The Scholar Journal for Sciences & Technology sjst

ISSN (Print) 2794-7629 Received 07/11/2022
ISSN (Online) 2794- 4549 Accepted 20/12/2022

FULL PAPER

Store Data in DNA

Prepared by

.Aouragh Karim
Cyber Security Management
Charisma University – UK
Aouragh.karim@yahoo.com

Abstract

Magnetic or optical data storage systems that currently contain this volume of digital zeros and ones will not last more than a century at most. Moreover, running data centers consumes huge amounts of energy. In short, we are on the verge of a serious data storage problem, and it will only get worse over time. To solve this problem, we present in this article a better method represented in the technology of storing information and data in DNA known as "DNA", which stores genetic traits in each cell, as this method will become in the future a new way of storing information, given that the field of nuclear DNA It is the way organisms store and maintain genetic information over hundreds of years.

Key words: DNA, magnetic or optical storage systems, data centers, genetic information.

المستخلص:

لن تتمكن أنظمة التخزين المغناطيسية أو الضوئية للبيانات والتي تحتوي حاليًا على هذا الحجم من إشارات الصفر والواحد الرقمية من الاستمرار لأكثر من قرن على أقصى تقدير. وفوق ذلك، فإن تشغيل مراكز البيانات يستهلك كميات هائلة من الطاقة. باختصار، نحن على وشك مواجهة مشكلة خطيرة في تخزين البيانات، وستزداد حدةً بمرور الوقت. ولحل هذه المشكلة، نقدم في هذه المقالة وسيلة أفضل تتمثل في تكنولوجيا تخزين المعلومات، والبيانات في الحمض النووي المعروف اختصارا باسم "دي إن إيه" والذي يخزّن الصفات الوراثية في كل خلية، حيث ستصبح هذه الوسيلة مستقبلاً مخزن المعلومات الجديد باعتبار أن حقل الحمض النووي هو الوسيلة التي تخزّن بها الكائنات المعلومات الوراثية وتحتفظ بها على مر مئات السنين.

الكلمات المفتاحية: الحمض النووي ، أنظمة التخزين المغناطيسية أو الضوئية ، مراكز البيانات ، المعلومات الوراثية ..

1- مقدمة:

كل يوم يزداد حجم البيانات التي نقوم بإنتاجها، سواء أكانت صورا أم أصواتا وموسيقى أو مقاطع فيديو نرفعها إلى شبكة الانترنت، أو ربما تكون بيانات تحتوي نسخا احتياطية، لكنها جميعا تحتاج بالضرورة إلى وسائط لتخزينها، وهذه الوسائط في طريقها إلى النفاد.

بحسب معهد ماساتشوستس للتكنولوجيا MIT فإن الإنتاج الحالي للبيانات يبلغ زيتابايت واحد (مليون تيرابايت – أو مليار غيغابايت) كل سنة، وفي سنة 2025 سيصبح هذا الرقم متاخما لـ 160 زيتابايت في السنة، ولا يمكن للتقنيات الموجودة بين أيدينا حاليا أن توفر مساحة تخزين كافية لاستيعاب هذا الحجم الهائل من البيانات.

ومع التزايد الرهيب في حجم هذه البيانات تكبر التحديات الخاصة بحلول التخزين، ورغم أن العلماء حاولوا خلال السنوات الماضية إيجاد بديل لأنظمة التخزين المغناطيسية القديمة). (http://www.idc.com/downloads/، عبر استخدام أنظمة تخزين ضوئية للبيانات، فإن الاستهلاك الكبير في الطاقة يجعل إمكانية نجاح هذه الفكرة محل شك، فالخلايا الحيّة تستعمل سلاسل الحمض النووي DNA لتشفير كمية كبيرة جدا من المعلومات ونقلها من جيل لآخر. وعلى سبيل المقارنة يمكن لخيط دي إن إيه واحد أن يتسع لما يتسع له قرصان مدمجان CD، وغرام واحد فقط منها قادر على حفظ زيتابايت بكامله، أي إجمالي إنتاج العالم السنوي الحالي من البيانات، وهوما يجعله أسلوبًا جديدًا للتخزين ومكانًا آمنًا لحفظ البيانات التي تتضخم وتزداد يومًا بعد يوم (C. T. . Clelland, V. Risca, and C. Bancroft,1999 : 399:533–534)
(G. M. Church, Y. Gao, and S. Kosuri, 2012: 337)

(N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust,B. Sipos, and E. Birney, 2013

2- مشكلة البحث:

مع استحالة الاعتماد الكلي على أساليب التخزين التقليدية مع التزايد الكبير في حجم المعلومات، ومن المعروف وبيولوجيًّا أن ال DNA هو المكان الذي يُستخدَم لحفظ البيانات والمعلومات في حياتنا، لكن هل يمكن استخدامه لتخزين البيانات الرقمية.

3- فرضيات البحث:

1. استخدام ال DNA في تخزين البيانات الرقمية.
2. استبدال أنظمة التخزين المغناطيسية و الضوئية بشريط الحمض النووي.

4- أهداف البحث:

الآن و في الوقت الحاضر يضاف إلى كل يوم أكثر من 2,5 مليون جيجابايت إلى وسائط تخزين الملفات العالمية وهدا الكم الهائل من الملفات سيشكل لنا اكبر مشكل، وحدات التخزين العالمية و محركات الأقراص

للبيانات بشكل مماثل للطرق التي صنعها الإنسان، مثلما نجحت من قبل في نقل البيانات الوراثية، تقترح أيضًا مخطط ترميز جديد التي توفر إمكانية التحكم في التكرار.

- دراسة (G. M. Church, Y. Gao, and S. Kosuri. 2012)قامت بتطويرِ طريقةٍ لتشفير المعلومات الرقمية تعتمدُ على الحمض النووي DNA ، وكتابة كتابًا بحجم 5.27 ميغابت باستخدام شرائح الحمض النووي الدقيقة، وقراءة الكتاب باستخدام تسلسل الحمض النووي.

7- تخزين البيانات على الحمض النووي:

مبدئيًا يجب أن نعرف بشكل مبسط كيف يتم تخزين البيانات على الحاسب الآلي أو وسائط التخزين الإلكترونية

1.7- طريقة تخزين البيانات على وسائط التخزين الإلكترونية:

إنّ جميع البيانات المخزّنة على وسائط التخزين سواءً كانت محركات الأقراص الصلبة أو الأقراص القابلة للإزالة USB أو بطاقات SD وغيرها، تُخزَّن على شكل سلسلةٍ من البتات، والمعروفة باسم الأرقام الثنائية. هذه الأرقام الثنائية تمتلك القيمة 1 أو 0. ويجب تحويل جميع أنواع البيانات مثل الصور والمستندات والصوت والفيديو إلى سلاسلٍ ثنائيةٍ حتى نتمكن من تخزّينها.

البت هو أصغر عنصرٍ في البيانات، وهو يمثل الأرقام الثنائية. أما البايت هو تجميعٌ أساسيٌّ للبتات التي يعمل بها الكمبيوتر كوحدةٍ واحدةٍ. عادةً، يتكون البايت من ثمان بتات. وبالتالي، فإن سعة ذاكرة

المتطورة لا تستطيع احتواء كل هذا الكم الهائل من البيانات المتزايدة في المستقبل. لذلك يهدف البحث إلى تطوير هذه الوحدات إلى تقنية تخزين بيانات في الحمض النووي DNA لمواكبة أحجام البيانات الضخمة .

5- أهمية البحث:

تعتبر الأهمية العلمية للدراسة كون نحن بحاجة إلى حلول جديدة لتخزين هذه الكميات الهائلة من البيانات التي يقوم العالم بإنتاجها وتجميعها، وخاصة البيانات الأرشيفية، والحمض النووي الرايبوزي أكثر كثافة بألف مرة من ذاكرة الفلاش، وهناك خاصية أخرى مثيرة وهي أنه بمجرد أن تصنع بوليمر الحمض النووي، فإنه لا يستهلك أي طاقة، ويمكنك كتابة الحمض النووي ثم تخزينه إلى الأبد .

6- الإطار النظري والدراسات السابقة:

1.6-الإطار النظري:

كما أوردنا سابقاً ومع التزايد الرهيب في حجم هذه البيانات، والتي جميعها تحتاج بالضرورة إلى وسائط لتخزينها ،وهذه الوسائط في طريقها إلى النفاد. لذا يعتقد العديد من العلماء أن الحل البديل يكمن في الجزيء الذي يحتوي على معلوماتنا الجينية، أي الحمض النووي (DNA)، والذي يمكن تطويره لتخزين كميات هائلة من المعلومات بكثافة عالية جدا و لكن باعتماد طرائق مختلفة و ميزات مختلفة.

2.6- الدراسات المرجعية:

- دراسة (Bornholt et al, 2016) تقدم بنية للحمض النووي DNA القائم على تخزين المعلومات الوراثية و إمكانيات استخدام هذه التقنية لتخزين ونقلها

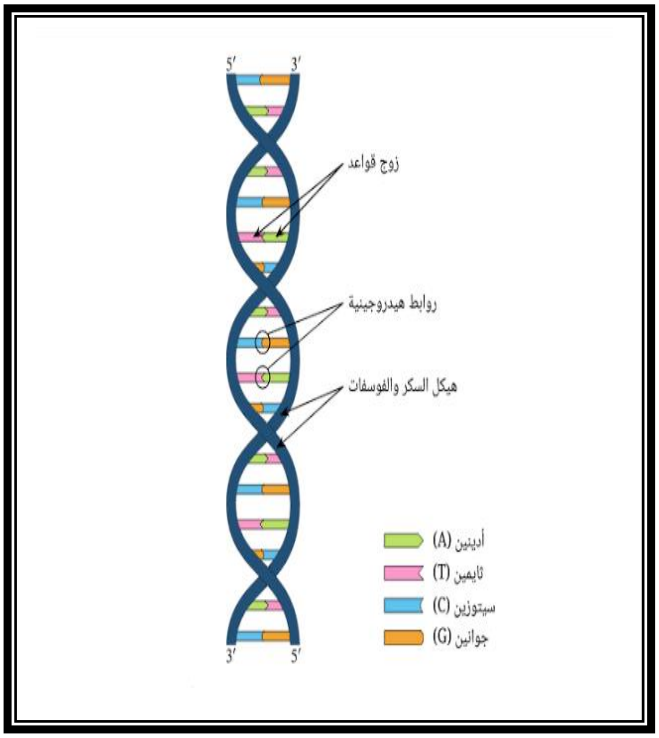
تخزن الوسائط التقليدية مثل محركات الأقراص الصلبة أو أقراص DVD البيانات الرقمية عن طريق تغيير

الخصائص المغناطيسية والكهربائية أو الضوئية لمادة لتخزين 0 و1 (النظام الثنائي)

(http://www.idc.com/downloads/)

لتخزين البيانات في الحمض النووي، فإن المفهوم هو نفسه، ولكن العملية مختلفة. جزيئات الحمض النووي هي سلاسل طويلة من جزيئات أصغر، تسمى النيوكليوتيدات ucleotidesn (مركب يدخل في بناء الأحماض النووية) ويتكون من :

- الأدينين Adenine
- السيتوزين Cytosine
- الثايمين Thymine
- الجرانين Guanine



شكل 1: نموذج للحمض النووي DNA.

وهي القواعد النيتروجينية الأربعة المكونة للأحماض

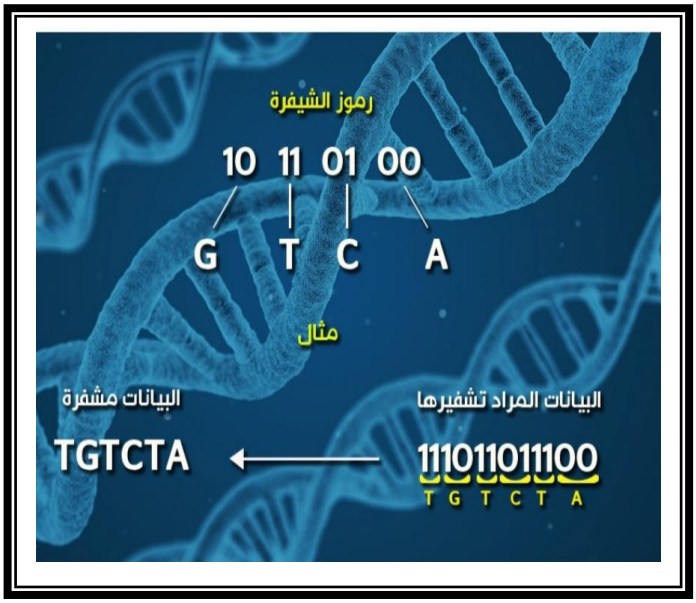
الكمبيوتر وأجهزة التخزين الثانوية يتم التعبير عنها من حيث البايتات.

يخزن الكمبيوتر البيانات كأرقام؛ فالحروف تصبح أرقامًا. على سبيل المثال لتمثيل الأعمال الكاملة لشكسبير التي هي حوالي 1250 صفحة مطبوعة، نحتاج إلى 40 مليون بت، مع بايت واحد لكل حرفٍ، أي بإجمالي خمسة ميغابايت (5 ميغابايت) لتخزينها، وأيضًا الصور الفوتوغرافية تُحوّل إلى مجموعةٍ كبيرةٍ من الأرقام التي تشير إلى لون وسطوع كل بكسل، والأرقام العشرية تحوّل إلى أرقام ثنائية. فالأرقام من 0 إلى 8 تبدو مثل هذه الأرقام الثنائية: 0000، 0001، 0010، 0011، 0100، 0101، 0110، 0111، 1000. ومع أن الأرقام الثنائية طويلةً جدًا، ولكنها تمكننا من تخزين أي قيمةٍ كسلسلةٍ من العناصر التي تكون صحيحة (1) أو خاطئة (0).

يخزن الكمبيوتر البيانات عن طريق وجود أو عدم وجود إشارات إلكترونيّة أو مغناطيسيّة في دارات الكمبيوتر أو في وسائط التخزين ويدعى هذا التمثيل ثنائي الحالة، لأن الكمبيوتر ووسائط التخزين يستطيعان تميز حالتين محتملتين فقط ، إما on أو off.

2.7- طريقة تخزين البيانات على الحمض النووي:

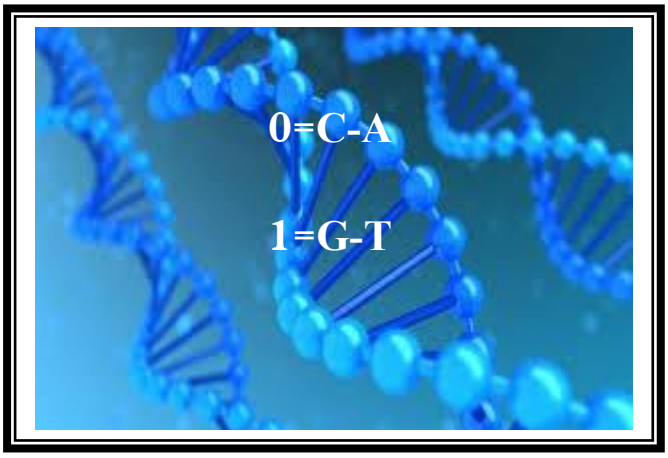
الحمض النووي DNA خلق لتخزين المعلومات الجينية ، والمخططات لبناء البروتينات والاحتفاظ بها , و يتم ترميز كل ميزة ووظيفة موجودة في التنوع الواسع للحياة على الأرض في سلسلة من الأحماض النووية التي توفر تعليمات للخلية.



شكل 2: الطريقة الأولى لتشفير البيانات على الحمض النووي DNA.

- **الطريقة الثانية:** يمكن أن تكون كالتالي:

يرمز الحرفان C–A للصفر وG–T للواحد، وبذلك يصبح الحمض النووي كأي مصفوفة مكونة من أصفار ووحدات مثل أي ملف آخر، وبدلًا من أن يكون ملف التشفير في الحاسوب 10011101110 سوف يكون في الحمض النووي GAAGTGAGGTC، بنفس آلية العمل وكفاءة القدرة على تخزين الكم الهائل من المعلومات.



النووية وعادة ما يتم كتابتها كـ A و C وT و G وتحتوي كل خلية بشرية نموذجية على جينوم يتكون من حوالي 6 مليار زوج قاعدي للحمض النووي المزدوج الحلزوني ، ينقسم إلى 23 مجموعة من الكروموسومات (3 مليار زوج من الحمض النووي تقابل الكروموسومات في كل نصف المجموعة). يشفر الحمض النووي في هذه الكروموسومات حوالي 1.6 جيجابايت من المعلومات.

وبدلاً من إنشاء تسلسلات من 0 و 1 قيمتين فقط متوفرة مع البيانات الثنائية (واحد أو صفر) ، كما هو الحال في الوسائط الإلكترونية ، يستخدم الحمض النووي تسلسل النوكليوتيدات للتخزين يأخذ إحدى القيم الأربع .

هناك عدة طرق للقيام بذلك، ولكن الفكرة العامة هي تخصيص أنماط البيانات الرقمية لنواة النيوكليوتيدات.

- **الطريقة الأولى:**

على سبيل المثال، يمكن أن يكون 00 مساوياً لـ A و 01 مساوياً لـ C و 10 مساوياً لـ G و 11 مساوياً لـ T لتخزين صورة ، على سبيل المثال ، نبدأ بتشفيرها كملف رقمي ، مثل JPEG. هذا الملف في جوهره هو سلسلة طويلة من 0 و 1 لنفترض أن البتات bits الثمانية الأولى من الملف هي 111011011100 ؛ يمكن كسرها في أزواج – 01 11 10 00 – التي تتوافق مع C– T– A–G. هذا هو الترتيب المناسب للتعامل مع النيوكليوتيدات لتشكيل سلسلة الحمض النووي.

بعد العثور على حفرية له.

يوجد الحمض النووي على شكل سلاسل يمكن قراءتها وكتابتها ونسخها بدقة وبسهولة. إحدى الميزات الأساسية للحمض النووي هي أنه مخزن كثيف للمعلومات إلى حد لا مثيل له من قبل أي وسيط تخزين آخر من صنع الإنسان، فيمكن ضغط 181 زيتابايتا من البيانات المتوقع أن ينتجها البشر بحلول عام 2025 في حجم مماثل لكرة تنس الطاولة باستخدام تقنية تخزين الحمض النووي. ويعتقد أن تخزين هذا القدر الكبير من المعلومات في الحمض النووي يمكن أن يكون على بعد عقد من الزمن فقط.

10- التوصيات:

تتمثل أبرز التحديات الحالية التي تعوق التوسع في تخزين المعلومات في الحمض النووي في التكاليف المرتفعة وسرعة الكتابة والقراءة، والتي تحتاج إلى أن تنخفض كثيرا، كي تنافس بقوة طرق التخزين التقليدية بالوسائط الإلكترونية.

وبالرغم من أن عملية حفظ البيانات في الحمض النووي لحفريات على سبيل المثال، تعد عملية مكلفة تتجاوز ألف دولار لكل ميغابايت من البيانات، وبناءً على النتائج التي تم التوصل إليها توصي الدراسة العلماء والباحثين على تطوير طرق وأدوات التشفير والقراءة واستعادة البيانات من الحمض النووي، و التي تساعد على تخفيض التكاليف بالتدريج.

شكل 3: الطريقة الثانية لتشفير البيانات على الحمض النووي DNA.

3.7 إعادة قراءة البيانات:

تتم هذه العملية عن طريق استخدام آلة تسلسل sequencing machine مثل التي تستخدم فى تحليل الحمض النووي الجيني فى الخلايا لتحديد البصمة الوراثية، حيث يتم تحليل كل جزئ إلى تسلسل من الحروف ثم بعد ذلك يتم فك تشفير الحروف طبقاً لما يقابلها من النظام الثنائي 0,1 بالترتيب، ولكن هذه العملية قد تقوم بتدمير الحمض النووي، وهنا تأتى أهمية النسخ الاحتياطية backup copies من خيوط الحمض النووي ،حتى إن تلفت تلك النسخ فالحل بسيط وهو القيام بعملية نسخ مستمر لإعادة ملئ التخزين مثلما يحدث بشكل طبيعي مع الـ DNA داخل الخلية.

8- مناقشة النتائج:

نلاحظ عموماً من النتائج السابقة أنه هناك عدة طرق للتشفير البيانات على الحمض النووي DNA و يمكن اختيار أياً منها لكن في النهاية لابد من استخدام الطريقة ذاتها لفك التشفير.

9- الخلاصة:

يمكن لغرام واحد من الحمض النووي (DNA) أن يخزن كل المعارف التي يولدها البشر في عام واحد، ويحافظ على هذه البيانات سليمة لآلاف السنين، فقد تمكن العلماء من التعرف على التسلسل الجيني الكامل لحصان عاش قبل أكثر من 500 ألف عام

Reference

[1] C. T. Clelland, V. Risca, and C. Bancroft.(1999) **Hiding messages in DNA microdots**. *Nature*, 399:533–534, 1999.

[2] G. M. Church, Y. Gao, and S. Kosuri.(2012) **Next-generation digital information storage in DNA. Science**, 337(6102):1628, 2012

[3] IDC. Where in the world is storage. http://www.idc.com/downloads/where_is_storage_ infographic_243338.pdf, 2013.

[4] J. Bornholt et al.(2016) **"A DNA-Based Archival Storage System**," Proc. 21st Int'l Conf.
Architectural Support for Programming Languages and Operating Systems (ASPLOS),2016,pp.637–649.

[5] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust,B. Sipos, and E. Birney.(2013) **Towards practical, high-capacity, lowmaintenance information storage in synthesized DNA**. *Nature*,494:77–80, 2013.