CSC 334: TOPICS IN COMPUTATIONAL BIOLOGY

"Algorithms for Genomic Data"

Fall 2015

Smith College

Instructor: Prof. Sara Sheehan

Outline: 9/9

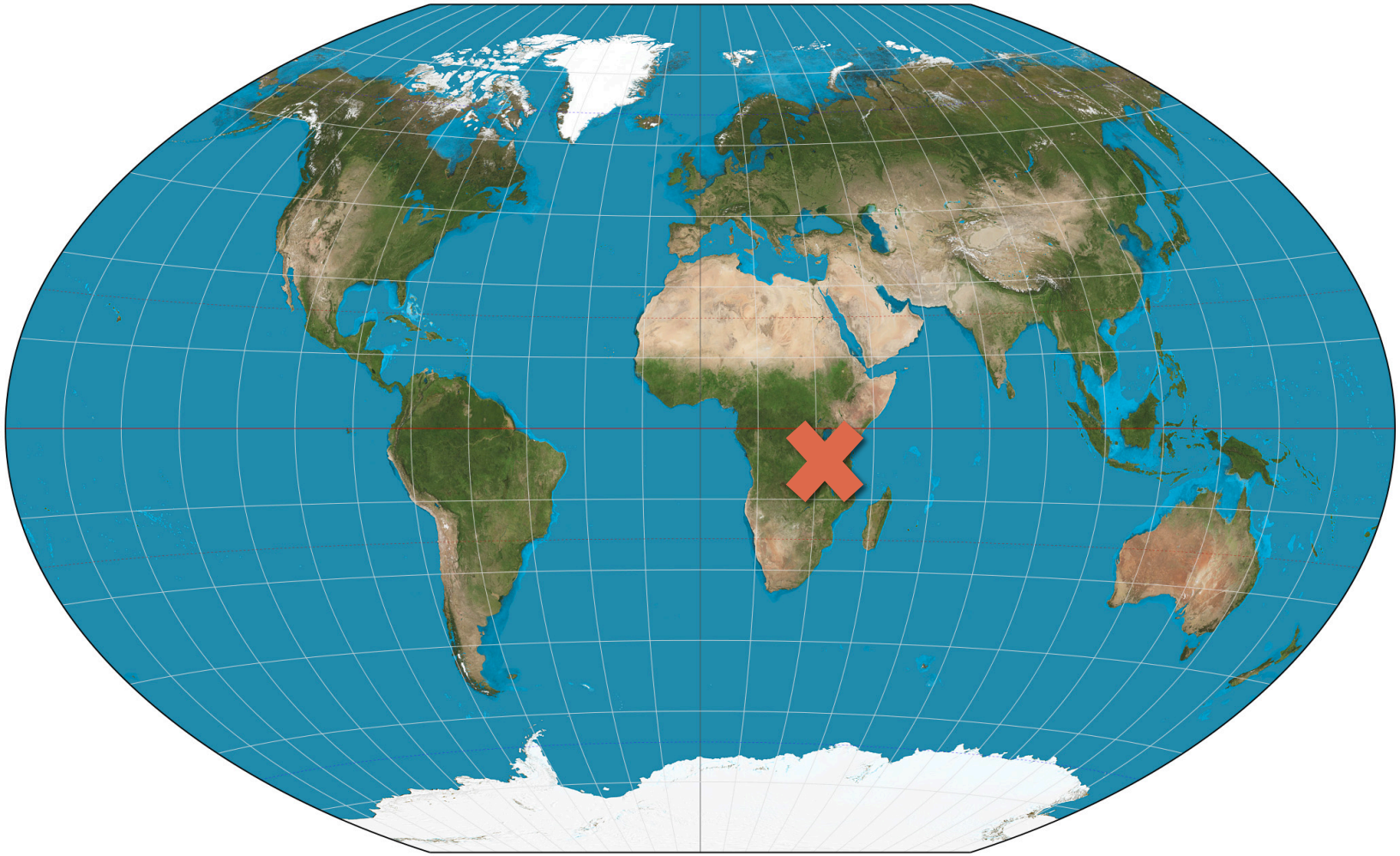
- Introductions
- Computational biology overview
- Syllabus
- Genetics basics

What is computational biology?

Learning from data

How did humans move out of Africa?

Origin of modern humans



What is computational biology?

Obtaining the data

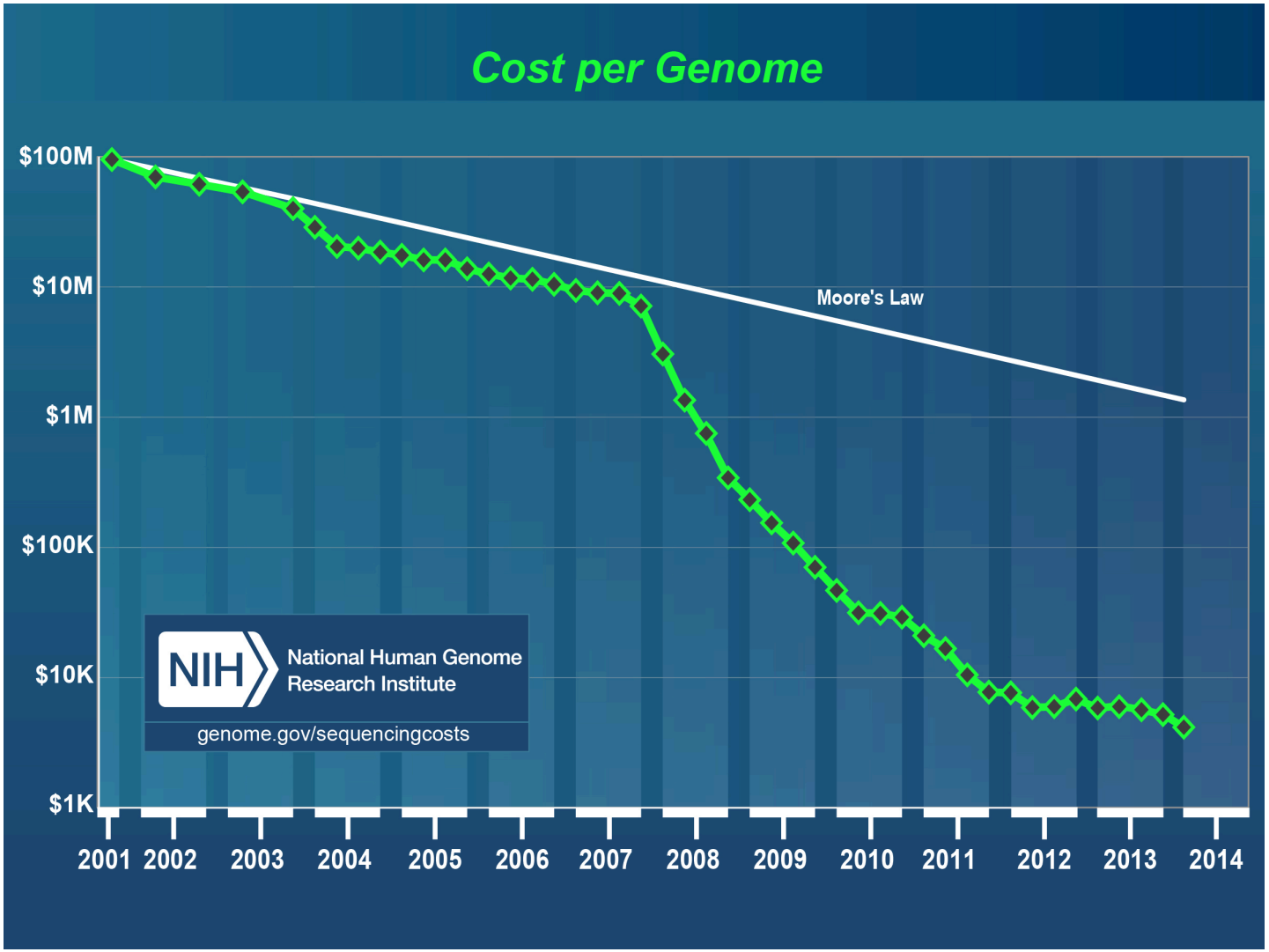
Learning from data

How did humans move out of Africa?

Obtaining the data

Illumina

Getting cheaper and cheaper



What is computational biology?

Obtaining the data

Sequence genomes
Base calling
Error correction
Genome assembly
$$$

Learning from data

How did humans move out of Africa?

What is computational biology?

Modeling

Migration
Population size changes
Admixture
Natural Selection

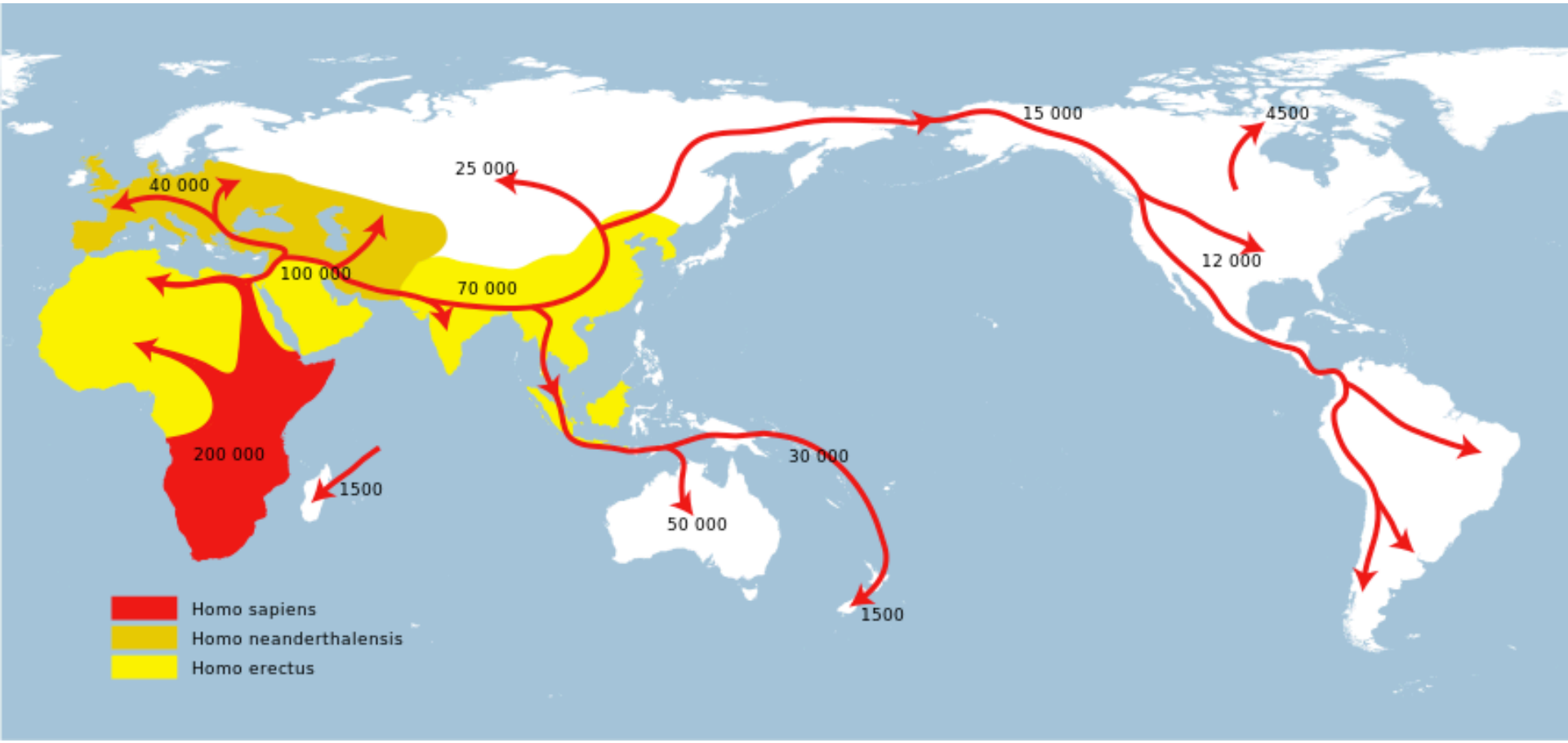
Obtaining the data

Sequence genomes
Base calling
Error correction
Genome assembly
$$$

Learning from data

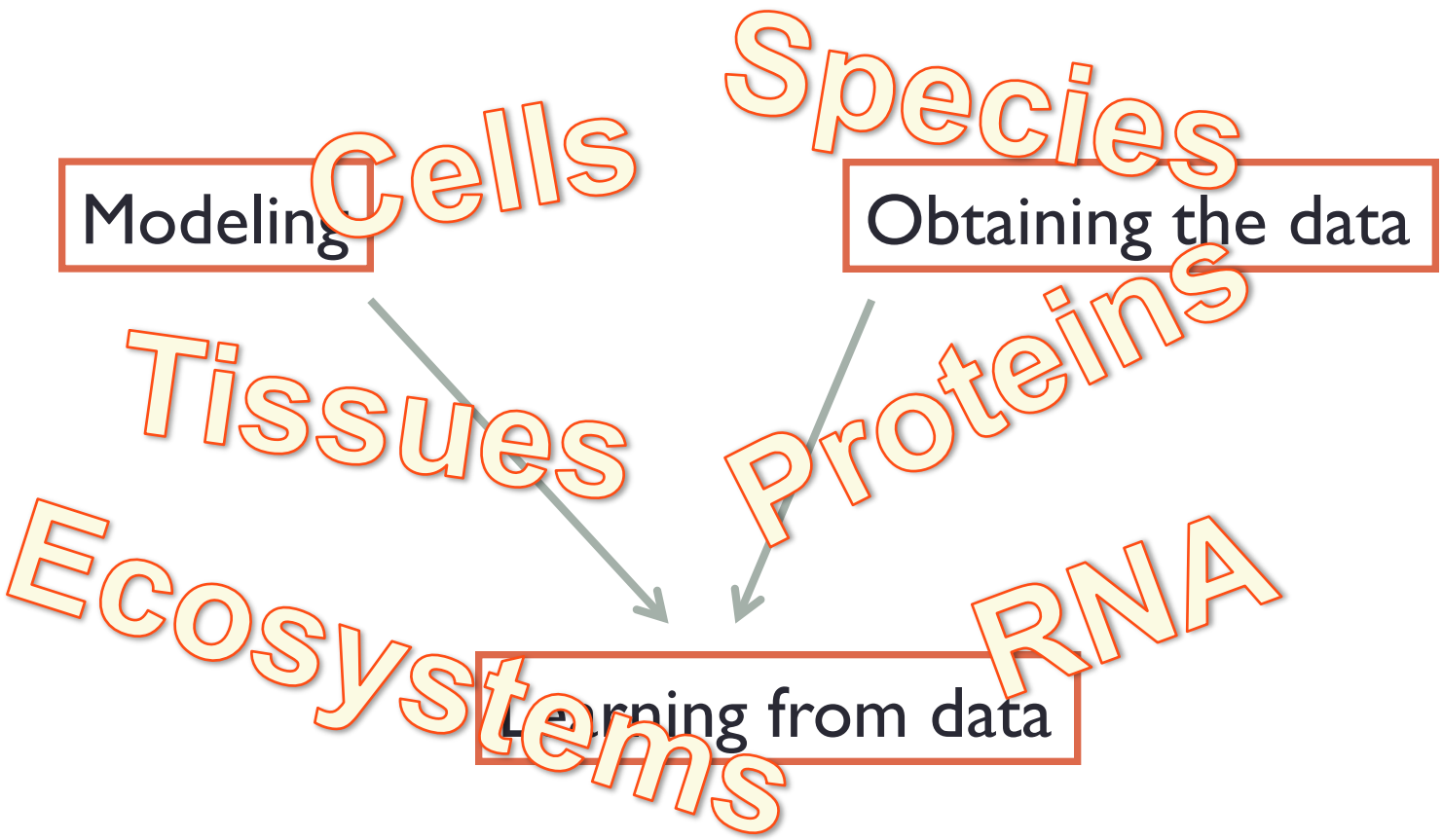
How did humans move out of Africa?

Human migration out of Africa: hypothesis



NordNordWest

What is computational biology?



This class:

Modeling

Population models
Statistics
Phylogenetic models

Obtaining the data

Algorithms after sequencing

DNA

Learning from data

Historical questions

Syllabus

Course goals

- Appreciate biologically interesting questions
- Understand and implement algorithms that can be used to solve them
- Use existing tools on genomic data
- Read and understand current computational biology literature
- Communicate ideas effectively
  - In front of an audience
  - In discussion
  - On paper
- Be a non-discriminating and independent learner

Topics (tentative)

- Next-generation sequencing
- Genome assembly
- Alignment (string matching)
- Genome storage (BWT)
- Phylogenetics (tree building)
- Fitch's algorithm
- Sequence diversity and comparative genomics
- PCA in genomics
- Natural selection
- Population growth and decay
- GWAS

Prerequisites

- CSC 111: Introduction to Computer Science
- Math 111: Calculus 1
- (helpful but not required) CSC 212: Data Structures

Assignments

- Mix of programming, pencil-and-paper, reading
- Roughly weekly in the beginning
- Move into research papers and projects
- Python (any language for the final project)
- Submitted through Moodle

http://cs.smith.edu/~ssheehan/fall15/csc334/home.html

Getting help

- Office hours (Monday: 4-6pm, 355 Ford Hall)
- Fellow students
- Tutoring at the Spinelli Center for Quantitative Learning

Honor code

- Collaboration encouraged!
- Please cite:
  - student collaborators
  - online resources
  - books and research papers
- Course goal: learn how to use available resources

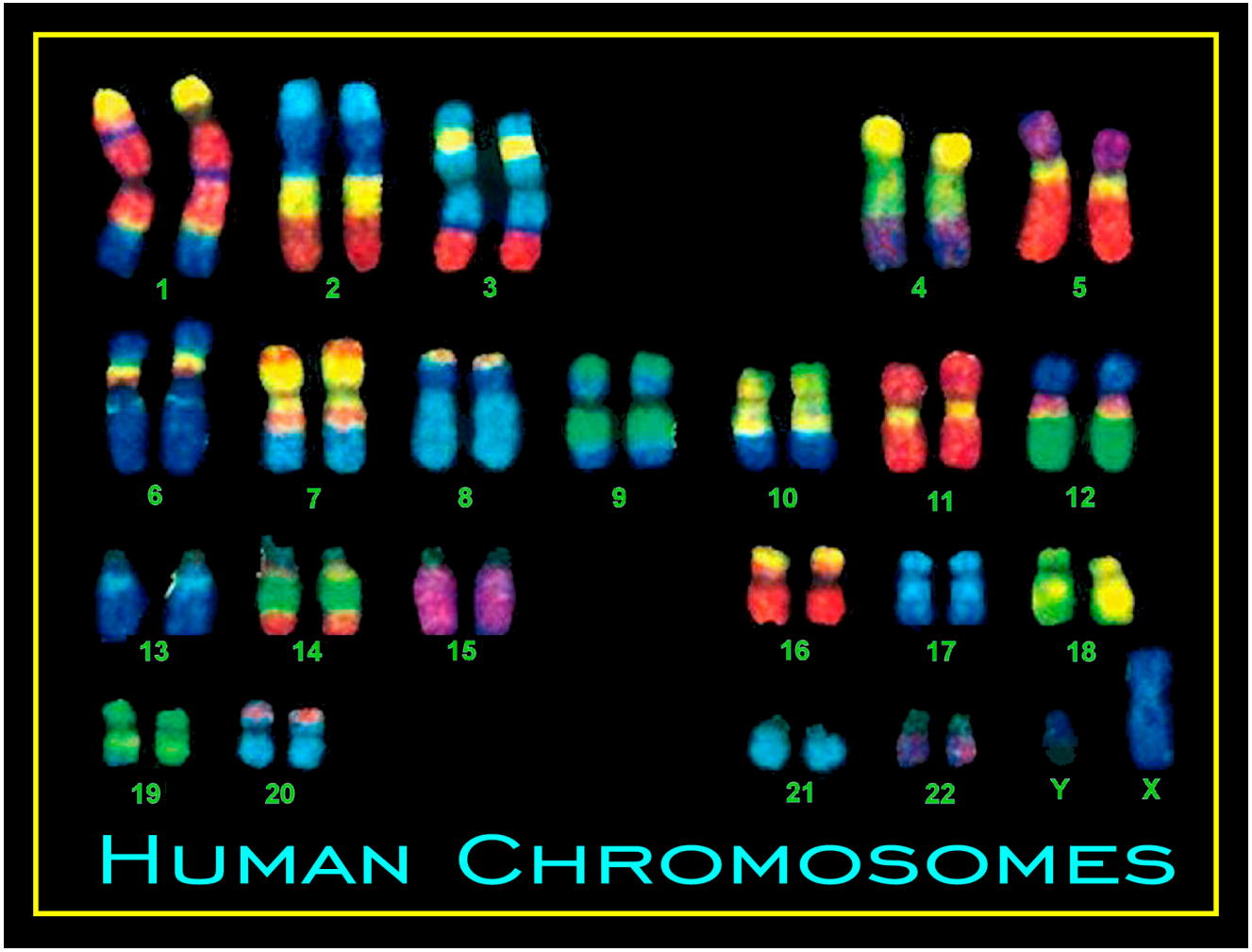
Assessment

- Homeworks: 35%
- Midterm assessment: 20% (Oct 23, tentative)
- Project presentation: 20% (last week of classes)
- Project writeup: 15% (due Dec 18, tentative)
- Participation: 10%

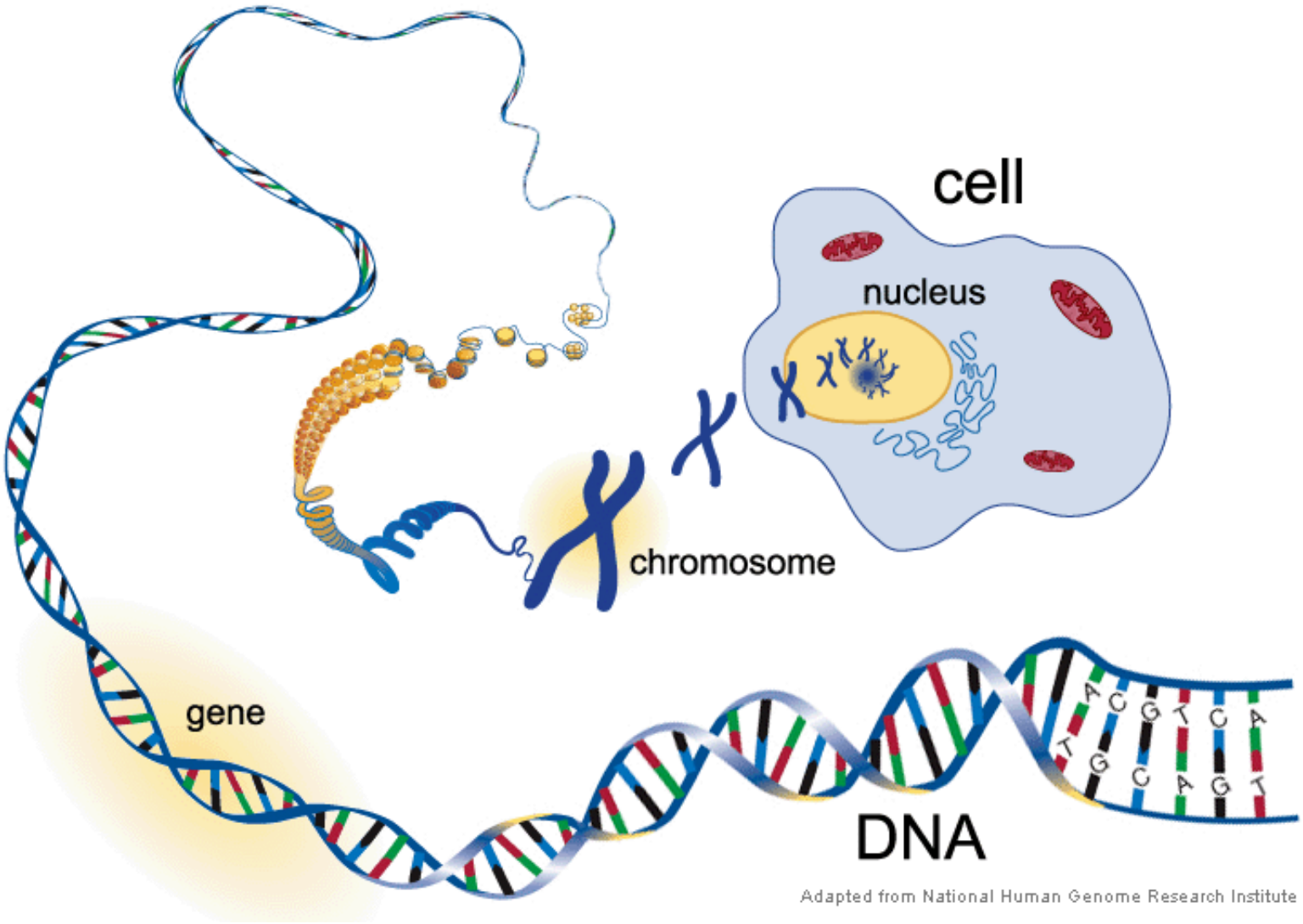
What is DNA?

What forces shape DNA?

Chromosomes



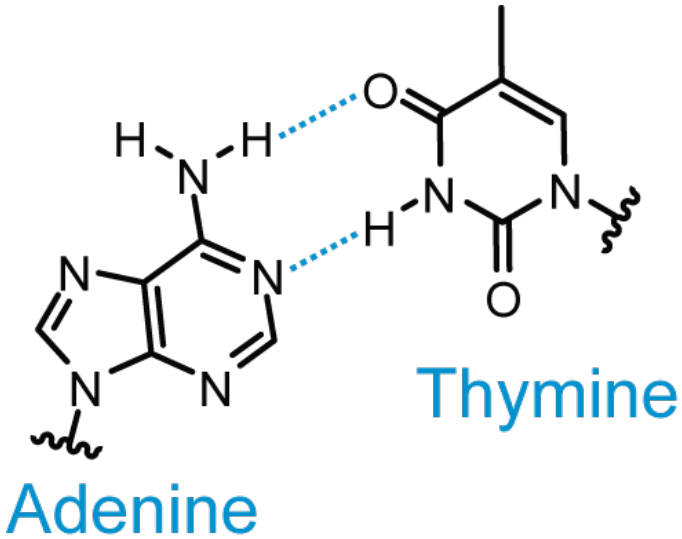
Zooming in on a chromosome

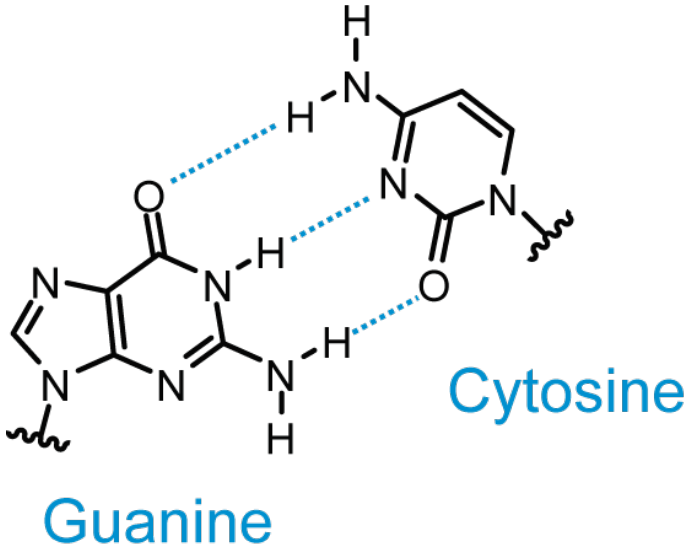


Adapted from National Human Genome Research Institute

Base-pairing

- “A” with “T”
- “G” with “C”
- Humans: 3 billion base pairs





Not only humans…

Chimp

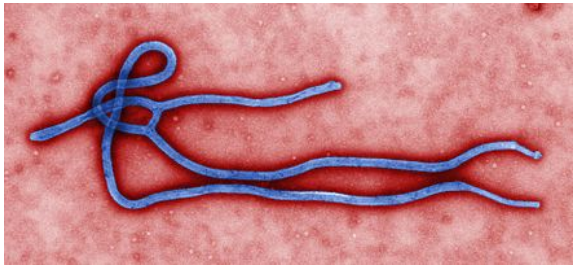
Melitaea cinxia

Buffalo

Images: wikipedia

Maize

Chinese liver worm



Ebola

Loa loa (eye worm)

DNA as a string

GCCTAGCTAGGTTACGTACG

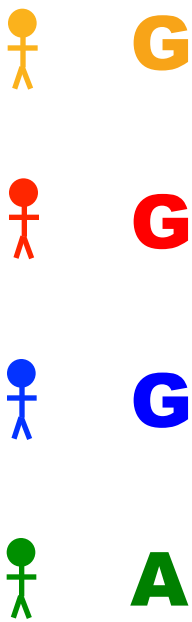
GCCTAGCTAGGTTACGTACG

GCCTAGCTAGGTTACGTACG

ACCTAGCTAGGTTACGTACG

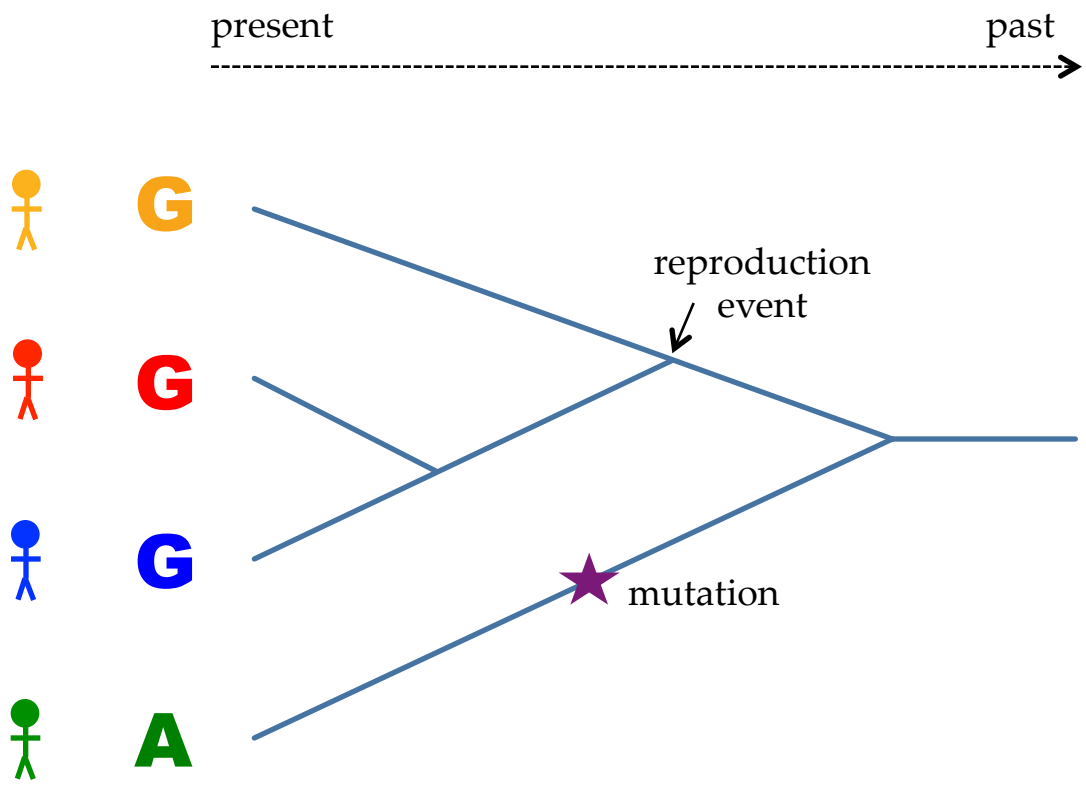
How often will we see a difference?

SNP: single-nucleotide polymorphism



Why are they different at this base?

Mutation



Next-generation sequencing applications

- Personalized genomics
- Tree of life (phylogenetics)
- Genomic forces: selection, recombination
- Disease association studies