

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

## Welcome to Biol-360 Bioinformatics

### Lec'01'slides

# Your challenge: Cure a patient

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology  
Summary

- 30 month old boy, seemed normal until 15 months old
- Extremely severe inflammatory gastrointestinal disease (seems like Crohn's disease but much worse)
- Numerous surgeries failed
- Standard drug treatments failed
- Failing to grow from malabsorption. **Will likely die.**

**You're the doc. How do you find the underlying cause?**

Think creatively & write **several** ideas on paper.

# Welcome to Biol-360 Bioinformatics

## Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

- “**Job market projections are extremely positive** for bioinformatics professionals”
- “According to the U.S. Bureau of Labor Statistics, **bioinformatics is a particularly vibrant new area of work, and job opportunities in this area are expected to have the highest growth**”
- What is the market like for computational biologists?  
“**There is a huge demand**, but not that many people who marry a deep understanding of molecular biology with the ability to think in terms of algorithms. **It's really a seller's market for people with these skills.** As a result, I'm in the enviable position of getting to pick and choose what I want to work on.”

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology  
Summary

## 1 Welcome

## 2 Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

Summary

# Three driving forces behind bioinformatics

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

- Massive volumes of DNA sequence data
- Gene and protein conservation between species
- Systems biology

# Three driving forces behind bioinformatics

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

- Massive volumes of DNA sequence data
- Gene and protein conservation between species
- Systems biology

# Three driving forces behind bioinformatics

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

- Massive volumes of DNA sequence data
- Gene and protein conservation between species
- Systems biology

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

## ② Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

Summary



# DNA sequence data is accumulating exponentially

Welcome

Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

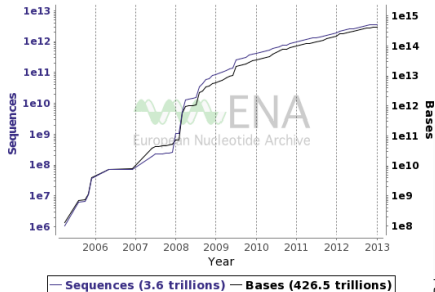
Gene are conserved through evolution

Systems biology

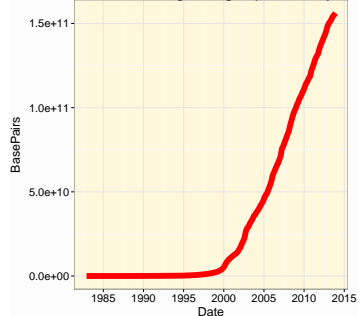
Summary

## Sequence Read Archive (SRA) Growth

28-Jan-2013



## Genbank is growing exponentially



# 1st generation DNA sequencing reaction

Welcome

Three driving forces behind bioinformatics

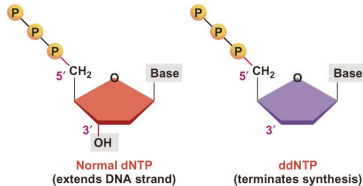
Massive volumes of DNA sequence data

Gene are conserved through evolution

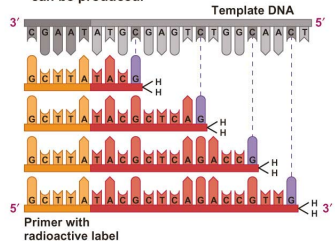
Systems biology

Summary

(a) ddNTPs terminate DNA synthesis.



(b) Using ddNTPs, daughter strands of different length can be produced.



Called:

- Dideoxy sequencing
- Chain termination sequencing
- Sanger DNA sequencing

# Lasers and computers read the DNA sequence

Welcome

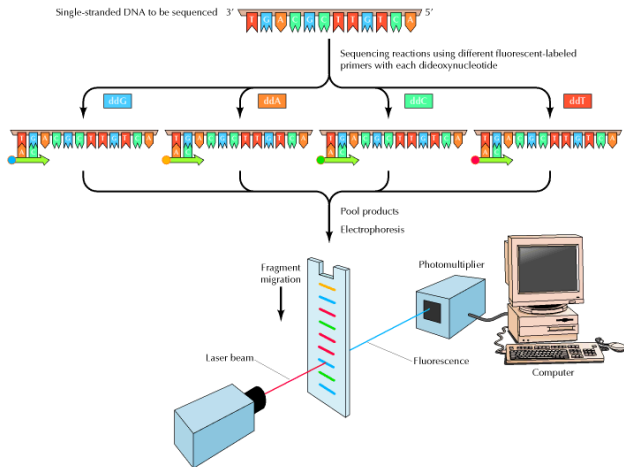
Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

Summary



# DNA sequencer produces chromatograms

## Welcome

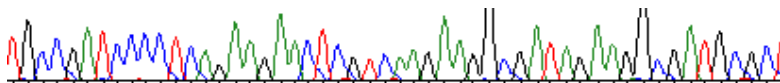
Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

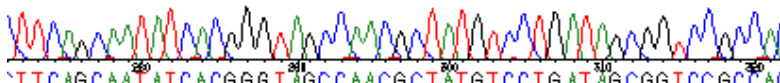
Gene are conserved through evolution

Systems biology

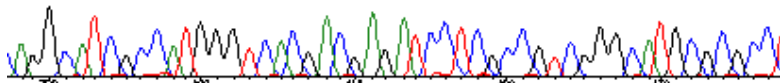
Summary



T G C C G A T C C C C T C A G A A G A A C T C G T C A A G A A G G C G A T A G A A G G C G A T G C G C



T T C A G C A A T A T C A C G G G T A G C C A A C G C T A T G T C C T G A T A G C G G T C C G C C A



A G G C A T C G C C A T G G G T C A C G A C G A G A T C C T C G C C G T C G G G C A T G C G C G C C

# 1<sup>st</sup> generation DNA sequencing robot

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary



# Many DNA sequencing robots

Welcome

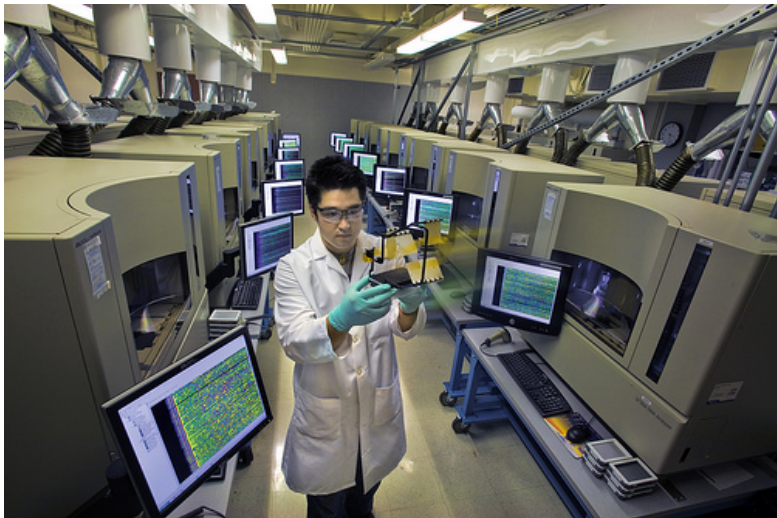
Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary



# Illumina next-generation sequencing I

## Welcome

Three driving forces behind bioinformatics

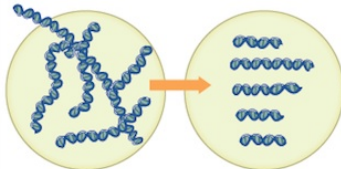
Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

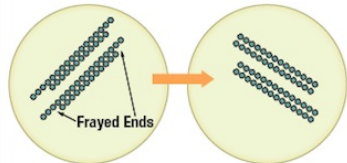
Summary

## 1 SONICATION



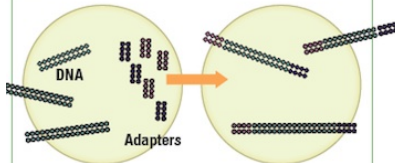
Genomic DNA is fragmented into 100-500 base pair fragments by sonication to create a library.

## 2 FRAGMENT END REPAIR



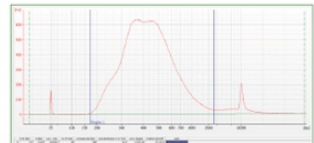
Sonication creates frayed DNA ends which must be blunted or repaired.

## 3 A-TAILING AND ADAPTER LIGATION



Adapters are ligated to each end of the A-tailed DNA fragment.

## 4 QC CHECK



The electropherogram shows the size and concentration of the final library. This library size also confirms the ligation of adapters.

# Illumina next-generation sequencing II

## Welcome

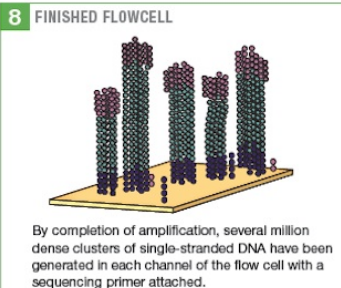
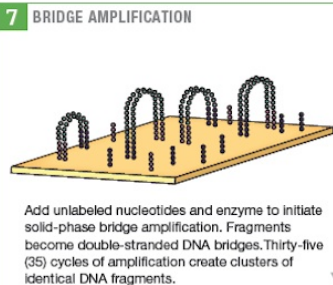
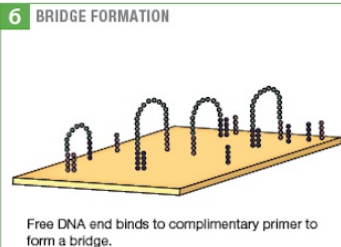
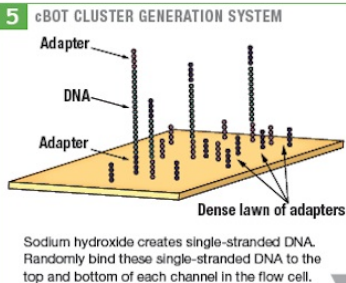
Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

Summary





# Illumina next-generation sequencing III

Welcome

Three driving forces behind bioinformatics

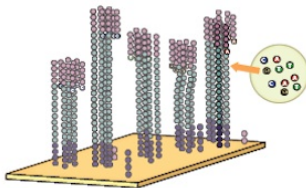
Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

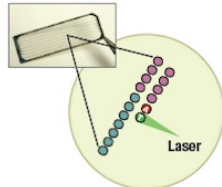
Summary

## 9 DNA SEQUENCING



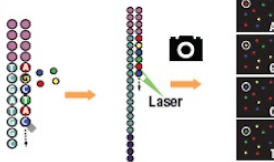
To initiate the first sequencing cycle and determine the first base, all four labeled reversible terminators and DNA polymerase enzyme are first added. Only one base can incorporate at a time.

## 10 BASE CALLING



Lasers excite the fluorescent tags and the images are captured via CCD camera. The identity of the first base in each cluster is recorded, and then the fluorescent tag is removed.

## 11 SEQUENCING-BY-SYNTHESIS



In the first cycle, the first base is incorporated. Its identity is determined by the signal given off and then recorded. In subsequent cycles, the process of adding sequencing reagents, removing unincorporated bases and capturing the signal of the next base to identify is repeated.

## 12 DUAL FLOW CELLS



Once the top surface of the flow cell channel has been scanned, the imaging step is repeated on the bottom surface.

Illumina

# Next-generation sequencing is **massively parallel**

## Welcome

Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

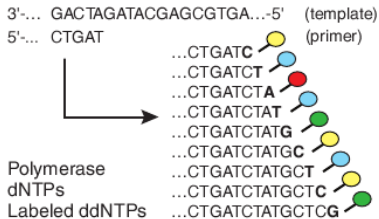
Gene are conserved through evolution

Systems biology

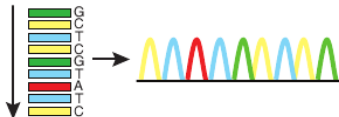
Summary

## Sanger sequencing

### Cycle sequencing

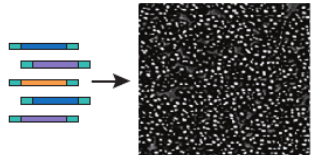


### Electrophoresis (1 read/capillary)



## Next-generation sequencing

### Generation of polony array



### Cyclic array sequencing ( $>10^6$ reads/array)



What is base 1? What is base 2? What is base 3?

# Next-generation sequencing produces enormous amounts of data

Welcome

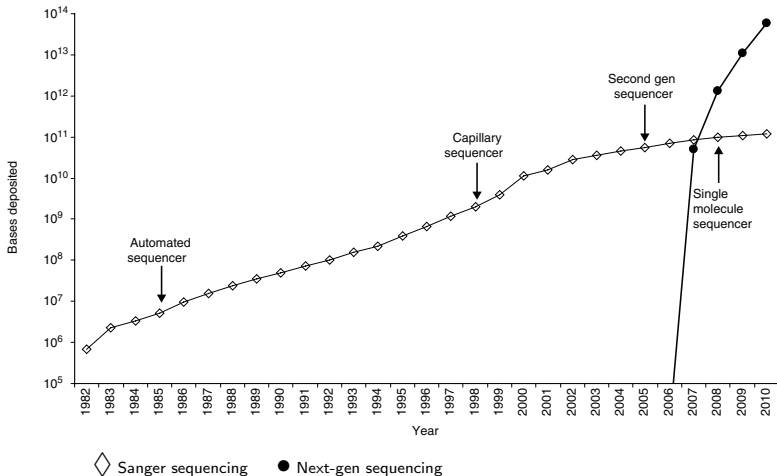
Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

Summary



# DNA sequencing is incredibly cheap and routine

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology  
Summary



## Rates for Partners HealthCare Users

### Sequencing

Order format	Price per sample	Difficult template (per sample)
Tubes or partial plate	\$2.95	\$3.42
96 samples in plate format	<b>\$1.95</b>	\$2.42

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology  
Summary

## Los Angeles Times

What a \$1,000 genome could mean for medicine

January 10, 2012

By Eryn Brown, Los Angeles Times

Two biotechnology firms said Tuesday that they would soon offer machines capable of *sequencing a human genome in about a day, at a cost of less than \$1,000.*

Illumina Inc. and Life Technologies Corp. both announced that their new sequencers would come to market in the second half of this year.

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

## ② Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

Summary

# Evolution conserves sequence and 3-D protein structure

Welcome

Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

Summary

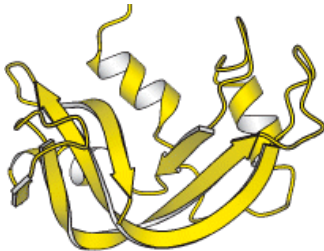
## Bovine ribonuclease sequence

GCAGAACTG-CCTTC--TCTCTCTCAGACATCAAAGTAGAGACCCAGGTT

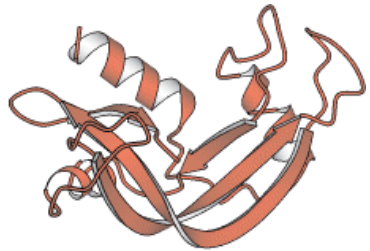
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

GCAGAACTGGCCTTCCATCTCTCTCAGACACCAAGCTGCAGATCCAGGCT

## Human ribonuclease sequence



**Bovine ribonuclease**



**Human ribonuclease**

# We can infer evolution from DNA sequences

## Welcome

Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

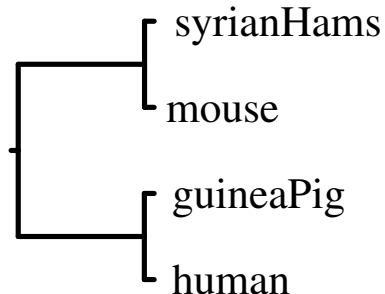
Summary

```

mouse      (M)  TGTAGAGCACAA---ATCAGAGACTGGCAGGATTTGCAAGAAGGAGAGT
syrianHams (S)  TGTAGAGCGT-----TCAGAGATGGACTGGACCTGCAAGAAGGAGAGT
human      (H)  TGTAGAGCAGCAAATTTTCAGAGAAGGCCTGAACTTACAAGAAGGAGAAT
guineaPig  (G)  TGTGGAGCACCG---TTCAGAAGTCGTTTGAACCTTGCAAGAAGGAGAAT
                ***  ****                *****      *   * *   * *****
    
```

DNA sequence  
differences  
(relative units)

	H	M	G	S
H	0.0	3.6	0.3	3.6
M		0.0	4.7	0.3
G			0.0	3.8
S				0.0





Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

## ② Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

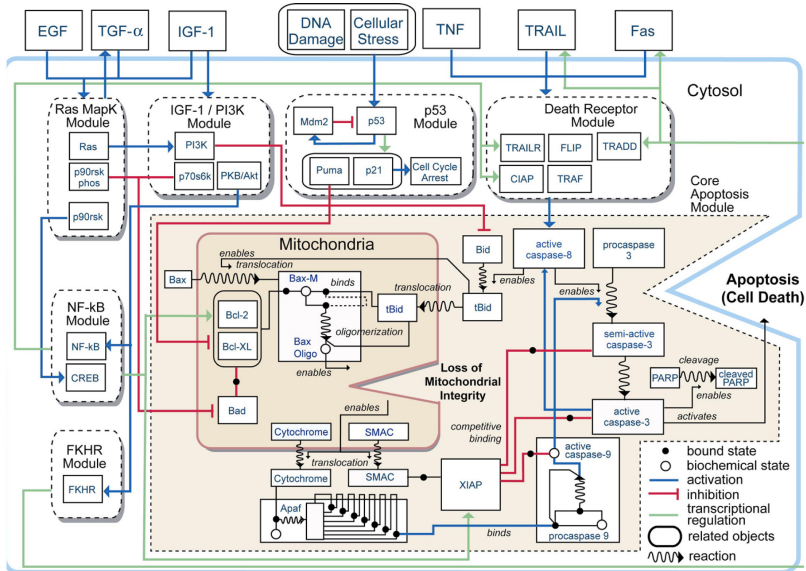
Summary

# Biologists want to understand complete systems

Welcome

Three driving forces behind bioinformatics

Massive volumes of DNA sequence data  
Gene are conserved through evolution  
Systems biology  
Summary



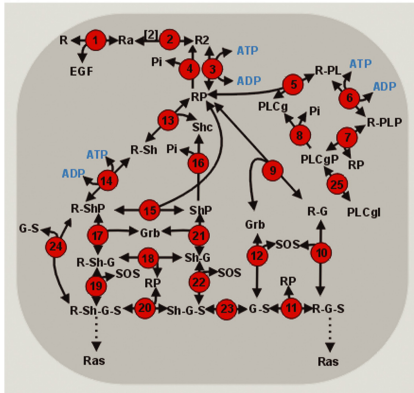
# Biologists want to model systems

## Welcome

### Three driving forces behind bioinformatics

Massive volumes of DNA sequence data  
Gene are conserved through evolution  
Systems biology  
Summary

(A)

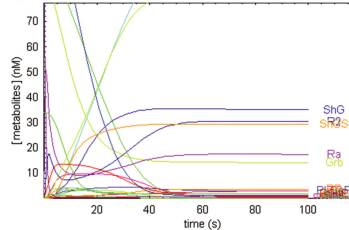


(B)

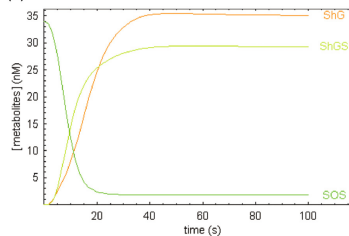
$$V4 = \frac{V4RP[t]}{K4 + RP[t]}$$

$$v13 = k13f RP[t] Shc[t] - k13b RSh[t]$$

(C)



(D)



# Biology is shifting from single genes to full systems

## Welcome

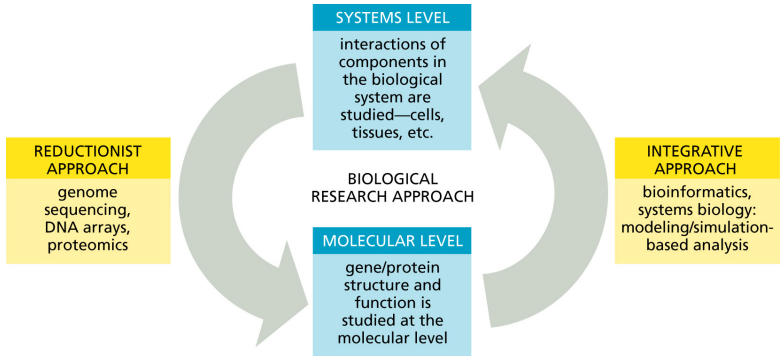
Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

Summary



Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

## ② Three driving forces behind bioinformatics

Massive volumes of DNA sequence data

Gene are conserved through evolution

Systems biology

Summary

# Three driving forces behind bioinformatics

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

- Massive volumes of DNA sequence data
- Gene and protein conservation between species
- Systems biology

# Three driving forces behind bioinformatics

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

- Massive volumes of DNA sequence data
- Gene and protein conservation between species
- Systems biology

# Three driving forces behind bioinformatics

Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

- Massive volumes of DNA sequence data
- Gene and protein conservation between species
- Systems biology



Welcome

Three driving  
forces behind  
bioinformatics

Massive volumes  
of DNA  
sequence data

Gene are  
conserved  
through  
evolution

Systems biology

Summary

- 1 Read Chapter 1 of Understanding Bioinformatics:  
<http://tinyurl.com/UB-sample-ch1-pdf>

*If this chapter is really unfamiliar, please contact me.*

- 2 Be ready to discuss this chapter on Thursday.
- 3 Slides for Thursday will posted before lecture.