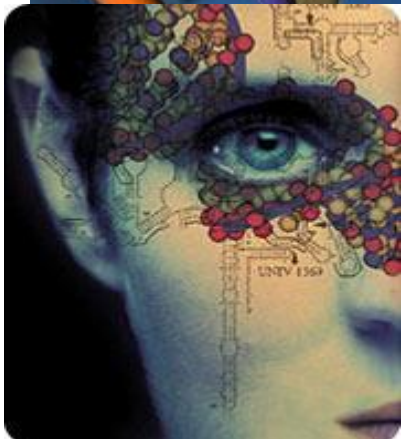
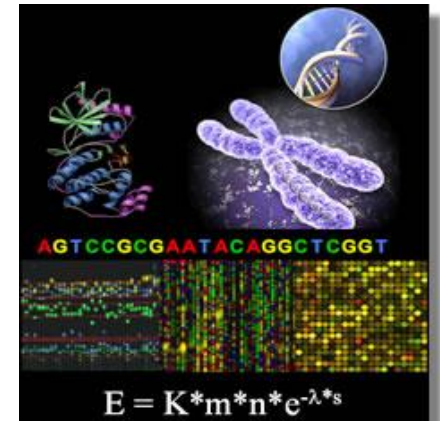


# المعلوماتية الحيوية

# Bioinformatics

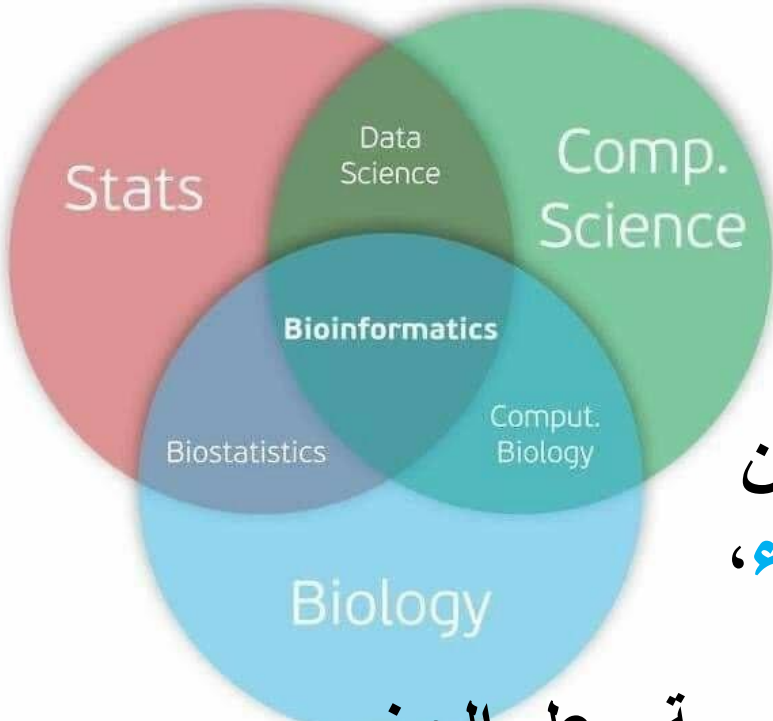


عصر جديد من العلم



# Bioinformatics المعلوماتية الحيوية

هي أحد هذه المجالات الجديدة نسبيًا، ذات التطبيقات الواسعة والمهمة. ومن المتوقع ازدياد الطلب مستقبلاً على الخبراء في هذا المجال.



هي تخصص علمي حديث النشأة، يجمع بين علوم المعلومات والرياضيات و علم الأحياء،

يُوفر الأدوات اللازمة لتحليل البيانات الجينومية «علم الجينوم

**Genomics**

والبروتيومية «علم البروتين - Proteomics لفهم النظم البيولوجية المعقدة وإيجاد الحلول.

# المعلوماتية الحيوية Bioinformatics

- أول استخدام لمصطلح المعلوماتية الحيوية Bioinformatics كان عام 1968، من قبل «مارغريت دايهوف – Margret Dayhoff»، والتي تعد رائدة هذا المجال.
- ولكن ظهر تعريفه للمرة الأولى عام 1978.
- نشأ هذا العلم وتطور بالتزامن مع نشأة الحواسيب وتطورها.
- حيث يشار إليه أيضاً باسم «علم الأحياء الحاسوبي – (computational biology)

# المعلوماتية الحيوية Bioinformatics

أدت التطورات التكنولوجية الحديثة إلى تفجر كميات هائلة من البيانات في مختلف العلوم الحيوية، إلى حد يتعدى معه تحليل تلك البيانات واستخراج النتائج منها بالعقل البشري المجرد.

وواكب ذلك تطور كبير في مجال المعلوماتية، مما أسهم في اندماج معظم هذه التقنيات مع التحليل تلك البيانات الضخمة، والتوصل إلى حلول علمية كثيرة أدت إلى ثورة علمية في العديد من القطاعات العلمية والصناعية، وفك غموض العديد من الأسرار العلمية الدفينة.

وتطلب ذلك التطور المعرفي انبثاق علم جديد هو علم المعلوماتية الحيوية Bioinformatics أو «بيوإنفورماتيكس»، بوصفه علماً يلتقي فيه النمو الهائل للتقنية الحيوية مع التطور المطرد لتقانة المعلومات.

وهو علم يدمج الحاسوب مع الرياضيات والجينوم والأحياء، ويتعامل مع الخلية والحمض النووي كما لو أنها حاسوب.

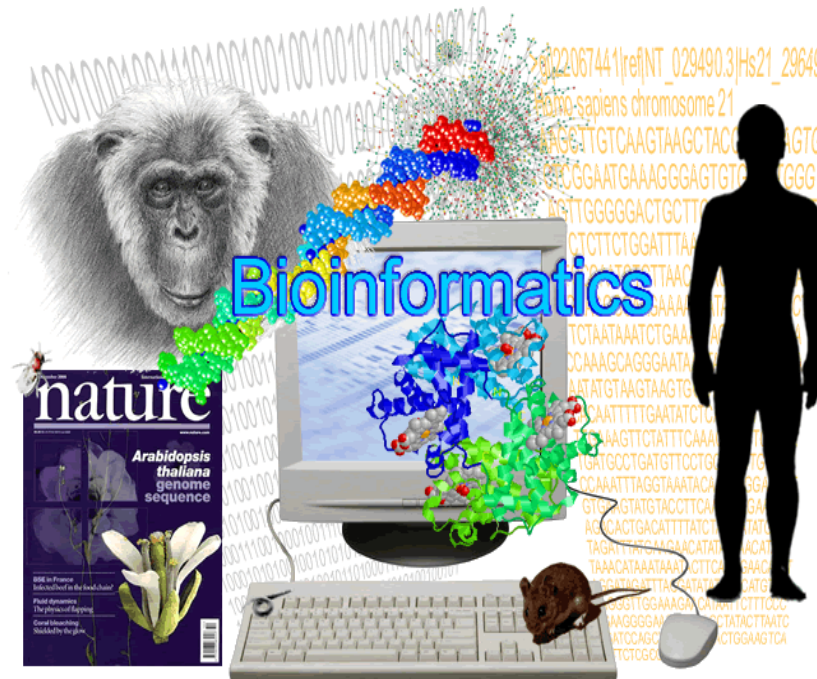
(8) (PDF) ملف العدد : المعلوماتية الحيوية - مجلة التقدم العلمي - العدد 97 إبريل 2017م. Available from:

[https://www.researchgate.net/publication/316456158\\_mlf\\_aldd\\_almlwmatyt\\_alhywy4](https://www.researchgate.net/publication/316456158_mlf_aldd_almlwmatyt_alhywy4)

- milt altqdm allmy - aldd 97 abryl 2017m [accessed Mar 11 2023].

# Bioinformatics المعلوماتية الحيوية

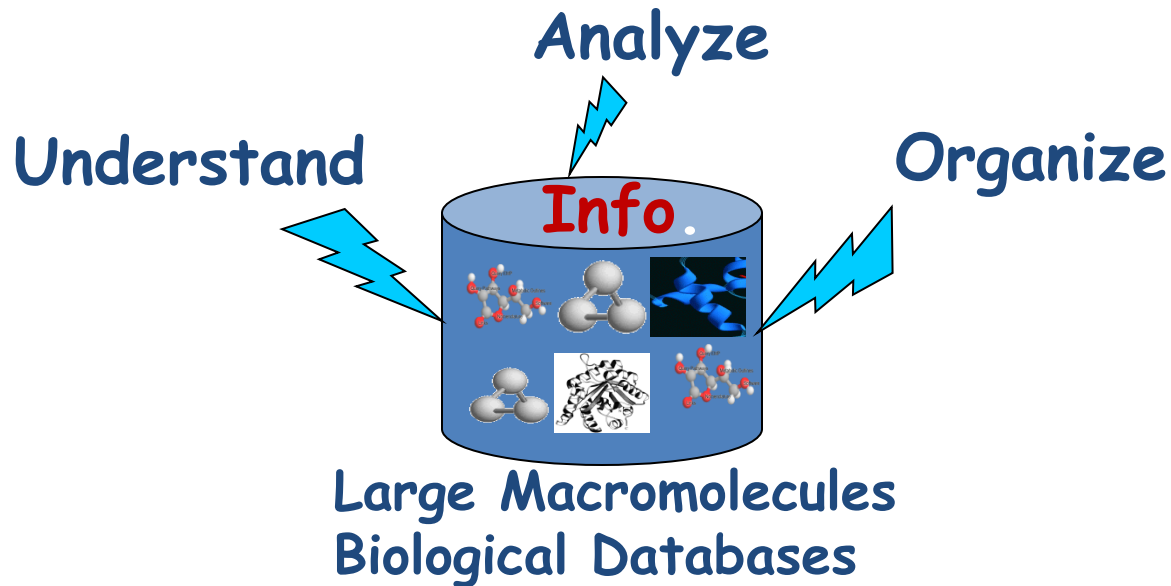
- المعلوماتية الحيوية هي تحليل المعلومات البيولوجية باستخدام الكمبيوتر و التقنيات الإحصائية. وهي العلم الذي يسعى لاستخدام وتطوير قواعد البيانات و الخوارزميات الحاسوبية لتسهيل وتعزيز الأبحاث البيولوجية.
- Bioinformatics is the analysis of biological information using computers and statistical techniques. It is the science that seeks to use and develop databases and computer algorithms to facilitate and enhance biological research.**



# Bioinformatics المعلوماتية الحيوية

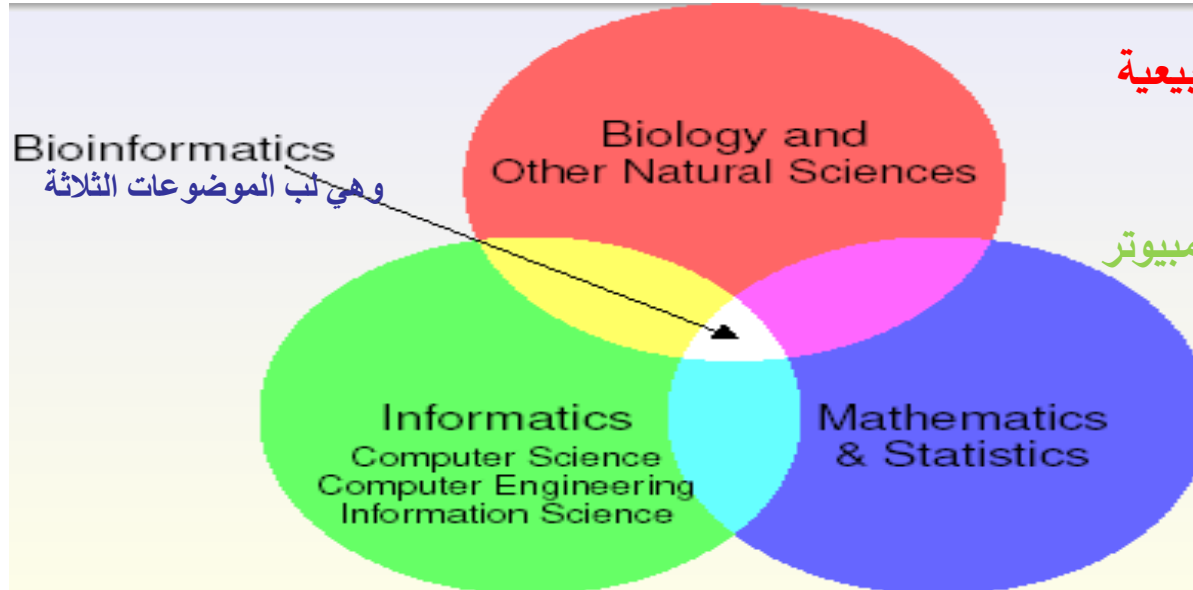
## What is Bioinformatics?

The use of computers to **collect**, **analyze**, and **interpret** biological information at the molecular level.



• يُعرف العلم المختص بذلك بالمعلوماتية الحيوية *Bioinformatics* أو علم الأحياء الحاسوبي ( البيولوجيا الحاسوبية ) **computational biology** هو استخدام أحدث تقنيات الرياضيات التطبيقية، المعلوماتية informatics، الإحصاء، و علوم الحاسب لحل مشكلات بيولوجية.

• Competent science knows so bio-Informatics and known in English Bioinformatics or computational biology (computational biology) computational biology is the use of the latest techniques applied mathematics, informatics, statistics, and computer science to solve biological problems.



1. البيولوجي والعلوم الطبيعية

2. الرياضيات والإحصاء

3. المعلوماتية وعلوم الكمبيوتر

# المعلوماتية الحيوية آفاق واعدة في تشخيص الأمراض

❖ تؤدي المعلوماتية الحيوية دورا مهما في مجال تشخيص وعلاج أمراض الإنسان ويطلق على المعلوماتية الحيوية المستخدمة في مجال الطب «المعلوماتية الحيوية الطبية».

❖ ساهم التطور الهائل في معرفة وتحليل تتابعات الجينوم أو لسلاطات المرض المختلفة عن طريق أدوات المعلوماتية الحيوية في تطوير طرق تشخيص الأمراض، وإنتاج وتطوير الأدوية والتنبؤ بالأمراض المتوقع حدوثها، وإنتاج اللقاحات الفعالة.



# المعلوماتية الحيوية تطورات في فهم العلاجات

❖ أدى استخدام المعلوماتية الحيوية في دراسة وتحليل الجينوم البشري إلى تقدم هائل في فهم العلاقة بين المريض والاستجابة للعلاج، إذ يوجد تباين في درجة استجابة المرضى للعلاج بسبب اختلافات في الجينوم الخاص بهم.

❖ ومن خلال التعرف على هذه الاختلافات على مستوى الجينوم البشري باستخدام طرق المعلوماتية الحيوية، وبخاصة مع التقدم الهائل في تقنيات قراءة تتابع الجينوم البشري، فإن الطبيب لم يعد بحاجة لتجربة وتغيير أكثر من دواء للمريض للتوصل للعلاج المناسب.

❖ باستخدام نتائج تحليل الجينوم للمريض فيمكن للطبيب وصف الدواء والجرعة المناسبة منذ البداية.

❖ كما أدى تحليل تتابعات الجينوم البشري إلى الربط بين وجود طفرات/تغيرات في الجينوم والتنبؤ باحتمالات الإصابة المتوقعة بالأمراض.



➤ من المعروف أن اللقاحات تستخدم للوقاية من الأمراض، لكن من مشكلات إنتاج اللقاح التغير المستمر في تركيب سلالات الممرض، بحيث يؤدي ظهور أي طفرة في الممرض إلى عدم تعرف اللقاح على هذه السلالة من الممرض، ومن ثمّ لا يتأثر الممرض باللقاح المستخدم.

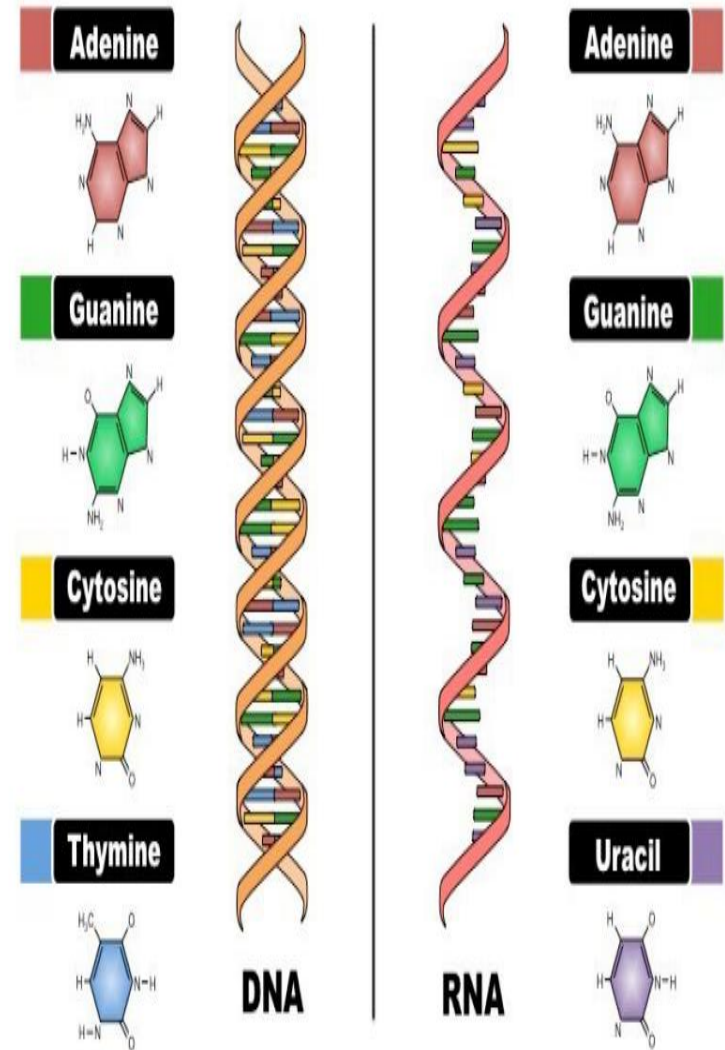
➤ باستخدام المعلوماتية الحيوية يمكن رصد أمكنة التباين في تركيب جينوم الممرض، ومن ثمّ نستطيع فهم الكائن الممرض بصورة أكبر، ويصبح لدينا القدرة على اختيار اللقاح من مناطق على الجينوم تتميز بالثبات وعدم وجود طفرات فيها، كما يمكننا استخدام لقاحات تمثل مناطق على الجينوم متشابهة لكل سلالات الممرض، وبذلك يمكن استخدام لقاح واحد لفترات طويلة لمقاومة أي سلالات متوقعة من الممرض، وبذا يمكن مقاومة الممرض والسلالات المحتمل وصولها في أي وقت.

# المعلوماتية الحيوية الشفرة الوراثية و أدوات معرفتها

- أدى التطور في مجال تقانة المعلومات والمعلوماتية الحيوية إلى اكتشاف أسرار الشفرة الوراثية لمختلف الكائنات الحية
- وتحدد المعلومات الوراثية تاريخ تطور الكائنات وعلاقتها بالكائنات الأخرى، وكيفية نقل المعلومات إلى الصفات الظاهرة و الوظائف البيولوجية.
- 1990 وضع مجموعة من علماء الوراثة البشرية مشروع الجينوم البشري الذي كان الدافع إلى التطور السريع في علم الوراثة الذي يسمى حالياً علم البيولوجيا الجزيئية. وأدى ذلك التطور إلى استحداث تخصصات جديدة مثل المعلوماتية الحيوية وعلم الجينوم وعلم البروتين. هذه العلوم تجمع بين الأدوات الإحصائية والتقنيات الحاسوبية عن طريق توظيف برامج الحاسوب لتنظيم وتحليل البيانات الوراثية و البروتينية.
- أصبح مجال المعلوماتية الحيوية ضروريا لعلم البيولوجيا والطب والصيدلة<sup>11</sup> والدراسات البيئية والأدلة الجنائية.

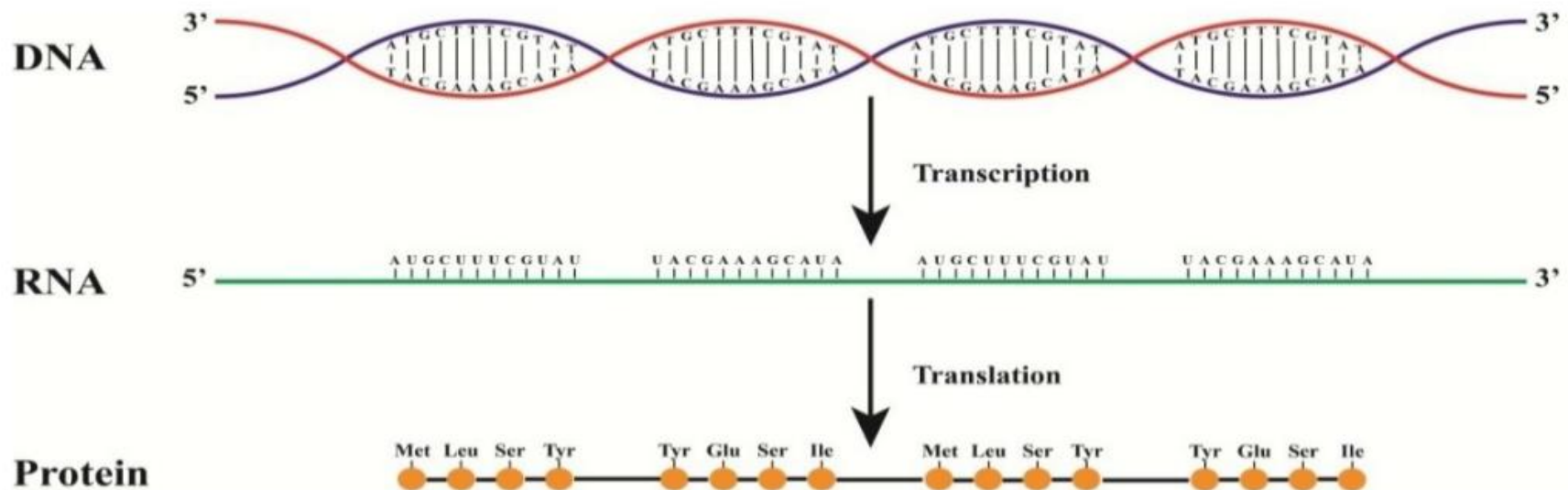
# Bioinformatics works at:

- **DNA level:**
  - DNA sequence alignment;
  - gene prediction;
  - gene evolution;...
- **RNA level:**
  - Study of gene expression;
  - transcription mechanism;
  - post-transcription modification;...



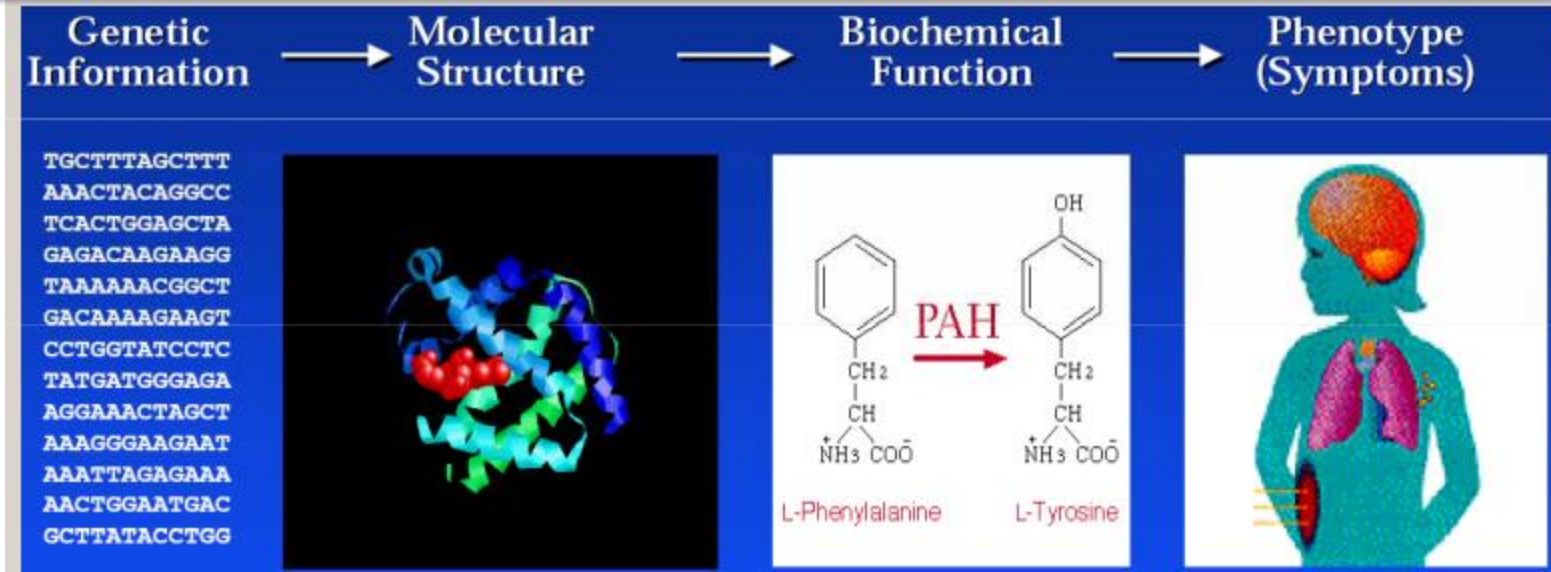
# • Protein level:

- protein 2D and 3D structure prediction;
- protein active site prediction;
- protein-protein interactions;
- protein-DNA interactions;...



ترجمة RNA إلى بروتين بعد نسخه من شريط DNA. المصدر: (nature.com)

# أهمية المعلوماتية الحيوية



تظهر أهمية المعلوماتية الحيوية من خلال:

1- ازدياد الاهتمام بها على كافة الأصعدة

2- دخولها المؤسسات التعليمية العالية المستوى

مثل أ- الدراسات الأولية

ب- الدراسات العليا

ج- مجالات البحوث العلمية على مختلف مشاربها،

د- خصصت لها الأقسام العلمية الخاصة بها وكذلك الكليات

# تطبيقات المعلوماتية الحيوية

تطبيقات المعلوماتية الحيوية واسعة وهي تتسع يوماً بعد آخر

1- في مجالات الدراسات الأكاديمية أصبحت ضرورة لا بد منها،

2- تحليل تركيب الحوامض الأمينية

3- تحليل البروتينات،

4- دراسة التراكيب المختلفة بالمقارنة

5- دراسة عمليات الاصطفاف لغرض تصنيفها

6- التنبؤ بوظائف التراكيب المختلفة للبروتينات والاحماض النووية بالاعتماد على

التركيب فضلا عن تحديد Alignments وتحديد Gene expression profile التحليل

الوظيفي مثل تحديد نسق التعبير الجيني

7- التنبؤ بتداخلات البروتينات مع بعضها

## What skills are needed?

- Well-grounded in one of the following areas:
  - Computer science
  - Molecular biology
  - Statistics



# Bioinformatics المعلوماتية الحيوية

- **The computer and molecular databases** are a necessary, integral part of this entire process.



# قاعدة البيانات البيولوجية Biological database

- هي قاعدة بيانات بيولوجية كبيرة،
- عبارة عن هيئة منظمة للمعلومات الثابتة، وعادة ما ترتبط البرمجيات الحاسوبية المصممة بإمكانيات التحديث ، والاستفسار ، واسترجاع عناصر من البيانات المخزنة داخل النظام.

- Is a large biological database, an organized body of information fixed, usually associated with computer software designed potential update, inquiry, and retrieval of data elements stored within the system.

# قاعدة البيانات البيولوجية Biological database

- من أبسط أشكال قواعد البيانات قد يكون ملف واحد يحتوي على العديد من السجلات ، كل سجل يضم نفس المجموعة من المعلومات على سبيل المثال ، سجل متعلق بتسلسل الحمض النووي يحتوي على مدخلات مثل إدخال التسلسل مع وصف نوع الجزيء ، والاسم العلمي للكائن المصدر الذي عُزل منه الحمض النووي ، و بعض المراجع المهمة المرتبطة بالتسلسل.

. Of the simplest forms of databases may be a single file containing many records, each record includes the same set of information, for example, record related sequence DNA contains entries such as the introduction sequence with a description of the type molecule, and the scientific name for the object source, who isolate him DNA, and some important references associated sequentially.

# Comprehensive Databases

- وهي معلومات عن أي شيء متعلق بالبيولوجي .
- وهناك ثلاثة قواعد بيانات رئيسة :

## 1. المركز القومي لمعلومات البيوتكنولوجي :

- **National Center for Biotechnology Information (NCBI)**
- <http://www.ncbi.nih.gov>
- **NCBI is a resource for molecular biology information.**

## 2. المعمل الأوروبي للبيولوجيا الجزيئية :

### **European Molecular Biology Laboratory (EMBL)** •

- <http://www.embl.org>
- **EMBL provides biomolecular databases and bioinformatics tools.**

## 3. بنك بيانات الـ DNA الياباني :

- **Data Bank of Japan (DDBJ)**
- <http://www.ddbj.nig.ac.jp>
- **DDBJ is a DNA data bank in Japan.**

تشارك في البيانات لكن تختلف في الـ tools

## Databases in Bioinformatics

قواعد البيانات في المعلوماتية الحيوية

# Databases in Bioinformatics

## قواعد البيانات في المعلوماتية الحيوية

- Why?
  - The different types of databases
  - Database language: identifiers
  - Nucleotide sequence databases
  - Protein sequence databases
  - 3D structure databases
- لماذا؟
  - الأنواع المختلفة من قواعد البيانات
  - لغة قواعد البيانات: المعرفات
  - قواعد بيانات تسلسل النكليوتيد
  - قواعد بيانات تسلسل البروتين
  - قواعد البيانات للبناء ثلاثي الأبعاد

# Biological databases: Why?

## قواعد البيانات البيولوجية : لماذا؟

- **Make biological data available to scientists**

- جعل البيانات متاحة للعلماء البيولوجي

- Consolidation of data (gather data from different sources)

- توحيد البيانات (جمع البيانات من مصادر مختلفة)

- Provide access to large dataset that cannot be published explicitly (genome, ...)

- توفير إمكانية الوصول إلى مجموعة البيانات كبيرة التي لا يمكن نشرها بشكل موسع (الجينوم، ...)

- **Make biological data available in computer-readable format**

- إتاحة البيانات البيولوجية في شكل مقروء بالحاسوب

- Make data accessible for automated analysis

- إتاحة الحصول على البيانات للتحليل الآلي

**Bioinformatics: “a collective term for data compilation, organisation, analysis and dissemination”**

المعلوماتية الحيوية : "وهو مصطلح جمعي لتجميع البيانات وتحليلها وتنظيمها ونشرها"

# The different types of Databases in Bioinformatics

الأنواع المختلفة من قواعد البيانات

## *Data: (1) البيانات*

### أنواع البيانات: Type of data:

- **nucleotide sequences**

تسلسل النكليوتيد

- **protein sequences**

تسلسل البروتين

- **3D structures**

البناء الثلاثي الأبعاد

- **gene expression data**

بيانات التعبير الجيني

- **metabolic pathways**

مسارات الأيض



# أشكال البيانات (Data formats)

- تتواجد البيانات البيولوجية في عدة أشكال مختلفة ، تختلف في أسلوب إدخالها (annotation) والبرامج التي تقرأها . . . الخ .
- توجد tools لتحويل البيانات فيما بين الأشكال مثل BioPerl الذي يحول من شكل forma لآخر .
- ومما هو جدير بالذكر أن المجتمع العلمي يميل إلى شكل الـ XML لأنها سريعة (وهي فورمة للملفات الإلكترونية مثل hotmail) .

# 1. FASTA format

• هي بسيطة الشكل ، يتم كتابة سطر للتعريف يبدأ بـ “>” ويليه مباشرة اسم يتركب من عشرة حروف أو أقل ثم بقية السطر يمكن أن يتضمن مسافات وأى حروف . ثم يتم كتابة تتابع الـ DNA أو البروتين في صورة حرف واحد كرمز لكل حامض أميني وذلك في السطر التالي .  
تستخدم في كثير من قواعد البيانات والبرامج .

## • مثال:

- >sp|O53189|TIG\_MYCTU Trigger factor (TF) –  
Mycobacterium tuberculosis .
- MKSTVEQLSPTRVRINVEVPFAELEPDFQRAYKELAKQV  
RLPGFRPGKAPAKLLEARIGREAMLDQIVNDALPSRYGQ  
AVAESDVQPLGRPNIEVTKKEYGQDLQFTA EVDIRPKISP  
PDL SALT V SVDPIEIGEDDVDAELQSLRTRFGTLTAVDRP  
VAVGDVVSIDLSATVDGEDIPN
- الأحماض الأمينية لبروتين ما .



Search Nucleotide for [ ] Go Clear  
Display FASTA Show 5 Send to [ ]  
Range: from begin to end [ ] Reverse complemented strand Refresh

1: NM\_207667. Reports Mus musculus fibr...[gi:178557810] Links

>gi|178557810|ref|NM\_207667.3| Mus musculus fibroblast growth factor 14 (Fgf14), transcript variant 2, mRNA

TTTTTTTTCTTTTCTTTTTTTTTTTTTTCCCCCTTCTCTCCTAGGGGTACACAATGGCAGTC  
TTCTCTCGCTAAGATGAATCCGGCTTGCCCTCTGCAGATCCGCCCATCTTCAGTCCATTTAGTAAAAAGG  
CTATGATTTTCTTTGGGAICTTTTGTGCATTACTGTTCTCCCTGGTGTAGTCTGGTCTCGGTTGGGA  
TTTCTTTGCTTTTTTTTGTGGCTCATGTGAAAAAGGATTTTTTCTCTCCAACCTTTGGTCAIGAT  
CCAGTGGTATCAAGAGGGGATAAATCATTACCCTGGCCGACAAAAACAATCGCTGAGAAGTCTCAA  
GAGATATACCAGCTGAGGGGAAAAAAGCTGGGAGAAGATCCGGAAATATAACGTTTTTCTATGGTAAAC  
CGGTGCCCTCTCAGGAGAAGTGAATCAAGTATTATTATGCAACCAAGGGCCCTCTTTTCTCAG  
GGTGTCTAAGCTGCTGGGTTGCTTTTGGCCCAATCAATGTGGTTTCTTTGGAACTTTTCAGCAAAAGGA  
ACGCATATGCTGCAAGTGTCTTTGGCAAGAGTCTTAAGAAAAACAAGAACCCCACTGATCCCAGCTCA  
AGGCATAGTGACCAAGTATATGCAAGCAAGGCTACTACTTGCAGATGCACCCCGATGGAGCTCTCGA  
TGGAAACCAAGGATGACAGCACCAATCCACACTGTTCAACCTCATCCAGTGGGACTGCGGTTGTTGCC  
ATCCAGGGAGTGAAGACAGGTTGTACATAGCAATGAATGGAGAGGTTACCTCTACCCATCAGAACTTT  
TTACCCTGAATGCAAGTTAAAGAGTCTGTTTTGAAAACTATTATGTAATCTACTCATCCATGCTGTA  
CAGGCAACAGGAGTCTGGCAGAGCCTGGTTTTGGGATTAATAAGGAAGGGCAAGTTATGAAAGGAAAC  
AGATAAAGAAAAACCAACAGCAGCTCAATTTTACCCAAGCCATGGAAGTGGCATGTACCGAGAAC  
CATCCTTGCAATGATGTTGGTGAACAGTCCCGAAAGCTGGGAGCAGCCAAAGTACAAGTGCATC  
TGCAATAATGAATGGAGCAACCAAGTCAACAAGTCAAGACCATAGCCAGATCCTCAGGTGTTCTGA  
CTTCTGCTGCTGAGCAGTGGAGCGATTATCCTTACTAGACATTCCTGCTCTGTGGCGAAAAAGCC  
CCAGCAAGCAAGCTGATGCTTGTCTTATGCTGTCTCAGACCTCCTTCTTTGCAAGTGGATAAATCTCAAC  
CTGTGGTCCCCCTACACAAAGAACACCTGGACAAACCAAGCTAAACTCAGACCATGGAATGCCCTACCA  
GATATGGAATGCCTTTTTAATATCTTTCTGTGACTGTGACACTTCATGTGAATGACATACTTCACAAGT  
ACACTCAATACCTTGCTGCTGACAGCTACCCATAATCCTTTTTGAGTCCCTGTTTCAGCGAAATCCATGT  
GTTTAAAGTCAATTTGTAGCACAAATAAATGAGTAATTTCTAGTATAGTGTGTAACCTGTGCT  
ATCATGGATTTCTCTCTCCATTTTACAGGGCTGCTCGCTCCACTGTCTGTGACCTTTAGCAGGGAT  
TGTGTTGCTCTGAATCTTCAATGTTGCAAGTGGTTGCTTAGGAGAACAAATAGGGAATCAGGAAGCCTT  
CTAAACCTATTACTACAAATAGTATCTATAAAGATTAAAGATTGTTGTCTTTGGCTCACATTATTGATTA  
ACACATATATGTTCTGTCTAGTAGGAGACGCTGTGCTCCGAGGTAGGCATGGCCTGGTGGAAACAAG  
GGTCAAAACAAATCCATGGAAGCTCTCTATGATAGGTGTTGGCATTCCCTCTTAGTTTCTTTATGTTG  
GTGTTTTATACATATCACAAGCTTACTGGTAATGGTAACATTTGCCCTGGCCAGCAGCAAGACCCAC  
TTAGTTTTGAGAAAGTGGGTCAAAAGAAATCTGTAGGCCTGTGAGGCCTGATTAAAGTTTCATTTATTT  
TCATCAATTAATCCTCAATTTGGGAAAAAGAAATAAAAACCCAAATCATTACAACATAAAGAAATATGCT  
ACCTAAATCCATTTCCGGCCAACTCCTTCTATTTTTAAATGAAGATAACTCAGTGCCACAGCTGATTTGG

فورمة الفاستا

اولاً : الخرائط الكروموسومية .

GenBank

[Help](#)

- GenBank
- Submit
- Genomes
- WGS
- HTGs
- EST/GSS
- Metagenomes
- TPA
- TSA
- INSDC

## GenBank Overview

### What is GenBank?

GenBank<sup>®</sup> is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids Research, 2011 Jan;39\(Database issue\):D32-7](#)). There are approximately 126,551,501,141 bases in 135,440,924 sequence records in the traditional GenBank divisions and 191,401,393,188 bases in 62,715,288 sequence records in the WGS division as of April 2011.

The complete [release notes](#) for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

An example of a GenBank [record](#) may be viewed for a *Saccharomyces cerevisiae* gene.

### Access to GenBank

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with [Entrez Nucleotide](#), which is divided into three divisions: [CoreNucleotide](#) (the main collection), [dbEST](#) (Expressed Sequence Tags), and [dbGSS](#) (Genome Survey Sequences).
- Search and align GenBank sequences to a query sequence using [BLAST](#) (Basic Local Alignment Search Tool). BLAST searches CoreNucleotide, dbEST, and dbGSS independently; see [BLAST info](#) for more information about the numerous BLAST databases.
- Search, link, and download sequences programatically using [NCBI e-utils](#).

### GenBank Data Usage

The GenBank database is designed to provide and encourage access within the scientific community to the most up to date and

## GenBank Resources

- [GenBank Home](#)
- [Submission Types](#)
- [Submission Tools](#)
- [Search GenBank](#)
- [Update GenBank Records](#)

ندخل على البنك الجيني

ثانياً: تحديد تتابع جين معين

لتحديد تتابع جين معين نتبع الآتي  
الطريقة الأولى.

**Browse your genome**

Click on a chromosome to show

Genes ▾



**Find A Gene**

Search for

from  ▾



A challenge facing researchers today is that of piecing together and analyzing the plethora of data currently being generated through the Human Genome Project and scores of smaller projects. NCBI's Web site serves an an integrated, one-stop, genomic information infrastructure for biomedical researchers from around the world so that they may use these data in their research efforts. [More...](#)

**Genes and Human Health**

▶ **Gene Database**

A new database of genes and associated information is now available for searching in Entrez.

▶ **OMIM**

A guide to human genes and inherited disorders maintained by Johns Hopkins University and collaborators.

▶ **dbSNP**

A database of single nucleotide

▶ **dbGaP**

The database of Genotypes and

احدد أي جين مثل جين سرطان الثدي

GenBank

Gene HMGA2

Search

Help

GenBank

Submit

S

HTGs

EST/GSS

Metagenomes

TPA

TSA

INSDC

## GenBank Overview

## What is GenBank?

GenBank® is the NIH gene

[2011 Jan;39\(Database iss](#)

GenBank divisions and 191

The complete [release note](#):GenBank is part of the [Inte](#)

European Molecular Biolog

An example of a GenBank

## Access to GenBank

There are several ways to

- Search GenBank for

[CoreNucleotide](#) (the

- Search and align Ge

[CoreNucleotide](#), dbE

- Search, link, and dow

## GenBank Data Usage

Clone  
dbGaP  
dbVar  
Epigenomics  
Gene  
GEO DataSets  
GEO Profiles  
HomoloGene  
MedGen  
MeSH  
NCBI Web Site  
NLM Catalog  
OMIA  
OMIM  
PMC  
PopSet  
Probe  
Protein Clusters  
PubChem BioAssay  
PubChem Compound  
PubChem Substance  
PubMed Health  
SNP  
SRA  
Taxonomy  
ToolKit  
ToolKitAll  
ToolKitBook  
UniGene  
UniSTS

annotated collection of all publicly available DNA sequences ([Nucleic Acids Research](#), approximately 126,551,501,141 bases in 135,440,924 sequence records in the traditional 15,288 sequence records in the WGS division as of April 2011.

GenBank are available on the NCBI ftp site. A new release is made every two months. [GenBank Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the GenBank at NCBI. These three organizations exchange data on a daily basis.

*Saccharomyces cerevisiae* gene.

m GenBank.

notations with [Entrez Nucleotide](#), which is divided into three divisions:

Expressed Sequence Tags), and [dbGSS](#) (Genome Survey Sequences).

any sequence using [BLAST](#) (Basic Local Alignment Search Tool). BLAST searches

ntly; see [BLAST info](#) for more information about the numerous BLAST databases.

atically using [NCBI e-utilities](#).

## GenBank Resources

[GenBank Home](#)[Submission Types](#)[Submission Tools](#)[Search GenBank](#)[Update GenBank Records](#)

نحدد الجين المطلوب مثل الجين  
المسؤول عن الطول.



# الطريقة الثالثة

التنبؤ بالجين عن طريق الكمبيوتر

BLAST (1

(<http://ncbi.nih.gov/BLAST>)

هل الكائنات الأخرى بها تتابع مماثل (similar)؟  
وهل تتابع مماثل لجين أو بروتين معلوم؟