Introduction to Bioinformatics

biological data processes

Presented by Sadoon H. Abdullah

saadustego@gmail.com

Outline

- ✓What is bioinformatics?
- ✓ Research area to manage and analyze biological data
- ✓New Science
- ✓Biological data
- ✓Bioinformatics Today
- ✓Bioinformatics Flow Chart
- ✓Computational approaches to biological questions
- ✓Bioinformatics Applications.
- ✓What can I do as computer science?
- ✓Bioinformatics Applications for Computer Science

BioInformatics Group

What is bioinformatics?

The combination of biology and information technology. It is a branch of science that deals with the computer based analysis of large biological data.

biological data= DNA, RNA and Protein Sequence

Information technology applied to the management and analysis of biological data statistical tools and algorithms to analyze and determine relationships between biological sets, such as macromolecular sequences and structures, etc…

What is bioinformatics (contd…)?

Bioinformatics use disciplines of computer science such as AI, neural networks, genetic algorithms, dynamic programming, pattern recognition, data mining, machine learning , algorithms to accelerate and enhance biological research.

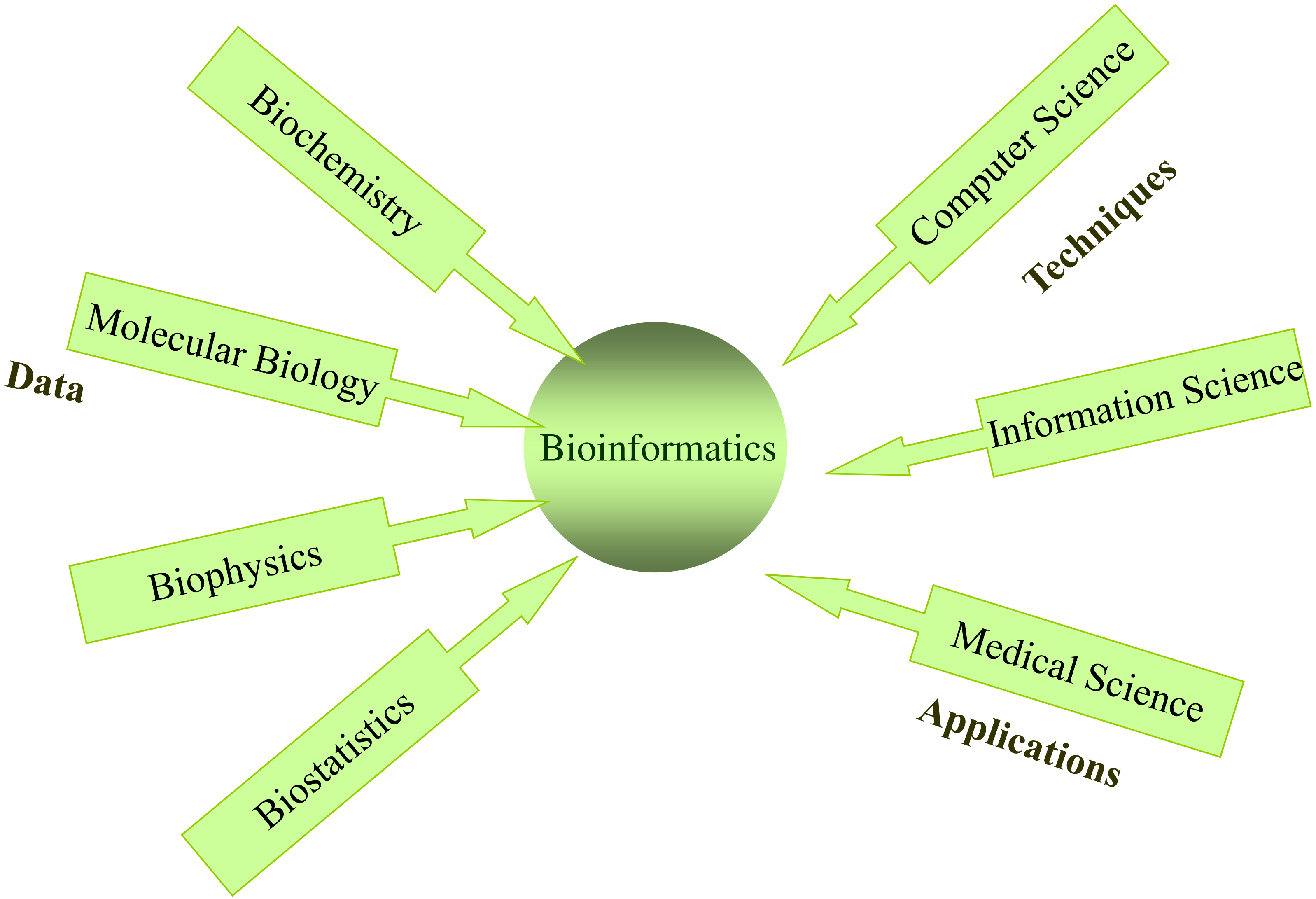
What is bioinformatics (contd...)?

Biologists
collect molecular data:
DNA & Protein sequences,
gene expression, etc.

Bioinformaticians
Study biological questions by
analyzing molecular data

Computer scientists
(+Mathematicians, Statisticians, etc.)
Develop tools, software's, algorithms
to store and analyze the data.

Research area to manage and analyze biological data



New Science

Most of the biologists don’t know computer science. Most computer scientists don’t know biology.

… There is a new science ----------------→ Bioinformatics

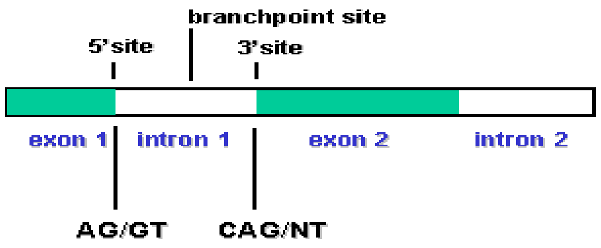
Biological data comes in many forms:

- Sequence
  - DNA and protein sequence
- Structure
  - RNA Secondary structure, protein secondary and 3D structure
- Graphs
  - Biological networks

Bioinformatics Today

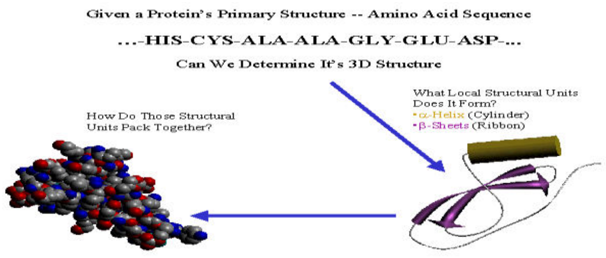
Sequence analysis

- Sequence alignment
- Structure and function prediction
- Gene finding



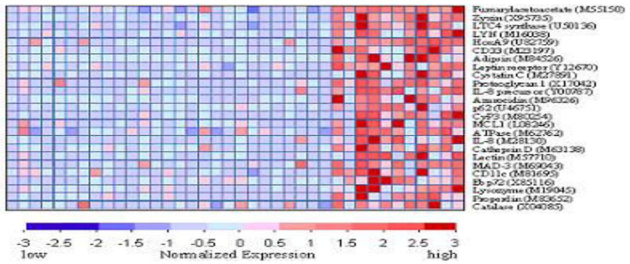
Structure analysis

- Protein structure comparison
- Protein structure prediction
- RNA structure modeling



Expression analysis

- Gene expression analysis
- Gene clustering



Bioinformatics Flow Chart

1. Sequencing

↓

1a. Analysis of nucleic acid(DNA,RNA) seq.

↓

2. Analysis of protein seq.

↓

3. Molecular structure prediction

↓

4. molecular interaction

↓

5. Gene & Protein expression data

↓

7. Drug Design

Computational approaches to biological questions

Biological question	Computational approach
• Understanding one genome	• String and graph based algorithms for sequence assembly
• Understanding many genomes	• Comparing multiple genomes using trees and hidden markov models
• Identifying causal genes for a disease	• Clustering\ etc..

Bioinformatics Applications

Some Simple Bioinformatics Applications

- Identify a gene in a genome
- Mutation location
- ORF
- Motif
- Predict function of unknown genes/proteins
- Retrieve/compare gene sequences (Alignments)

Bioinformatics Applications (contd ...)

1

How do we identify a gene in a genome?

A gene is characterized by several features (promoter, ORF…)

Bioinformatics Applications (contd …)

Q: Where are the genes in this genome?

1

CCACACCACACCCACACACCCACACACCACACCACACACCACACCACACCCACACACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCTGGCCAACCTGTCTCTCAACTT
CGTTACCCTGTCCCATTCAACCATACCACTCCGAACCACCATCCATCCCTCTACTTACTACCACTCACCCACCGTTACCCTCCAATTACCCATATCCAACCCACTGCCACTTACCCTACCATT T
CACCATACTGTTCTTCTACCCACCATATTGAAACGCTAACAAATGATCGTAAATAACACACACGTGCTTACCCTACCACTTTATACCACCACCACATGCCATACTCACCCTCACTTGTATACT
CTACAGTATATACCATCTCAAACTTACCCTACTCTCAGATTCCACTTCACTCCATGGCCCATCTCTCACTGAATCAGTACCAAATGCACTCACATCATTATGCACGGCACTTGCCTCAGCGG T
AACGCCCATCATTATCCACATTTTGATATCTATATCTCATTCGGCGGTCCCAAATATTGTATAACTGCCCTTAATACATACGTTATACCACTTTTGCACCATATACTTACCACTCCATTTATATAC A
TCCCCACAAAAATCACCTAAACATAAAAATATTCTACTTTTCAACAATAATACATAAACATATTGGCTTGTGGTAGCAACACTATCATGGTATCACTAACGTAAAAGTTCCTCAATATTGCA
AGAATATTTCGTACTTACACAGGCCATACATTAGAATAATATGTCACATCACTGTCGTAACACTCTTTATTCACCGAGCAATAATACGGTAGTGGCTCAAACTCATGCGGGTGCTATGATACAATTATATCTTATTTCCATTCCCAT
ATGCTAACCGCAATATCCTAAAAGCATAACTGATGCATCTTTAATCTTGTATGTGACACTACTCATACGAAGGGACTATATCTAGTCAAGACGATACTGTGATAGGTACGTTATTT ATAAT
TTTACGGTAATATAACTTATCAGCGGCGTATACTAAAACGGACGTTACGATATTGTCTCACTTCATCTTACCACCCTCTATCTTATTGCTGATAGAACACTAACCCCTCAGCTTTATTTCTAGTTACAGTTACACAAAAAACTATGC
CAACCCAGAAATCTTGATATTTTACGTGTCAAAAAATGAGGGTCTCTAAATGAGAGTTTGGTACCATGACTTGTAACTCGCACTGCCCTGATCTGCAATCTTGTTCTTAGAAGTGACGCATATTCTATACGGCCCGACGCGAC
GCGCCAAAAAATGAAAAACGAAGCAGCGACTCATTTTTATTTAAGGACAAAGGTTGCGAAGCCGCACATTTCCAATTTCATTGTTGTTTATTGGACATACACTGTTAGCTTTATTACCGTCCACGTTTTTTCTACAATAGTGTA
GAAGTTTCTTTCTTATGTTCATCGTATTCATAAAATGCTTCACGAACACCGTCATTGATCAAATAGGTCTATAATATTAATATACATTTATATAATCTACGGTATTTATATCATCAAAAAAAAAGTAGTTTTTTTATTTTATTTTGTTCGT
TAATTTTCAATTTCTATGGAAACCCGTTCGTAAAATTGGCGTTTGTCTCTAGTTTGCGATAGTGTAGATACCGTCCTTGGATAGAGCACTGGAGATGGCTGGCTTTAATCTGCTGGAGTACCATGGAACACCGGTGATCATTCT
GGTCACTTGGTCTGGAGCAATACCGGTCAACATGGTGGTGAAGTCACCGTAGTTGAAAACGGCTTCAGCAACTTCGACTGGGTAGGTTTCAGTTGGGTGGGCGGCTTGGAACATGTAGTATTGGGCTAAGTGAGCTCTG
ATATCAGAGACGTAGACACCCAATTCCACCAAGTTGACTCTTTCGTCAGATTGAGCTAGAGTGGTGGTTGCAGAAGCAGTAGCAGCGATGGCAGCGACACCAGCGGCGATTGAAGTTAATTTGACCATTGTATTTGTTTTG
TTTGTTAGTGCTGATATAAGCTTAACAGGAAAGGAAAGAATAAAGACATATTCTCAAAGGCATATAGTTGAAGCAGCTCTATTTATACCCATTCCCTCATGGGTTGTTGCTATTTAAACGATCGCTGACTGGCACCAGTTCCTC
ATCAAATATTCTCTATATCTCATCTTTCACACAATCTCATTATCTCTATGGAGATGCTCTTGTTTCTGAACGAATCATAAATCTTTCATAGGTTTCGTATGTGGAGTACTGTTTTATGGCGCTTATGTGTATTCGTATGCGCAGAATG
TGGGAATGCCAATTATAGGGGTGCCGAGGTGCCTTATAAAACCCTTTTCTGTGCCTGTGACATTTCCTTTTTCGGTCAAAAAGAATATCCGAATTTTAGATTTGGACCCTCGTACAGAAGCTTATTGTCTAAGCCTGAATTCA
GTCTGCTTTAAACGGCTTCCGCGGAGGAAATATTTCCATCTCTTGAATTCGTACAACATTAAACGTGTGTTGGGAGTCGTATACTGTTAGGGTCTGTAAACTTGTGAACTCTCGGCAAATGCCTTGGTGCAATTACGTAATTTT
AGCCGCTGAGAAGCGGATGGTAATGAGACAAGTTGATATCAAACAGATACATATTTAAAAGAGGGTACCGCTAATTTAGCAGGGCAGTATTATTGTAGTTTGATATGTACGGCTAACTGAACCTAAGTAGGGATATGAGAGT
AAGAACGTTCGGCTACTCTTCTTTCTAAGTGGGATTTTTCTTAATCCTTGGATTCTTAAAAGGTTATTAAAGTTCCGCACAAAGAACGCTTGGAAATCGCATTCATCAAAGAACAACTCTTCGTTTTCCAAACAATCTTCCCGA
AAAAGTAGCCGTTCATTTCCCTTCCGATTTCATTCCTAGACTGCCAAATTTTTCTTGCTCATTTATAATGATTGATAAGAATTGTATTTGTGTCCCATTCTCGTAGATAAAATTCTTGGATGTTAAAAAATTAAAGGGACTATATCT
AGTCAAGACGATACTGTCAGTAGCAGCGATGGCAGCGTGGCTTGTGGTAGCAACACTATCATGGTATCACTAACGTAAAAGTTCCTCAATATTGCAATTTGCTTGAACGGATGCTATTTCAGAATATTTCGTACTTACACAGG
CCATACATTAGAATAATATGTCACATCACTGTCGTAACACTCTTTATTCACCGAGCAATAATACGGTAGTGGCTCAAACTCATGCGGGTGCTATGATACAATTATATCTTATTTCCATTCCCATATGCTAACCGCAATATCCTAAAA
GCATAACTGATGCATCTTTAATCTTGTATGTGACACTACTCATACGAAGGGACTATATCTAGTCAAGACGATACTGTGATAGGTACGTTATTTAATAGGATCTATAACGAAATGTCAAATAATTTTACGGTAATATAACTTATCAGC
GGCGTATACTAAAACGGACGTTACGATATTGTCTCACTTCATCTTACCACCCTCTATCTTATTGCTGATAGAACACTAACCCCTCAGCTTTATTTCTAGTTACAGTTACACAAAAAACTATGCCAACCCAGAAATCTTGATATTTT
ACGTGTCAAAAAATGAGGGTCTCTAAATGAGAGTTTGGTACCATGACTTGTAACTCGCACTGCCCTGATCTGCAATCTTGTTCTTAGAAGTGACGCATATTCTATACGGCCCGACGCGACGCGCCAAAAAATGAAAAACGAA
GCAGCGACTCATTTTTATTTAAGGACAAAGGTTGCGAAGCCGCACATTTCCAATTTCATTGTTGTTTATTGGACATACACTGTTAGCTTTATTACCGTCCACGTTTTTCTAGCACCATATACTTACCACTCCATTTATGAATCAG
TACCAAATGCA

whole genome…

Where are the genes in this genome?

Answer

1

CCACACCACACCCACACACCCACACACCACACCACACACCACACCACACCCACACACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTC
GTTACCCTGTCCCATTCAACCATACCACTCCGAACCACCATCCATCCCTCTACTTACTACCACTCACCCACCGTTACCCTCCAATTACCCATATCCAACCCACTGCCACTTACCCTACCATTACCCTACCATCCACC**ATGACCTACTC**
ACCATACTGTTCTTCTACCCACCATATTGAAACGCTAACAAATGATCGTAAATAACACACACGTGCTTACCCTACCACTTTATACCACCACCACATGCCATACTCACCCTCACTTGTATACTGATTTTACGTACGCACACGGATG
CTACAGTATATACCATCTCAAACTTACCCTACTCTCAGATTCCACTTCACTCCATGGCCCATCTCTCACTGAATCAGTACCAAATGCACTCACATCATTATGCACGGCACTTGCCTCAGCGGTCTATACCCTGTGCCATTTACCC
ATAACGCCCATCATTATCCACATTTTGATATCTATATCTCATTCGGCGGTCCCAAATATTGTATAACTGCCCTTAATACATACGTTATACCACTTTTGCACCATATACTTACCACTCCATTTATATACACTTATGTCAATATTACAGAAAA
ATCCCCACAAAAATCACCTAAACATAAAAATATTCTACTTTTCAACAATAATACATAAACATATTGGCTTGTGGTAGCAACACTATCATGGTATCACTAACGTAAAAGTTCCTCAATATTGCAATTTGCTTGAACGGATGCTATTTC
AGAATATTTCGTACTTACACAGGCCATACATTAGAATAATATGTCACATCACTGTCGTAACACTCTTTATTCACCGAGCAATAATACGGTAGTGGCTCAAACTCATGCGGGTGCTATGATACAATTATATCTTATTTCCATTCCCATA
TGCTAACCGCAATATCCTAAAAGCATAACTGATGCATCTTTAATCTTGTATGTGACACTACTCATACG … TGATAGGTACGTTATTTAATAGGATCTATAACGAAATGTCAAATAATTT
TACGGTAATATAACTTATCAGCGGCGTATACTAAAACGGACGTTACGATATTGTCTCACTTCATCTTA … AACCCCTCAGCTTTATTTCTAGTTACAGTTACACAAAAAACTATGCCA

Protein coding sequence

ACCCAGAAATCTTGATATTTTACGTGTCAAAAAATGAGGGTCTCTAAATGAGAGTTTGGTACCATGACTTGTAACTCGCACTGCCCTGATCTGCAATCTTGTTCTTAGAAGTGACGCATATTCTATACGGCCCGACGCGACGCG
CCAAAAAATGAAAAACGAAGCAGCGACTCATTTTTATTTAAGGACAAAGGTTGCGAAGCCGCACATTTCCAATTTCATTGTTGTTTATTGGACATACACTGTTAGCTTTATTACCGTCCACGTTTTTTCTACAATAGTGTAGAA
GTTTCTTTCTTATGTTCATCGTATTCATAAAATGCTTCACGAACACCGTCATTGATCAAATAGGTCTATAATATTAATATACATTTATATAATCTACGGTATTTATATCATCAAAAAAAAAGTAGTTTTTTTATTTTATTTTGTTCGTTAAT
TTTCAATTTCTATGGAAACCCGTTCGTAAAATTGGCGTTTGTCTCTAGTTTGCGATAGTGTAGATACCGTCCTTGGATAGAGCACTGGAGATGGCTGGCTTTAATCTGCTGGAGTACCATGGAACACCGGTGATCATTCTGGTC
ACTTGGTCTGGAGCAATACCGGTCAACATGGTGGTGAAGTCACCGTAGTTGAAAACGGCTTCAGCAACTTCGACTGGGTAGGTTTCAGTTGGGTGGGCGGCTTGGAACATGTAGTATTGGGCTAAGTGAGCTCTGATATCA
GAGACGTAGACACCCAATTCCACCAAGTTGACTCTTTCGTCAGATTGAGCTAGAGTGGTGGTTGCAGAAGCAGTAGCAGCGATGGCAGCGACACCAGCGGCGATTGAAGTTAATTTGACCATTGTATTTGTTTTGTTTGTTA
GTGCTGATATAAGCTTAACAGGAAAGGAAAGAATAAAGACATATTCTCAAAGGCATATAGTTGAAGCAGCTCTATTTATACCCATTCCCTCATGGGTTGTTGCTATTTAAACGATCGCTGACTGGCACCAGTTCCTCATCAAATA
TTCTCTATATCTCATCTTTCACACAATCTCATTATCTCTATGGAGATGCTCTTGTTTCTGAACGAATCATAAATCTTTCATAGGTTTCGTATGTGGAGTACTGTTTTATGGCGCTTATGTGTATTCGTATGCGCAGAATGTGGGAATG
CCAATTATAGGGGTGCCGAGGTGCCTTATAAAACCCTTTTCTGTGCCTGTGACATTTCCTTTTTCGGTCAAAAAGAATATCCGAATTTTAGATTTGGACCCTCGTACAGAAGCTTATTGTCTAAGCCTGAATTCAGTCTGCTTTA
AACGGCTTCCGCGGAGGAAATATTTCCATCTCTTGAATTCGTACAACATTAAACGTGTGTTGGGAGTCGTATACTGTTAGGGTCTGTAAACTTGTGAACTCTCGGCAAATGCCTTGGTGCAATTACGTAATTTTAGCCGCTGAG
AAGCGGATGGTAATGAGACAAGTTGATATCAAACAGATACATATTTAAAAGAGGGTACCGCTAATTTAGCAGGGCAGTATTATTGTAGTTTGATATGTACGGCTAACTGAACCTAAGTAGGGATATGAGAGTAAGAACGTTCG
GCTACTCTTCTTTCTAAGTGGGATTTTTCTTAATCCTTGGATTCTTAAAAGGTTATTAAAGTTCCGCACAAAGAACGCTTGGAAATCGCATTCATCAAAGAACAACTCTTCGTTTTCCAAACAATCTTCCCGAAAAAGTAGCCG
TTCATTTCCCTTCCGATTTCATTCCTAGACTGCCAAATTTTTCTTGCTCATTTATAATGATTGATAAGAATTGTATTTGTGTCCCATTCTCGTAGATAAAATTCTTGGATGTTAAAAAATTAAAGGGACTATATCTAGTCAAGACGAT
ACTGTCAGTAGCAGCGATGGCAGCGTGGCTTGTGGTAGCAACACTATCATGGTATCACTAACGTAAAAGTTCCTCAATATTGCAATTTGCTTGAACGGATGCTATTTCAGAATATTTCGTACTTACACAGGCCATACATTAGAAT
AATATGTCACATCACTGTCGTAACACTCTTTATTCACCGAGCAATAATACGGTAGTGGCTCAAACTCATGCGGGTGCTATGATACAATTATATCTTATTTCCATTCCCAT**ATGCTAACCGCAATATCCTAAAAGCATAACTGATGC**
ATCTTTAATCTTGTATGTGACACTACTCATACGAAGGGACTATATCTAGTCAAGACGATACTGTGATAGGTACGTTATTTAATAGGATCTATAACGAAATGTCAAATAATTTTACGGTAATATAACTTATCAGCGGCGTATACTA
AAACGGACGTTACGATATTGTCTCACTTCATCTTACCACCCTCTATCTTATTGCTGATAGAACACTAACCCCTCAGCTTTATTTCTAGTTACAGTTACACAAAAAACTATGCCAACCCAGAAATCTTGATATTTTACGTGTCAA
AAAATGAGGGTCTCTAAATGAGAGTTTGGTACCATGACTTGTAACTCGCACTGCCCTGATCTGCAATCTTGTTCTTAGAAGTGACGCATATTCTATACGGCCCGACGCGACGCGCCAAAAAATGAAAAACGAAGCAGCGACT
CATTTTTATTTAAGGACAAAGGTTGCGAAGCCGCACATTTCCAATTTCATTGTTGTTTATTGGACATACACTGTTAGCTTTATTACCGTCCACGTTTTTTCTAGCACCATATACTTACCACTCCATTTATGAATCAGTACC

Bioinformatics Applications (contd …)

Healthy Individual 2

>gi|28302128|ref|NM_000518.4| Homo sapiens hemoglobin, beta (HBB), mRNA

ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGA
GGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGC
AGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATG
CTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGC
TCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGAT
CCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTCA
CCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCACAAGTATCA
CTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACTAAACT
GGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGC

>gi|4504349|ref|NP_000509.1| beta globin [Homo sapiens]

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG
AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH

Bioinformatics Applications (contd …)

Diseased Individual (Mutation)

2

>gi|28302128|ref|NM_000518.4| Homo sapiens hemoglobin, beta (HBB), mRNA

ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGA

GGTGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGC

AGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATG

CTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGC

TCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGAT

CCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTCA

CCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCACAAGTATCA

CTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACTAAACT

GGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGC

>gi|4504349|ref|NP_000509.1| beta globin [Homo sapiens]

MVHLTPVEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG

AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVAN

ALAHKYH

Random Sample

whole Data

3

atgaccgggatactgataccgtatttggcctaggcgtacacattagataaacgtatgaagtacgttagactcggcgc

acccctatttttgagcagatttagtgacctggaaaaaaatttgagtacaaacttttccgaatactgggcataag

tgagtatccctgggatgacttttgggaacactatagtgctctcccgatttttgaatatgtaggatcattcgccagggtccga

gctgagaattggatgaccttgtaagtgttttccacgcaatcgcgaaccaacgcggacccaaaggcaagaccgataaaggaga

tcccttttgcggtaatgtgccgggaggctggttacgtagggaagccctaacggacttaatggcccacttagtccacttatag

gtcaatcatgttcttgtgaatggattttaactgagggcatagaccgcttggcgcacccaaattcagtgtgggcgagcgcaa

cggttttggcccttgttagaggcccccgtactgatggaaactttcaattatgagagagctaatctatcgcgtgcgtgttcat

aacttgagttggtttcgaaaatgctctggggcacatacaagaggagtcttccttatcagttaatgctgtatgacactatgta

ttggcccattggctaaaagcccaacttgacaaatggaagatagaatccttgcatttcaacgtatgccgaaccgaaagggaag

ctggtgagcaacgacagattcttacgtgcattagctcgcttccggggatctaatagcacgaagcttctgggtactgatagca

Implanting Motif AAAAAAAGGGGGGG

atgaccgggatactgatAAAAAAAAGGGGGGGggcgtacacattagataaacgtatgaagtacgttagactcggcgccgccg

acccctattttttgagcagatttagtgacctggaaaaaaaatttgagtacaaaacttttccgaataAAAAAAAAGGGGGGGa

tgagtatccctgggatgacttAAAAAAAAGGGGGGGtgctctcccgattttgaatatgtaggatcattcgccagggtccga

gctgagaattggatgAAAAAAAAGGGGGGGtccacgcaatcgcgaaccaacgcggacccaaaggcaagaccgataaaggaga

tcccttttgcggtaatgtgccgggaggctggttacgtagggaagccctaacggacttaatAAAAAAAAGGGGGGGcttatag

gtcaatcatgttcttgtgaatggatttAAAAAAAAGGGGGGGgaccgcttggcgcacccaaattcagtgtgggcgagcgcaa

cggttttggcccttgttagaggcccccgtAAAAAAAAGGGGGGGcaattatgagagagctaatctatcgcgtgcgtgttcat

aacttgagttAAAAAAAAGGGGGGGctggggcacatacaagaggagtcttccttatcagttaatgctgtatgacactatgta

ttggcccattggctaaaagcccaacttgacaaatggaagatagaatccttgcatAAAAAAAAGGGGGGGaccgaaagggaag

ctggtgagcaacgacagattcttacgtgcattagctcgcttccggggatctaatagcacgaagcttAAAAAAAAGGGGGGGa

Where is the Implanted Motif?

atgaccgggatactgataaaaaaaaggggggggcgtacacattagataaacgtatgaagtacgttagactcggcgccgccg

acccctattttttgagcagatttagtgacctggaaaaaaaatttgagtacaaaacttttccgaataaaaaaaaaggggggga

tgagtatccctgggatgacttaaaaaaaaggggggggtgctctcccgatttttgaatatgtaggatcattcgccagggtccga

gctgagaattggatgaaaaaaaagggggggtccacgcaatcgcgaaccaacgcggacccaaaggcaagaccgataaaggaga

tcccttttgcggtaatgtgccgggaggctggttacgtagggaagccctaacggacttaataaaaaaaaggggggcttatag

gtcaatcatgttcttgtgaatggatttaaaaaaaaggggggggaccgcttggcgcacccaaattcagtgtgggcgagcgcaa

cggttttggcccttgttagaggcccccgtaaaaaaaagggggggcaattatgagagagctaatctatcgcgtgcgtgttcat

aacttgagttaaaaaaaaagggggggctggggcacatacaagaggagtcttccttatcagttaatgctgtatgacactatgta

ttggcccattggctaaaagcccaacttgacaaatggaagatagaatccttgcataaaaaaaaagggggggaccgaaagggaag

ctggtgagcaacgacagattcttacgtgcattagctcgcttccggggatctaatagcacgaagcttaaaaaaaaaggggggga

Implanting Motif AAAAAAGGGGGG with Four Mutations 3

atgaccgggatactgatAgAAgAAAGGttGGGggcgtacacattagataaacgtatgaagtacgttagactcggcgccgccg
acccctatttttgagcagatttagtgacctggaaaaaaaatttgagtacaaacttttccgaatacAAtAAAAcGGcGGGa
tgagtatccctgggatgacttAAAAtAAtGGaGtGGtgctctcccgatttttgaatatgtaggatcattcgccagggtccga
gctgagaattggatgcAAAAAAAGGGattGtccacgcaatcgcgaaccaacgcggacccaaaggcaagaccgataaaggaga
tcccttttgcggtaatgtgccgggaggctggttacgtagggaagccctaacggacttaatAtAAtAAAGGaaGGGcttatag
gtcaatcatgttcttgtgaatggatttAAcAAtAAGGGctGGgaccgcttggcgcacccaaattcagtgtgggcgagcgcaa
cggttttggcccttgttagaggcccccgtAtAAAcAAGGaGGGccaattatgagagagctaatctatcgcgtgcgtgttcat
aacttgagttAAAAAAtAGGGaGccctggggcacatacaagaggagtcttccttatcagttaatgctgtatgacactatgta
ttggcccattggctaaaagcccaacttgacaaatggaagatagaatccttgcatActAAAAAGGaGcGGaccgaaagggaag
ctggtgagcaacgacagattcttacgtgcattagctcgcttccggggatctaatagcacgaagcttActAAAAAGGaGcGGa

What can I do as computer science?

In the beginning should be proficient in one of the programming languages that can be programmed : Matlab is the easiest of the programming languages because it has contain functions (tools) that can be use. We may need a language that deals with databases as (Foxpro,Oracl ,..).

The programmer can do the following:

- Bioinformatics use certain training of computer science such as AI, neural networks, genetic algorithms, dynamic programming to accelerate and enhance biological research.

What can I do as computer science (contd…)?

- One of the most basic operations in bioinformatics involves searching for similarities, or homologies, between a newly sequenced piece of DNA and another sequenced DNA segments from various organisms can be use computer techniques to solve this operations.

- Correspondingly, there is a need to think about changes in hardware and software configurations of the present day computers, as they would very soon become redundant. Watch for days in near future when a multi-processor computer would be available for the price of a personal computer…

What can I do as computer science (contd…)?

the programmer can do the following (specific)

- ✓ Detection of the location of mutations depending on the **neural network (genetic programming).**
- ✓ **Statistic** pattern (codon) in a sequence .
- ✓ Distinguish patterns (**recognize**)   for different sequence .
- ✓ Comparison of two or more sequences based on **neural networks.**
- ✓ Sequence alignment based on **artificial intelligence.**

What can I do as computer science (contd…)?

- ✓ Design an **integrated** system to handle a different database (fast and ease)
- ✓
- ✓ Comparison of two different amino acid structures (**image processing**)
- ✓ **Design** protein and nucleic acid structural
- ✓ **Image processing** for output microarray
- ✓ **Simulation** protein-protein interaction.
- ✓ Etc……

Some Simple Bioinformatics Applications (Research's) for Computer Science

Bioinformatics Applications (Research's) for Computer Science

DNA Sequence Assembly using Particle Swarm Optimization

Ravi Shankar Verma
National Institute of Technology
Raipur, India

Vikas Singh
ABV- Indian Institute of Information Technology and management, Gwalior, India

Sanjay Kumar
National Institute of Technology
Raipur, India

AI

Applications of artificial intelligence in bioinformatics: A review

Zoheir Ezziane

College of Information Technology, P.O. Box 14143, Dubai, United Arab Emirates

A Genetic Algorithm Approach to Solving DNA Fragment Assembly Problem

Shu-Cherng Fang,[*,†] Yong Wang, and Jie Zhong

Operations Research and Industrial Engineering, North Carolina State University, Raleigh, NC 27695-7906, USA

Bioinformatics Applications (Research's) for Computer Science

Data and text mining

Bioimage informatics: a new area of engineering biology

Hanchuan Peng
Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, USA

Received on June 6, 2008; revised on July 1, 2008; accepted on July 2, 2008
Advance Access publication July 4, 2008
Associate Editor: Jonathan Wren

Image processing

The role of Digital Image Processing and pattern recognition in the Bioinformatics."

Dr. Evelio Luis Báez Pérez, Centro de Bioplantas,Universidad de Ciego de Avila

Bioinformatics Applications (Research's) for Computer Science

International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 4, August 2011

DNA Lossless Differential Compression Algorithm Based on Similarity of Genomic Sequence Database

Heba Afify[1], Muhammad Islam[1] and Manal Abdel Wahed[1]

[1]Department of Systems and Biomedical Engineering, Cairo University, Egypt
hebaaffify@yahoo.com, Manalaw2003@yahoo.com

Big Data in Bioinformatics & the Era of Cloud Computing

Prakash Nemade[1], Heena Kharche[2]

[1]*(Department of Bioinformatics, Maulana Azad National Institute of Technology, Bhopal MP, India)*
[2]*(Department of Computer Science & Engineering, IES-IPS Academy, Indore MP, India)*

Data management

Bioinformatics Applications (Research's) for Computer Science

Annals of the University of Craiova, Mathematics and Computer Science Series
Volume 37(3), 2010, Pages 147–151
ISSN: 1223-6934

Aspects of DNA Cryptography

Calina Popovici

Security

Good Practice in (Pseudo) Random Number Generation for Bioinformatics Applications

David Jones, UCL Bioinformatics Group

(E-mail: d.jones@cs.ucl.ac.uk)

(Last revised May 7th 2010)

Using Cloning Technique for hiding secret message

Sadoon H.Abdullah
Ass. Lecturer
University of Mosul/ Collage of Science/Biology Dep.

Thanks

BioInformatics Group