محاضرة 2

تطبيقات الحاسب الآلي في الميكروبيولوجي

Bioinformatics in MICROBIOLOGY

الأستاذ الدكتور

مظهر العيسوي الشريف

دكتوراه الميكروبيولوجي والبيوتكنولوجي – جامعة كيل – ألمانيا

رئيس قسم الميكروبيولوجي – كلية الزراعة – جامعة سوهاج

2020

مميزات دراسات المعلوماتية الحيوية: في التعامل مع الكم الهائل والمتسارع من البيانات والمعلومات الواردة من العلوم الحيوية فانه لا يمكن اقصاء هذا المجال للعديد من الاسباب منها :

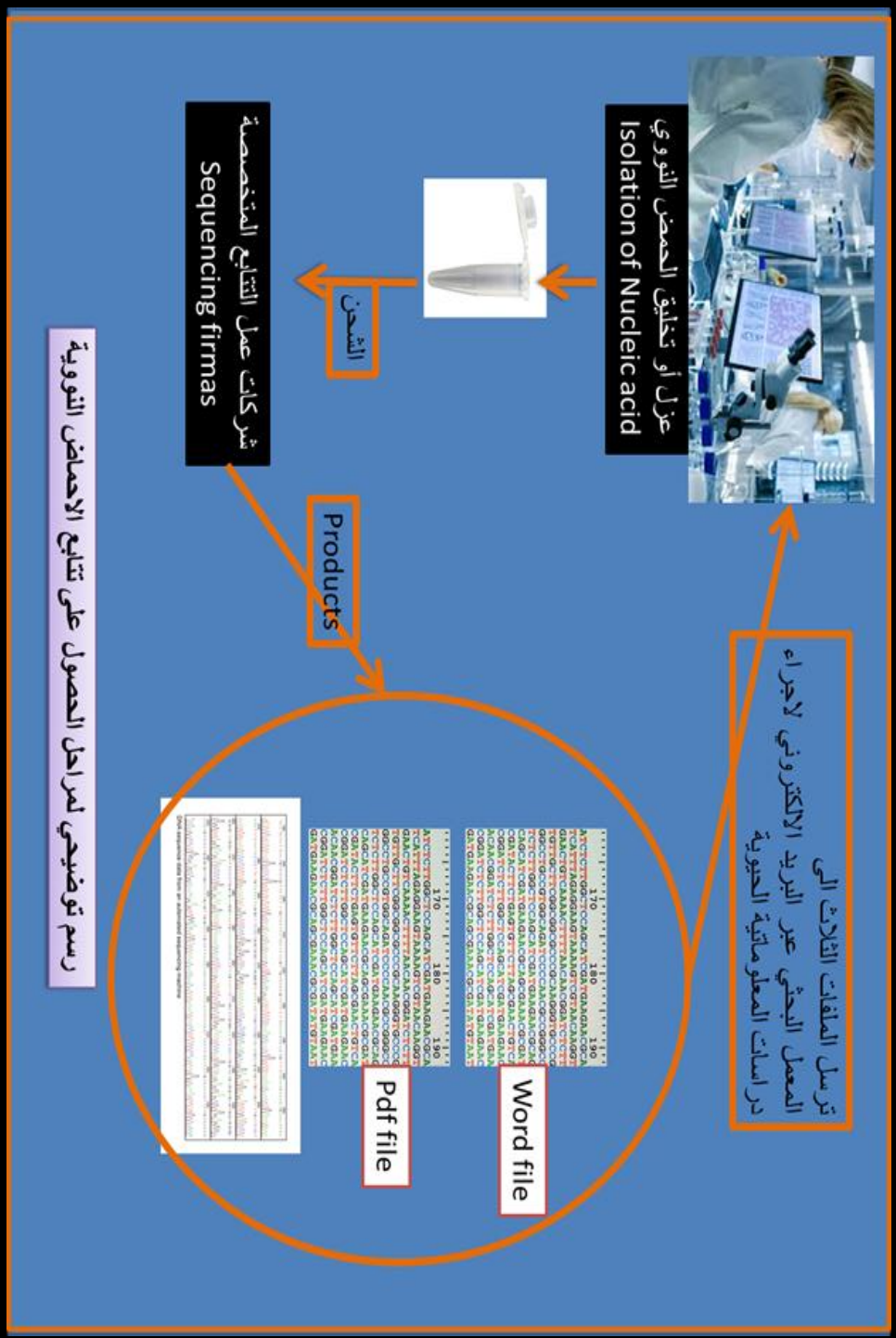
1 - توفير الوقت بشكل كبير جداً 2 -توفير الجهد (بل واستحالة انجاز المطلوب بالطرق النمطية) 3 - توفير النفقات التي قد تتكبدها المعامل والتجارب والاستعاضة عنها بشكل محاكاه وتشابه ونمذجة على الكمبيوتر 4 - دقة النتائج بشكل يصل الى 100٪ 5 -امكانية التعرف على الكثير من الميكروبات المجهولة بمجرد معرفة تتابع حمضها النووي 6 - دراسة درجات القرابة بين الكائنات 7 - العمل في شكل منظومة دولية متوافقة تصب جميعها وتستفيد كذلك في ومن نفس المستودعات الدولية 8 - له استخدامات تطبيقية هامة في مجالات الزراعة والصناعة والطب وغيرها..

ماهو دور المعلوماتية الحيوية؟ What does bioinformatics do?

- تتعامل **المعلوماتية الحيوية** مع البيانات البيولوجية الهامة مثل : (i)تسلسل القواعد Sequences في DNA, RNA و (ii) تسلسل الأحماض الأمينية amino acids (aa) في حالة البروتينات.

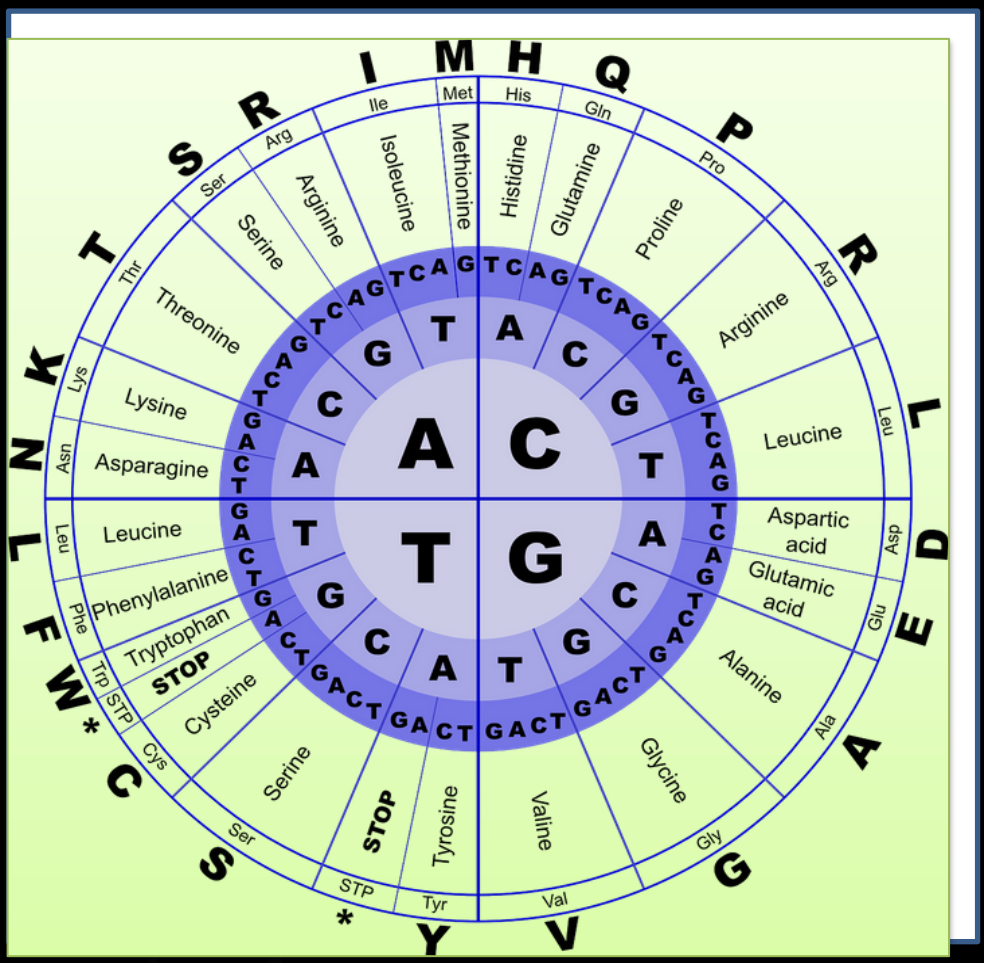
(أ) **التسلسل** (**أو التتابع**) **Sequences** يمكن أن يجرى من خلال تقنيات معملية معينة في مراكز البحوث أو شركات علمية تجارية دولية ومحلية متخصصة في هذا الشأن، حيث يتم اعداد عينة الـ **DNA or RNA** بتركيزات محددة (ميكروجرامات/ ميكروليتر) وبكميات محددة (بضع ميكروليترات (**5-20µl**)) في انابيب ابيندورف **(100µl-Eppendorf tube)** طبقاً لبروتوكول الشركة المختصة بإجراء هذا التتابع. والمادة الوراثية المرسلة، الى تلك الشركة، اما ان تكون مستخلصة مباشرة من الميكروب، أو أن تكون مخلقة لجزء معين من الـ **DNA or RNA** بتقنية الـ **PCR** . يتم ارسال تلك العينات بطرق الشحن أو البريد، لتقوم الشركة باجراء دراسة التتابع وارساله بعد ذلك عبر البريد الالكتروني للجهة خلال بضع أيام ويكون ذلك على شكل ثلاثة ملفات: (**1**) ملف **word** (يتم استخدام هذا الملف لسهولة النسخ والادراج والتعديل)، (**2**) ملف **pdf**، (**3**) ملف على شكل رسم منحنيات ومضية متتالية (شكل بياني)، حيث يمكن الرجوع الى هذا الرسم في حالة الشك في أي من التتابعات. كما هو موضح بالشكل التالي:

-



- معظم البيانات المتاحة في **المعلوماتية** الحيوية تتكون اساساً من تتابع DNA وكما

- هو معروف انها تتكون من أربع قواعد هي A, G, C, and T



- على المستوى الخلوي، تتابع القواعد في الـ DNA يمكن أن تكون لها وظيفة تشفيرية coding sequence أو ان لا يكون لها تلك الوظيفة التشفيرية non-coding sequence كما سنوضح لاحقاً باذن الله.

- الجين الواحد قد يحتوي آلاف قواعد الـ DNA ، وان الكروموسوم البكتيري النمطي يحتوي ملايين القواعد.

There can be several thousand bases in a gene, and several millions bases in a typical bacterial genome.

- **بالنسبة للبروتين**، فإن تسلسلها عبارة عن تتابع من الأحماض الأمينية (amino acids (aa)) وهي عبارة عن 20 حامض أميني مختلف تشكل كل البروتينات المعروفة (تقريبا). وكما هو موضح بالشكل الدائري أدناه، فإن لكل حمض أميني رمزين: الأول مختصر من حرف واحد مثل M للحمض الاميني الميثيونين Methionine، والثاني: من ثلاثة أحرف مثل Met لنفس الحمض الاميني السابق. ويبدو أن الاستخدام الاكثر تداولاً هو الأكثر إختصاراً (اي الحرف الواحد) لأن ذلك يوفر المساحة والجهد في التعامل خاصة مع الارقام الفلكية التي تشملها الأحماض الأمينية في دراسات المعلوماتية الحيوية.

البروتين في حجمه المتوسط يتكون من نحو 300 حمض أميني. وهناك الملايين من البروتينات معروفة الترتيب والتركيب من الأحماض الامينية

There are millions of known protein sequence

- وكما هو معروف، فإن التركيب الجيني (الجينوتايب) the genotype أي الـ DNA، وما يضفي اليه من تسلسل الأحماض الأمينية وعددها التي تشكل نوع البروتينات وهو بالتالي ما يحدد الشكل الظاهري والوظيفي للخلية والكائن الحي. وهذا ما تنص عليه الـ Central Dogma in biology
- من أهم أهداف المعلوماتية الحيوية هو الفهم الأفضل لحياة الخلية وكيف تؤدي وظائفها المختلفة على المستوى تحت الخلوي؛ حيث أن الوظائف الخلوية يتم انجازها بواسطة البروتينات والتي ترجع في قدرتها الوظيفية الى تتابع واعداد الاحماض الأمينية بها.

ومن هنا يتضح أن المعلوماتية الحيوية يمكن حصرها في تحت مجالين رئيسيين

Bioinformatics consists of two subfields:

1) تطوير برامج وآليات الحاسب الآلي وكذلك قاعدة البيانات

The development of computational tools and db

ملحوظة هامة

database (USA) or databank (UK) = db (the global abbreviation)

التطوير يشمل وضع نظم السوفتوير Software for sequence، التحليل التركيبي والوظيفي؛ أي تحليل التتابع molecular sequence analysis ، والتحليل التركيبي molecular structural analysis، والتحليل الوظيفي الجزيئي molecular functional analysis

2) تطبيقات هذه البرامج وقاعدة البيانات لوضع معرفة افضل للنظم الحيوية.

هذان المجالان يكمل بعضهما بعضاً.

مراحل المعلوماتية الحيوية:

1) جمع الاحصائيات من التجارب والبيانات البيولوجية

Collect Statistics from biological data

2) تصميم نموذج حاسوبي Build a computational modeling

3) حل مشكلات النموذج Solve a computational modeling problem

4) اختبار وتقييم النموذج Test and evaluate a computational algorithm

- أيضاً من خلال المعلوماتية الحيوية يمكن دراسة التركيب الجزيئي Molecular structures، والتعرف على الجينات وبيانات التعبير الجيني Expression

data ودراسة الـ Bibliographic data، ودراسات التنبؤ بوظائف الجينات المكتشفة حديثاً

.Gene identification & functional annotation

-من خلال المعلوماتية الحيوية يمكن التنبؤ بالشكل الظاهري للخلية

use bioinformatics to predict cellular phenotypes

ويتم ذلك اذا كانت تلك الجينات الجديدة مشابهة في التتابع مع جينات مسجلة بالفعل في قاعدة البيانات.

وهذا بالطبع يوفر الكثير من الوقت والجهد والمال؛ حيث أن عالم الميكروبيولوجي يمكن له من خلال هذا المجال ان يتوقع الشكل الظاهري أو الأيضي للميكروب دون الحاجة الى إجراء الدراسات المعملية وهذا ما يعد من اعظم فوائد المعلوماتية الحيوية.

The microbiologists can determine the phenotypes for the microbe without culture methods and microscopy

ويتجلى أهمية ذلك بشكل خاص عند دراسة البكتريا الاجبارية الغير قابلة للتنمية المعملية unculturable bacteria حيث يمكن دراستها من خلال بياناتها الرقمية، كذلك يمكن التنبؤ ولكن لا يجب المغالاة في هذا الاعتقاد؛ اذ أن

الدراسات المعملية تظل ضرورية لتأكيد دراسات الكمبيوتر (*In silico* studies). كما أن الدراسات المعملية بالغة الاهمية لتغذية قاعدة البيانات التي يتم عليها القياس في دراسات التشابه والتنبؤ التي تتم باستخدام الحاسب الآلي، وهنا يأتي التأكيد على انه لا يمكن الاعتماد بشكل مطلق على دراسات المعلوماتية الحيوية فلا يمكن الاستغناء كليةً عن الدراسات المعملية والواقعية.

ملاحظات

In vivo= in a living thing *على الكائن الحي أو البيئة الحقيقية*

In vitro= in Lab or test tube *دراسات مختبرية*

In silico= in computur دراسات على الكمبيوتر

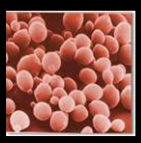
أمثلة لطول الجينوم

Temperate *S. thermophilus* phage TP-J34

about 45.6 kb linear DNA, comprising 61 putative *orfs*

E. coli **4.6 x 10^6 nucleotides**

– **Approx. 4,000 genes**



Yeast 15 x 10^6 nucleotides

– **Approx. 6,000 genes**

Human 3 x 10^9 nucleotides

– **Approx. 30,000 genes**