HIV and population genetics

Please sit in row H or forward
Opportunities for summer research in biology

For more information go to: http://goo.gl/lwvNF1
Acquired immune deficiency syndrome (AIDS)

The human immunodeficiency virus (HIV) is identified as the cause of AIDS

*Science, 20 May 1983*

**Isolation of human T–cell leukemia virus in acquired immune deficiency syndrome (AIDS)**


**Isolation of a T–lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)**

F Barre–Sinoussi, JC Chermann, F Rey, MT Nugeyre, S Chamaret, J Gruest, C Dauguet, C Axler–Blin, F Vezinet–Brun, C Rouzioux, W Rozenbaum, L Montagnier
HIV epidemic in perspective

- As of the end of 2015
  - 36.7 million people with HIV
  - 35 million deaths since beginning of pandemic
  - 1.1 million died in 2015

http://www.who.int/mediacentre/factsheets/fs360/en/
Infection rates vary by region
Question: why has HIV killed so many people?
Topics for today

• A note on homework
• Introduction to HIV
• HIV evolution inside a patient (the four forces, part 1)
Coding homework assignments

- Submitted in same place as written (http://submissions.cs.hmc.edu/)
- CS5 python setup (Anaconda python 3). We will use matplotlib, scipy etc. in some assignments.
- Testcases, autograding, grade.py
- Pair programming ok
Introduction to HIV: the parts of the virus

Single stranded RNA genome (present in two copies)

Cone shaped capsid

Host derived envelope

Glycoprotein ‘spikes’

Assembled virion (viral particle)
Introduction to HIV: the replication cycle

HIV reverse transcriptase and integrase proteins insert its genetic material into the DNA of the host cell.

Host transcription and translation produce new viral proteins.
Introduction to HIV: the virus targets helper T-cells of the immune system
The immune system in brief

- Innate immune system (many kinds of cells)
- Adaptive immune system
  - B cells (B lymphocytes)
  - Cytotoxic T cells (cytotoxic T lymphocytes)
  - Helper T cells (helper T lymphocytes)
Plotting the course of an infection: helper T-cells over time

HIV infection causes a drop in helper T-cell counts over time

The immune system’s response to HIV

Antibody protein

HIV virion

Antibodies bound to glycoproteins on the surface of the virus.

Cytotoxic T-cell of immune system

HIV infected cell

Dead cell

Cellular immune response attacks cells with HIV provirus.
...however HIV evolves to evade the patient’s immune response

The AIDS phase

• Weight loss
• Severe fatigue
• Fever / night sweats
• Nausea/vomiting/diarrhea
• Secondary infections

http://hondurasaids-photographyproject.info/galleries/the-hospice/
Secondary infections are typically what kills AIDS patients.
Topics for today

• A note on homework
• Introduction to viruses and HIV
• HIV evolution inside a patient (the four forces, part 1)
HIV evolution inside a patient

- Some terms
  - Population: individuals from the same strain living inside a single patient
  - Allele: a genetic variant

Two alleles from an HIV gene (rev)

...CAAUUAUGGAUGAUUUUAUGUAGGAUCUGAC...
...CAAUUAUGGAUGAUUUUAUGUAGGAUCAGAC...
A single founder virus replicates to produce a huge population

- Population in blood: \(~300\) million virions
- Daily turnover: \(~100\) million new virions
Four forces that can change allele frequencies in populations

• Mutation
• Genetic drift
• Natural selection
• Migration
Four forces that can change allele frequencies in populations

- **Mutation**
- Genetic drift
- Natural selection
- Migration
Mutations in HIV

...AACAGACAUACAGGAAAAAGUAUACGGAU...

...AACAGACAUACAGGAGAAAGUAUACGGAU...
HIV (like other RNA viruses) has a high mutation rate

~ $10^{-5}$ mutations per nucleotide in HIV

~ $10^{-8}$ mutations per nucleotide in humans

~ $10^{-9}$ mutations per nucleotide in *Drosophilla*

~ $10^{-10}$ mutations per nucleotide in *E. coli*
1. What is the expected number of mutations in the whole genome each time the HIV virus replicates?

2. How many mutations occur in a typical patient in a day?

3. Many advantageous mutations that occur in HIV are single nucleotide mutations. From a given viral genome sequence, how many possible single nucleotide mutations are there?

4. What do the above considerations say about the ability of HIV to evolve as it tries to overcome the immune system?

What is a single nucleotide mutant?

A sequence: \(\text{ACCGT}\)

Example single nucl. mutant: \(\text{ACCTT}\)

Example double nucl. mutant: \(\text{CCCGA}\)
**Worksheet**  
(Rip it off from the back of your packet)

**Name:**

HIV by the numbers. Note: the HIV genome is $10^4$ nucleotides long.

1. What is the expected number of mutations in the whole genome each time the HIV virus replicates?

   \[10^{-5} \times 10^4 = 1/10\]

2. How many mutations occur in a typical patient in a day?

   \[1/10 \times 100 \text{ million} = 10 \text{ million}\]

3. Many advantageous mutations that occur in HIV are single nucleotide mutations. From a given viral genome sequence, how many possible single nucleotide mutations are there?

   \[3 \times 10^4\]

4. What do the above considerations say about the ability of HIV to evolve as it tries to overcome the immune system?

   HIV can evolve to evade the immune system because it produces a very large number of mutations, among which some advantageous ones are likely to arise.

---

**What is a single nucleotide mutant?**

- **A sequence:** ACCGT
- **Example single nucl. mutant:** ACCTT
- **Example double nucl. mutant:** CCGA
Partial sequences from a single HIV patient

...AACAGACATACAGGAAAAAAGTATACGGAT...
...AACAGACATACAGGAGAAAGTATACGGAT...
...AACAGGCATACAGGAAAAAAGTATACGGAT...
...AACAMAATACAAAGAAAAAAGTATACGAAT...
Four forces that can change allele frequencies in populations

• Mutation
• **Genetic drift**
• Natural selection
• Migration
A simple model of allele frequency changes due to genetic drift

- Two alleles
- Fixed population size
- Organisms have only one copy of each sequence
- Asexual reproduction
- Make next generation by sampling with replacement
- No new mutation
Generation

1

A1
A2
A2
A1

2

A2

A2
Repeatedly sample until next generation is same size as first.
Generation

1

A2
A1
A2
A2
A1

2

A1
A2
A2
A2
A1

3

A1

A1
Generation

1

2

3
Genetic Drift

![Graph showing the frequency of A1 allele over generations with a popsize of 100. The x-axis represents generations ranging from 0 to 60, and the y-axis represents the frequency of the A1 allele ranging from 0 to 1.0. The graph displays fluctuations in allele frequency over time.]
Genetic Drift

Frequency of A1 allele vs. Generations

Popsize 100

A1 allele fixed
Genetic Drift

- A1 allele fixed
- Popsize 100
- A1 allele lost

The graph shows the frequency of the A1 allele over generations with different colored lines representing different populations. As generations progress, some lines show the allele frequency decreasing, indicating the possible loss of the A1 allele. The y-axis represents the frequency of the A1 allele, and the x-axis represents generations.
Hand in your worksheets please!

(and be sure you put your name on it)