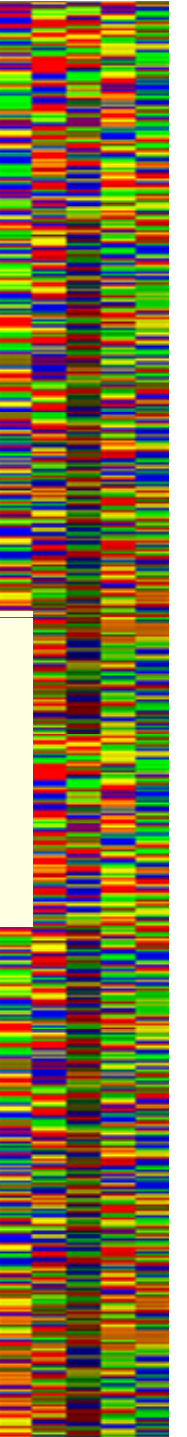
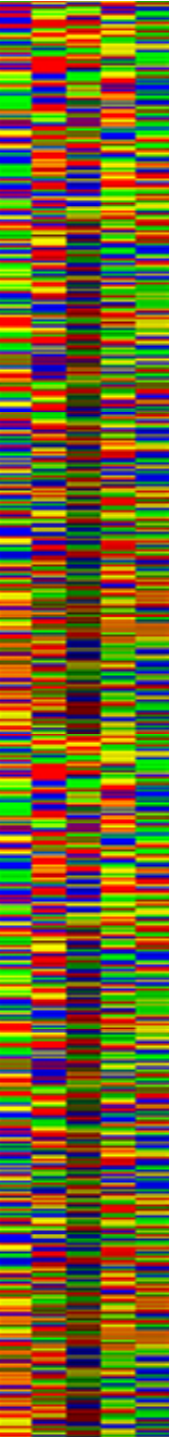

Introduction To Bioinformatics ***Biol 478 & 595B***



-
- ***For Undergrads: BIOL 478 / CS 478 / STAT 490B (3 credits)***
 - ***For Grads: BIOL 595B (4 credits)***

 - ***Coordinator: Michael Gribskov***
Gribskov@purdue.edu
Lilly G-233, 494-6933
office hours by appointment
 -
 - ***Co-Instructor:***
Daisuke Kihara
DK; Biol. Sci. & CS; dkihara@purdue.edu
 -
 - ***Grad TA***
Yifeng (David) Wang
Biol. Sciences; yang41@purdue.edu
 -



Regular lectures: Unless indicated on the schedule below all classes will be regular lectures. Readings for the next lecture will be announced at the beginning of each lecture. See course materials below for a description of the text and background references.

Tutorials: Practical demonstrations of bioinformatic analyses using available programs.

Homework: Homework assignments, generally weekly. Handed out on Monday, due on the Friday (four days later) indicated on the schedule.

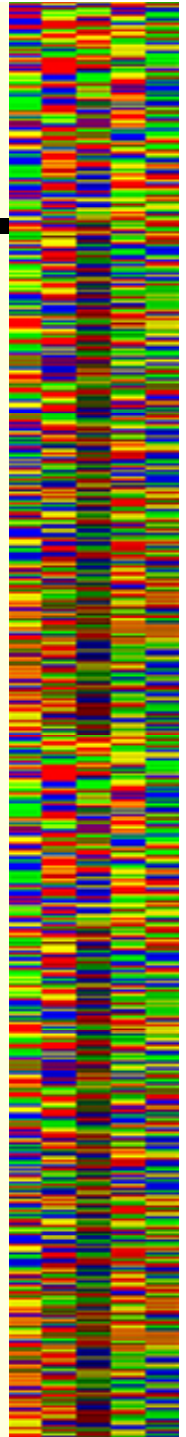
Mini-projects: There will be three long homeworks or mini-projects during the semester. Each mini-project will involve a multidisciplinary approach, containing both biological and computational analyses. The due dates for each mini-project will be listed on the schedule. The assignment for each will be handed out 2-3 weeks before they are due.

Evening exams: There will be a midterm in the evening of Monday, 29 September from 7-9 pm in (TBA). The comp time will be provided by having no required lecture during class time on that day, instead class time on 29 September will be used for an optional review session. The exam is open book, open notes, but no calculators and no computers. It will cover material through Friday, 26 September.

Quizzes: There will be two quizzes. Each quiz will last 30 minutes. The time before the quiz will be used as catch-up time for the lectures and time for a question and answer review of the material. You will be able to make a single sheet of notes (one side) for reference, but otherwise the quizzes will be closed book, closed notes and no calculators or computers.

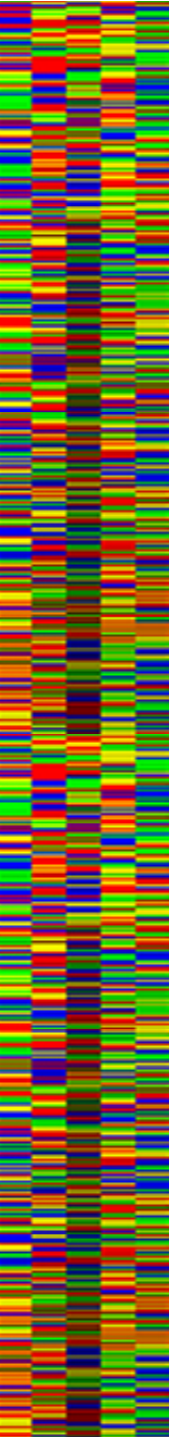
Assessment and Grading: Grades in the course will be based on one mid-term and one final exam, together worth a total of 40% of the course grade. The balance of the course grade (60%) will be based on homework assignments, mini-projects and quizzes. The breakdown of points follows (note that this may be tweaked slightly):

Activity	Points
Midterm	100
Final	300 (See note below)
Homework	200 (20 points each)
Quizzes	100 (50 points each)
Mini-projects	300 (100 points each)
Total	1000

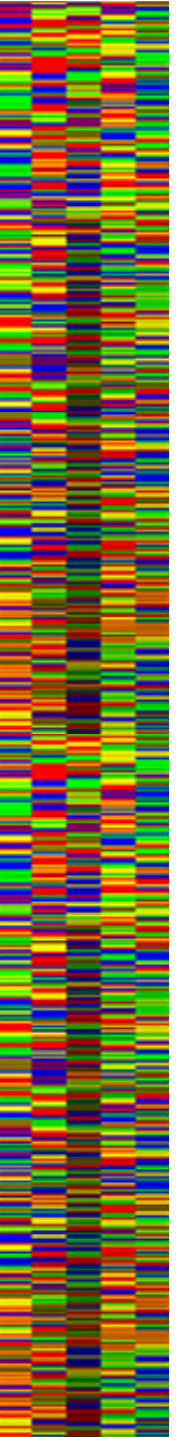


Policies - Academic behavior

- ***Academic dishonesty of any kind (cheating, plagiarism, fabrication of data, improper collaboration, etc.) is not tolerated and is grounds for failing the course (grade F) and notification of University administration for further disciplinary action.***
- ***All assignments will be explicitly labeled for individual versus group effort; groups will be instructed as to the rules for collaboration.***
- ***All questions about course policy and administration should be directed to the Course Coordinator.***



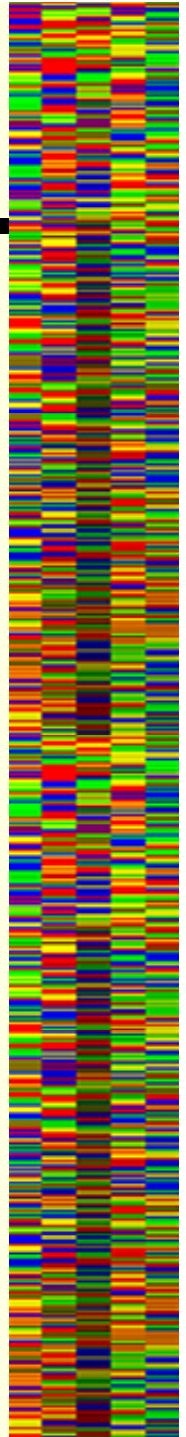
-
- ***Module 1: Genomics***
DNA and protein sequence analysis
 - ***Module 2: Evolution and Phylogenetics***
 - ***Module 3: Systems Biology***
 - ***Module 4: Protein structure***



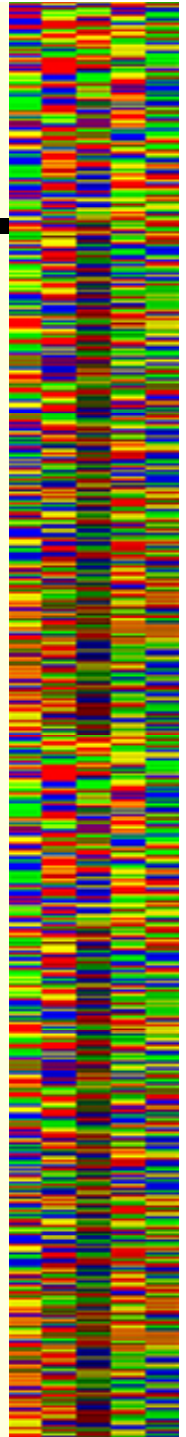
Genomics

Genomics – Study of the whole nucleotide sequence of an organism

```
TTCTTTCTGTGGTGCTTTACAAAATCTTGAACCTTCTGGAATTGATCAAGCAGATAGACG
AACGAAATACTGGAATAACAGTTAAAGATCGTGCTGCTTTTAAAAAATTTTAGAAGCT
ACCAAACAAAGCAAATTCAGTGTATTGCACCTAATTGCCAAAAACAAGTCTCTCCTTT
ACAATATTCGAAAAATAATAACTTTATATATAATTCCGGTACTACAAAGGGTATAGTTT
TGGATAACAGGCATGTGTTTAATATCTTACAAAATCTTCCACAAACGTTTAAATTATTG
TTAACCCCTTCGAATGCTCATCAAATCGTATCTCCCGAAAATGTCTTTTATGCTAATAG
TATCTTACTTCCACCACATAATCTACGAACTATCAATGTTTATGATGGTCAGGTTACGA
GTTTGTTAACAAGTGATTTGAATCTGATAATCGAAGAGTTGCTAATAATGAGACAAAT
GCAAAAATACAAAAAATCTTGGATTCTATCGATAACAGCCGAGGTGCCAATCCATATGC
TACAAATAAAAAGCTTACTTTGGATACTTTGACAGGTGGACACTCAAAGAATCTTATT
TGCGAAGTTATATTAATGGCAAACGTATTCCTGAGACTGCCAGAGCTGTAATCGAACCC
TCTATGAATAAAAAGCTTACTTTGAAAGTACCATCTTACATTTTAAACAAGTTAAGAGA
TGTTGTCTTTTATAATCACGTTACGAAAGATAACATACTCAAAGTCTTCAAACGAAC
AAGCTTTTCTAACATATATCAAAGTGATCATAATTCTGAAAATCCTTATATGGTTTAT
GATTTAGCACAGAAGAATGGATATTTAACCTTGGCTCCTAATTTCCGGTGATATTTTCGA
AAAAAGGAAAGAGGAAGGTGGTTTTGTAAGTATTTGCAGACATCCATCTATCTGGTTAA
CTAATATCCAATCTGGTATAATAAAAAGATCAGAAGGGTTTACTATTAACATCCCAACC
ACAATTTGCACATCTTTTAAATGCTGATTTTATGATGGAGATGAGATGACAATATATTCTTT
CAAATCCCATGTGCCAATCTCGAACAAGCTTTGATTATGAACTCACGAAATCTCTTCA
AAAATTTCTATAACAAGCAATCCAATGTTCCGGCTTGGTCCAAGATCAAATACCAGCCTTG
AATAAGTTATATAGACGACAAAATTATACATATAACGATGCGTTGGTGATTTTAGGACA
ATTCGGATTTCTGTTAACACCTGGAAAAGATAATTATACCGGAAAAGATATACTTTCTT
GTGTATTCCCAAACATTATACACTCAAAGGAATTGTTGAAAATGGCGAACTTATTTTG
```



Genomics



Genomics

What is genomics good for?

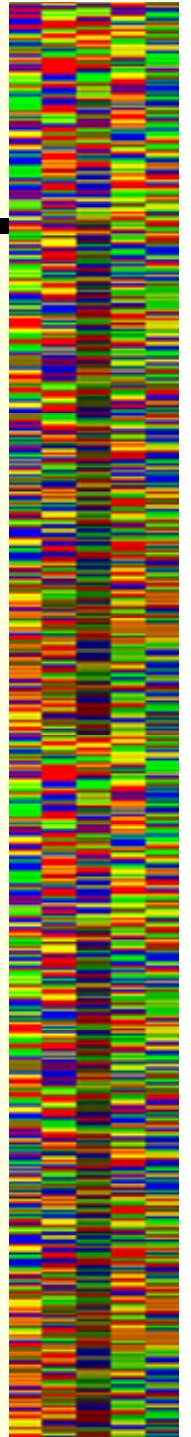
- ***Forensics***
- ***Medical diagnostics***
 - Genetic diseases
 - Drug response/sensitivity
- ***Identifying new diseases***
 - Outbreak
 - Andromeda strain



Genomics

Key technologies

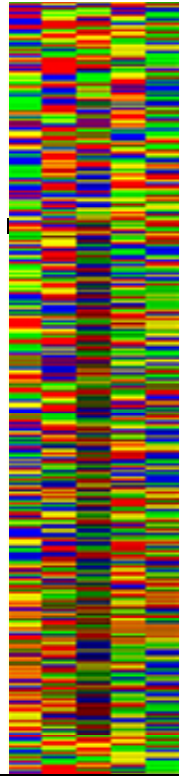
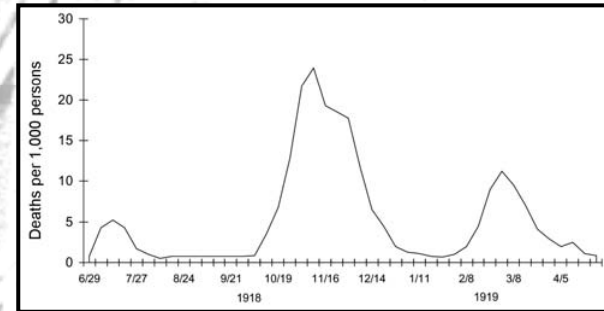
- *Methods to isolate very small amounts of DNA from environment or tissue (PCR – polymerase chain reaction)*
- *Automated machines to rapidly determine DNA base sequence (DNA sequencer)*
- *Reasoning on trees*



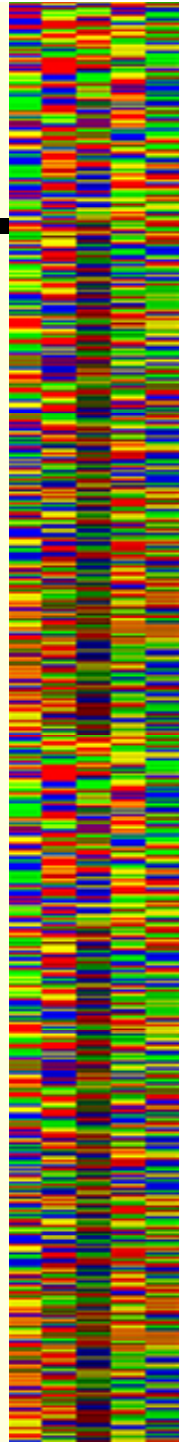
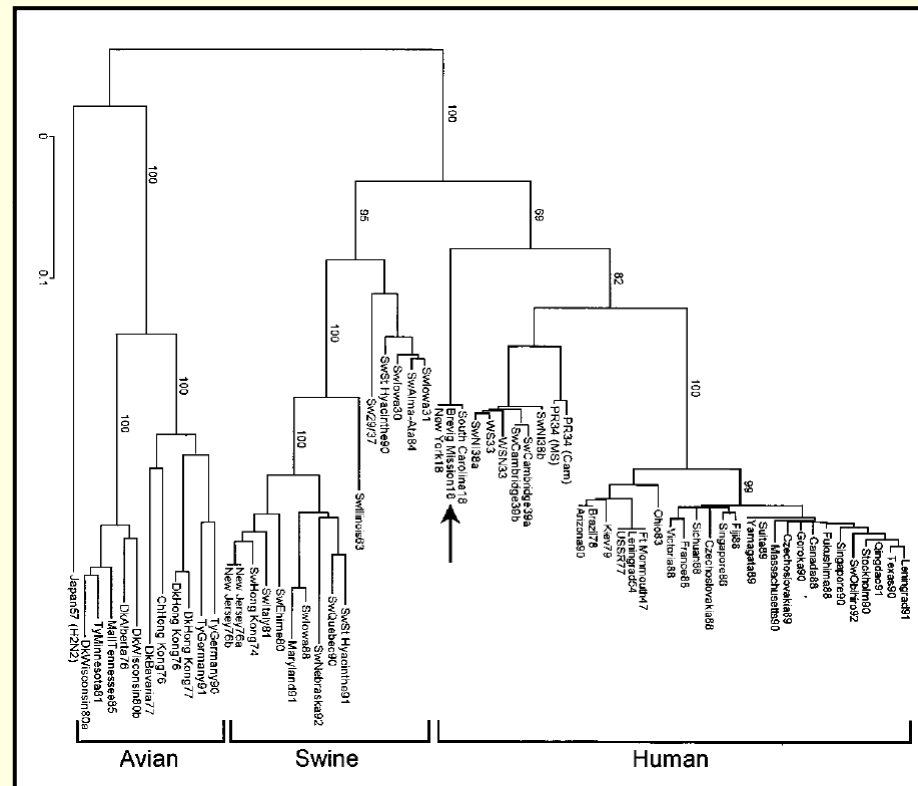
Genomics

1918 "Spanish" flu pandemic

- ***Appeared suddenly***
- ***So different from other flu strains that there was no resistance ~ 2% mortality***
- ***25-30% of world population infected***
- ***Killed 40-100 million worldwide, 675,000 US (in 4 months)***
 - Today a similar pandemic would kill in the 100 millions
 - Similar pandemic today would cause 2 million US deaths
- ***Using modern techniques, sequences have been obtained from pathology samples, frozen tissues, etc***



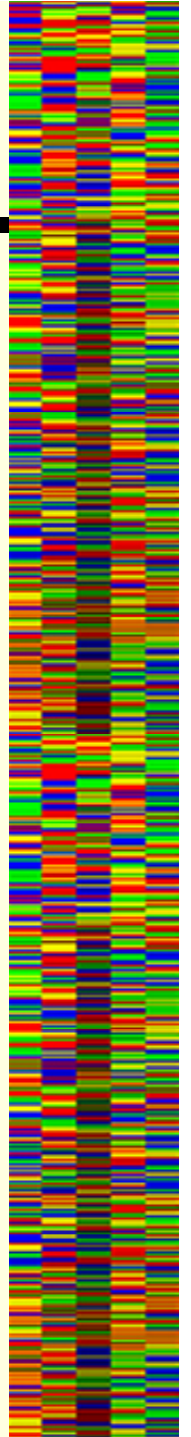
Genomics



Genomics

Mitochondrial inheritance

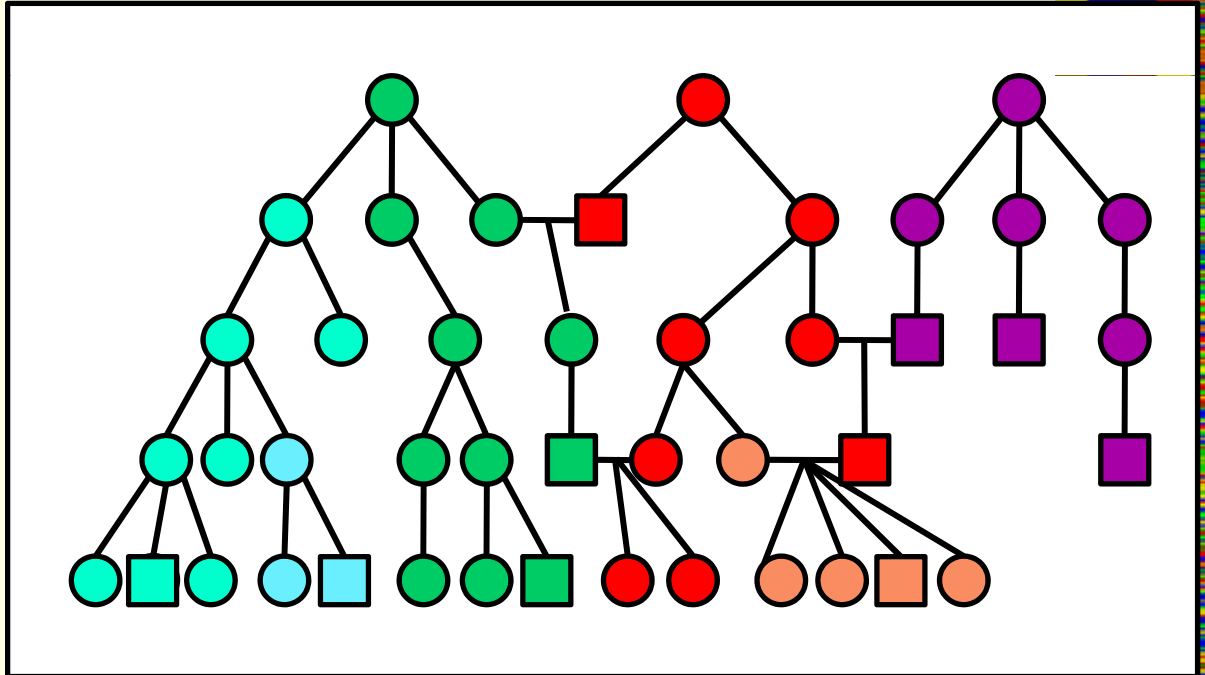
- *Mitochondria have their own DNA and replicate separately from nuclear DNA*
- *Mitochondria are inherited from the mother only*
- *Mitochondria have very limited recombination or transposable elements*
- *Good for making trees in the 10 Kyr to 1 Myr range*



Genomics

Gain and loss of mitochondrial lineages

- ***Due purely to chance, some lineages die out***
 - no children
 - no female children
- ***Random mutations gradually change sequence***
- ***Since changes are small, you can tell the relationships between new and old forms***

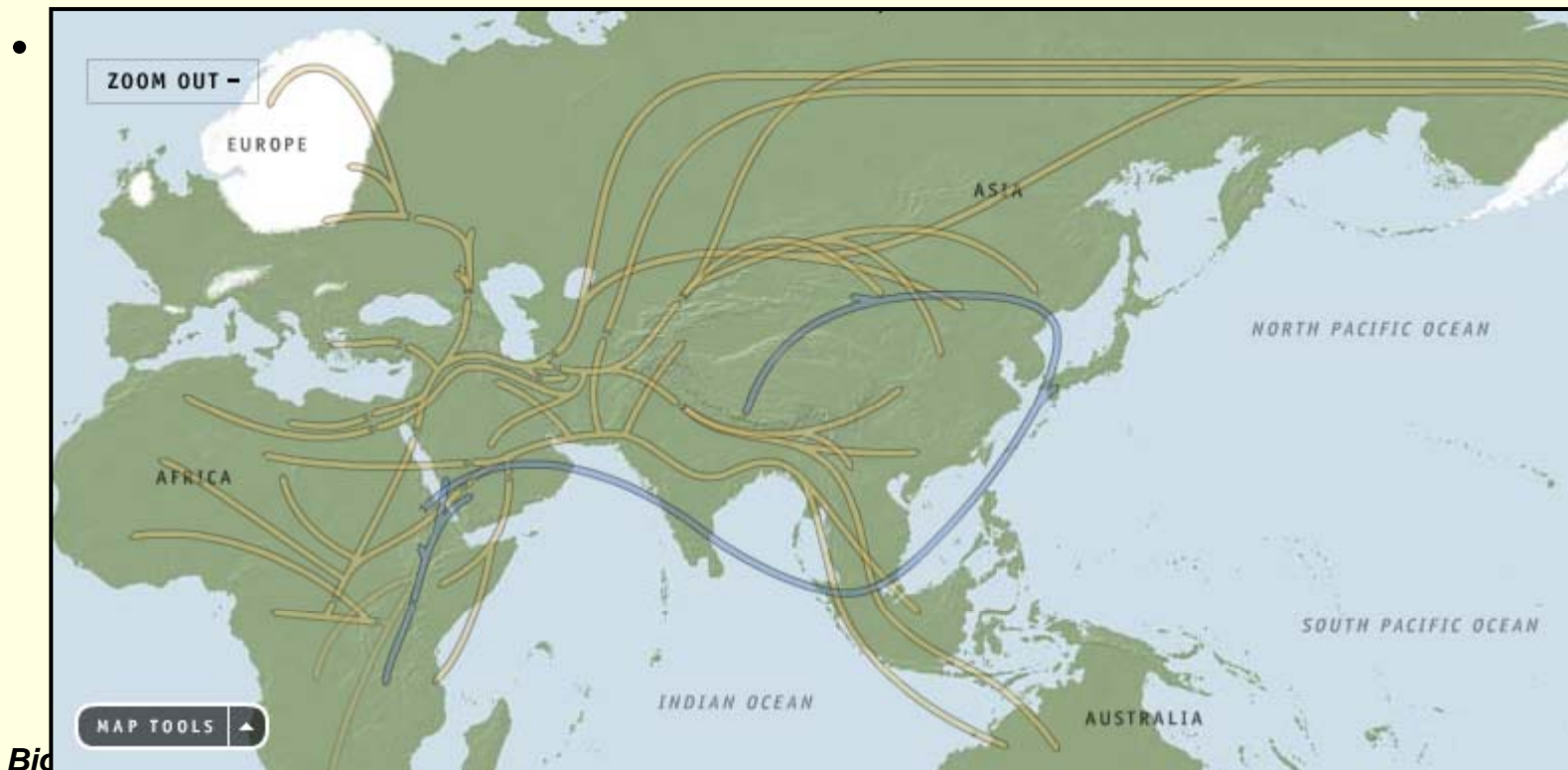


Genomics



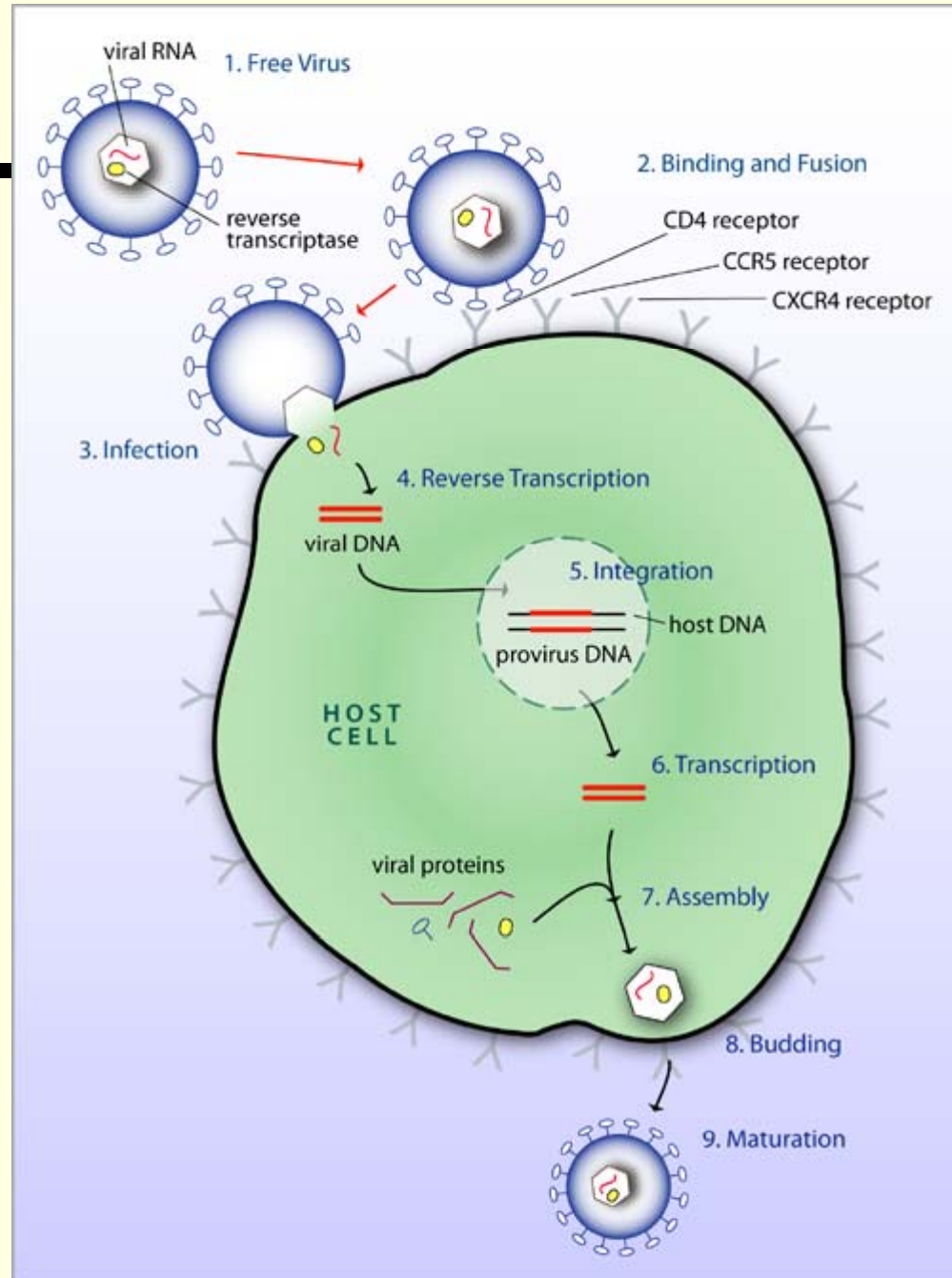
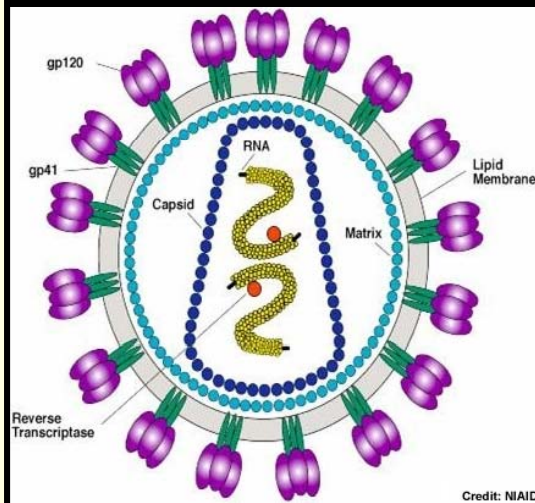
The Genographic Project

<https://www3.nationalgeographic.com/genographic/>



Genomics

HIV infection



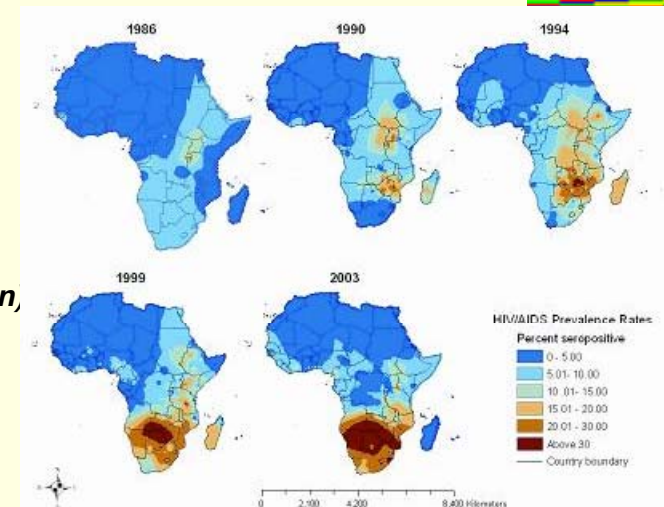
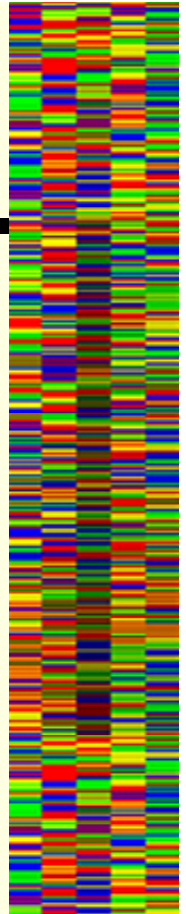
Genomics

Identification of AIDS/HIV

- 1978 - Gay men in the US and Sweden -- and heterosexuals in Tanzania and Haiti -- begin showing signs of what will later be called AIDS.
- 1980 - 31 known deaths identified in US
- 1981 - CDC reports 5 young homosexual men treated for *Pneumocystis carinii* at 3 different LA hospitals have with multiple infections including cytomegalovirus, 2 die
- 1981 - 26 cases of Kaposi Sarcoma identified in the last 30 months among gay males, 8 died within 2 years
- 1987 - first drugs
- 1981 - deaths US 234
- 1982 - deaths US 853
- 1983 - deaths US 2,304
- 1984 - deaths US 4,300
- 1985 - deaths US 2960 cumulative 16,301
- 1986 - deaths US
- 1987 - deaths US 4,100
- 1988 - deaths US 4,900
- 1989 - deaths US 14,500
- 1990 - deaths US 18,500
- 1991 - deaths US 20,500
- 1992 - deaths US 23,400
- 1993 - deaths US 41,900
- 1994 - deaths US 32,300
- 1995 - deaths US 48,400
- 1996 - deaths US 35,000
- 1997 - death count US 21,400, worldwide 6.4 million, 22 million infected
- 2004 - death count US 17,557, cumulative 524,000 (0.2 % of US population,
- 2006 - approximately 39 million infected worldwide (0.5% of population)

AIDS is defined, linked to blood transfusion
retrovirus isolated (HTLV-III, LAV)
HIV sequenced (HTLV-III)
first blood test

first drugs

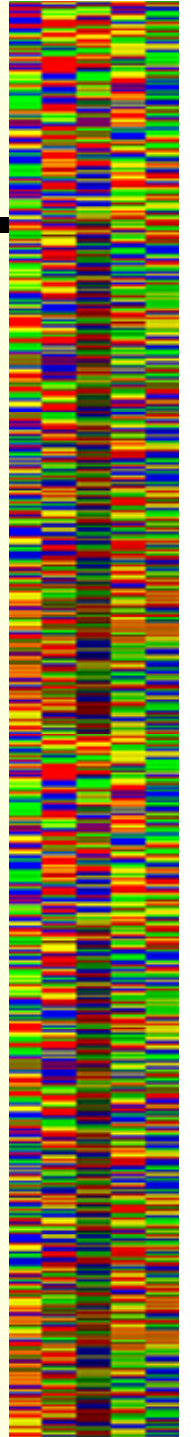


Genomics

Doctors on trial

- **1998 - HIV-1 and HepC infections reported in children at Al-Fateh Hospital, Benghazi, Libya**
 - About half (218/418) of children infected
 - Palestinian doctor and five Bulgarian nurses accused of deliberately or negligently infecting children
 - Gaddafi accuses the health workers of acting on orders from the CIA and the Israeli secret service, Mossad
- **May 2004 Libyan court sentences all five to death by firing squad**
- **Libya demands \$10,000,000 compensation, EU agrees to \$250,000**
- **November 2005 Supreme Court verdict delayed - riots in Benghazi**
- **December 2006 Supreme court confirms verdict: death by firing squad for deliberately infecting children**

- **Scientific evidence at trial**
 - Benghazi HIV Committee indicates that the first case of AIDS among the children at the Al-Fateh Children's Hospital was diagnosed in June 1997, a year before the Bulgarian nurses even began working at the Libyan hospital
 - Libyan physician who refused to be identified confirmed that the HIV strain that had infected the children, while rare, was definitely African
 - Vittorio Colizzi and Luc Montagnier co-discoverer of HIV, testified in the first trial that the HIV virus was active in the hospital before the Bulgarian nurses began their contracts there in 1998. (Based on evaluation of more than 200 infected children)



Genomics

Doctors on trial

- ***Conclusive study published in December 2006***
 - de Oliveira et al., Molecular Epidemiology: HIV-1 and HCV sequences from Libyan outbreak, Nature 444, 836-837 (2006)

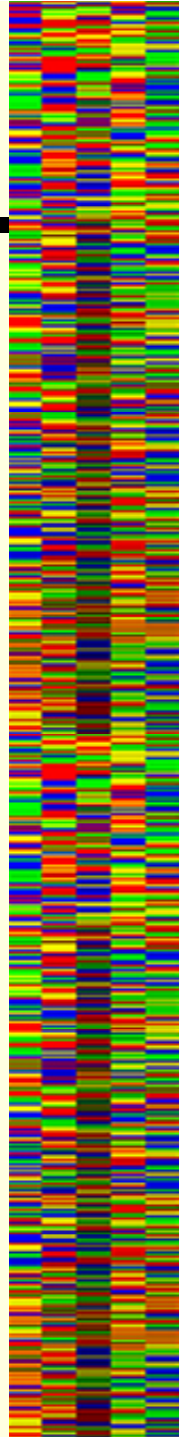
- ***Conclusions***

- HIV isolates form a single monophyletic cluster (clade) with west african isolates

"the probability that the clusters from the Al-Fateh Hospital originated after that time was almost zero"

"Al-Fateh Hospital had a long-standing infection-control problem"

"HIV-1 and HCV strains responsible were being spread and transmitted among individuals attending the hospital before March 1998, indicating that many of the transmissions giving rise to the infection clusters must have already occurred before the foreign medical staff arrived."



Genomics

Doctors on Trial

