

Course Overview

CSCI4181/6802 Bioinformatics Algorithms
Finlay Maguire (finlay.maguire@dal.ca)

Why am I teaching this course?

Maguire Lab Overview

Genomic Epidemiology

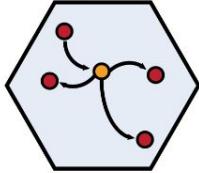
Sequencing Data



Clinical Data



Epidemiological Data



Affiliations:



Computer Science
Community Health & Epidemiology
Institute of Comparative Genomics



Sunnybrook
HEALTH SCIENCES CENTRE
Sunnybrook Research Institute
Shared Hospital Laboratory



Public Health Alliance for
Genomic Epidemiology

Maguire Lab Overview

Genomic Epidemiology

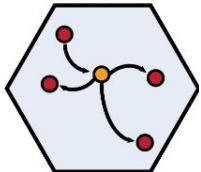
Sequencing Data



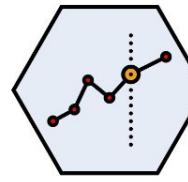
Clinical Data



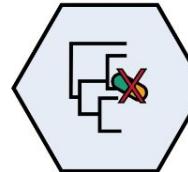
Epidemiological Data



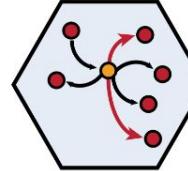
Epidemiological Modelling



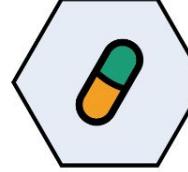
Pathogen Surveillance



Outbreak Prevention



Diagnostics & Prescribing



Affiliations:



Computer Science
Community Health & Epidemiology
Institute of Comparative Genomics



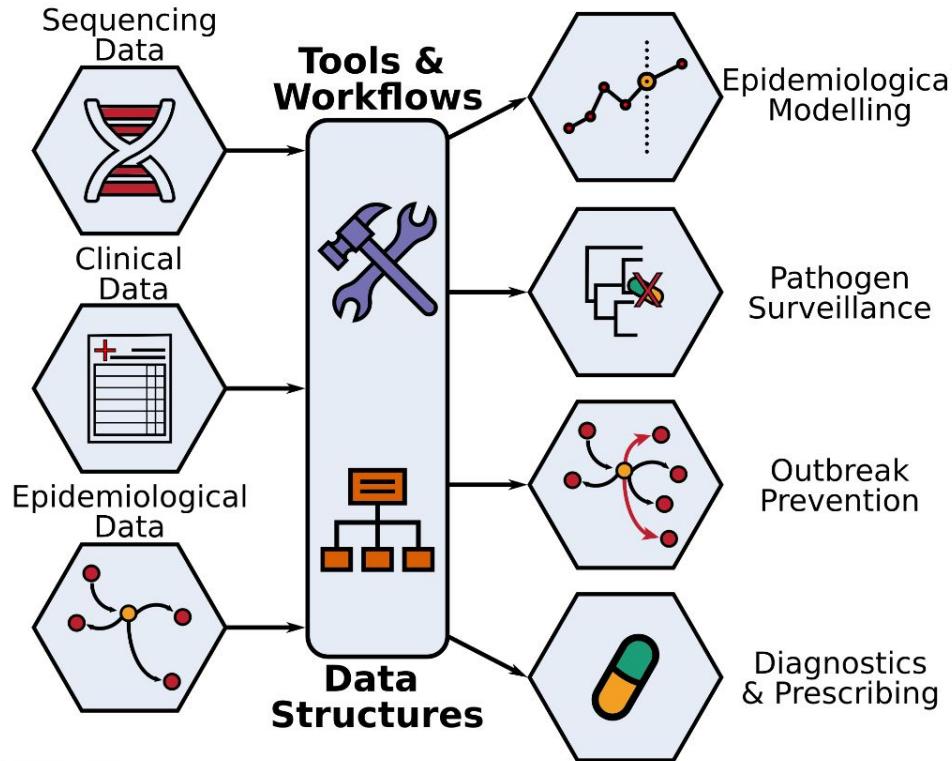
Sunnybrook
HEALTH SCIENCES CENTRE
Sunnybrook Research Institute
Shared Hospital Laboratory



Public Health Alliance for
Genomic Epidemiology

Maguire Lab Overview

Genomic Epidemiology



Affiliations:



Computer Science
Community Health & Epidemiology
Institute of Comparative Genomics



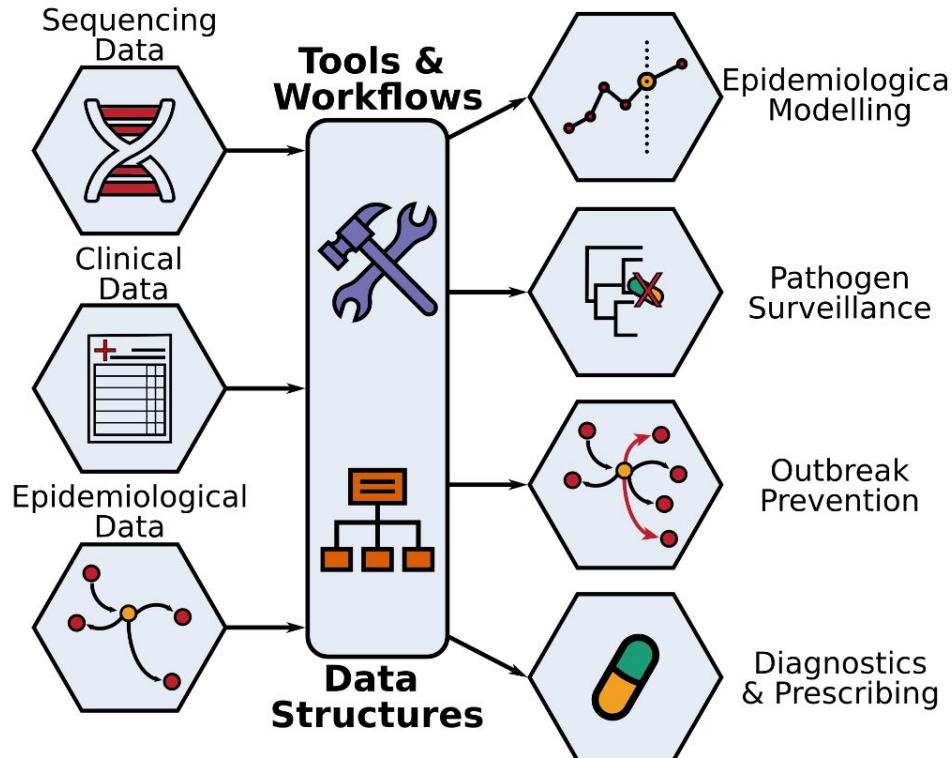
Computer Science
Community Health & Epidemiology
Institute of Comparative Genomics



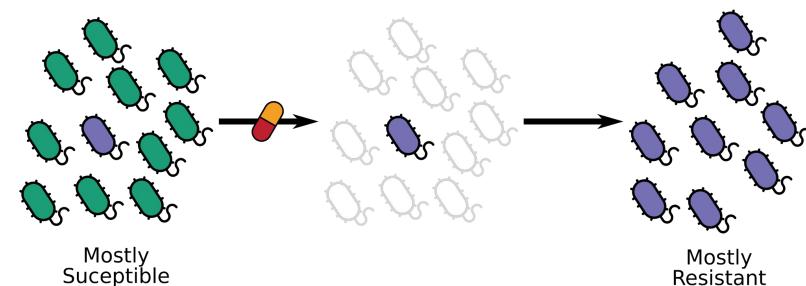
Public Health Alliance for
Genomic Epidemiology

Maguire Lab Overview

Genomic Epidemiology



Antimicrobial Resistance



Affiliations:



Computer Science
Community Health & Epidemiology
Institute of Comparative Genomics



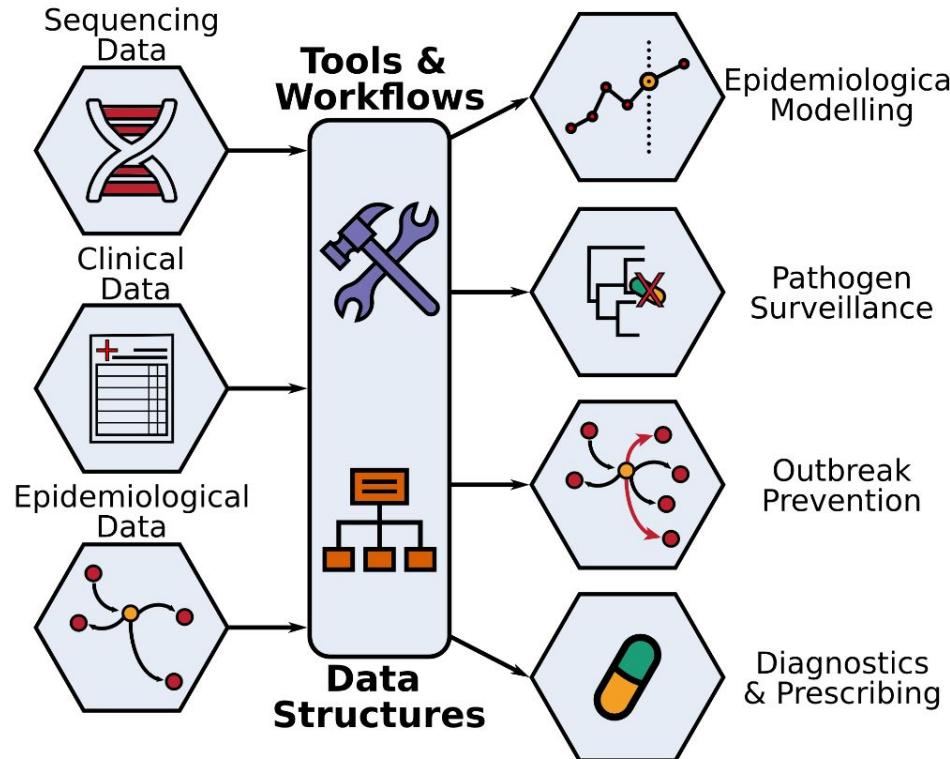
Sunnybrook Research Institute
Shared Hospital Laboratory



Public Health Alliance for
Genomic Epidemiology

Maguire Lab Overview

Genomic Epidemiology



Affiliations:



Computer Science
Community Health & Epidemiology
Institute of Comparative Genomics

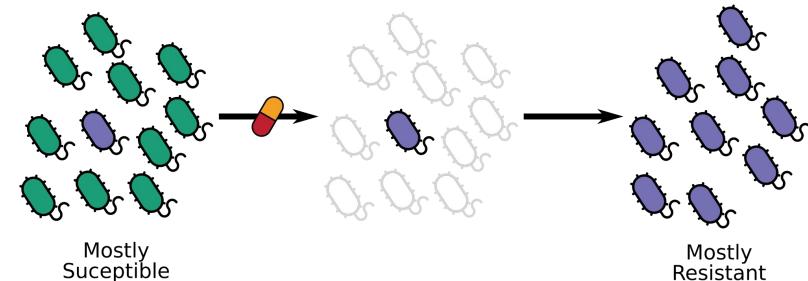


Sunnybrook Research Institute
Shared Hospital Laboratory

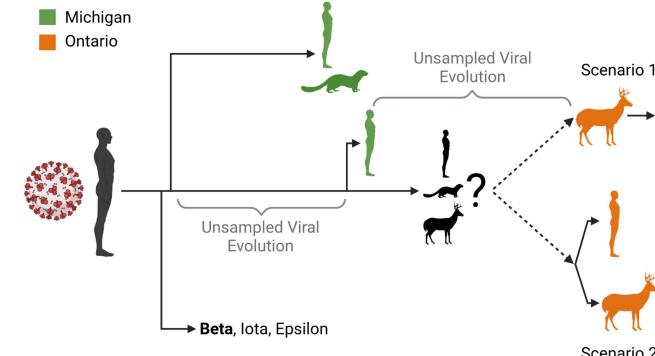


Public Health Alliance for
Genomic Epidemiology

Antimicrobial Resistance



Zoonotic Viruses



2020

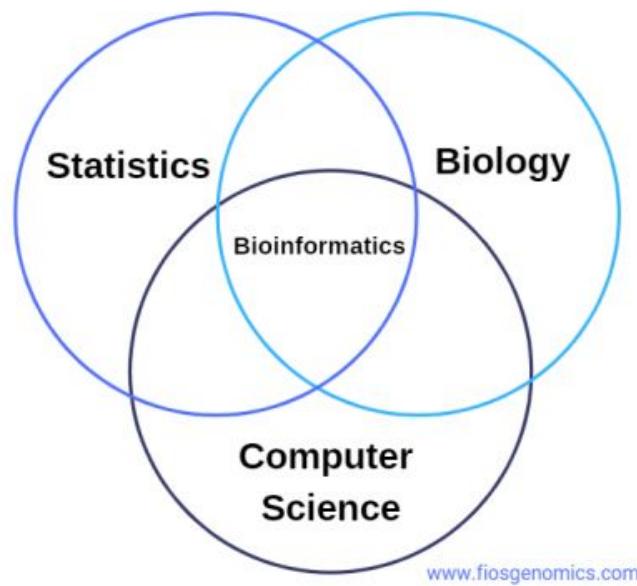
2021

2022

What is bioinformatics?

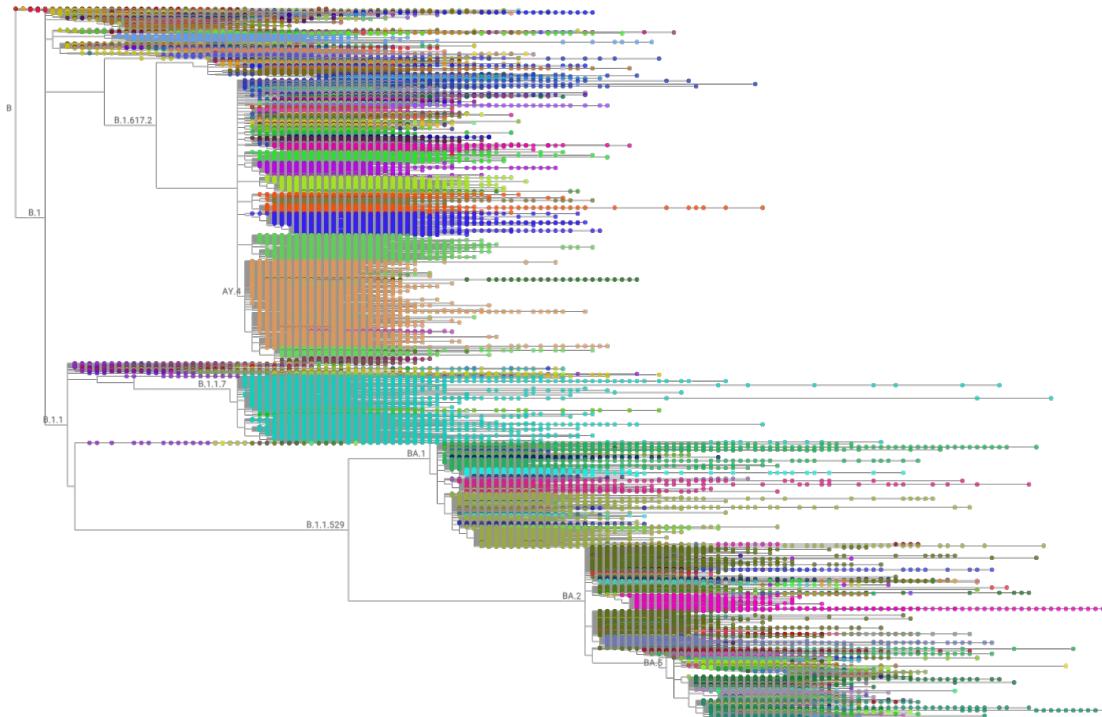
Using **computers** to understand **biology**

Using **computers** to understand **biology**



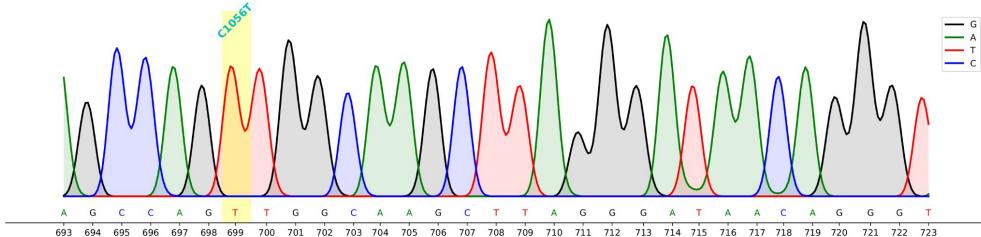
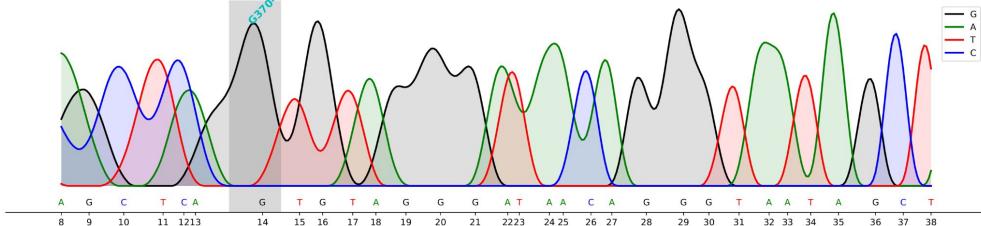
Why do we care about algorithms in
bioinformatics?

Lots of data



Cov2tree: taxonum 13.4 million SARS-CoV-2 Genomes

Lots of messy data



<https://github.com/y9c/cfutils>

Book2 - Excel

	A	B	C	D	E
1	MX1	HDAC5			
2	FZD1	MYC			
3	1-Mar	IL8			
4	PSEN2	1-Dec			
5	RBPJ	WNT5B			
6	PTPRN2	WNT6			
7	15-Sep	INF2			
8	CUL1	AGO2			
9					

<https://cosmosmagazine.com/science/excel-autocorrect-errors-still-plague-genetic-research/>

Lots of messy **biological** data

Biological data is special:

Lots of messy **biological** data

Biological data is special:

- Data is shaped by evolution i.e., “**what works at a given time**” not “what is pretty and easy to understand”

Lots of messy **biological** data

Biological data is special:

- Data is shaped by evolution i.e., “**what works at a given time**” not “what is pretty and easy to understand”
- Correct answer is often not obvious - we need **statistics**

Lots of messy **biological** data

Biological data is special:

- Data is shaped by evolution i.e., “**what works at a given time**” not “what is pretty and easy to understand”
- Correct answer is often not obvious - we need **statistics**
- Can’t agree on many of the objects and questions!

Lots of messy **biological** data

Biological data is special:

- Data is shaped by evolution i.e., “**what works at a given time**” not “what is pretty and easy to understand”
 - Correct answer is often not obvious - we need **statistics**
 - Can’t agree on many of the objects and questions!
- Philosophy of Biology** infuses what we do in many interesting ways

What are we actually going to learn?

1. |

2. |

|

|

3. |

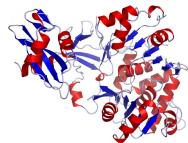
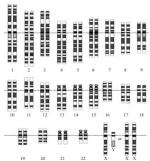
4. |

5. |

|

|

1. Introduction: Biological foundations



2.

1

2

3.

1

4.

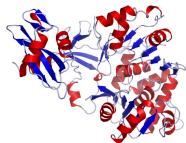
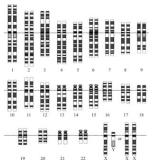
1

5.

1

2

1. Introduction: Biological foundations

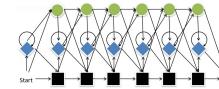
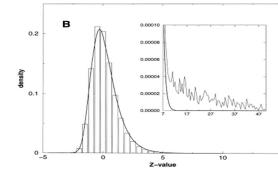


2. Homology: Comparing biological sequences

K-ELQRAASLTIEV

KDEGQK--SLVIDV

Amino acid	A	W	G	H	E	
A	0	-5	-10	-15	-20	-25
W	-5	2	-3	-8	-13	-18
G	-10	-3	19	14	9	4
H	-15	-8	14	17	20	15
E	-20	-13	9	14	18	24
A	-25	-18	4	10	13	19



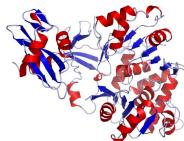
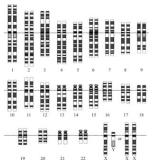
0	6: \$googo l
1	3: go1\$go o
2	0: googo1 \$
3	5: l\$goog o
4	2: ogol\$g o
5	4: ol\$goo g
6	1: oogol\$ g

3.

4.

5.

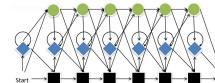
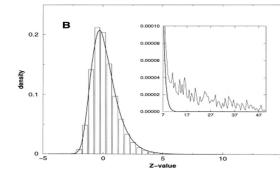
1. Introduction: Biological foundations



2. Homology: Comparing biological sequences

K-ELQRAASLTIEV

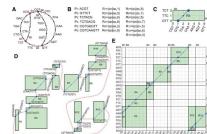
KDEGQK--SLVIDV



0	6	\$googo
1	3	gol\$go
2	0	googol \$
3	5	l\$goog
4	2	ogol\$g
5	4	ol\$goo
6	1	oogol\$

3. Assembly: Recovering genomes from sequencing data

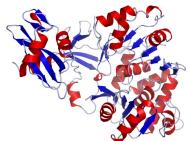
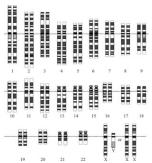
Reference	CCTAAAGAGATCCGGCTCTTAQCGGATAATAACAGCGGAATCTAAGCGGATATGCCAACACAG
Reads	CCTAAAGAGATCCGGCTCTTAQCGGATAATAACAGCGGAATCTAAGCGGATATGCCAACACAG GAGCGGATTCCTC ATTCGCGGATTC GCGTATCGGGG CTTACGGGGATAT TAGGGGGATAT TATAATAGAGCG GCGGAAATCT GAACTTGTACG CGGGATTTACG TCTTGATGGG CGGGAAATTCG AGCGGCAATGGC GCGGAAATTCG AATTCGCGGAC AATTCGCGGAC
Contig	CCTAAAGAGATCCGGCTCTTAQCGG CTTACGGGGATATAACAGCGGAATCTAAGCGG



4.

5.

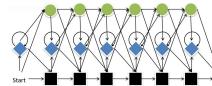
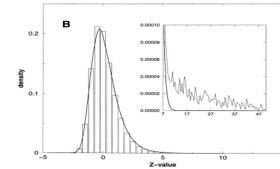
1. Introduction: Biological foundations



2. Homology: Comparing biological sequences

K-ELQRAASLTIEV
KDEGQK--SLVIDV

Amino acid	V	A	W	G	H	E
A	0	-5	-10	-15	-20	-25
A	-5	2	-3	-8	-13	-18
W	-10	-3	19	14	9	4
H	-15	-8	14	17	20	15
E	-20	-13	9	14	18	24
A	-25	-18	4	10	13	19

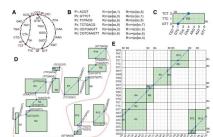


0	6: \$googo l
1	3: go1\$go o
2	0: googo1 \$
3	5: l\$goog o
4	2: ogo1\$g o
5	4: ol\$goo g
6	1: oogo1\$ g

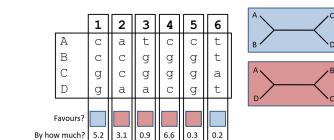
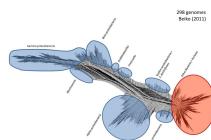
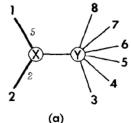
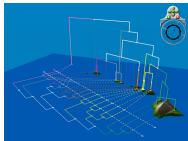
3. Assembly: Recovering genomes from sequencing data

Reference: CCTAGAGATCCCGCTTAA
Reads:
CCTAGAGATCCCGCTTAA
ATCGGGATATA
ATCGGGATATA
GCTCTTAA
CTTAA
TAGGGATATA
TATAATACAGCC
ACAGCAACT
GATTTCTTAC
GATTTCGGAT
GGGATTCGAC
AGGATTCGAC
TTGGCAGCACG

Contig: CCTAGAGATCCCGCTTAA
CTTAA
TAGGGATATAATACAGCCGATCTTACGG
TCTTACGGGATTCGACAG

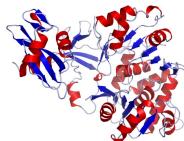
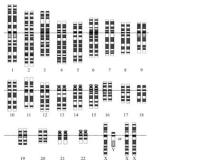


4. Phylogenetics: Inferring evolutionary relationships



5.

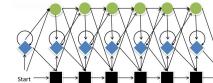
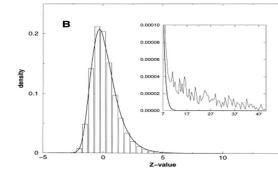
1. Introduction: Biological foundations



2. Homology: Comparing biological sequences

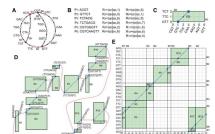
**K-ELQRAASLTIEV
KDEGQK--SLVIDV**

AWKING	A	W	G	H	E
vs.					
KIRKERA	0	-5	-10	-15	-20
A	-5	2	-3	-8	-13
W	-10	-3	19	14	9
H	-15	-8	14	17	20
E	-20	-13	9	14	18
A	-25	-18	4	10	13
					19

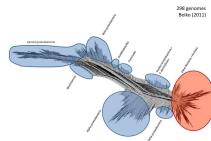
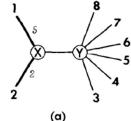
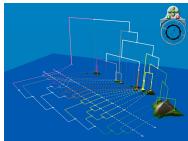


0	6	\$googo
1	3	gol\$go
2	0	googol
3	5	l\$goog
4	2	ogol\$g
5	4	ol\$goo
6	1	oogol\$

3. Assembly: Recovering genomes from sequencing data

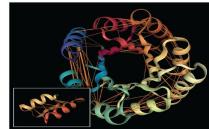
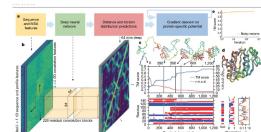
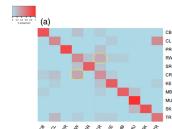
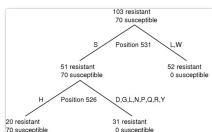


4. Phylogenetics: Inferring evolutionary relationships

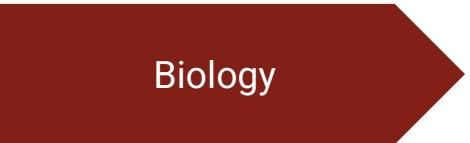


	1	2	3	4	5	6	
A	c	a	t	g	c	t	A
B	c	c	g	g	c	D	B
C	g	c	g	g	g	a	C
D	g	a	c	c	g	t	D
Favour?							
By how much?	5.2	3.1	0.9	6.6	0.3	0.2	

5. Machine Learning: Encoding and using biological data in machine learning



General Overarching Theme



Biology

Underlying messy
world of the (mostly
microbial) biology

General Overarching Theme

Biology

Biological Data

Underlying messy
world of the (mostly
microbial) biology

Biological sequence
data we measure or
extract from the
biology.

General Overarching Theme

Biology

Biological Data

Questions

Underlying messy
world of the (mostly
microbial) biology

Biological sequence
data we measure or
extract from the
biology.

1. What does this sequence do?
2. Has this pathogen acquired a new resistance gene?
3. Where did this pathogen come from?
4. Will this bacteria be resistant to this antibiotic?

General Overarching Theme

Biology

Biological Data

Questions

Models

Underlying messy
world of the (mostly
microbial) biology

Biological sequence
data we measure or
extract from the
biology.

1. What does this sequence do?
2. Has this pathogen acquired a new resistance gene?
3. Where did this pathogen come from?
4. Will this bacteria be resistant to this antibiotic?

1. Similar sequences do similar things
2. Genomes contain mobile elements
3. Similar sequences may be related to one another via evolution
4. Parts of genome are associated with observed phenotype

General Overarching Theme

Biology

Biological Data

Questions

Models

?

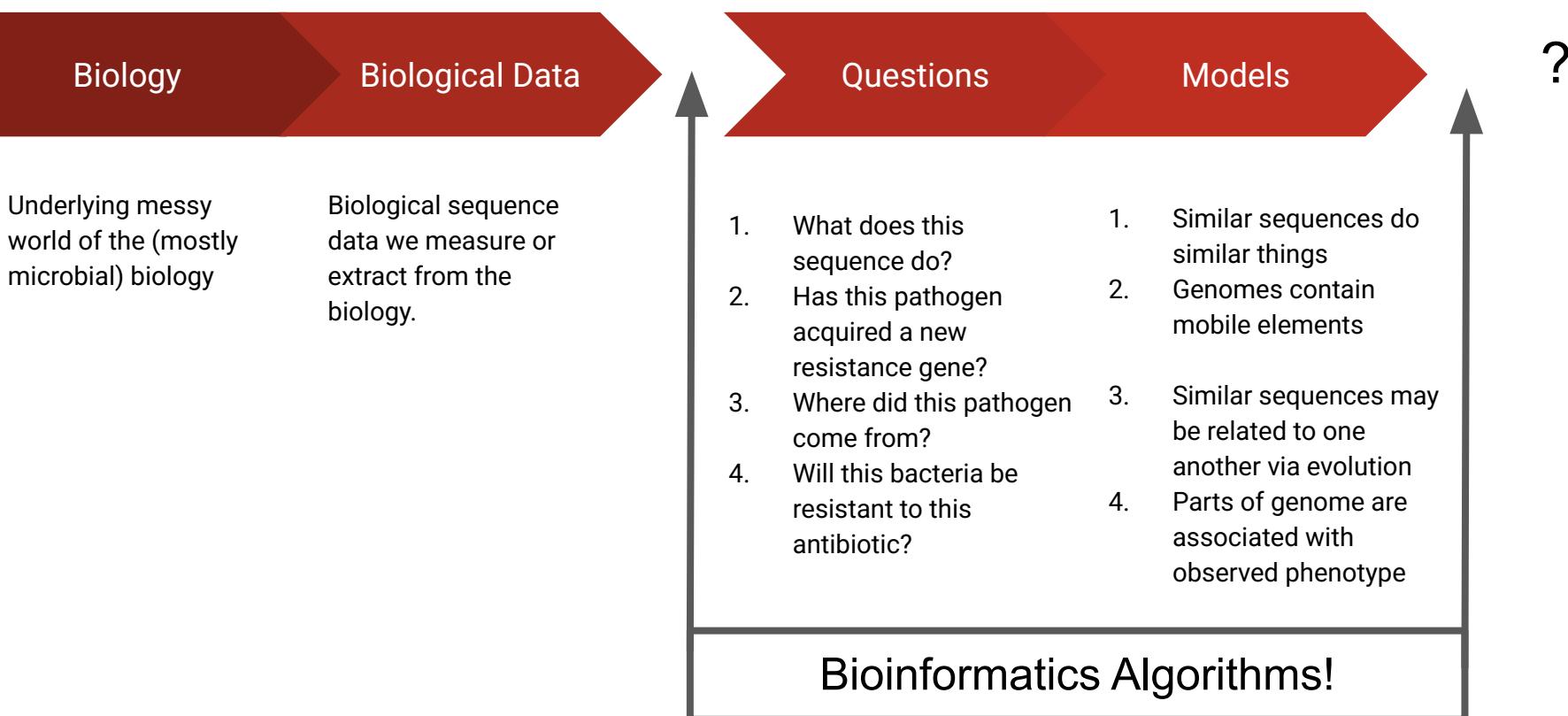
Underlying messy
world of the (mostly
microbial) biology

Biological sequence
data we measure or
extract from the
biology.

1. What does this sequence do?
2. Has this pathogen acquired a new resistance gene?
3. Where did this pathogen come from?
4. Will this bacteria be resistant to this antibiotic?

1. Similar sequences do similar things
2. Genomes contain mobile elements
3. Similar sequences may be related to one another via evolution
4. Parts of genome are associated with observed phenotype

General Overarching Theme



How are we going to learn?

Lectures, practical assignments, and a paper review

- 22 lectures (every Monday and Wednesday)
- 4 Practical Assignments (due via Brightspace 1 week after release)

Lectures, practical assignments, and a paper review

- 22 lectures (every Monday and Wednesday)
- 4 Practical Assignments (due via Brightspace 1 week after release)
 1. Alignment and Distant Homology (15%)
 2. Genome Assembly (15%)
 3. SARS-CoV-2 Phylogenetics (15%)
 4. Prediction of AMR Phenotypes (15%)
- Paper Review

Lectures, practical assignments, and a paper review

- 22 lectures (every Monday and Wednesday)
- 4 Practical Assignments (due via Brightspace 1 week after release)
 1. Alignment and Distant Homology (15%)
 2. Genome Assembly (15%)
 3. SARS-CoV-2 Phylogenetics (15%)
 4. Prediction of AMR Phenotypes (15%)
- Paper Review
 - Review of recent bioinformatics paper (selection due 2026/03/21)
 - Written Summary due 2026/04/08 (20%)
 - Oral Presentation 2026/04/04 to 2026/04/08 (20%)

Other Class Logistics

- Details on course website
- Contact for office hours
- **TA: Sneha Murthy**
- Assignment must be submitted via Brightspace as .docx or .pdf and named:

“BannerID_LastName_AssignmentX.{pdf,docx}”

The screenshot shows the course website for CSCI4181/6802 Bioinformatics Algorithms, Winter 2025-2026. The header features the Dalhousie University logo, the course title, and the semester. Below the header is a navigation bar with links for HOME, SCHEDULE, LECTURES, ASSIGNMENTS, and PAPER REVIEW.

CSCI4181/6802 Bioinformatics Algorithms / Winter 2025-2026

Course Description

Bioinformatics uses computational and statistical approaches to tackle questions of biological function and evolution. The goal of Algorithms in Bioinformatics is to introduce key applications of algorithms, data structures, and encodings to the analysis of large biological data sets. A recurring theme throughout the course will be the disconnect between algorithmic beauty and the horrifying realities of biological data. Every statistical model is violated and every classification comes with an asterisk, as we struggle with even the most basic concepts of 'gene' and 'species', and the challenges of understanding events that happened across ~3.5 billion years. In spite of these challenges, in this age of massive data sets we stand to learn a good deal if the computational tools we use are efficient, robust, properly validated, and correctly applied. The course covers major challenge areas in bioinformatics, each focused on an aspect of DNA or protein sequence analysis. The goal in each case is to define an overarching problem, and then explore different approaches that have been applied to solving that problem, with an emphasis on the match (or mismatch) between the algorithm and the underlying biological system.

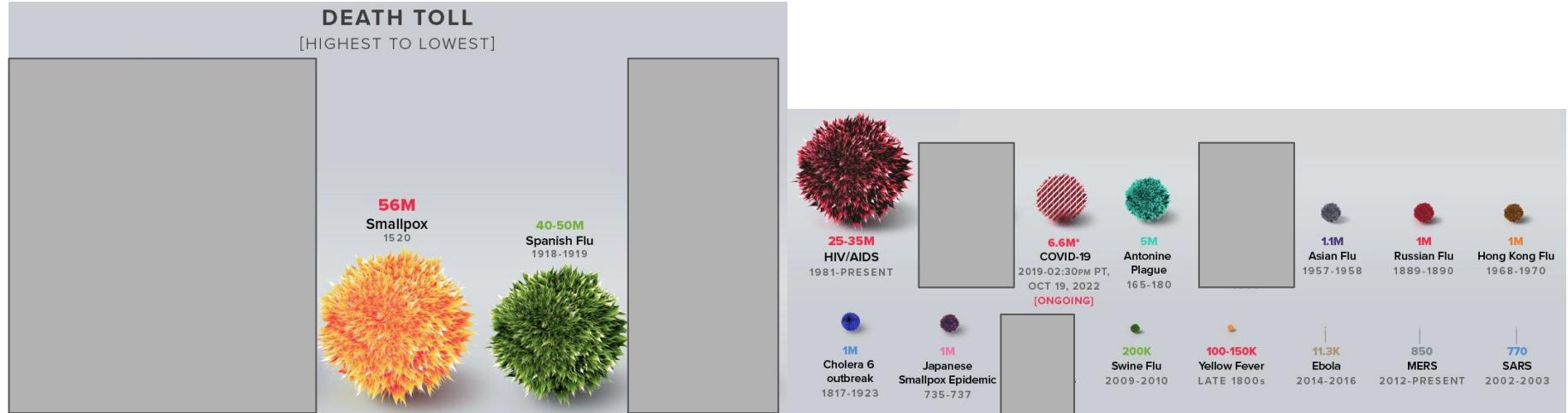
2026 Course Details

- Monday & Wednesday: 14:35-15:55, 2176 McCain Arts Social Sciences
- Office: 4242 Mona Campbell Building, Studley Campus
- Email: finlay.maguire@dal.ca
- Office Hours: Email for appointments
- BrightSpace for assignment/paper review submission.
- Syllabus
- Majority of materials used in this class were originally created by Prof. Robert Beiko and modified (to greater and lesser degrees) by myself.

https://maguire-lab.github.io/bioinformatics_algorithms_2026

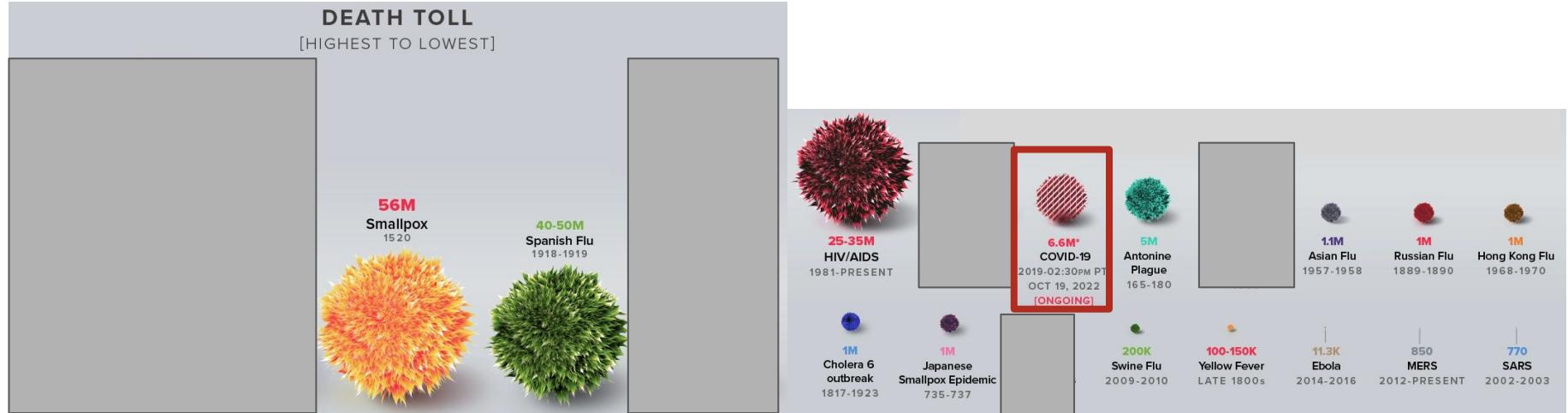
Why does any of this matter?

Pandemics matter to humans



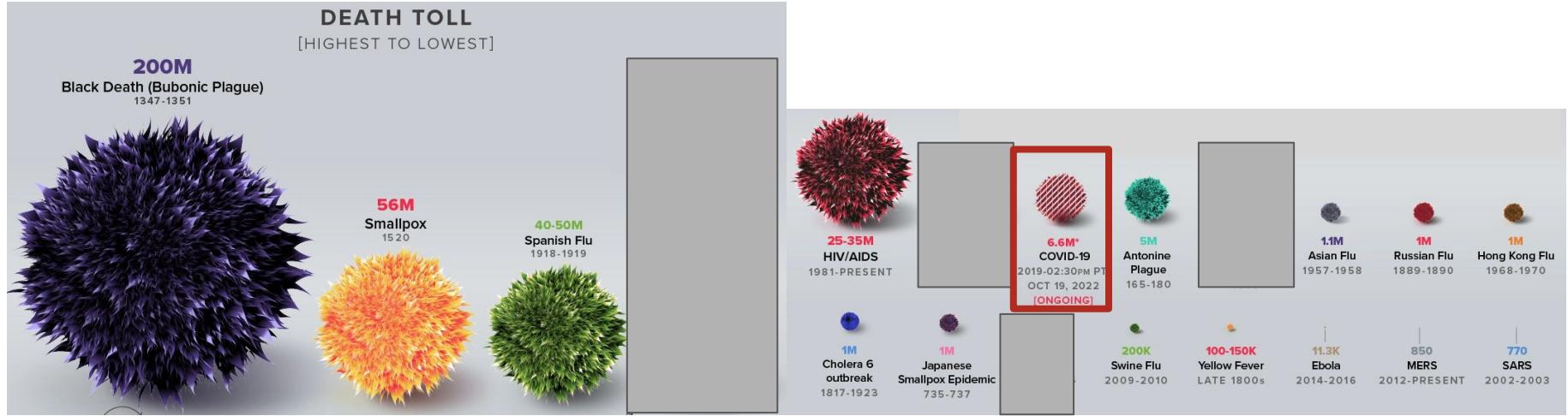
<https://www.visualcapitalist.com/history-of-pandemics-deadliest/>

Pandemics matter to humans



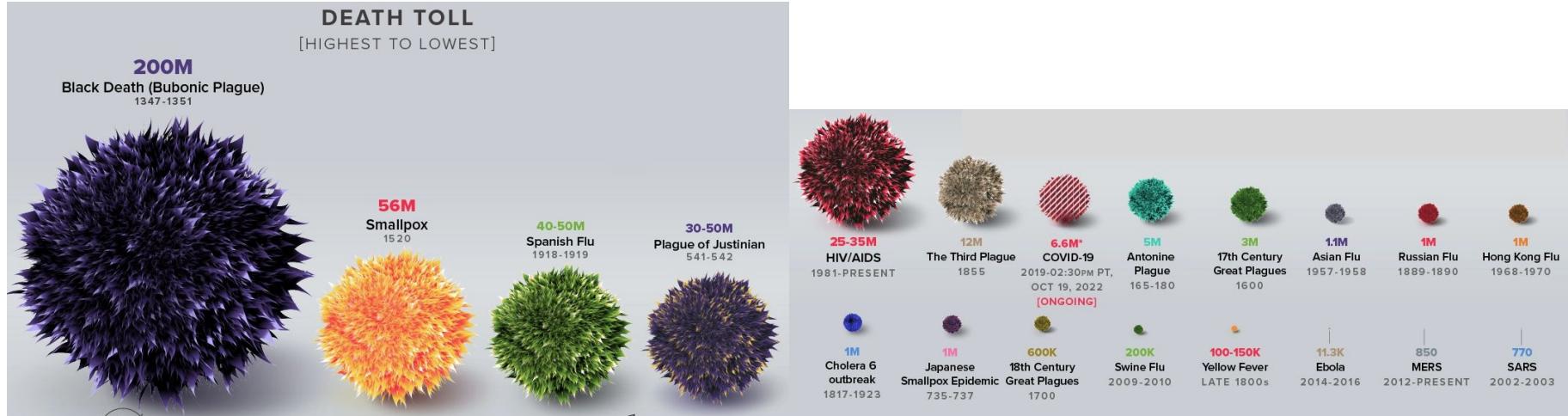
<https://www.visualcapitalist.com/history-of-pandemics-deadliest/>

Pandemics matter to humans



<https://www.visualcapitalist.com/history-of-pandemics-deadliest/>

Pandemics matter to humans



<https://www.visualcapitalist.com/history-of-pandemics-deadliest/>

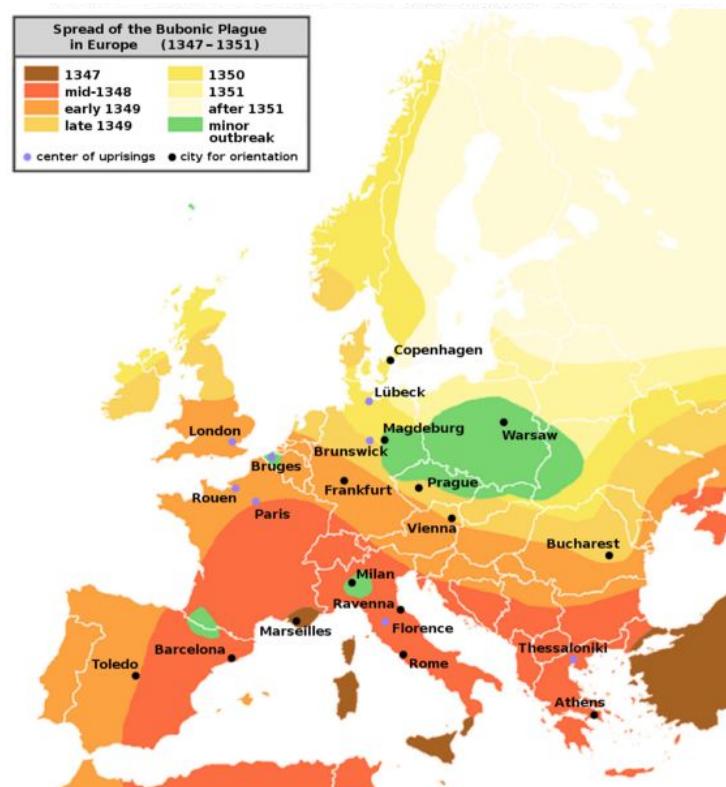
Plague had a huge impact on human history

- Justinian Plague (541–549)



Plague had a huge impact on human history

- Justinian Plague (541–549)
- Black Death (1347-1351 / 1330s to 1830s)



https://commons.wikimedia.org/wiki/File:Bubonic_plague_map.PNG

Plague had a huge impact on human history

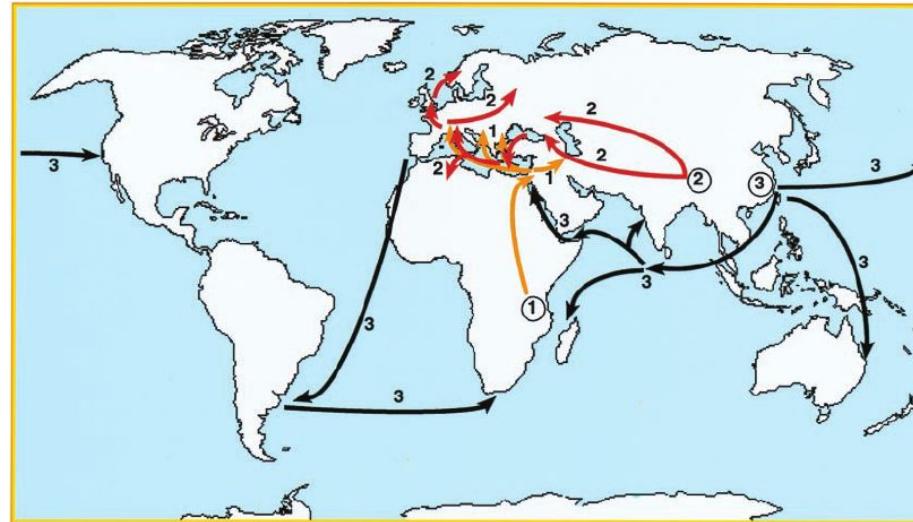
- Justinian Plague (541–549)
- Black Death (1347-1351 / 1330s to 1830s)



https://commons.wikimedia.org/wiki/File:Lord_haue_mercy_on_London.jpg

Plague had a huge impact on human history

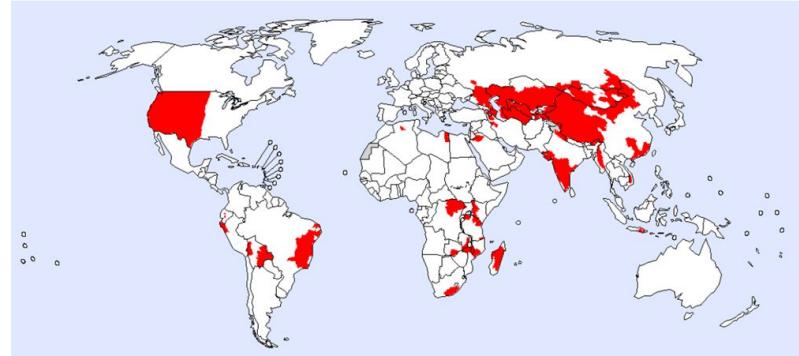
- Justinian Plague (541–549)
- Black Death (1347-1351 / 1330s to 1830s)
- 3rd Plague (1855-1960 / Today)



Plague had a huge impact on human history

- Justinian Plague (541–549)
- Black Death (1347-1351 / 1330s to 1830s)
- 3rd Plague (1855-1960 / Today)

Global distribution of natural plague foci
as of March 2016



Areas* with potential plague natural foci based
on historical data and current information

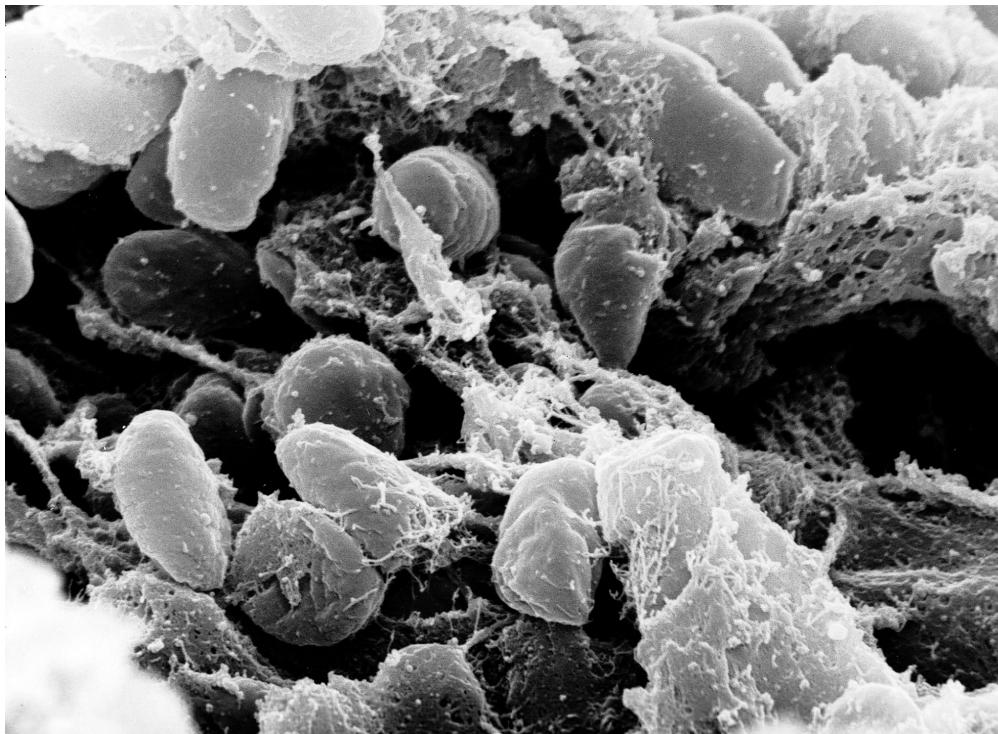
* First administrative level representation

Source: WHO/PED, as of 15 March 2016

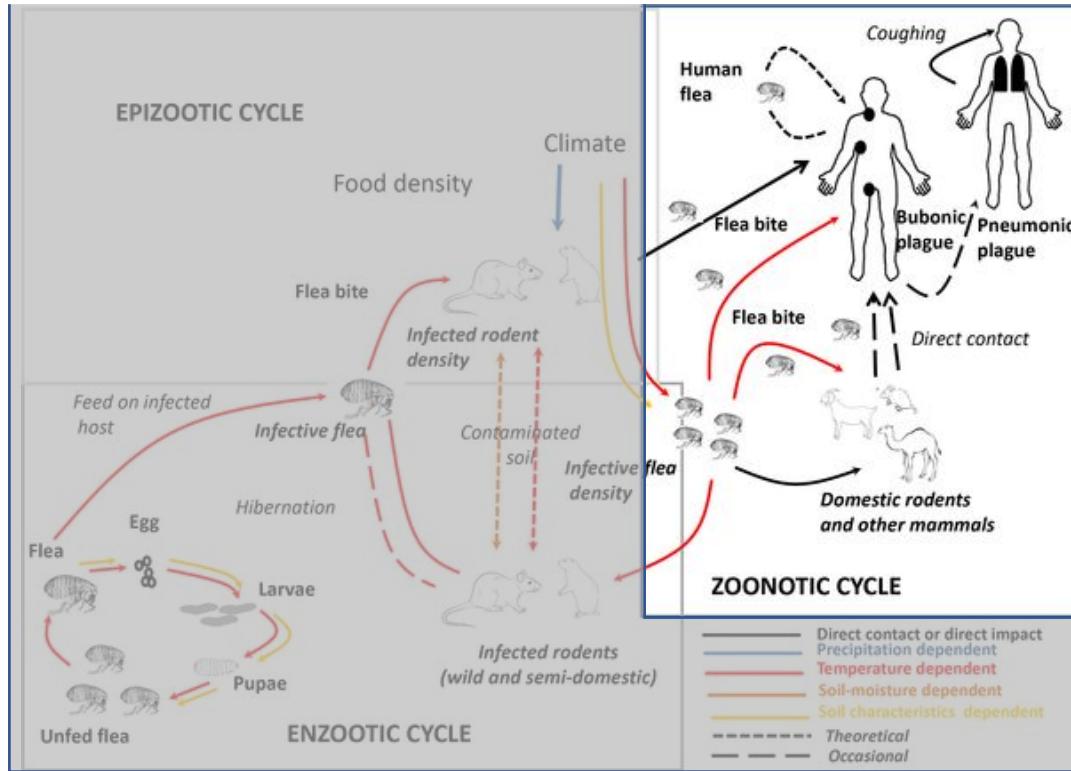
The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.
© WHO 2016. All rights reserved



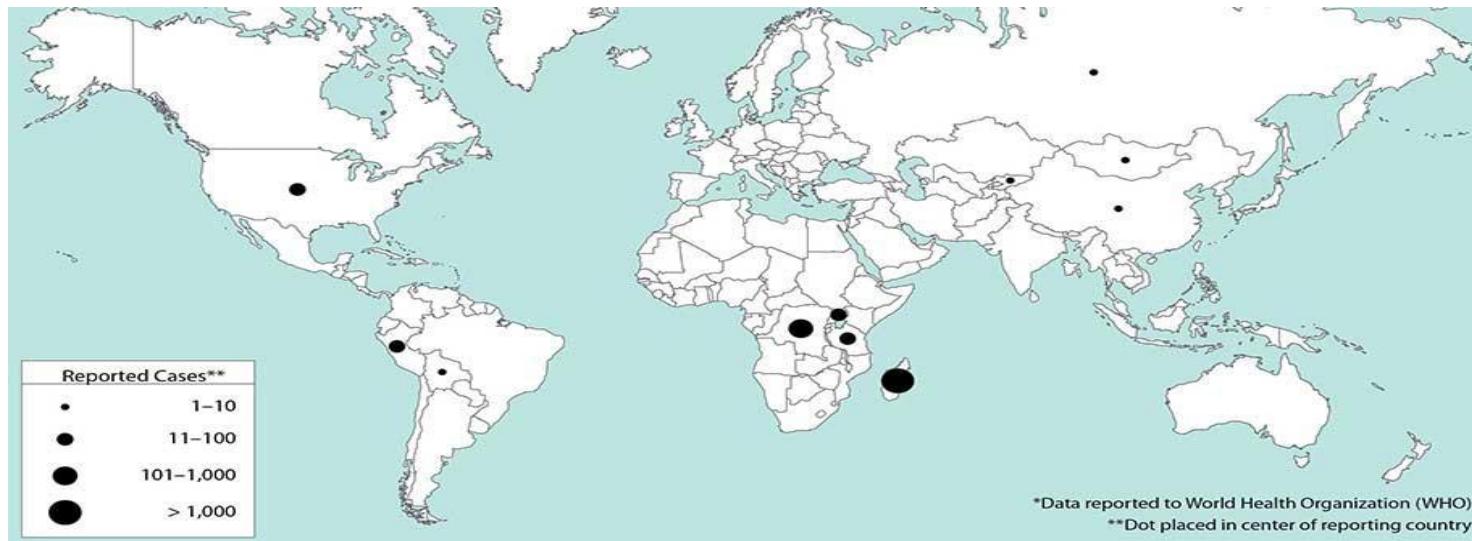
Bubonic plague is caused by *Yersinia pestis*



Complex life cycle including fleas and rodents

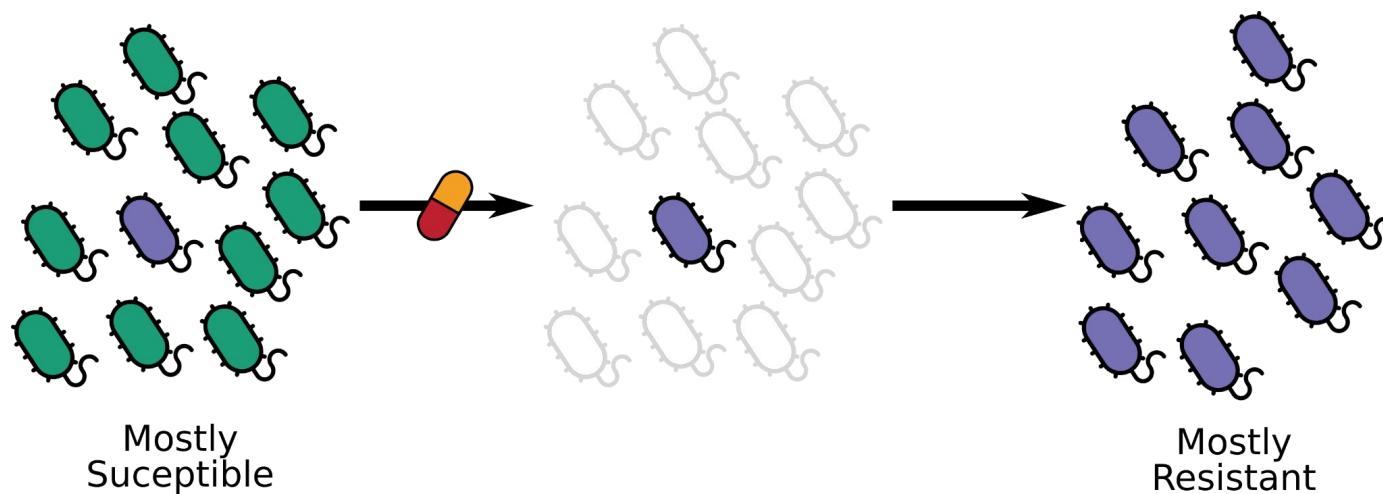


Still causes disease but is treatable



Fatality rate (untreated) = **40%-70%**
Fatality rate (treated) = **5%-15%**

Becoming less treatable



MULTIDRUG RESISTANCE IN
YERSINIA PESTIS MEDIATED BY
A TRANSFERABLE PLASMID

MARC GALIMAND, PH.D., ANNIE GUIYOULE, GUY GERBAUD,
BRUNO RASAMANANA, M.D., SUZANNE CHANTEAU, PH.D.,
ELISABETH CARNIEL, M.D., PH.D.,
AND PATRICE COURVALIN, M.D.

Plasmid-mediated doxycycline resistance in a *Yersinia pestis* strain isolated from a rat

Nicolas Cabanel ^a, Christiane Bouchier ^b, Minoarisoa Rajerison ^c, Elisabeth Carniel ^{a,*}

Transferable Plasmid-Mediated
Resistance to Streptomycin in a
Clinical Isolate of *Yersinia pestis*

Annie Guiyoule, * Guy Gerbaud, * Carmen Buchrieser, *
Marc Galimand, * Lila Rahalison, † Suzanne Chanteau, †
Patrice Courvalin, * and Elisabeth Carniel^{*}
†Institut Pasteur, Paris, France; and [‡]Institut Pasteur,
Antananarivo, Madagascar

Can see in real-time: Megaplate Experiment

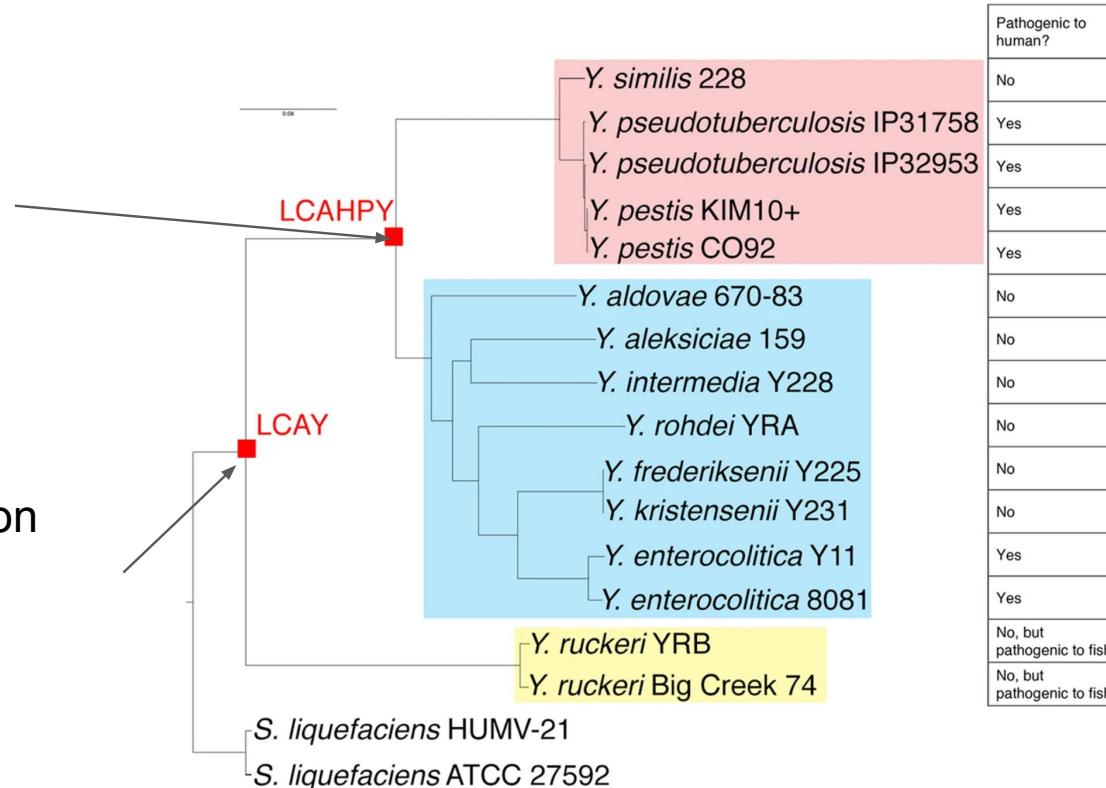


Need to know about a pathogen to stop it!

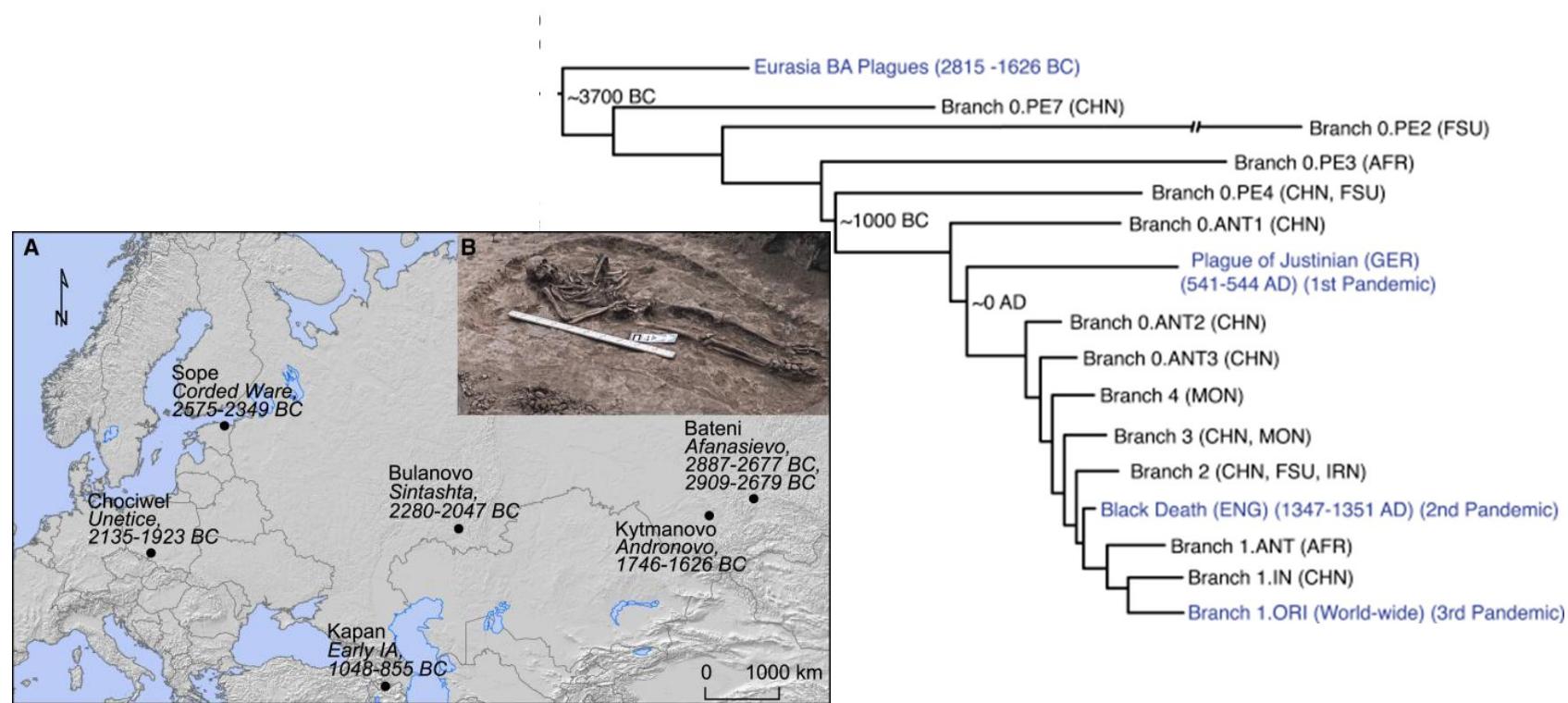
How did *Y. pestis* evolve?

Last common ancestor of human pathogenic *Yersinia*

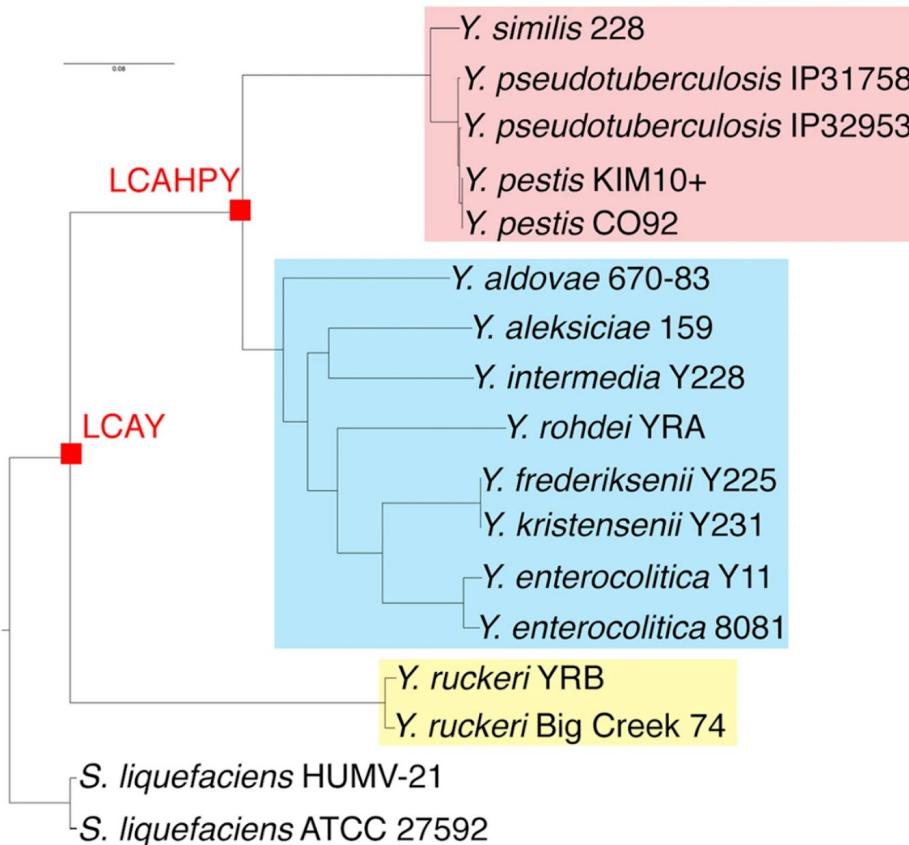
Last common ancestor of *Yersinia*



How old is *Y. pestis*?

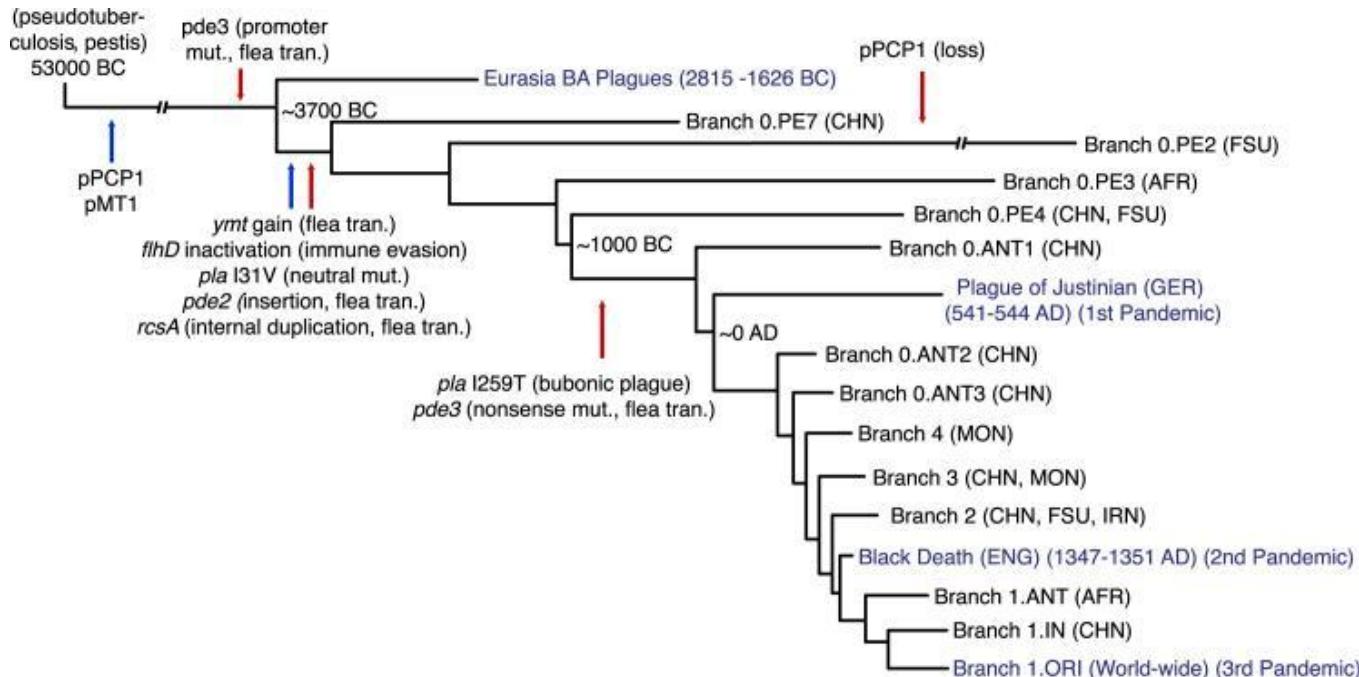


What makes it so virulent?



Pathogenic to human?	Presence of pYV plasmid?	Presence of CRISPR spacer to become immune to pYV or pYE854 plasmid?	Presence of <i>inv</i> homolog and N-terminal?	Number of copy of <i>ail</i> homolog?
No	No	Yes	Yes	4
Yes	Yes	No	Yes	4
Yes	Yes	Immune to pYV of <i>Y. enterocolitica</i>	Yes	4
Yes	Yes	Immune to pYV of <i>Y. enterocolitica</i>	No	4
Yes	Yes	Immune to pYV of <i>Y. enterocolitica</i>	No	4
No	No	No	No	1
No	No	No	No	1
No	No	No	No	1
No	No	No	No	1
No	No	Yes	No	1
No	No	Yes	No	1
Yes	Yes	No	Yes	2
Yes	Yes	No	Yes	2
No, but pathogenic to fish	No	No	No	1
No, but pathogenic to fish	No	No	No	1

How did it become so deadly?



How do we know this?

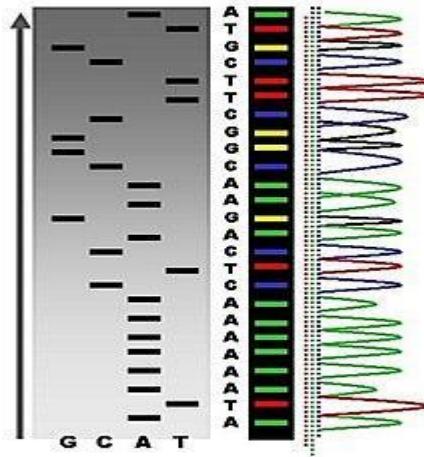


Molecular biology

How do we know this?



Molecular biology

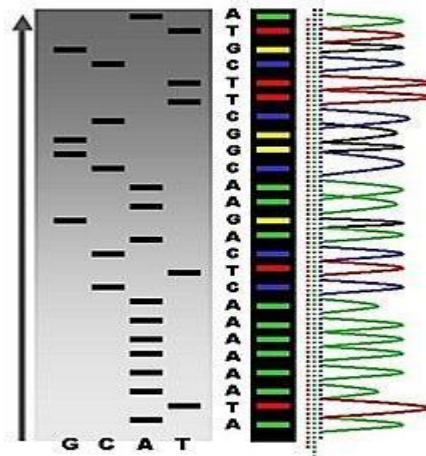


DNA sequencing

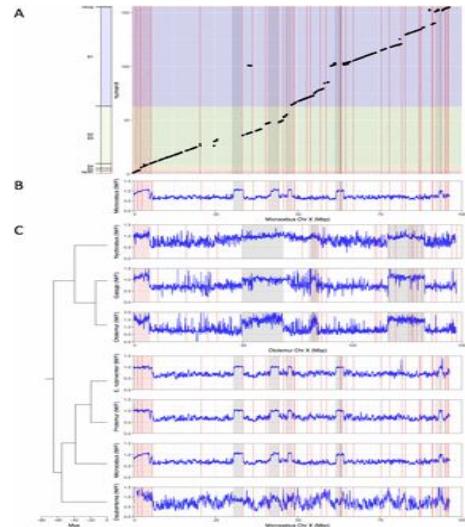
How do we know this?



Molecular biology



DNA sequencing



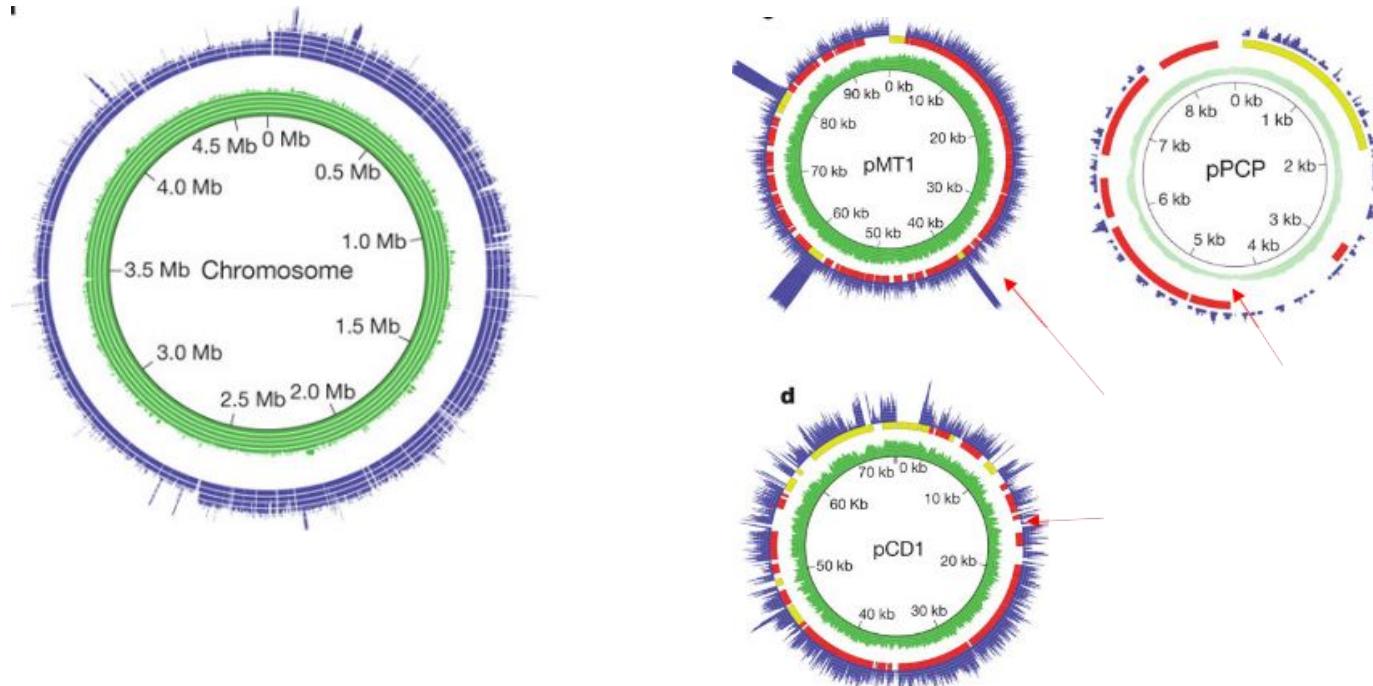
Bioinformatics!

Sequencing the Black Death genome

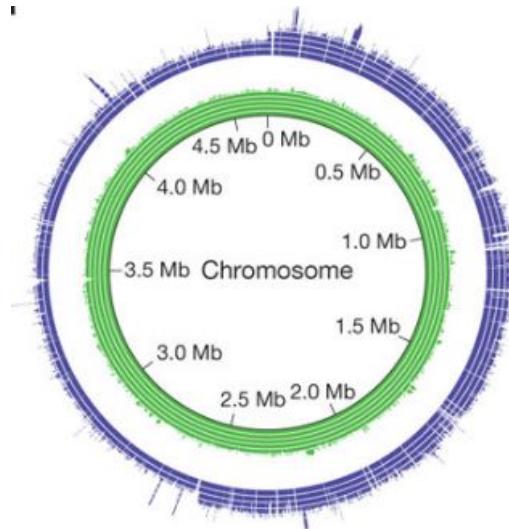


East Smithfield
burying ground

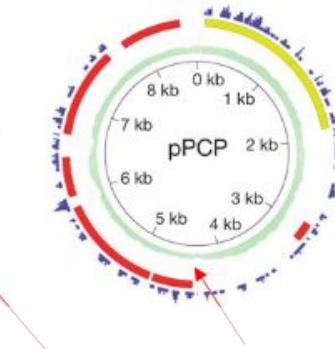
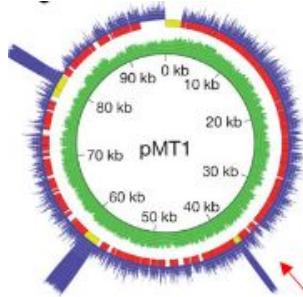
Need assembly algorithms to get genome



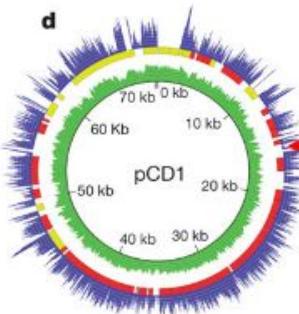
Need homology algorithms to decipher genome



Main Chromosome



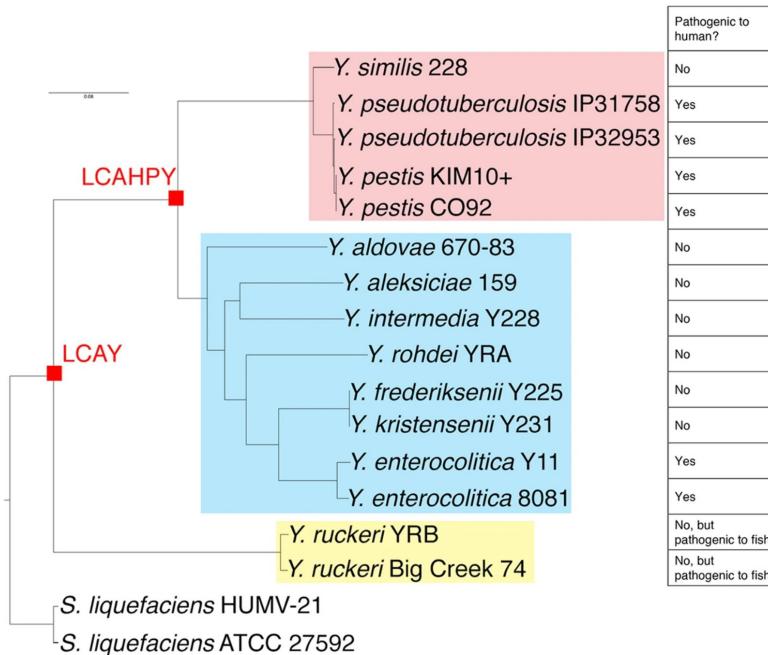
F1 (antigen; vaccine target)
Pla (host invasion)



Antihost proteins and cellular delivery mechanism

Plasmids

Need phylogenetic algorithms to trace evolution



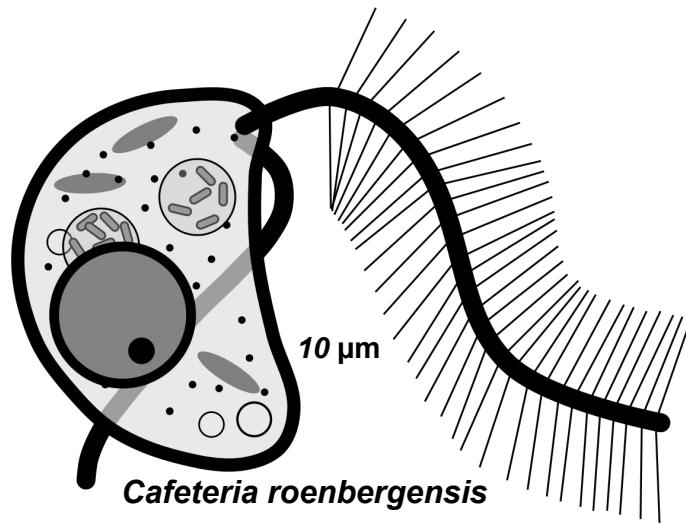
Summary!

- **Genome** data can tell us a lot about how things like plague works.
- Impossible to do this manually.
- Bioinformatics is using **computers** to understand **biology**
- **Algorithms** are vital to doing this effectively
- We need **bioinformatic algorithms** to solve important problems including in human health

01: Life at Resolution: Organisms, Genomes, Sequences, and so on

CSCI4181/6802 Bioinformatics Algorithms
Finlay Maguire (finlay.maguire@dal.ca)

Even microbial life covers a large range



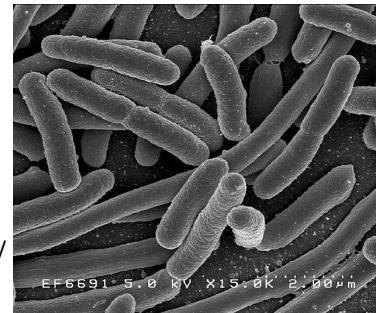
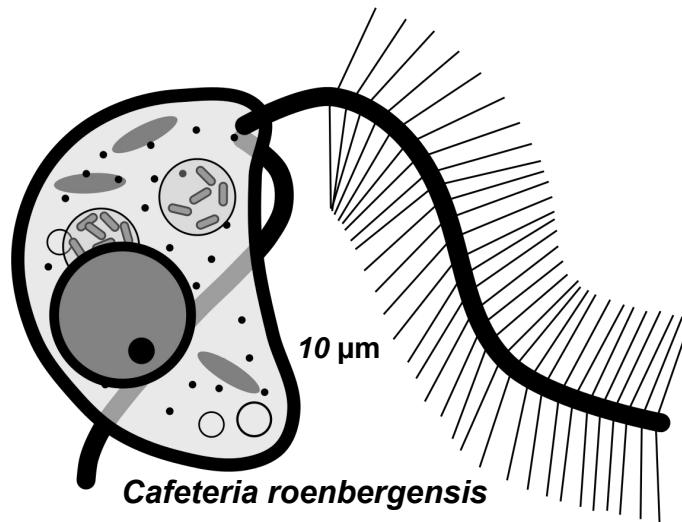
Cafeteria roenbergensis

https://www.cam.ac.uk/sites/www.cam.ac.uk/files/styles/content_885x432/public/news/research/news/mitochondria.jpg?itok=CIFFE3yjc

<https://commons.wikimedia.org/wiki/File:CafeteriaRoenbergensis.jpg>

https://upload.wikimedia.org/wikipedia/commons/f/f7/Giant_virus_CoV_with_its_virophage_Mavirus.png

Even microbial life covers a large range



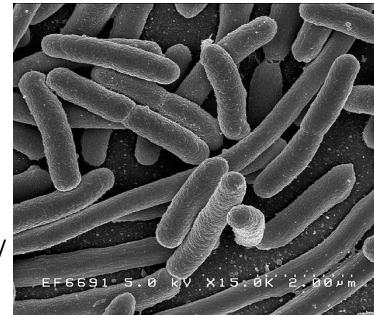
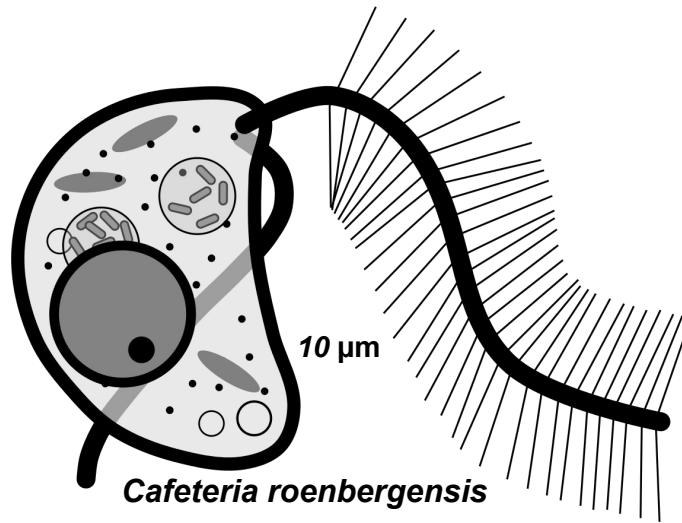
Escherichia coli

https://www.cam.ac.uk/sites/www.cam.ac.uk/files/styles/content_t-885x432/public/news/research/news/mitochondria.jpg?itok=CIFE3yjc

<https://commons.wikimedia.org/wiki/File:CafeteriaRoenbergensis.jpg>

https://upload.wikimedia.org/wikipedia/commons/f/f7/Giant_virus_CoV_with_its_virophage_Mavirus.png

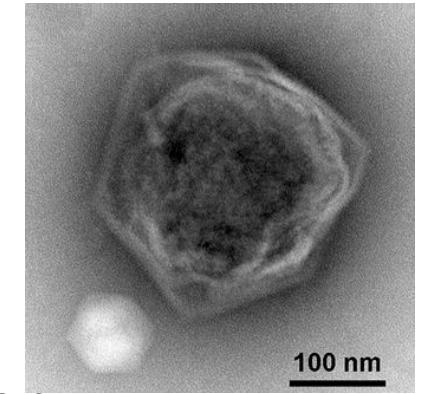
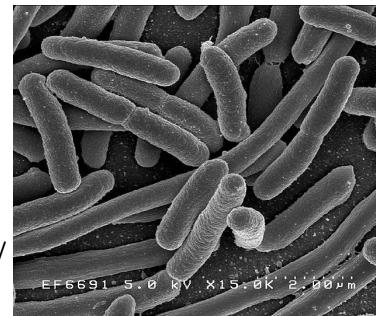
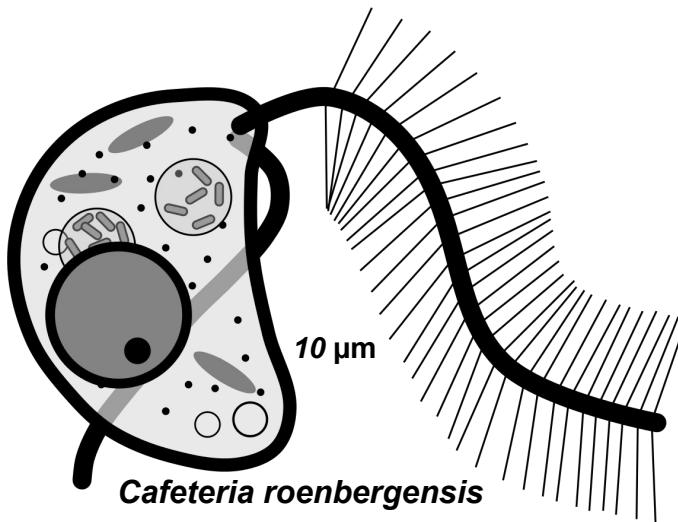
Even microbial life covers a large range



https://www.cam.ac.uk/sites/www.cam.ac.uk/files/styles/content_t-885x432/public/news/research/news/mitochondria.jpg?itok=CIFE3yjc

https://upload.wikimedia.org/wikipedia/commons/f/f7/Giant_virus_CoV_with_its_virophage_Mavirus.png

Even microbial “life” covers a large range



https://www.cam.ac.uk/sites/www.cam.ac.uk/files/styles/content_t-885x432/public/news/research/news/mitochondria.jpg?itok=CIFE3yjc

https://upload.wikimedia.org/wikipedia/commons/f/f7/Giant_virus_CoV_with_its_virophage_Mavirus.png

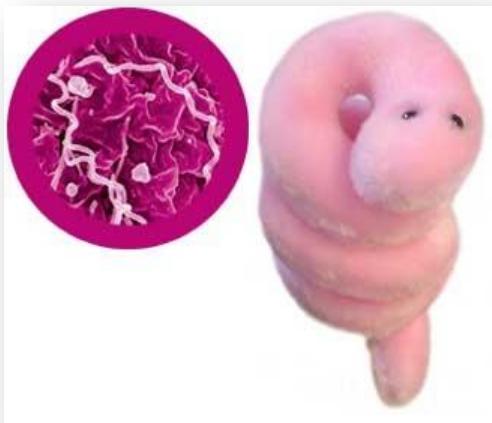
Overview

1. All living organisms have several key essential **properties**
2. Life can be viewed as a **hierarchical structure** with many levels of organization from **genome** (including genomic elements) to the **biosphere**
3. The levels we cannot observe with the naked eye are as (or more) **diverse** as the levels we can observe

Essential properties of an organism

Cellularity

Unicellular



Treponema pallidum
(www.teachersource.com)

Multicellular



Caenorhabditis elegans (959 cells)
(www.ucl.ac.uk)

Essential properties of an organism

Reproduction

Sexual



Tetrahymena thermophila
(www.isleepinadrawer.com)

Asexual

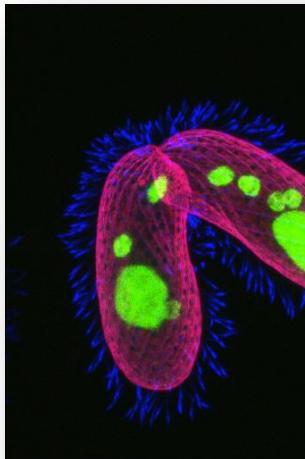


Amoeba proteus
(www.teachnet.ie)

Essential properties of an organism

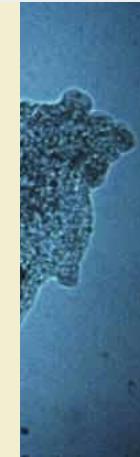
Reproduction

Sexual

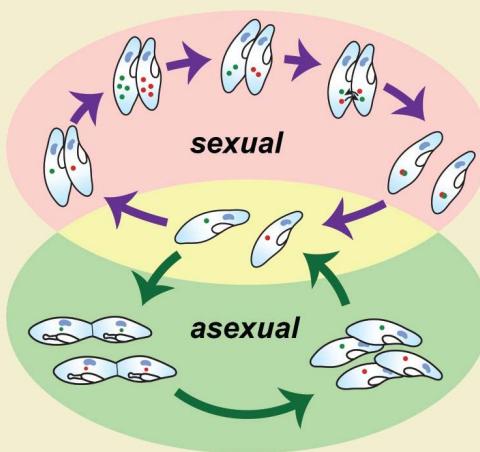


Tetrahymena
www.isleepinadrawer.com

Asexual



Amoeba
www.teachnet.ie



Essential properties of an organism

Biochemical Processes and Pathways, such as...

Fermentation



Toxin degradation



Antibiotic synthesis



Photosynthesis



<http://en.wikipedia.org/wiki/File:Kimchi.jpg>

http://en.wikipedia.org/wiki/File:NOVAMOXIN_antibiotic.jpg

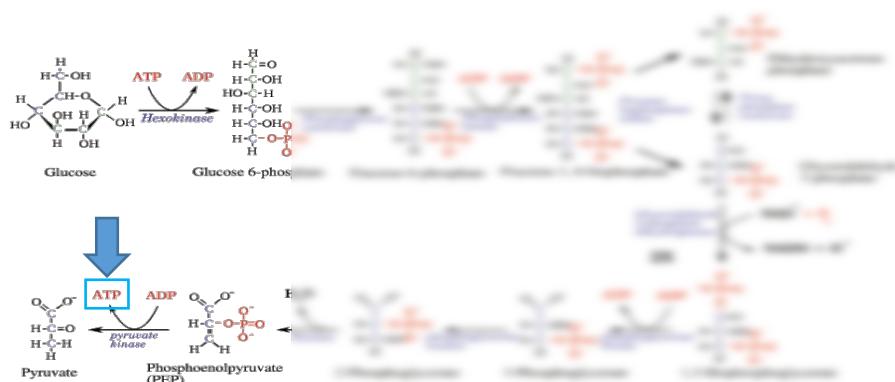
https://www.thestar.com/news/2007/01/28/sydney_tar_ponds_to_be_buried.html

https://commons.wikimedia.org/wiki/File:Prochlorococcus_marinus.jpg

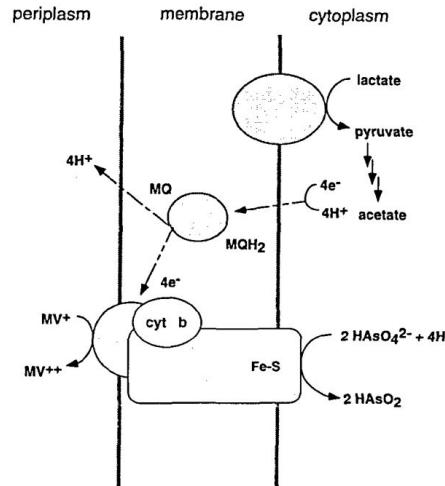
Essential properties of an organism

Respiration

Glucose - boring!



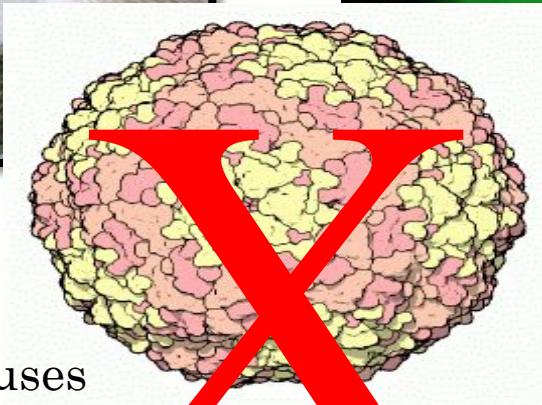
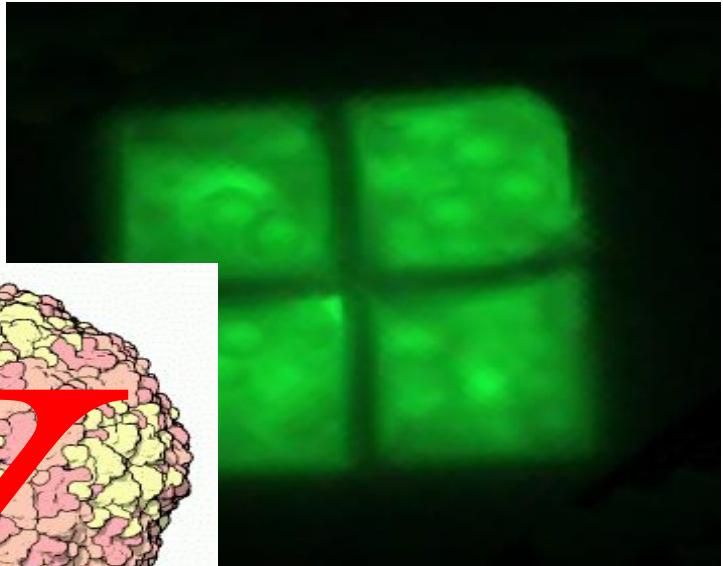
Arsenic - interesting!





Eukaryotes

Prokaryotes
(Bacteria and Archaea)



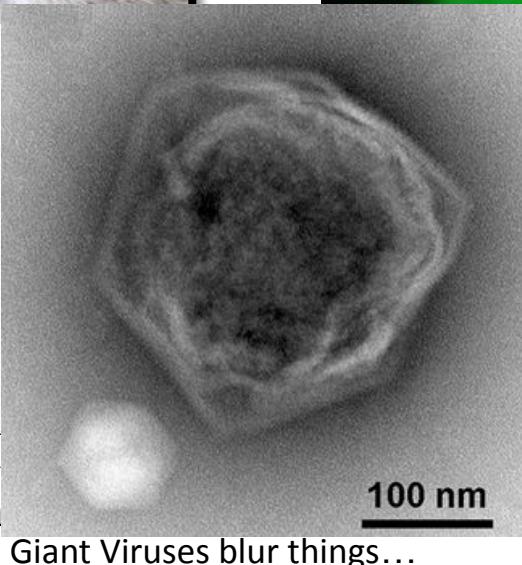
Viruses
(nonliving)
(some have RNA genomes!)

Photo: R.Beiko
http://en.wikipedia.org/wiki/File:Haloquadratum_walsbyi00.jpg
<http://www.rcsb.org/pdb/101/motm.do?momID=20>

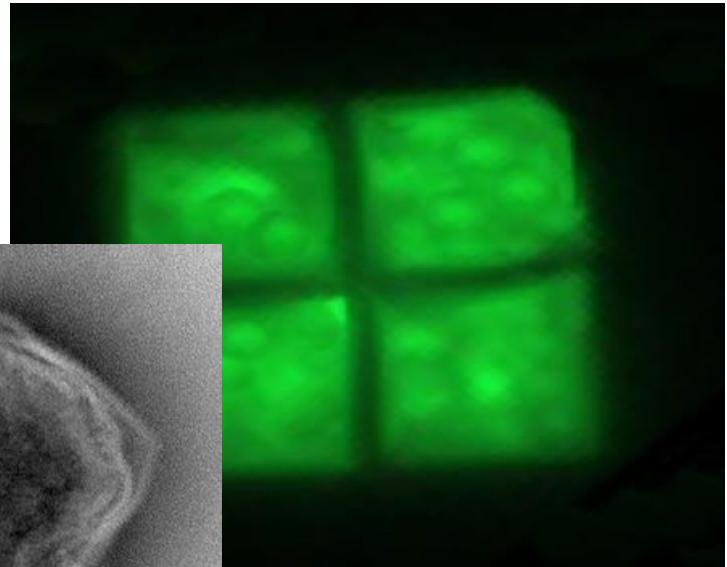


Eukaryotes

Prokaryotes
(Bacteria and Archaea)



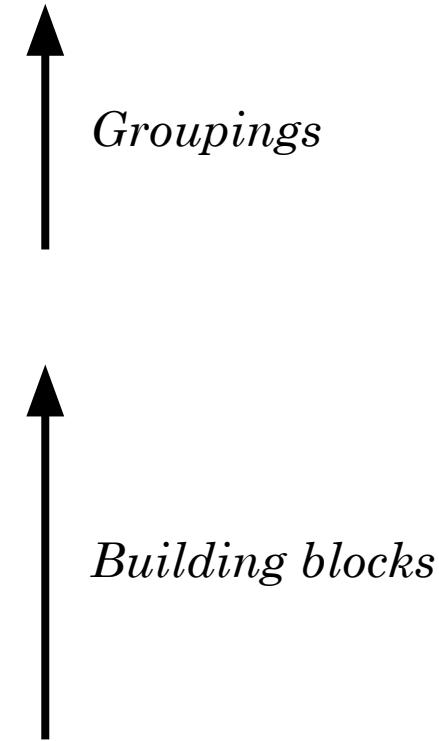
Giant Viruses blur things...



Viru
(non
(som

Photo: R.Beiko
http://en.wikipedia.org/wiki/File:Haloquadratum_walsbyi00.jpg
<http://www.rcsb.org/pdb/101/motm.do?momID=20>

Biosphere
Communities and Ecosystems
Populations
Organisms
Cells
Pathways and Systems
Proteins
Genes
Genomes



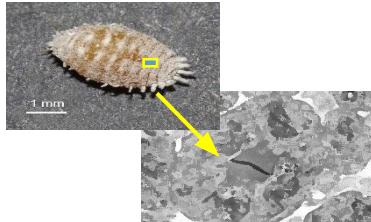
Genome:

The complete set of heritable genetic material

(DNA for all known cellular organisms)

Your Genome and You

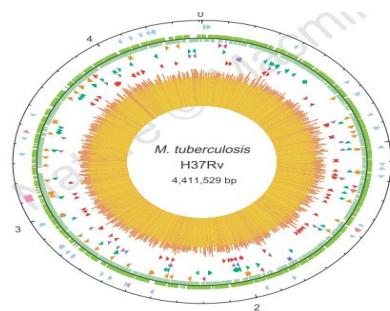
*Tremblaya
princeps*



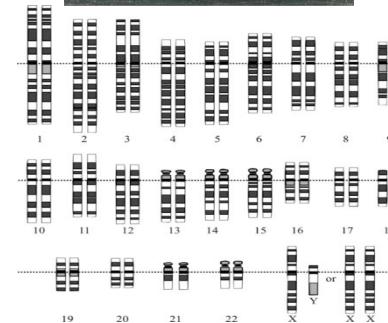
O

1 chromosome
110 genes
138,931 nucleotides

*Mycobacterium
tuberculosis H37Rv*

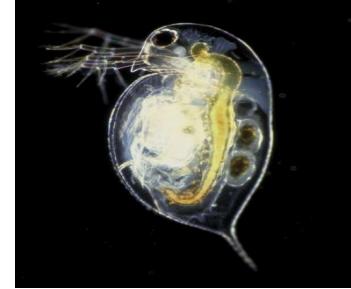


1 chromosome
4,000 genes
4.4 million nucleotides



23 chromosomes
+ mitochondrion
20,000 genes
3.1 billion nucleotides
(times two!)

Daphnia pulex



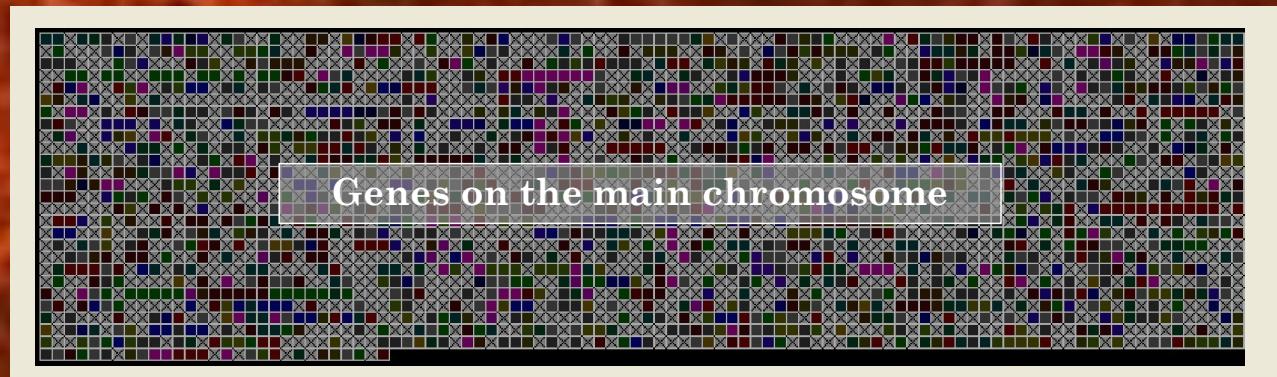
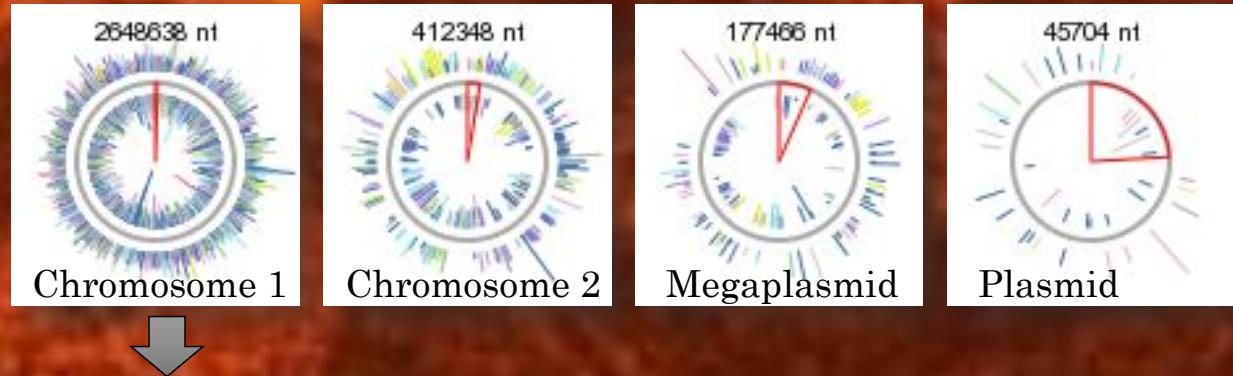
12 chromosomes
+ mitochondrion
31,000 genes
200 million nucleotides

Paris japonica



?? chromosomes
??? genes
150 billion nucleotides

The genome of *Deinococcus radiodurans*



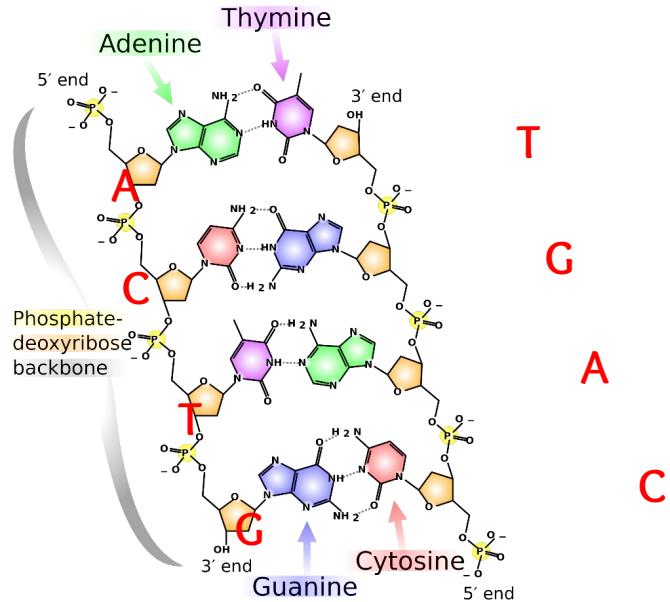
Gene
order

chael Daly

http://www.genomenewsnetwork.org/articles/07_02/deinococcus.shtml

Deinococcus genome paper: Science. Nov 19, 1999; 286(5444): 1571–1577.

The DNA sequence of a gene



5' - ATGC GTT ACTTC GAAATGGCAACCCACTCGGGGACTTCCTCCAACGGTTGA- 3'
3' - TACGCAATGAAGCTTACCGTTGGGTGAGCCCCTGAAGGAGGTTGCCAACT- 5'

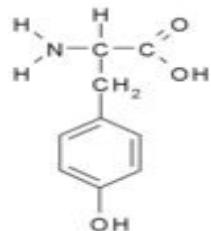
DNA to protein

DNA is read in triplets

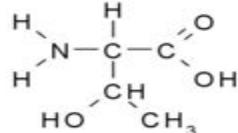
ATG CGT TAC TTC GAA ATG GCA ACC CAC TCG GGG ACT TCC TCC AAC GGT TGA



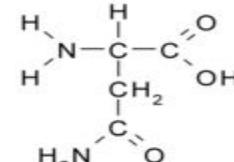
M A Y F E M A T H S G T S S N G *



Tyrosine



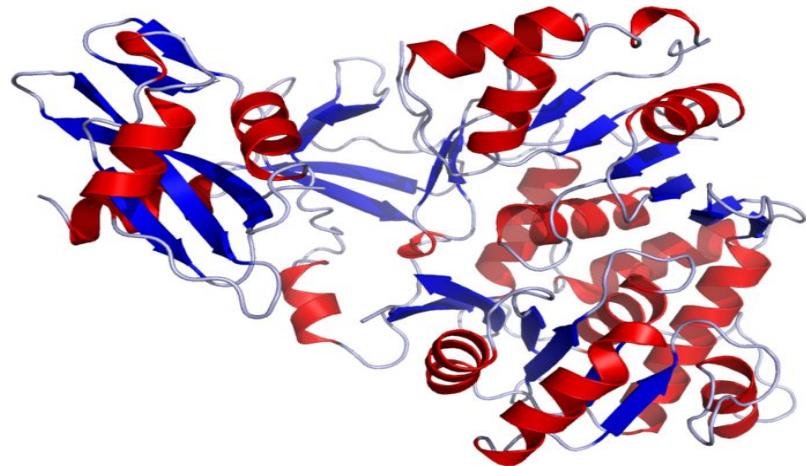
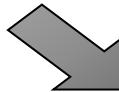
Threonine



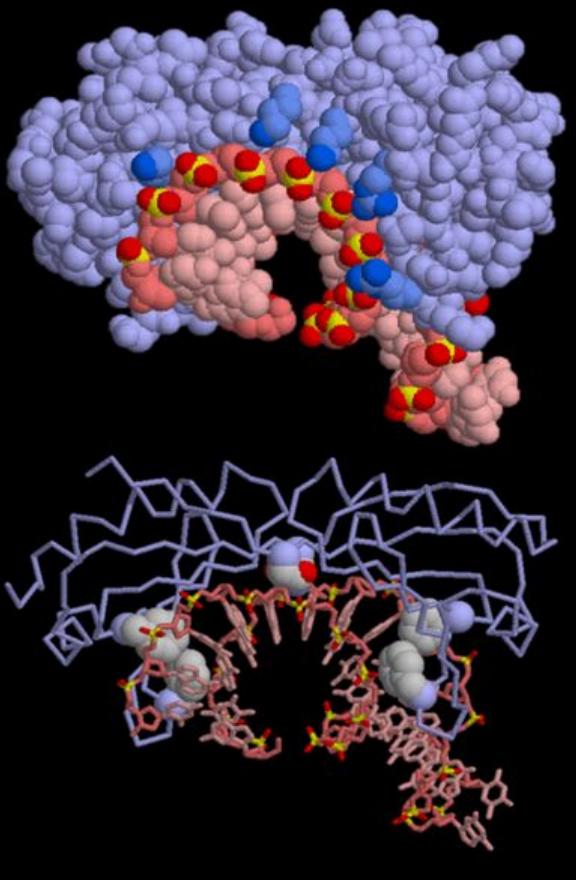
Asparagine

Protein sequence and structure

M A Y F E M A T H S G T S S N G *



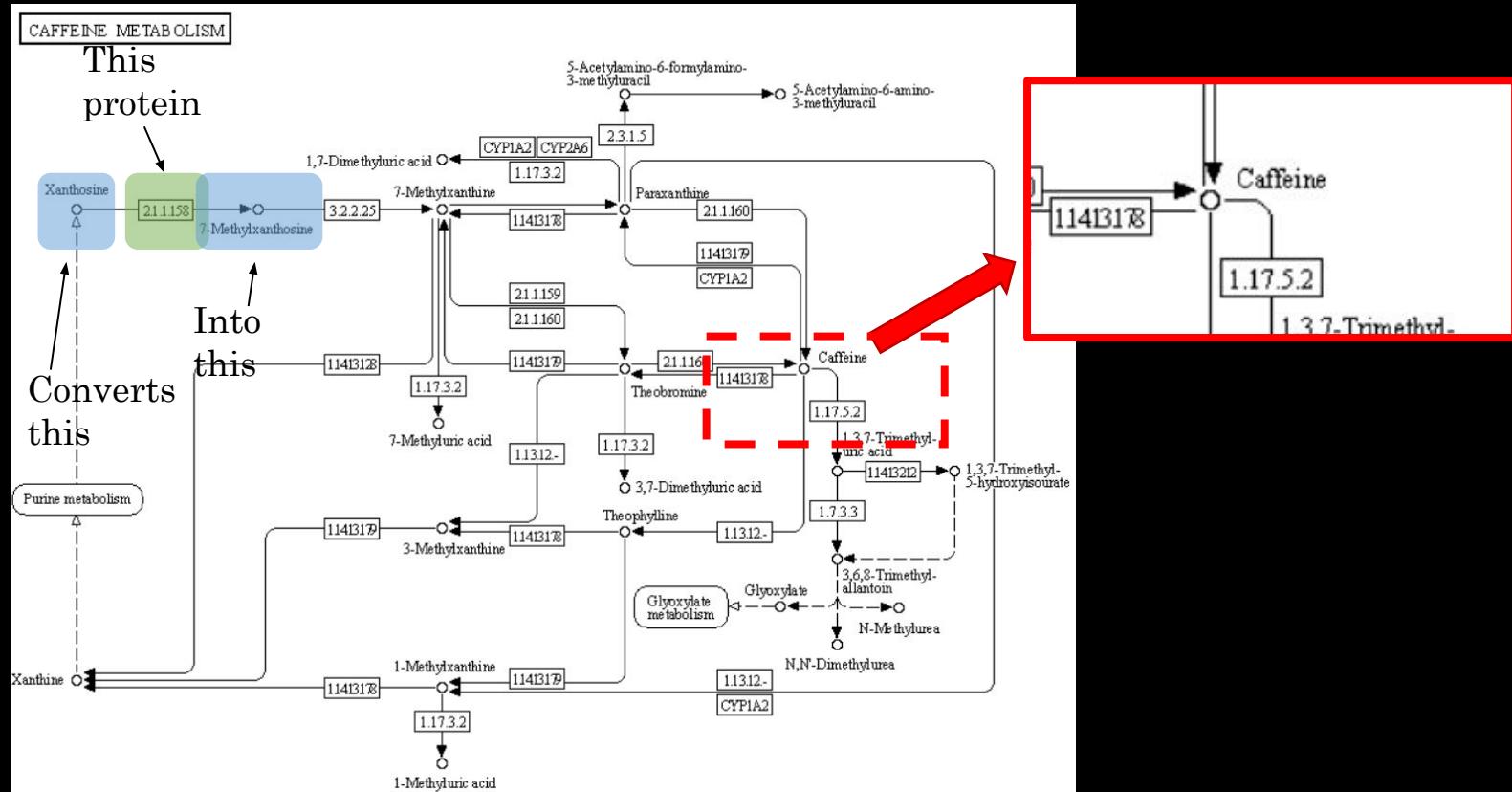
A DNA-protein complex

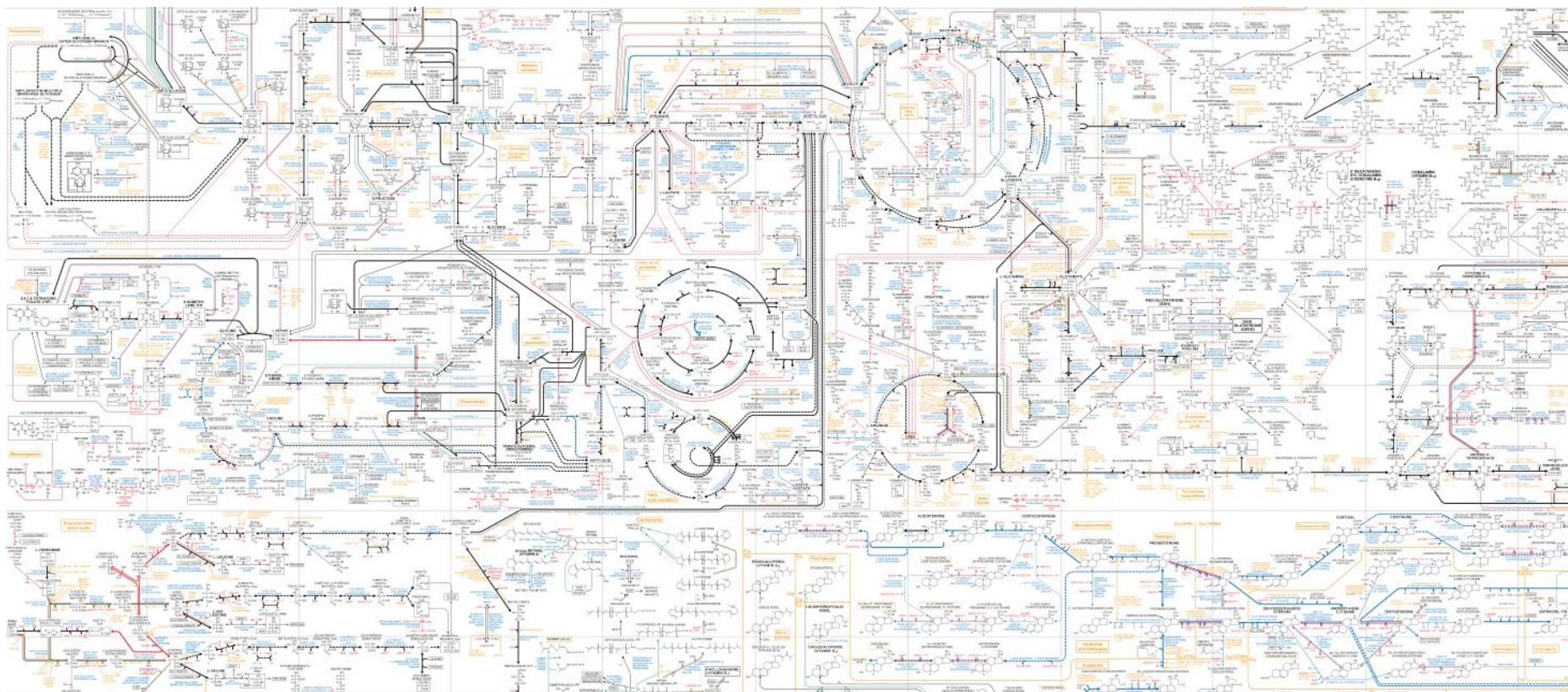


←
DNA-binding protein
("TATA-box binding protein")

←
DNA (note the recurring pattern;
yellow = phosphate)

Metabolism – Proteins working together





Populations



Population

“...A subset of individuals of one species that occupies a particular geographic area and, in sexually reproducing species, interbreeds.”

Populations exhibit genetic variation – this variation may be important! (or it might not)

Variation within a population

Each of these is a DNA sequence from a different Atlantic salmon (*Salmo salar*)

A close-up photograph of an Atlantic salmon's head and upper body, showing its scales, fins, and mouth.

```
CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA  
CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA  
CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA  
CCAGGAGTAGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA  
CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA  
CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA  
CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA  
CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA  
CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA  
CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA  
CCAGGAGTTGAGCCTGCGGCTTAATTTGACTCAACACGGAAACCTCACCCGGCCCCGGA
```

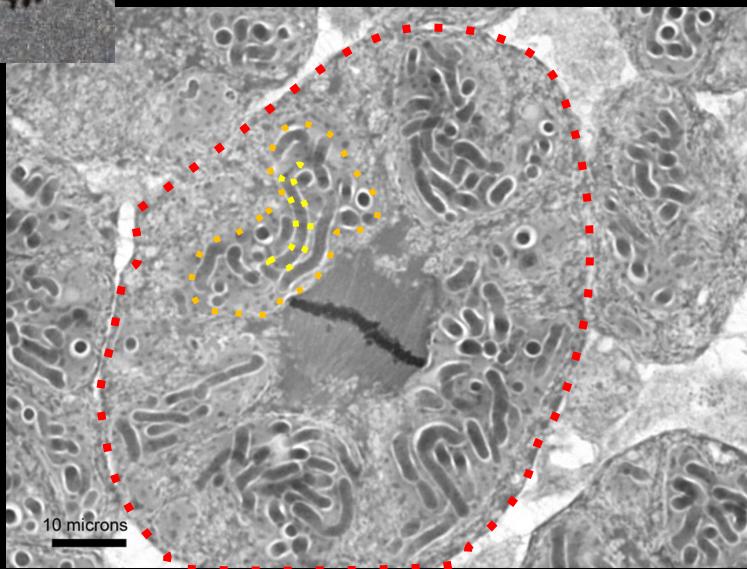
Communities and Ecosystems



Bacterial Symbionts – Not Your Usual Ecosystem



Mealybug (*Pseudococcus calceolariae*)



Pseudococcus calceolariae
(mealybug)

Candidatus “Tremblaya princeps”

Sodalis glossinidius
(-ish)

Overview

1. All living organisms have several key essential **properties**
2. Life can be viewed as a **hierarchical structure** with many levels of organization from **genome** (including genomic elements) to the **biosphere**
3. The levels we cannot observe with the naked eye are as (or more) **diverse** as the levels we can observe

Working out the cause of the plague

- **1546 Fracastro:** disease from invisible “seeds”
- **1683 Van Leeuwenhoek:** microscopic creatures
- **1857 Pasteur:** microbes are living creatures that don't spontaneously appear
- **1876 Koch:** microbes cause disease (Koch's Postulates)
- **1894 Shiabasaburō & Yersin:** Isolation of microbe in 3rd Plague
- **1896 Simon:** Isolation of microbe in fleas

