

Genomics & Bioinformatics

Gen711/811

<https://git.io/v6AZB>

ANNOUNCEMENTS

- Email: Matthew.MacManes@unh.edu
- Office hours:
 - Tuesday/Friday 8-9am in 189 Rudman, or by appt.
 - Schedule at <http://genomebio.org/office-hours/>

ANNOUNCEMENTS

- Lecture: (you've figured this out)
- Lab: Friday 1-3 and 3-5 in Hewitt 301
- Canvas: Grades and other unimportant things
- Website: <https://git.io/v6AZB>
- Slack: Find it via the website

GOALS OF THE CLASS

- Cloud Computing
- UNIX
- Don't be afraid of the command line
- Assemble genomes and transcriptomes
- Map billions of sequence reads
- Annotate genomes
- Learn some biology
- Have fun!

SYLLABUS

- <https://git.io/v6AZB>





Intro to Bioinformatics

**Lecture 1
Aug 29, 2016**

GENOMICS

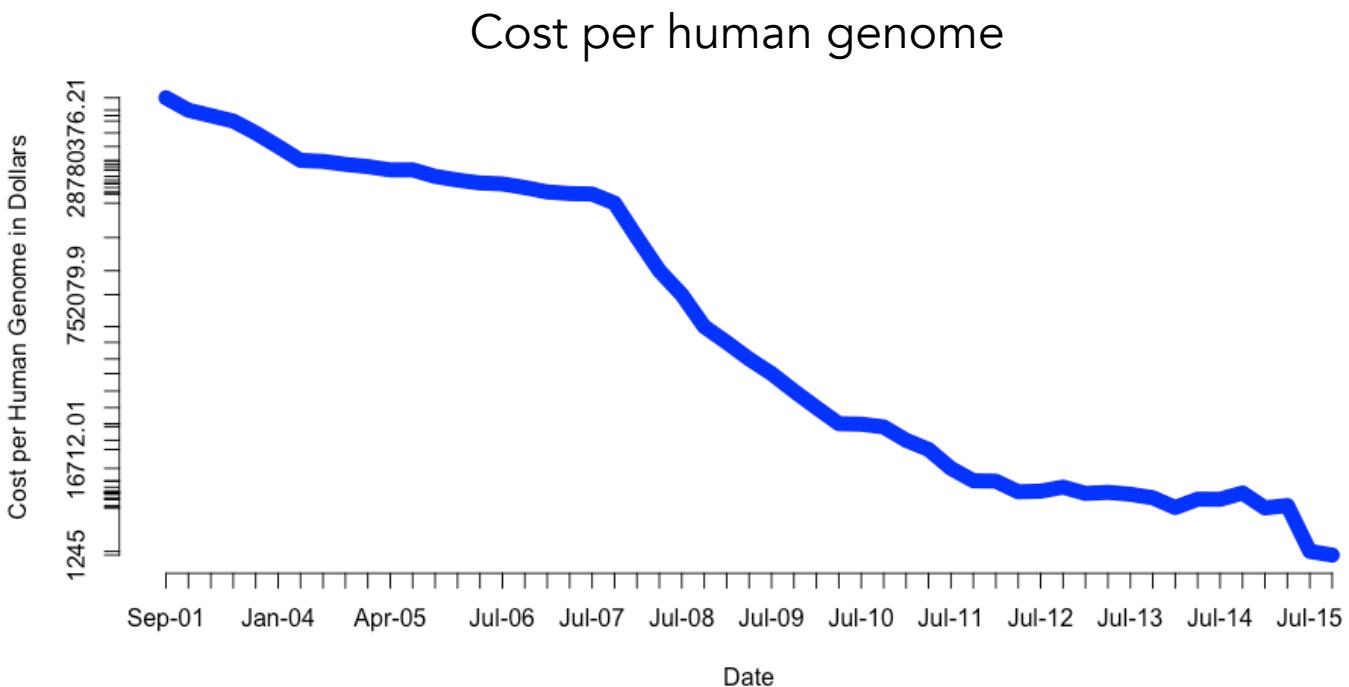
- What is genomics?

GENOMICS

- What is bioinformatics?

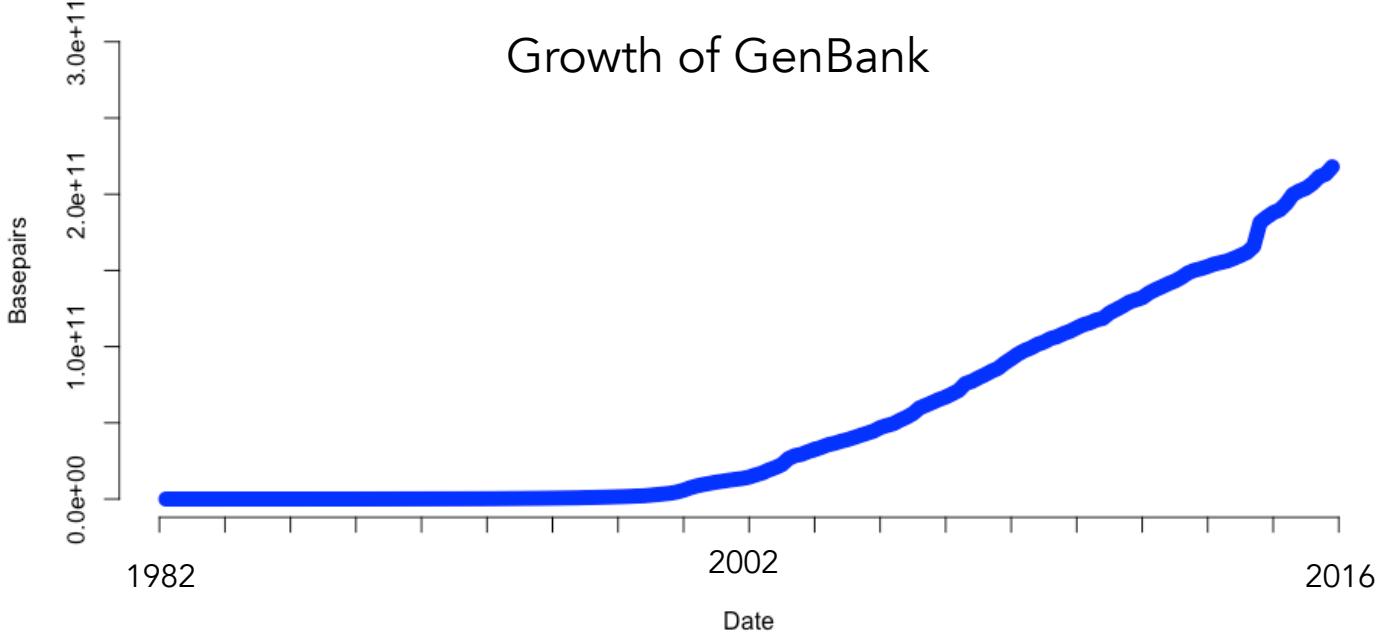
BIOINFORMATICS

- Why now??



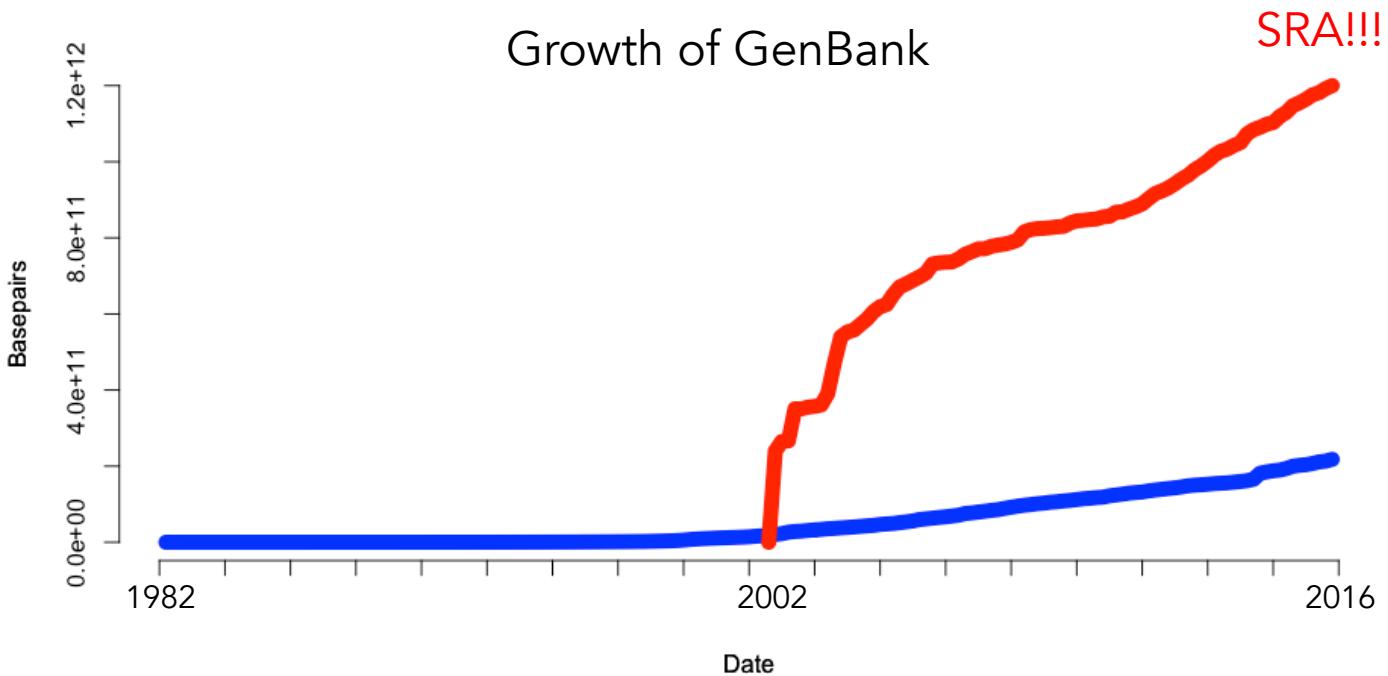
BIOINFORMATICS

- Why now??



BIOINFORMATICS

- Why now??



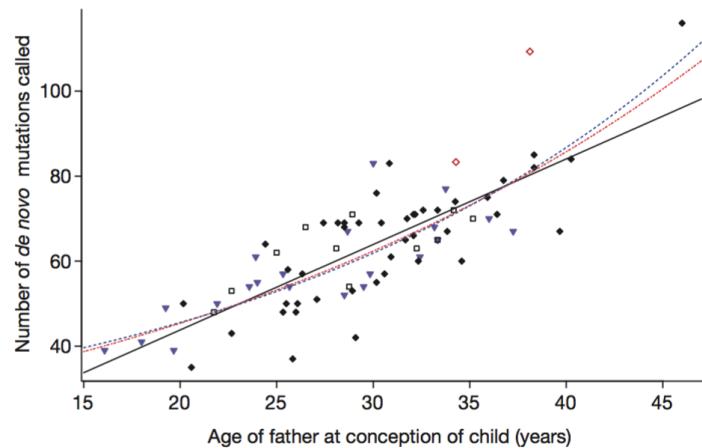
BIOINFORMATICS

- Why now??



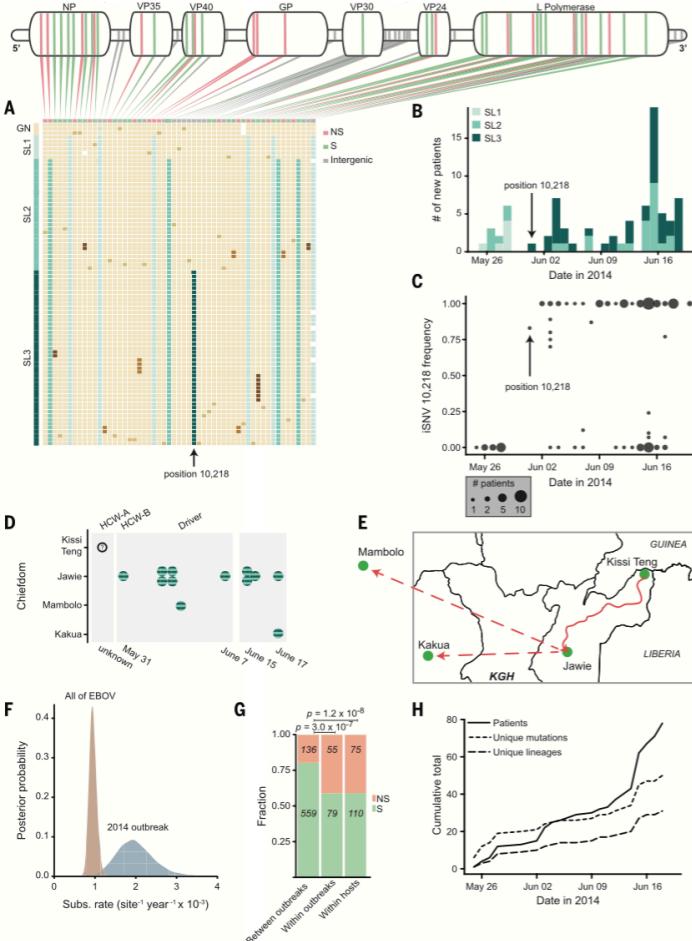
Rate of *de novo* mutations and the importance of father's age to disease risk

Augustine Kong¹, Michael L. Frigge¹, Gisli Masson¹, Soren Besenbacher^{1,2}, Patrick Sulem¹, Gisli Magnusson¹, Sigurjon A. Gudjonsson¹, Asgeir Sigurdsson¹, Aslaug Jonasdottir¹, Adalbjorg Jonasdottir¹, Wendy S. W. Wong³, Gunnar Sigurdsson¹, G. Bragi Walters¹, Stacy Steinberg¹, Hannes Helgason¹, Gudmar Thorleifsson¹, Daniel F. Gudbjartsson¹, Agnar Helgason^{1,4}, Olafur Th. Magnusson¹, Unnur Thorsteinsdottir^{1,5} & Kari Stefansson^{1,5}



Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak

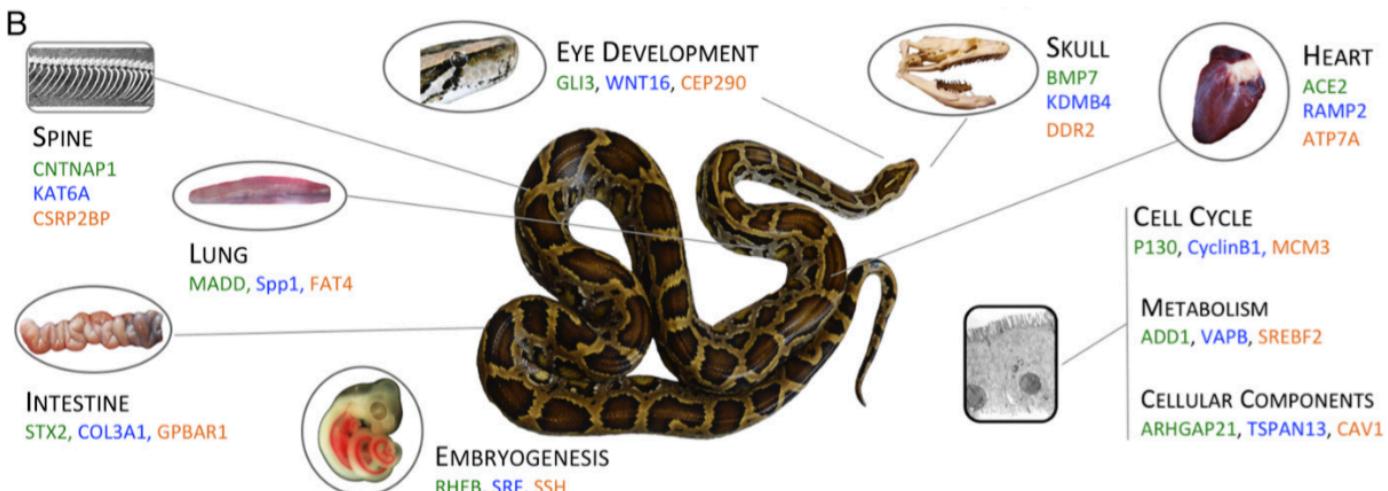
Stephen K. Gire,^{1,2*} Augustine Goba,^{3*}† Kristian G. Andersen,^{1,2,4†} Rachel S. G. Sealton,^{2,4§} Daniel J. Park,^{2§} Lansana Kamneh,³ Simbirie Jalloh,³ Mambu Momoh,^{3,5} Mohamed Fullah,^{3,5†} Gyts Dudas,⁶ Shirlee Wohr,^{1,2,7} Lina M. Moses,⁸ Nathan L. Yozwiak,^{1,2} Sarah Winnicki,^{1,2} Christian B. Matranga,² Christine M. Malboeuf,² James Qu,² Adrienne D. Gladden,² Stephen F. Schaffner,^{1,2} Xiao Yang,² Pan-Pan Jiang,^{1,2} Mahan Nekouli,^{1,2} Andres Colubri,¹ Molanya Ruth Coomber,⁷ Mbula Fomnie,^{9,‡} Alex Moigboi,^{7,‡} Michael Gbakpe,³ Fatima K. Kamara,³ Veronica Tucker,³ Edwin Konuwa,³ Sidiki Saffa,^{3,‡} Josephine Sellu,³ Abdul Azziz Jalloh,³ Alice Kovoma,^{3,‡} James Koninga,³ Ibrahim Mustapha,³ Kandeh Kargbo,³ Momoh Foday,³ Mohamed Yillah,³ Franklyn Kamneh,³ Willie Robert,³ James L. B. Massally,² Sinéad B. Chapman,⁷ James Bochicchio,² Cheryl Murphy,² Chad Nusbaum,² Sarah Young,² Bruce W. Birren,² Donald S. Grant,³ John S. Scheiffelin,⁸ Eric S. Lander,^{2,7,9} Christian Happi,¹⁰ Sahr M. Gevao,¹¹ Andreas Gnirke,^{2,§} Andrew Rambaut,^{6,12,13§} Robert F. Garry,^{8§} S. Humarr Khan,^{3,‡§} Pardis C. Sabeti^{1,2,†§}



The Burmese python genome reveals the molecular basis for extreme adaptation in snakes

Todd A. Castoe^{a,b}, A. P. Jason de Koning^{a,c}, Kathryn T. Hall^a, Daren C. Card^b, Drew R. Schield^b, Matthew K. Fujita^b, Robert P. Ruggiero^a, Jack F. Degner^d, Juan M. Daza^e, Wanjun Gu^f, Jacobo Reyes-Velasco^b, Kyle J. Shaney^b, Jill M. Castoe^{a,b}, Samuel E. Fox^g, Alex W. Poole^a, Daniel Polanco^a, Jason Dobry^h, Michael W. Vandewegeⁱ, Qing Li^j, Ryan K. Schott^k, Aurélie Kapusta^l, Patrick Minx^l, Cédric Feschotte^l, Peter Uetz^m, David A. Ray^{i,n}, Federico G. Hoffmann^{i,n}, Robert Bogden^h, Eric N. Smith^b, Belinda S. W. Chang^k, Freek J. Vonk^{o,p,q}, Nicholas R. Casewell^{q,r}, Christiaan V. Henkel^{p,s}, Michael K. Richardson^p, Stephen P. Mackessy^t, Anne M. Bronikowski^u, Mark Yandell^j, Wesley C. Warren^l, Stephen M. Secor^v, and David D. Pollock^{a,1}

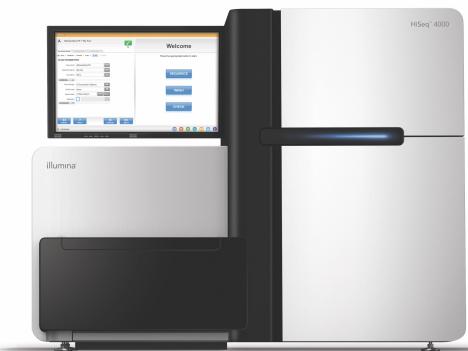
B



- Discover other stuff, too
 - Genes
 - Expression levels
 - Repetitive regions
 - Community composition
 - Methylation
 - ...

SEQUENCING PLATFORMS

- Illumina (@illumina)



SEQUENCING PLATFORMS

- Pacific Biosystems (@PacBio)



SEQUENCING PLATFORMS

- Oxford Nanopore (@nanopore)



<https://www.youtube.com/watch?v=CE4dW64x3Ts>

SEQUENCING PLATFORMS

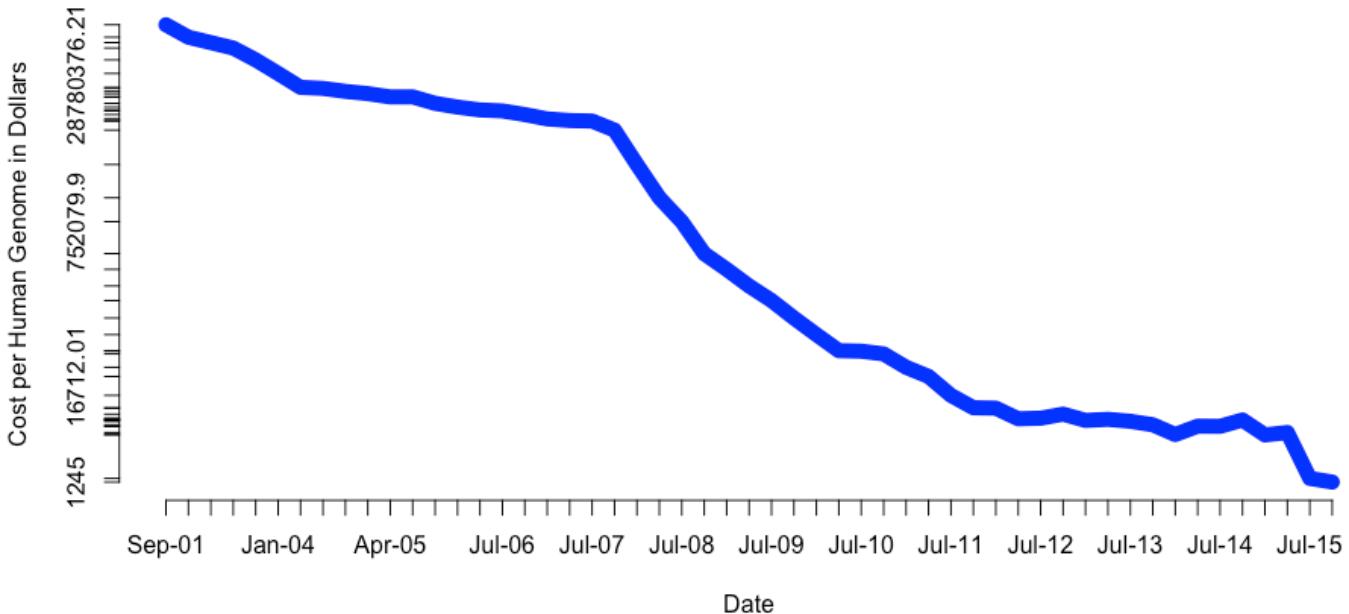
- 454 (@LifeTech)

SEQUENCING PLATFORMS

- Sanger

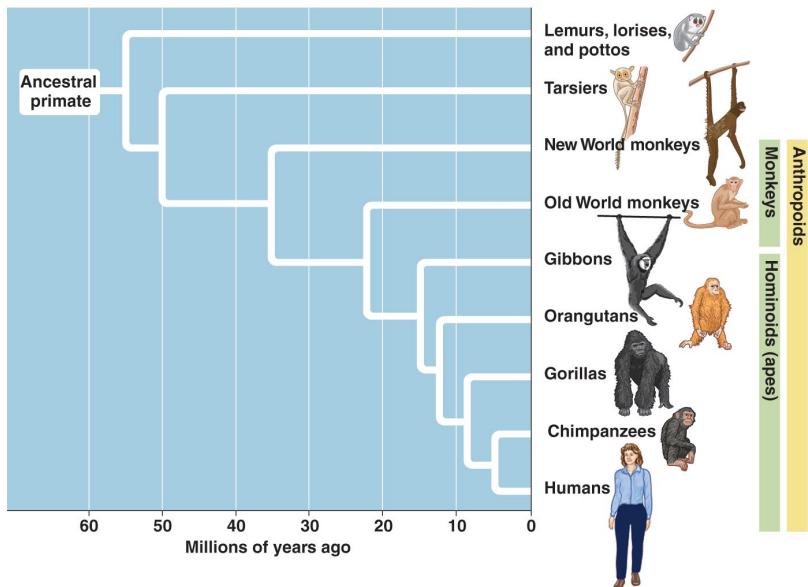
SEQUENCING PLATFORMS

Cost per human genome



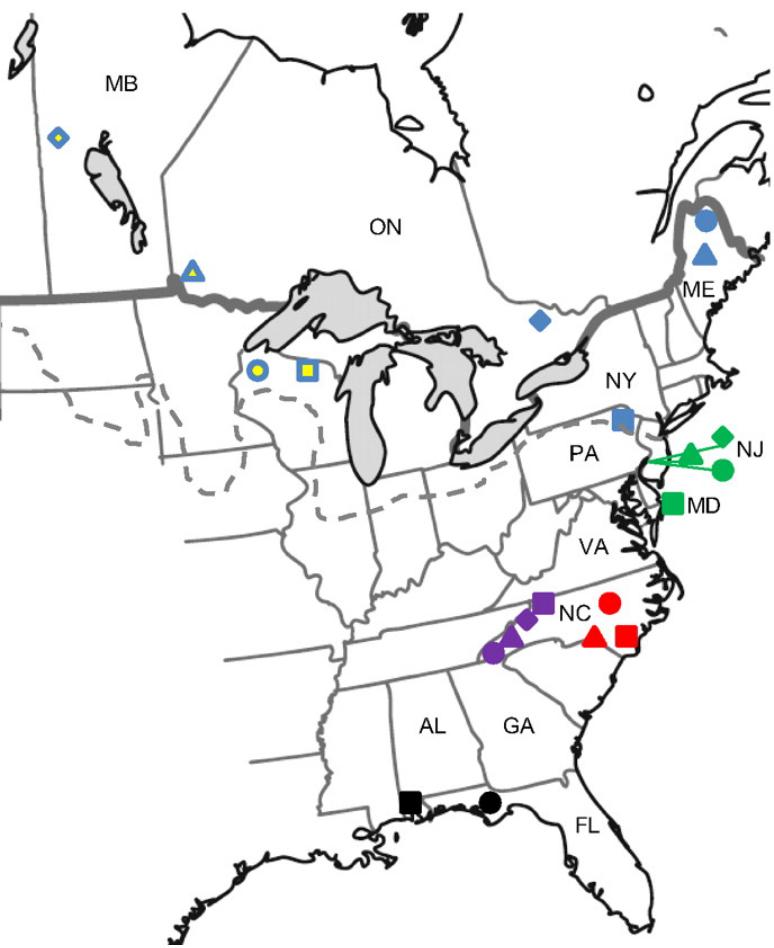
CONTEXT IS EVOLUTION

- Evolution

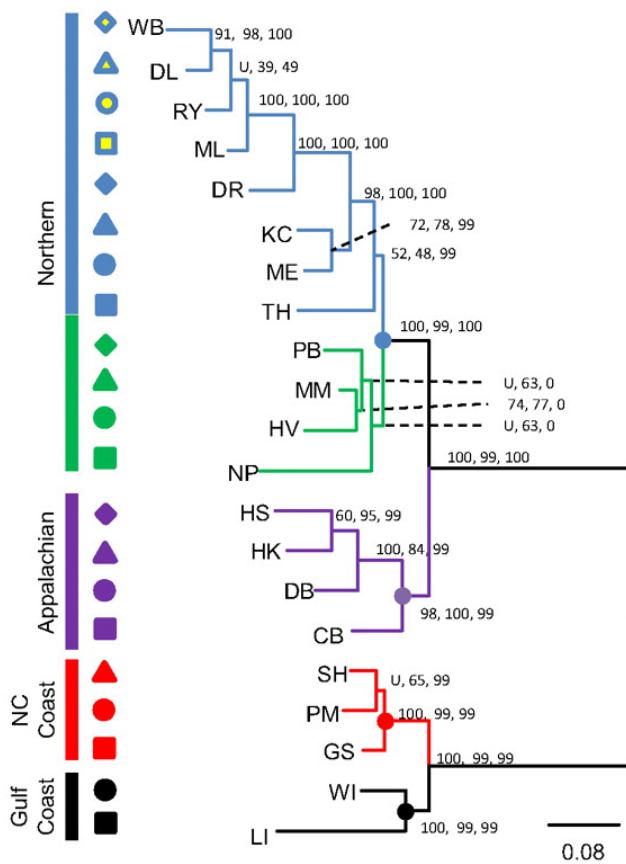


Copyright © 2009 Pearson Education, Inc.

A



B



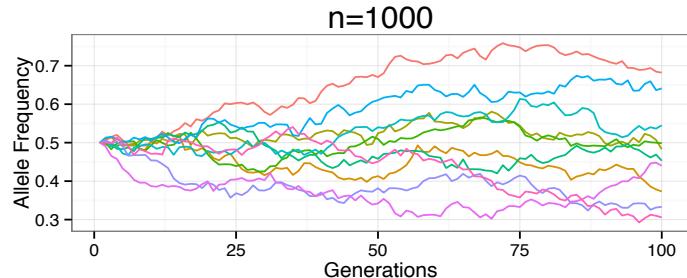
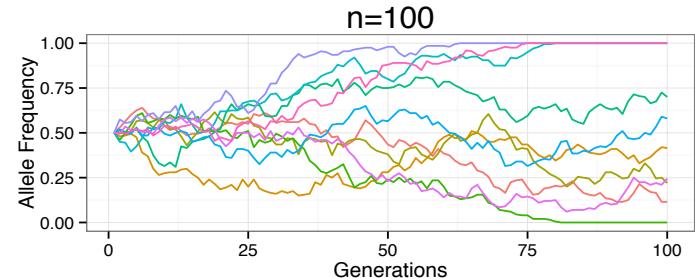
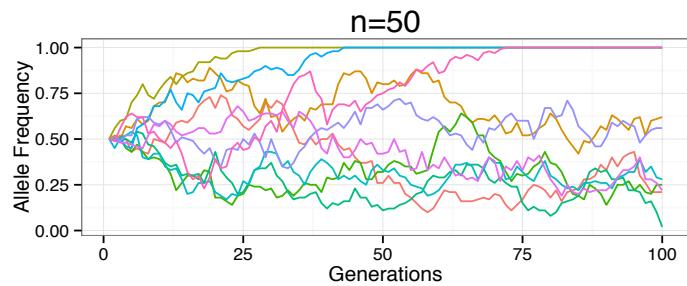
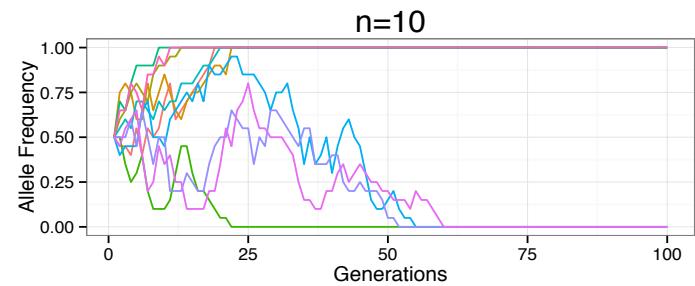
MECHANISMS OF EVOLUTION

- Gene Flow

MECHANISMS OF EVOLUTION

- Genetic Drift

MECHANISMS OF EVOLUTION



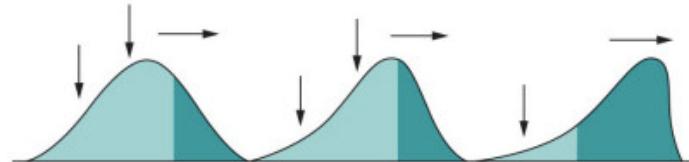
MECHANISMS OF EVOLUTION

- Mutation

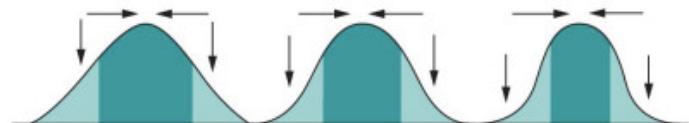
MECHANISMS OF EVOLUTION

- Natural Selection

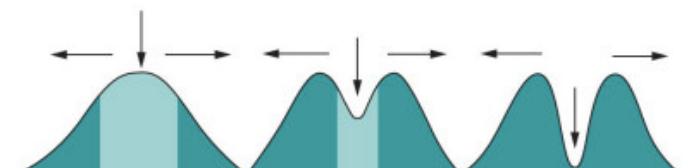
SELECTION



(a) Directional selection



(b) Stabilizing selection



(c) Disruptive selection

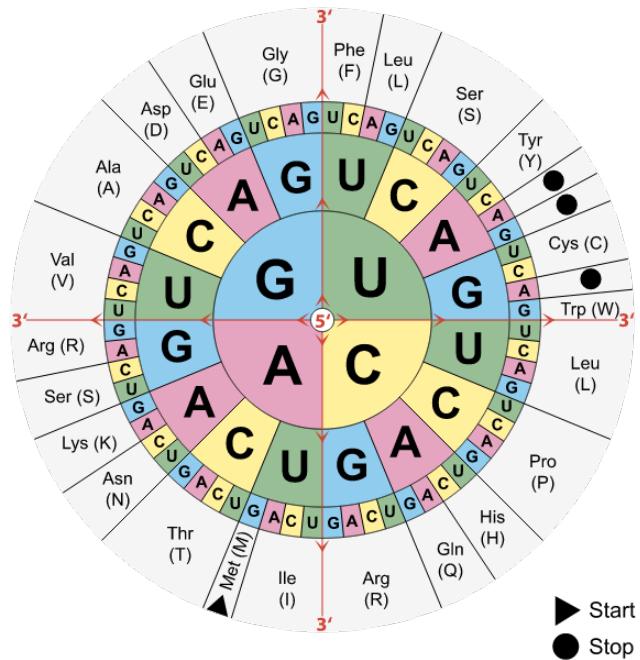
Copyright © 2009 Pearson Education, Inc.

SELECTION

- Synonymous change
- Non-synonymous

SELECTION

5'-AUGCAGGCAUGA-3'



SELECTION

- dN/dS

Finding Data & Pairwise Alignment

Lecture 2
Sept 9, 2014

FINDING DATA

- Genomics has embraced Open Science & Open Data

Genbank	ENA
Uniprot	dbSNP
Refseq	ClinVar
Swiss-prot	Rdp
Pfam	OMIM
Ensembl	UCSC GB
SRA	

FINDING CODE

- Genomics has embraced Open Science & Open Data

Github

[Biostars](#)

[Stackoverflow](#)

PAIRWISE ALIGNMENT

- Definition

PAIRWISE ALIGNMENT

- Homologous

- Paralogous

- Orthologus

PAIRWISE ALIGNMENT

- Alignment demo:
 - [mafft](#)

PAIRWISE ALIGNMENT

- What is aligner actually doing?

PAIRWISE ALIGNMENT

- Scoring Matrices