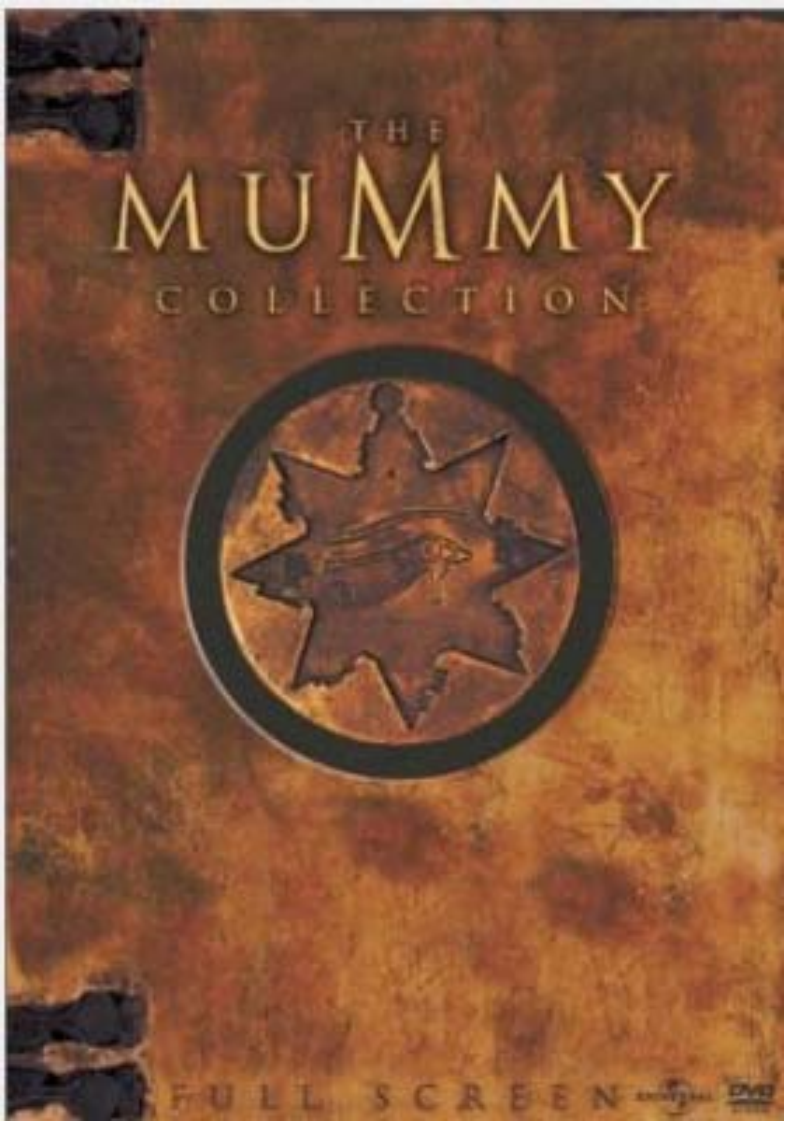
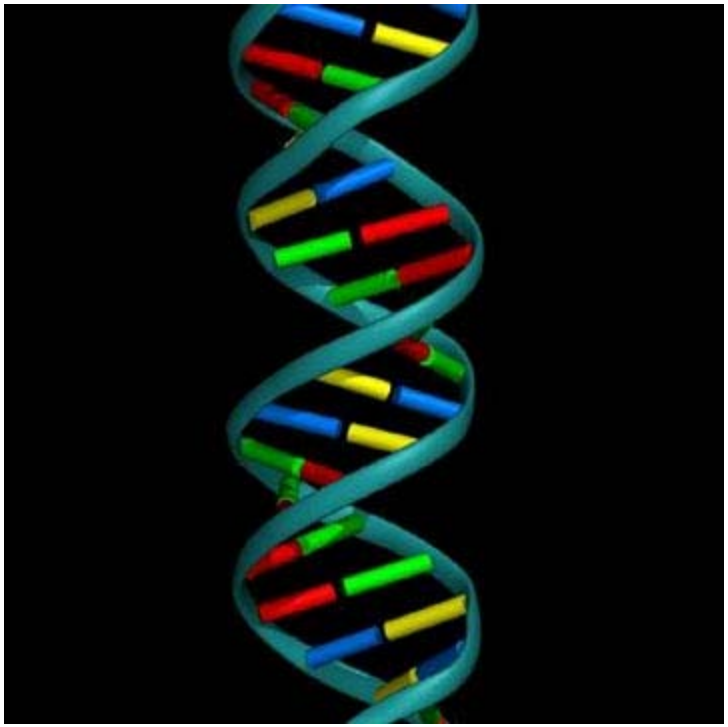
Hamilton

Fun with “Words”

Wei-Jen Chang

Book of Life





A, T, C & G

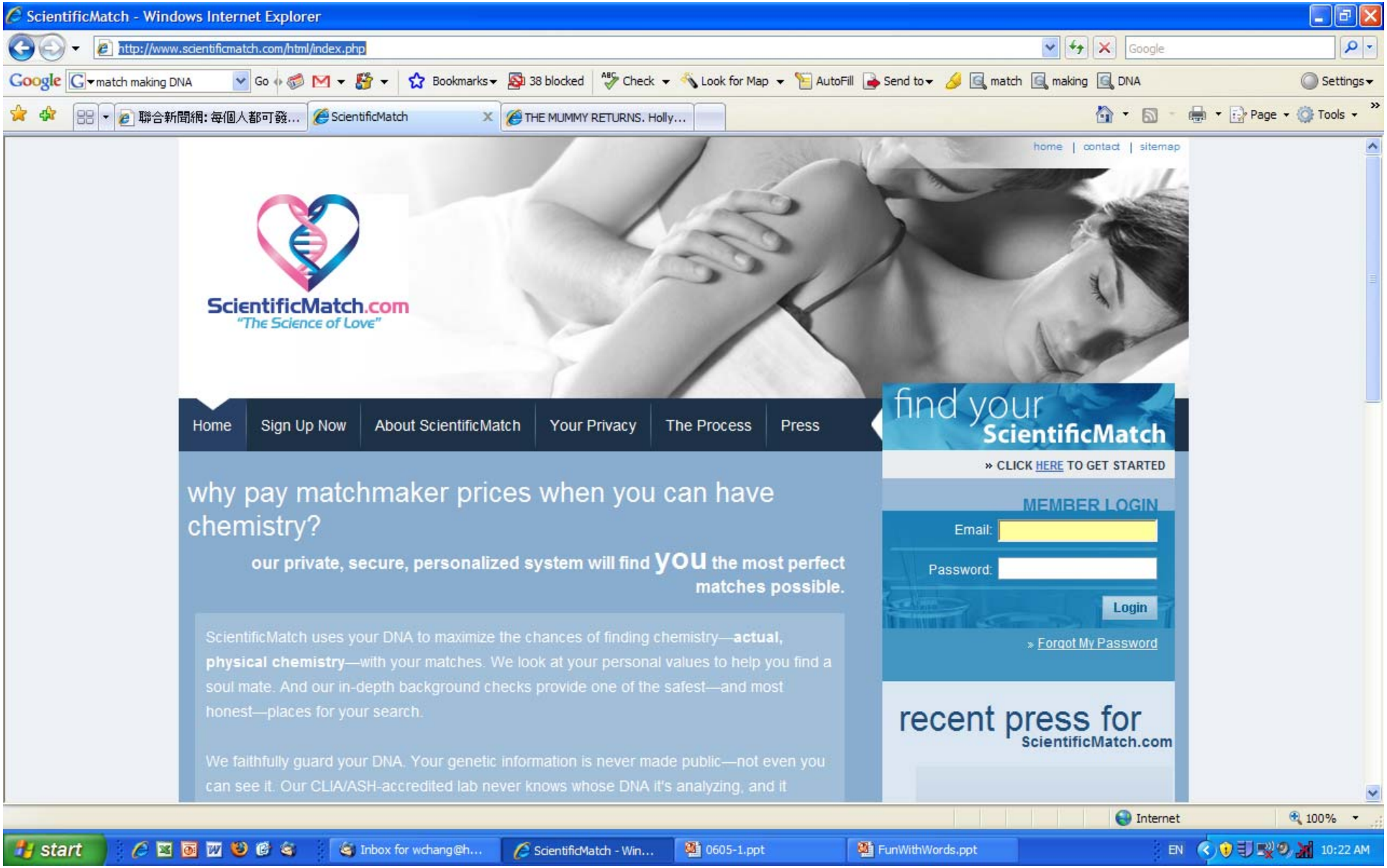
http://www.imdb.com/media/rm1092524288/tt0209163

http://www.sciencemuseum.org.uk/on-line/lifecycle/images/1-2-6-3-1-2-1-0-0-0-0.jpg

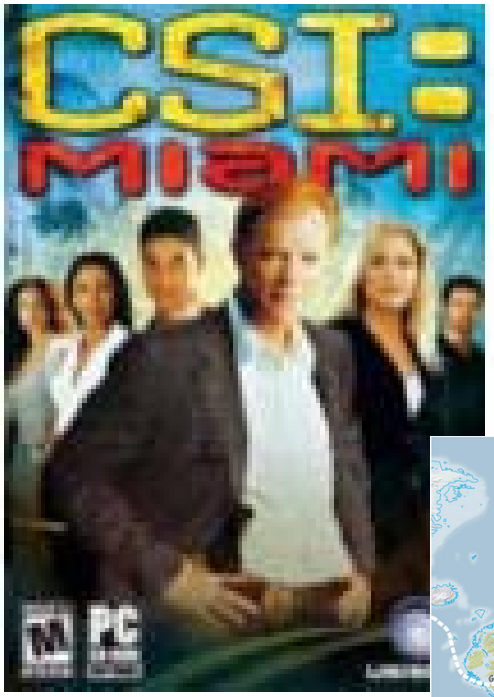
The Informatics Era

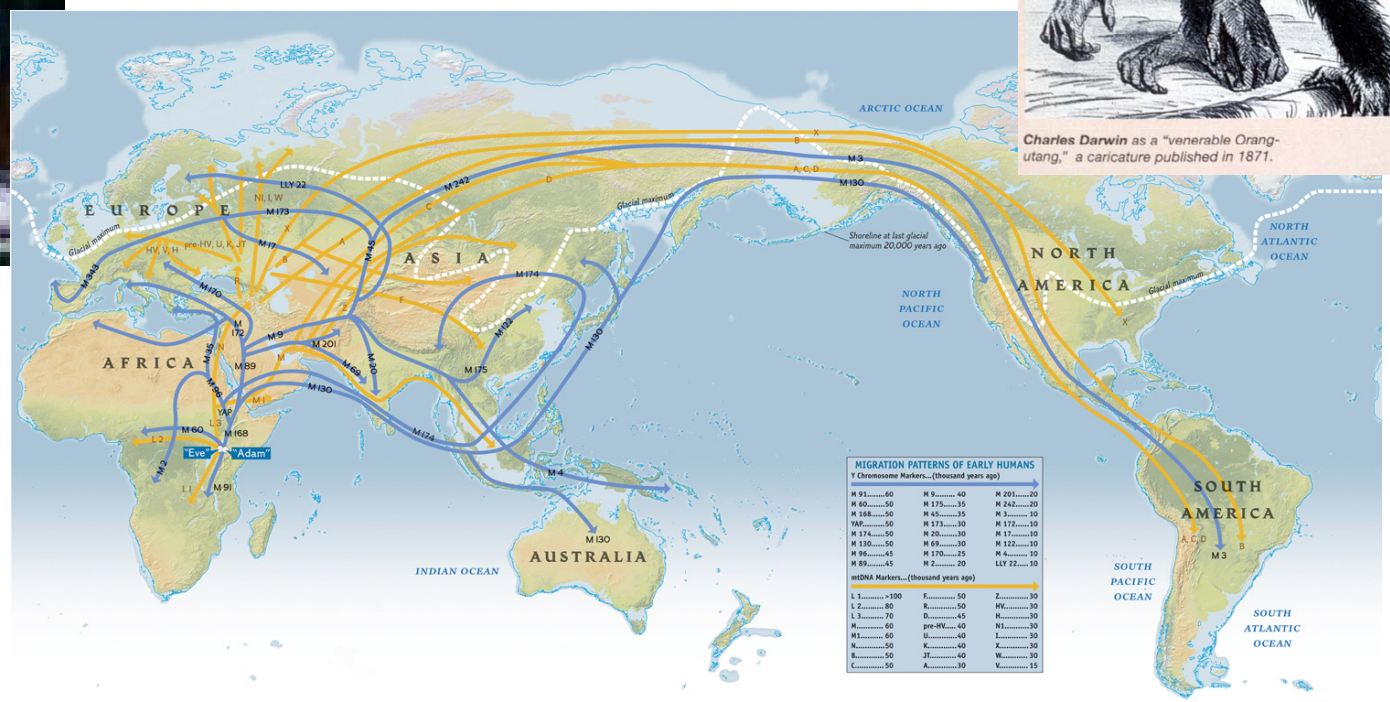
- Size of human genome (total of A, T, C & G):
- 5 yrs ago, you have to be a billionaire to sequence your genome.
- In 5 yrs, ~ 10,000 dollars/per person. Personalized genome, or iGenome (You think I am joking?).

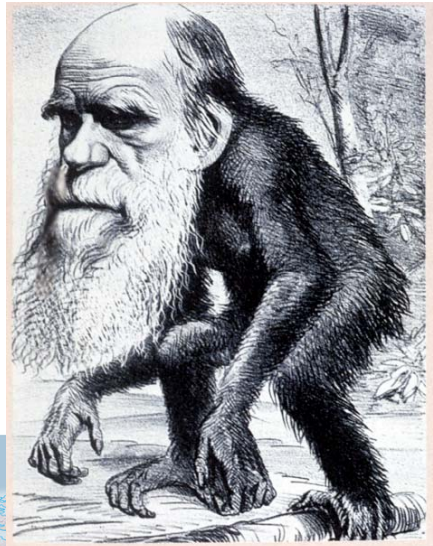
Matchmaking using DNA seq.



Other Applications of Genomes







Charles Darwin as a "venerable Orang-utang," a caricature published in 1871.

Our Goals Today

- Analyze sequence patterns of two viruses – HIV-1 and Human Papilloma Virus
- HIV1.txt and Papi68.txt
- And we are using Perl.

Print out your results on screen

- print “I am the King of the world!”;
- The thing that you want to say is flanked by “ ”.
- The semicolon **;** is required! It tells your computer that this is the end of the print command.

Numeric Variables

- Start with a dollar $ sign.

- $i = 5; #assign 5 to variable $i
- $y = 10; #assign 10 to variable $y
- print $i*$y;

- print “\n”; #This will print a line breaker

While Loop

- It’s like English – while (something is true), do this and that.

- While ($i < 5) {

```
        do this;
        do that;
        .
        .
        .
}
```

Print 1 to 10

```
$i = 1;
while ($i <= 10) {
        print $i.”\n”;
        $i = $i + 1;
}
```

Word match syntax

- $DNAseq =~ m/A/gi

This syntax will search matches (in this case is A) in $DNAseq ***g***lobally and case ***i***nsensitively.

The only thing you need to focus is the thing between two slashes. This is what you are searching for!

Repetitive Pattern

- (.)(.)\1\2 will search for something like:

ATAT

GAGA

- print $&; #will print the current match.