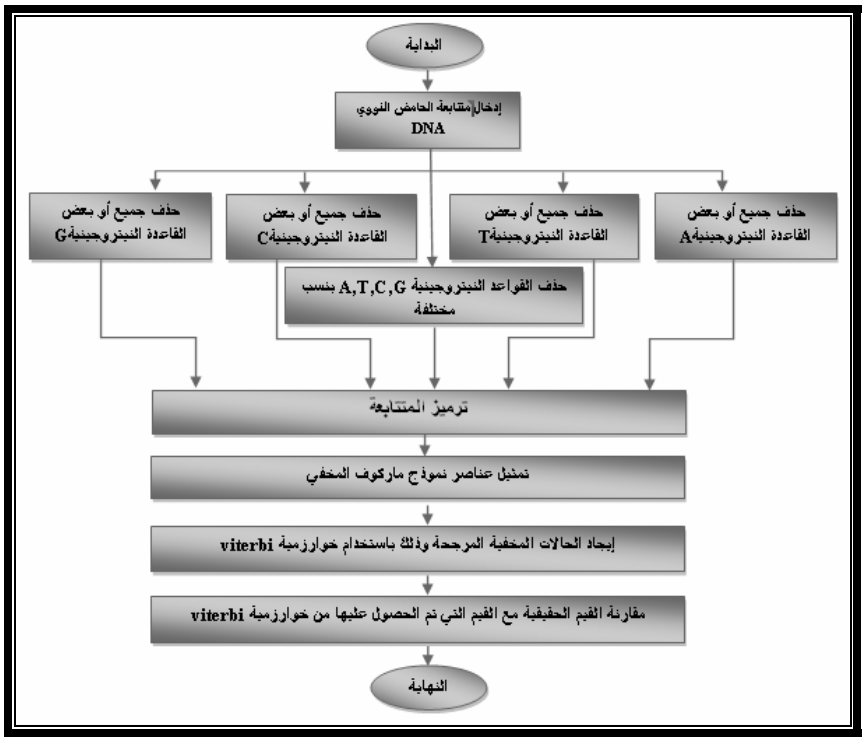
المرحلة الرابعة: حذف جميع القاعدة النيتروجينية C أو بعضها ووضع الحـرف M بدلاً منها.

المرحلة الخامسة: حذف جميع القاعدة النيتروجينية G أو بعضها ووضع الحرف M بدلاً منها.

الخطوة (4): تمثيل عناصر نموذج ماركوف المخفي (λ=(A ,B, π) حيث إن π تمثل متجه الحالة الابتدائية والذي أبعاده 1*N، N=4. أما A فتمثل مصفوفة الاحتمـالات الانتقالية بين الحالات المخفية والتي تكون أبعادها بشكل عام (N*N). أما B فتمثـل مصفوفة التوزيع الاحتمالي لرموز المشاهدات والتي أبعادها (N*M)، M=5.

الخطوة (5): استخدام خوارزمية viterbi لايجاد الحالات المخفية المرجحة.

الخطوة (6): تُقارن متتابعة الحالات الناتجة من الخطوة (5) مـع متتابعـة الحـالات الحقيقية. وبملاحظة مواقع القيم المفقودة يتم تقدير القواعد النيتروجينية المفقـودة بمـا يقابلها بمتتابعة الحالات الناتجة من الخطوة (5).



الشكل (2) : المخطط الانسيابي للخوارزمية المقترحة لتحديد نوعية القاعدة النيتروجينية المحذوفة.

نتائج التطبيق:

1- طبقت الخوارزمية (1) المقترحة على متتابعة الحامض النووي للإنسان الطبيعي التي طولها 16571، حيث أن أعداد القواعد النيتروجينية في هذه المتتابعة هي كالآتي: A=5113، T=4086، C=5192، G=2180. وقد حذفت نسب مختلفة من كل نوع من أنواع القواعد النيتروجينية الأربع ووضع بدلا منها الحرف M للدلالة على الطفرة. وبعد قراءتها ومقارنة متتابعة الحامض النووي الأصلية التي لا تحتوي على طفرات (قيم مفقودة) مع متتابعة الحامض

النووي التي تم إجراء تطفير عشوائي عليها ثم معالجتها باستخدام خوارزمية Viterbi فأعطت النتائج الآتية:

Transition Matrix=

$$\begin{pmatrix} 0.3118 & 0.2392 & 0.2924 & 0.1567 \\ 0.3348 & 0.2457 & 0.2942 & 0.1253 \\ 0.2958 & 0.2770 & 0.3426 & 0.0846 \\ 0.2822 & 0.1932 & 0.3286 & 0.1960 \end{pmatrix}.$$

Emission Matrix =

$$\begin{pmatrix} 0.9679 & 0 & 0 & 0 & 0.0321 \\ 0 & 0 & 0 & 0 & 1.0000 \\ 0 & 0 & 0.9682 & 0 & 0.0318 \\ 0 & 0 & 0 & 0.9606 & 0.0394 \end{pmatrix}.$$

mse_seq = 0.0406

sum_mutation = 4501

error = 415

Successful = 90.7798

2- حذفت كل من القواعد النيتروجينية الأربع A, T, C, G جزئياً وكلياً من متتابعة الحامض النووي، ووضع بدلها الحرف M. وبعد قراءتها ومعالجتها باستخدام خوارزمية Viterbi لإيجاد المشاهدات الأكثر احتمالا ثم مقارنة الأخيرة مع المشاهدات الأصلية تم الحصول على النتائج التي يلخصها الجدولان الآتيان.

الجدول (2-a): النسب المئوية للنجاحات في التعرّف على التغيرات في متتابعة الحامض النووي عند حذف قواعد نيتروجينية جزئياً.

نوع القاعدة النيتروجينية المحذوفة	مجموع القواعد المحذوفة	عدد الأخطاء	النسبة المئوية للنجاح (%)
A	65	2	96.9
T	73	2	97.3
C	57	2	96.5
G	77	1	98.7

الجدول (2-b): النسب المئوية للنجاحات في التعرّف على التغيرات في متتابعة الحامض النووي عند حذف قواعد نيتروجينية كلياً.

نوع القاعدة النيتروجينية المحذوفة	مجموع القواعد المحذوفة	عدد الأخطاء	النسبة المئوية للنجاح(%)
A	5113	0	100%
T	4086	0	100%
C	5192	0	100%
G	2180	0	100%

ومن النتائج الأخيرة تتضح الكفاءة الكبيرة التي يتمتع بها نموذج ماركوف المخفي في التعرّف على الطفرات الوراثية.

الاستنتاجات:

طرحت في هذه الورقة البحثية ثلاث طرائق إحصائية للتعـرف علـى عـدد التغيرات في سلسلة المادة الوراثية ،وعولجت هذه المسألة باعتبارها مسألة "تقدير قـيم مفقودة". الطريقة الاولى تعتمد على نمذجة متتابعة الحامض النووي باعتبـاره سلـسلة ماركوفية ذات أربع حالات، ومن المصفوفة الانتقالية المقدرة يمكـن التعـرف علـى الطفرات الوراثية. والطريقة الثانية تعتمد على نمذجة متتابعة الحامض النـووي وفـق نماذج الانحدار الذاتي للمتسلسلات الزمنية. في هذه الطريقة وجد ان المشكلة الاساسية تكمن في عدم امكانية الحصول على مرتبة معقولة لمتتابعة الحامض النووي باسـتخدام معايير المعلومات الشائعة. ومع ذلك فان التعرّف على الطفرات الوراثيـة بوسـاطة نموذج انحدار ذاتي ذي مرتبة مفروضة استطاع من تقدير الطفرات الوراثية وبنـسبة نجاح مقبولة. واما الطريقة الثالثة فهي تعتمد على خوارزمية مقترحة وظـف نمـوذج ماركوف المخفي لإيجاد الطفرات الوراثية. هذه الطريقة الاخيرة استطاعت من ايجـاد الطفرات الوراثية وبنسب نجاح كبيرة وصلت إلى 90% .

المصادر

1. الخياط ، باسل يونس ذنون (2010). **"النمذجة الماركوفية مع تطبيقات عملية"**، دار الكتب للطباعة والنشر، الموصل.

2. قاري، سمير بن حسن محمد و جبر، جميل فوزي جميل (2010). **" مدخل إلى الوراثة البشرية"** دار الفكر للطباعة والنشر , مكة المكرمة .

3. Aazami, Farshideh Einsele ،2008، **"Recognition of ultra low resolution, Anti-aliased text with small font sizes"**, Unpublished Ph.d. thesis, Scientarium informaticarum, Faculty of science, University of Fribourg, Switzerland.

4. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter ,P. (2002). "**Molecular Biology of the Cell**", 4th Ed , Garland Science, New York, USA

5. Drake, J.W. (1970).” **Molecular Basis of Mutation”**, Holdeh – Day.San Frameisco.

6. Fonzo, V. , Aluffi-Pentini, F., and Parisi, V،2007،" **Hidden Markov Models in Bioinformatics"**, Vol. 2, No. 1, Euro. Bio. Park, University di Roma, Roma, Italy.

7. Thomsen, R. (2001). **“ Evolving the Topology of Hidden Markov Models Using Evolutionary Algorithms ”** < thomsen @ daimi.au.dk>

8.Watson, James D and Berry, Andrew (2003). ” **DNA : the Secret of Life ‘’,** Alfred A. Knopf, New York.

الملحق

المصفوفة الانتقالية للنموذج الماركوفي ذي المرتبة الثالثة لمشاهدات الحامض النووي للإنسان الطبيعي .

إلى / من	A	T	C	G
AAA	0.2989	0.2107	0.3448	0.1456
AAT	0.3333	0.2312	0.3118	0.1237
AAC	0.3422	0.2281	0.3727	0.0570
AAG	0.2679	0.1914	0.3589	0.1818
ATA	0.3151	0.2247	0.2438	0.2164
ATT	0.3505	0.2810	0.2598	0.1088
ATC	0.3896	0.2125	0.2725	0.1253
ATG	0.1698	0.4151	0.2201	0.1950
ACA	0.1592	0.2377	0.2758	0.3274
ACT	0.4015	0.2044	0.2968	0.0973
ACC	0.2799	0.2529	0.3861	0.0811
ACG	0.3417	0.1750	0.3167	0.1667
AGA	0.3708	0.1292	0.3034	0.1966
AGT	0.3988	0.2209	0.2454	0.1350
AGC	0.3158	0.2246	0.4105	0.0491
AGG	0.3371	0.1657	0.3657	0.1314
TAA	0.3130	0.2934	0.3007	0.0929
TAT	0.2539	0.3313	0.2724	0.1424
TAC	0.2334	0.3660	0.3024	0.0981
TAG	0.2046	0.1815	0.3707	0.2432
TTA	0.3018	0.2622	0.2988	0.1372

TTT	0.2817	0.2500	0.3333	0.1349
TTC	0.2968	0.3355	0.2452	0.1226
TTG	0.2982	0.2544	0.2018	0.2456
TCA	0.3086	0.2967	0.3014	0.0933
TCT	0.2754	0.2656	0.3082	0.1508
TCC	0.2318	0.3883	0.2989	0.0810
TCG	0.2397	0.1570	0.3884	0.2149
TGA	0.2857	0.2328	0.2910	0.1905
TGT	0.4000	0.2400	0.2400	0.1200
TGC	0.3033	0.2623	0.3770	0.0574
TGG	0.2000	0.2100	0.4100	0.1800
CAA	0.3147	0.2263	0.3341	0.1250
CAT	0.3029	0.2476	0.3341	0.1154
CAC	0.3194	0.2467	0.3612	0.0727
CAG	0.2079	0.2475	0.3416	0.2030
CTA	0.2893	0.2950	0.2280	0.1877
CTT	0.3249	0.2114	0.3817	0.0820
CTC	0.3612	0.2129	0.3565	0.0694
CTG	0.3867	0.1934	0.2597	0.1602
CCA	0.2959	0.2808	0.3348	0.0886
CCT	0.3856	0.1845	0.3007	0.1292
CCC	0.2579	0.2975	0.3703	0.0744
CCG	0.2606	0.1972	0.3803	0.1620
CGA	0.3790	0.1855	0.2419	0.1935
CGT	0.3846	0.2692	0.2436	0.1026
CGC	0.3185	0.2739	0.3694	0.0382
CGG	0.3250	0.1625	0.3125	0.2000
GAA	0.1859	0.3417	0.2864	0.1859
GAT	0.2946	0.3125	0.2143	0.1786
GAC	0.2601	0.2832	0.3295	0.1272
GAG	0.2061	0.1985	0.3435	0.2519
GTA	0.2876	0.2353	0.2353	0.2418
GTT	0.2404	0.2788	0.3077	0.1731

GTC	0.2991	0.3178	0.3084	0.0748
GTG	0.2982	0.1579	0.3333	0.2105
GCA	0.2488	0.2679	0.2392	0.2440
GCT	0.3556	0.2889	0.2167	0.1389
GCC	0.2657	0.3100	0.3358	0.0886
GCG	0.3036	0.1786	0.3214	0.1964
GGA	0.2602	0.1707	0.2927	0.2764
GGT	0.2250	0.2875	0.3000	0.1875
GGC	0.2105	0.2697	0.3289	0.1908
GGG	0.2500	0.2361	0.3056	0.2083