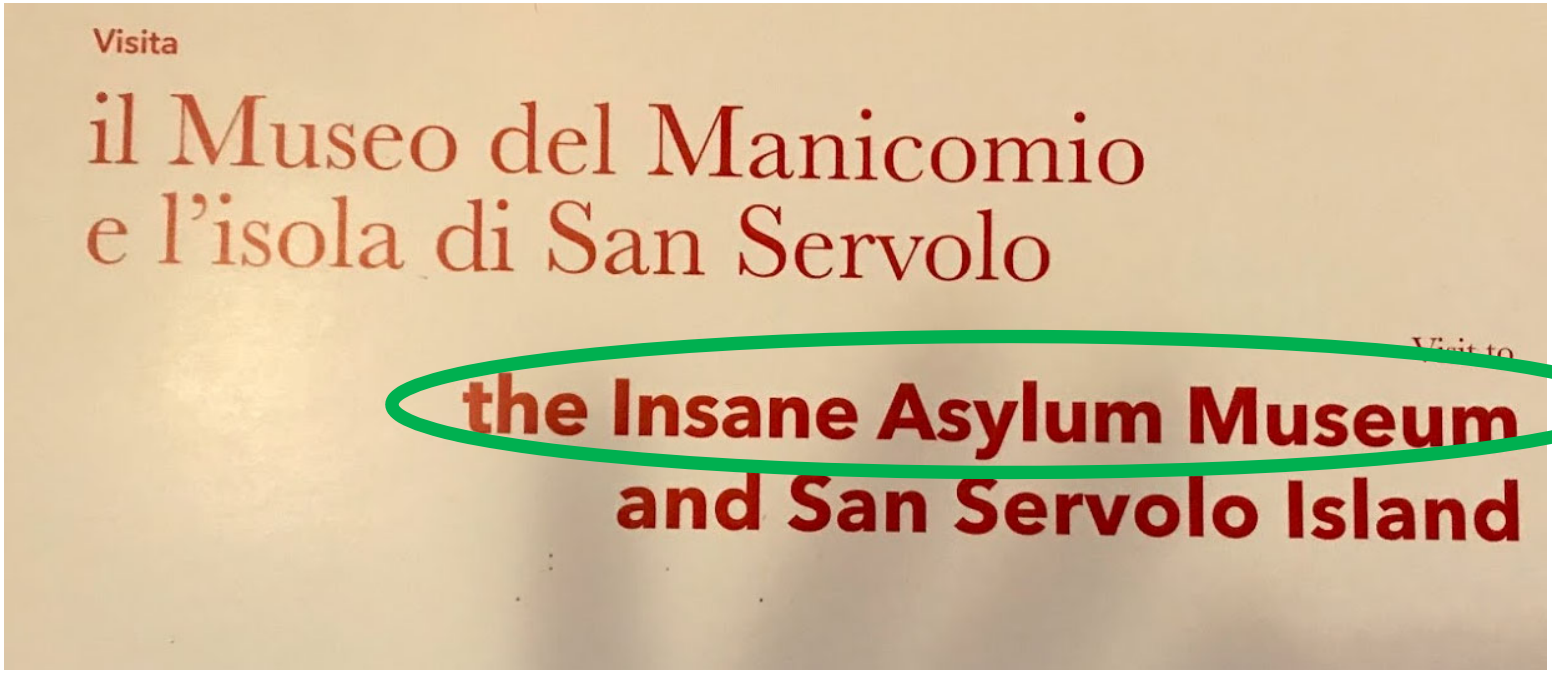
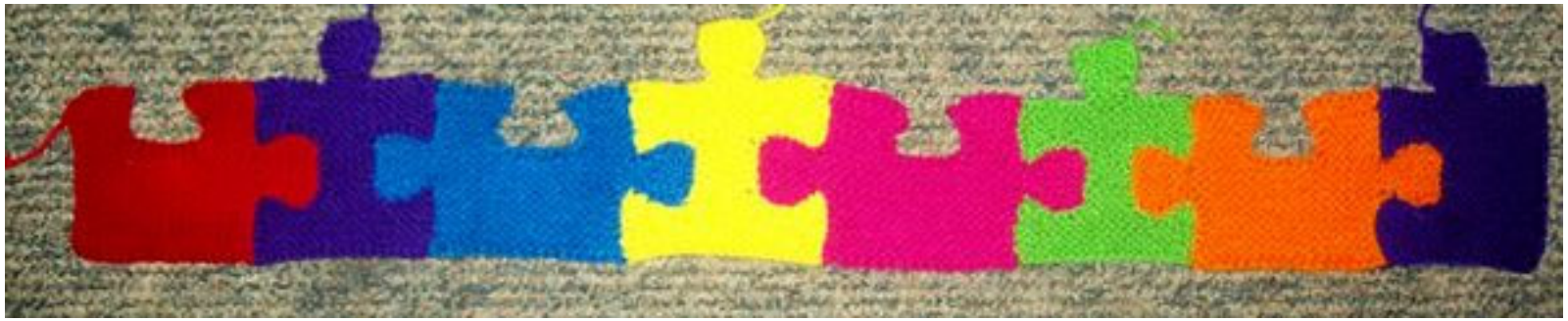

Visita
il Museo del Manicomio
e l'isola di San Servolo
Visit to
the Insane Asylum Museum
and San Servolo Island

Short Reads assemble into Contigs

Figure 5.1.



Genome Assembly

Whole-genome "shotgun" sequencing starts by copying and fragmenting the DNA

("Shotgun" refers to the random fragmentation of the whole genome; like it was fired from a shotgun)

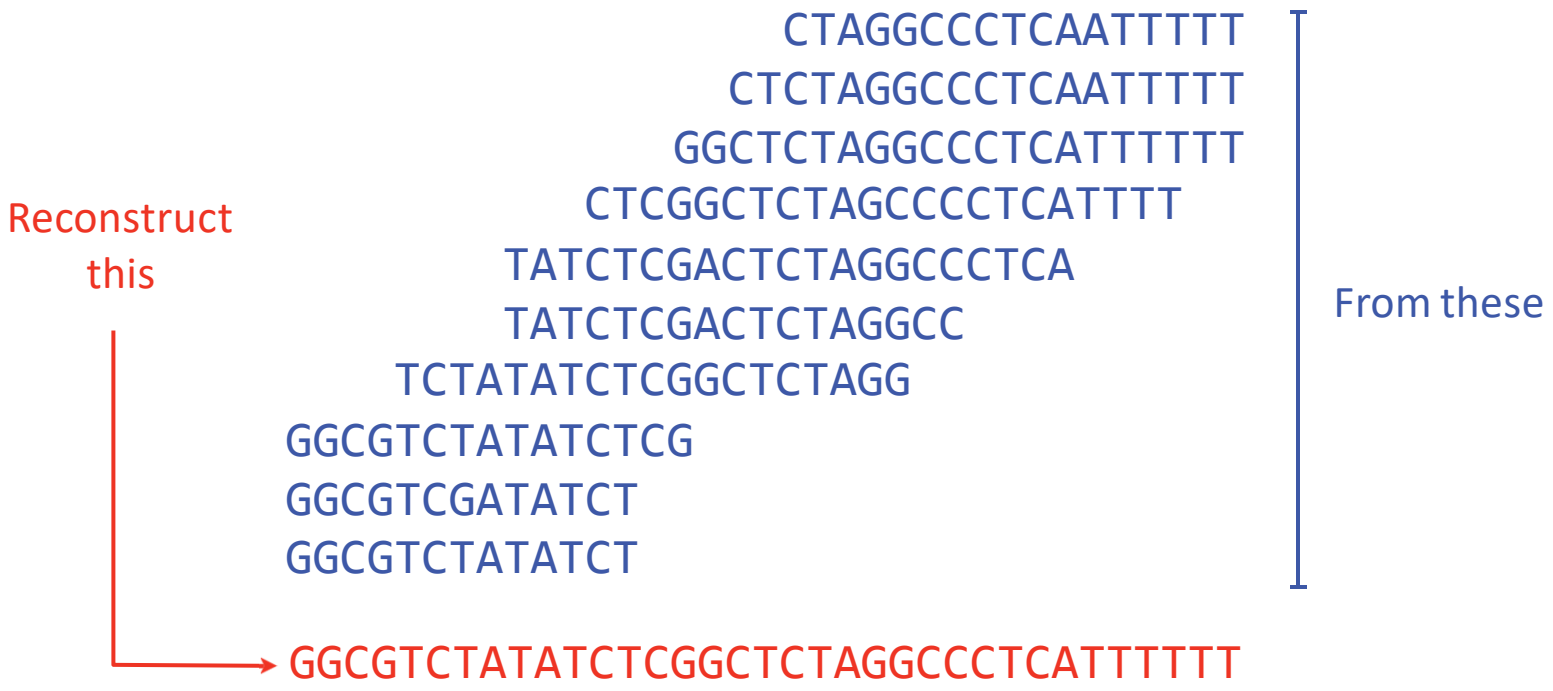
Input: GGCGTCTATATCTCGGCTCTAGGCCCTCATTTTTT

Copy:
GGCGTCTATATCTCGGCTCTAGGCCCTCATTTTTT
GGCGTCTATATCTCGGCTCTAGGCCCTCATTTTTT
GGCGTCTATATCTCGGCTCTAGGCCCTCATTTTTT
GGCGTCTATATCTCGGCTCTAGGCCCTCATTTTTT

Fragment:
GGCGTCTA TATCTCGG CTCTAGGCCCTC ATTTTTT
GGC GTCTATAT CTCGGCTCTAGGCCCTCA TTTTTT
GGCGTC TATATCT CGGCTCTAGGCCCT CATTTTTT
GGCGTCTAT ATCTCGGCTCTAG GCCCTCA TTTTTT

Assembly

Assume sequencing produces such a large # fragments that almost all genome positions are *covered* by many fragments...

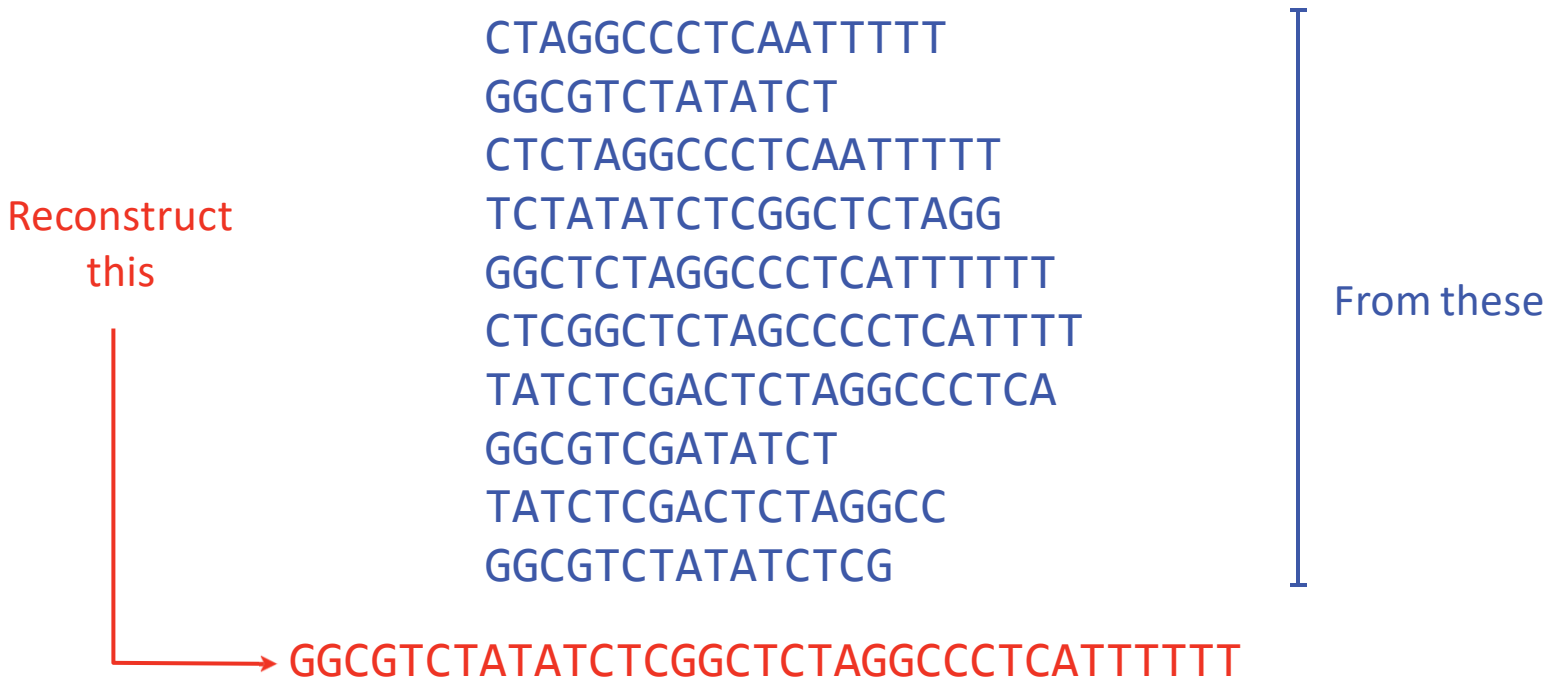


Courtesy of Ben Langmead. Used with permission.

http://www.langmead-lab.org/teaching-materials/

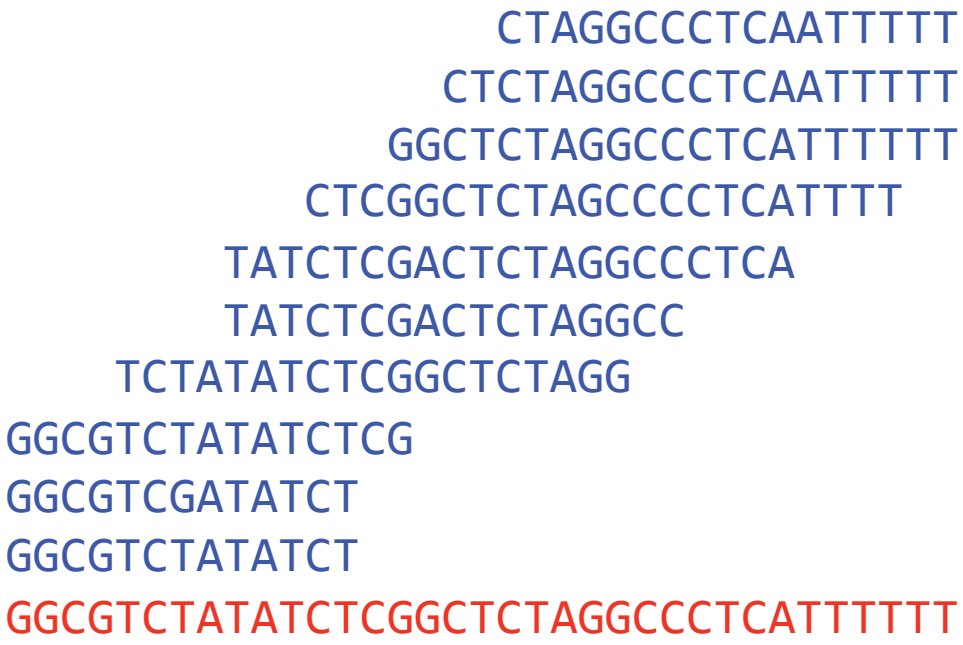
Assembly

…but we don't know what came from where



Assembly

Key term: *coverage*. Usually it's short for *average coverage*: the average number of reads covering a position in the genome.



177 nucleotides

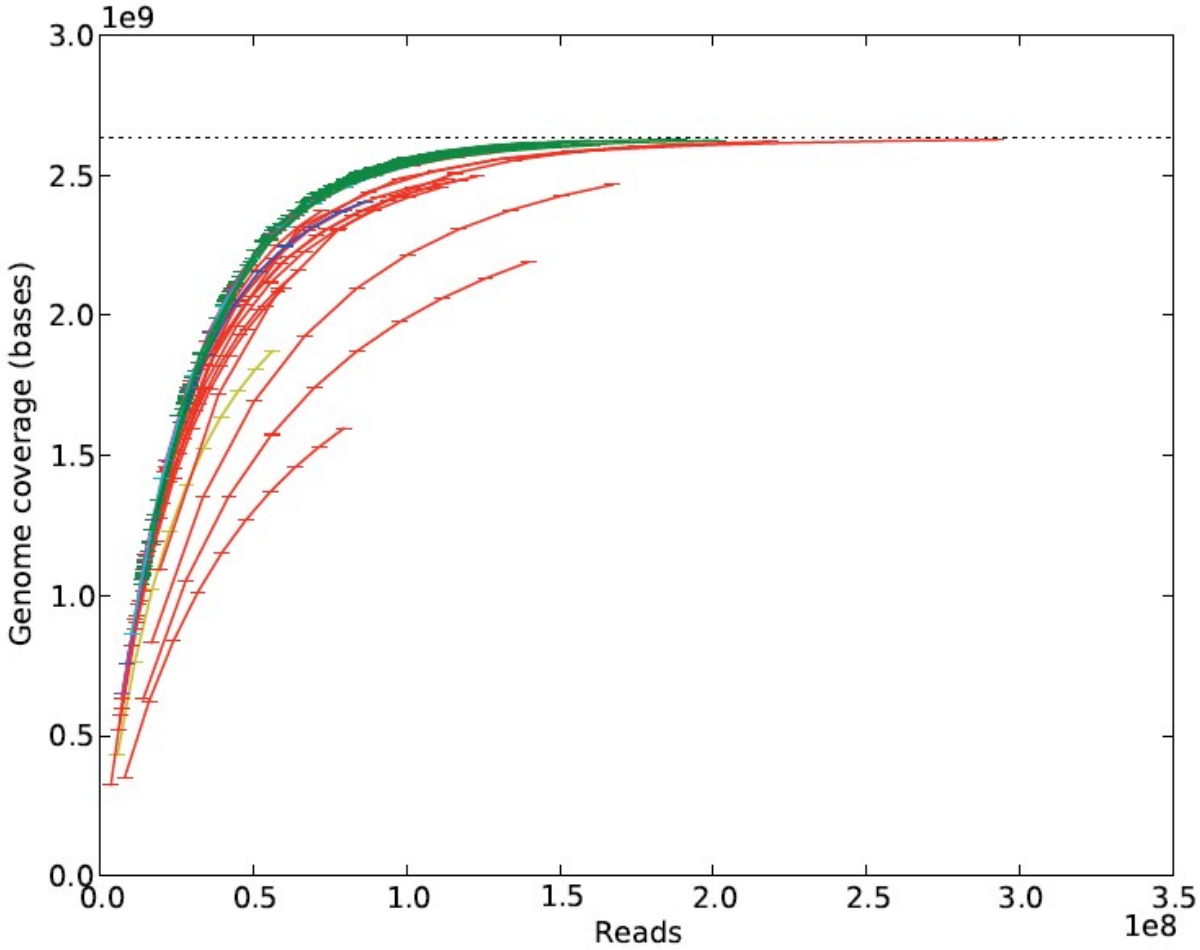
35 nucleotides

Average coverage = 177 / 35 ≈ 7x (7-fold coverage)

Estimating # of Uncovered Bases and # of Contigs

- G - genome size
- N - # of reads
- L - Length of read
- NL/G = reads/base = λ (coverage)
  - Poisson(O, λ) = $e^{-\lambda} =\sim$ probability a base is not covered
  - # of uncovered bases =~ $Ge^{-\lambda}$
  - # of contigs = #gaps= ~ $Ne^{-\lambda}$

Reads vs. coverage for 1000 Genomes Datasets



Adapted from Christopher Burge, David Gifford, and Ernest Fraenkel. *7.91J Foundations of Computational and Systems Biology.* Spring 2014. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.

How to assemble a genome?

Here we assume a “de novo” assembly without help from the previously assembled genomes

Who was de Bruijn?

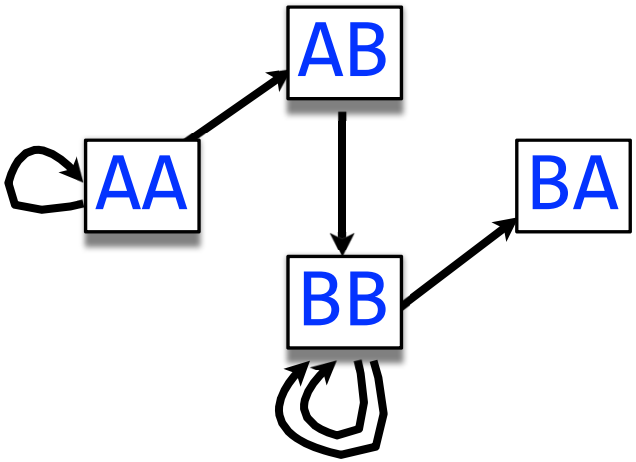
Nicolaas Govert de Bruijn (1918 – 2012) was a Dutch mathematician, noted for his many contributions in the fields of graph theory, analysis, number theory, combinatorics and logic

De Bruijn graph

genome: AAABBBBA

3-mers: AAA, AAB, ABB, BBB, BBB, BBA

L/R 2-mers: AA, AA AA, AB AB, BB BB, BB BB, BB BB, BA



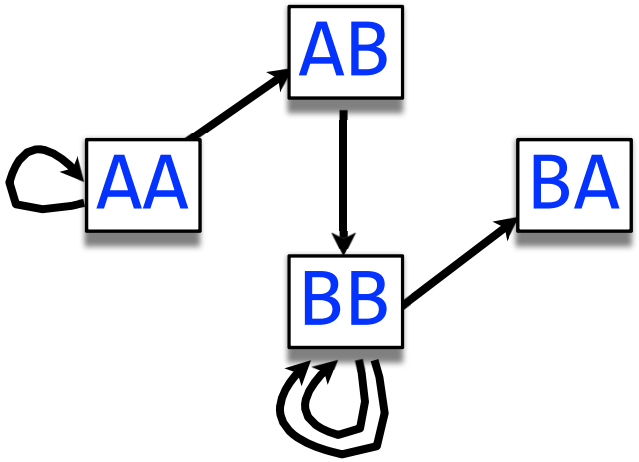
One edge per every *k*-mer

One node per distinct *k*-1-mer

Courtesy of Ben Langmead. Used with permission.

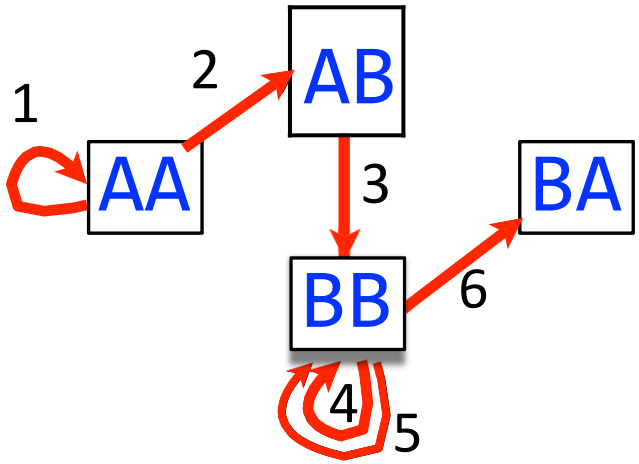
http://www.langmead-lab.org/teaching-materials/

De Bruijn graph



Walk crossing each edge exactly once gives a reconstruction of the genome

De Bruijn graph



AAABBBBA

Walk crossing each edge exactly once gives a reconstruction of the genome. This is an *Eulerian walk*.

Directed multigraphs

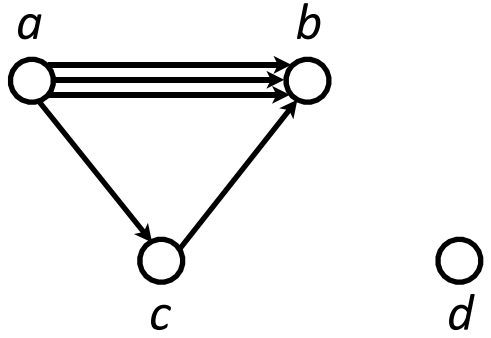
Directed **multigraph** $G(V, E)$ consists of set of *vertices, V* and **multiset** of *directed edges, E*

Otherwise, like a directed graph

Node's *indegree* = # incoming edges

Node's *outdegree* = # outgoing edges

De Bruijn graph is a directed multigraph



$V = \{ a, b, c, d \}$

$E = \{ (a, b), (a, b), (a, b), (a, c), (c, b) \}$

Repeated

Eulerian walk definitions and statements

Node is *balanced* if indegree equals outdegree

Node is *semi-balanced* if indegree differs from outdegree by 1

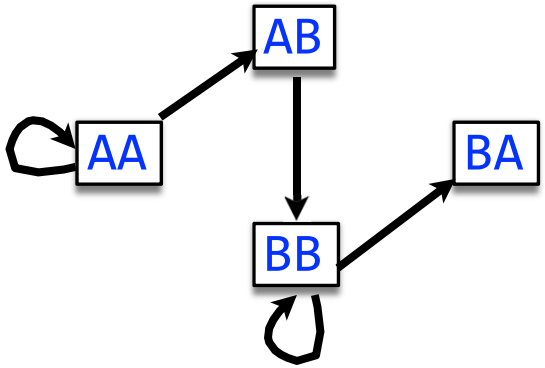
Graph is *connected* if each node can be reached by some other node

Eulerian walk visits each edge exactly once

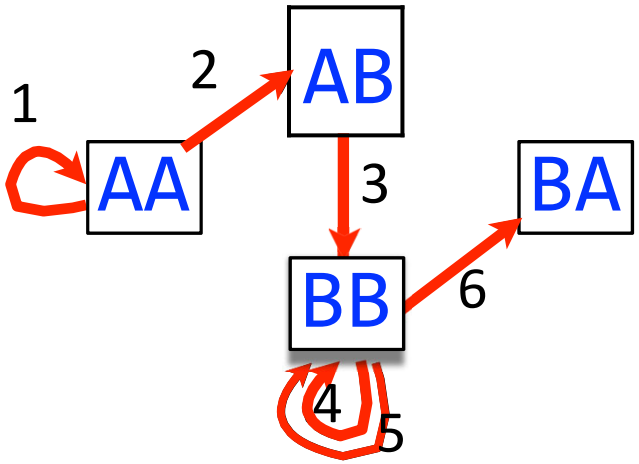
Not all graphs have Eulerian walks. Graphs that do are *Eulerian.*
(For simplicity, we won't distinguish Eulerian from semi-Eulerian.)

A directed, connected graph is Eulerian if and only if it has at most 2 semi-balanced nodes (one with outdegree-indegee=+1 and another with outdegree-indegree=-1 all other nodes are balanced

Jones and Pevzner section 8.8



De Bruijn graph



Assembled sequence

AAABBBBA

Is it Eulerian? Yes

Argument 1: AA → AA → AB → BB → BB → BB → BA

Argument 2: AA and BA are semi-balanced, AB and BB are balanced

Human Genome Project

Drew Sheneman, New Jersey -- The Newark Star Ledger, E-mail Drew.

Carl! I think I found the corner piece

Carl Hierholzer's (1840 –1871) algorithm

to find a Eulerian walk on a balanced graph

- If graph is not balanced – make it balanced by adding a link from end to start node
- Choose any starting vertex v, and follow a trail of edges from that vertex until returning to v.
- It is not possible to get stuck at any vertex other than v, because nodes are balanced. Hence, when the trail enters another vertex w, there must be an unused edge leaving w.
- The tour formed in this way is a closed tour, but may not cover all the vertices and edges of the initial graph.
- As long as there exists a vertex v that belongs to the current tour but that has adjacent edges not part of the tour, start another trail from v, following unused edges until returning to v, and join the tour formed in this way to the previous tour.
- Hierholzer's algorithm takes linear time

Edge-disjoint loops are a problem: multiple Eulerian paths

graph can have multiple Eulerian walks, only one of which corresponds to original superstring

Right: graph for ZABCDABEFABY, k=3

Alternative Eulerian walks:

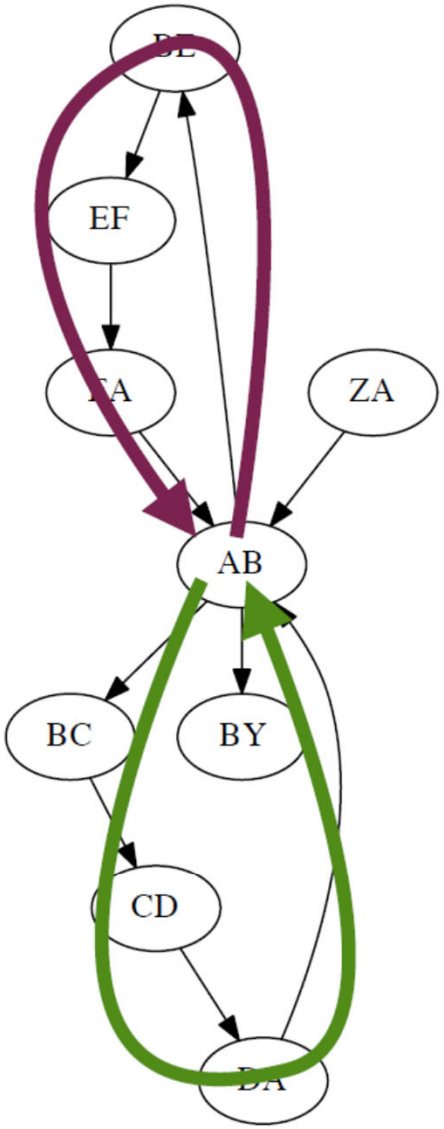
ZA → AB → BE → EF → FA → AB → BC → CD → DA → AB → BY

ZA → AB → BC → CD → DA → AB → BE → EF → FA → AB → BY

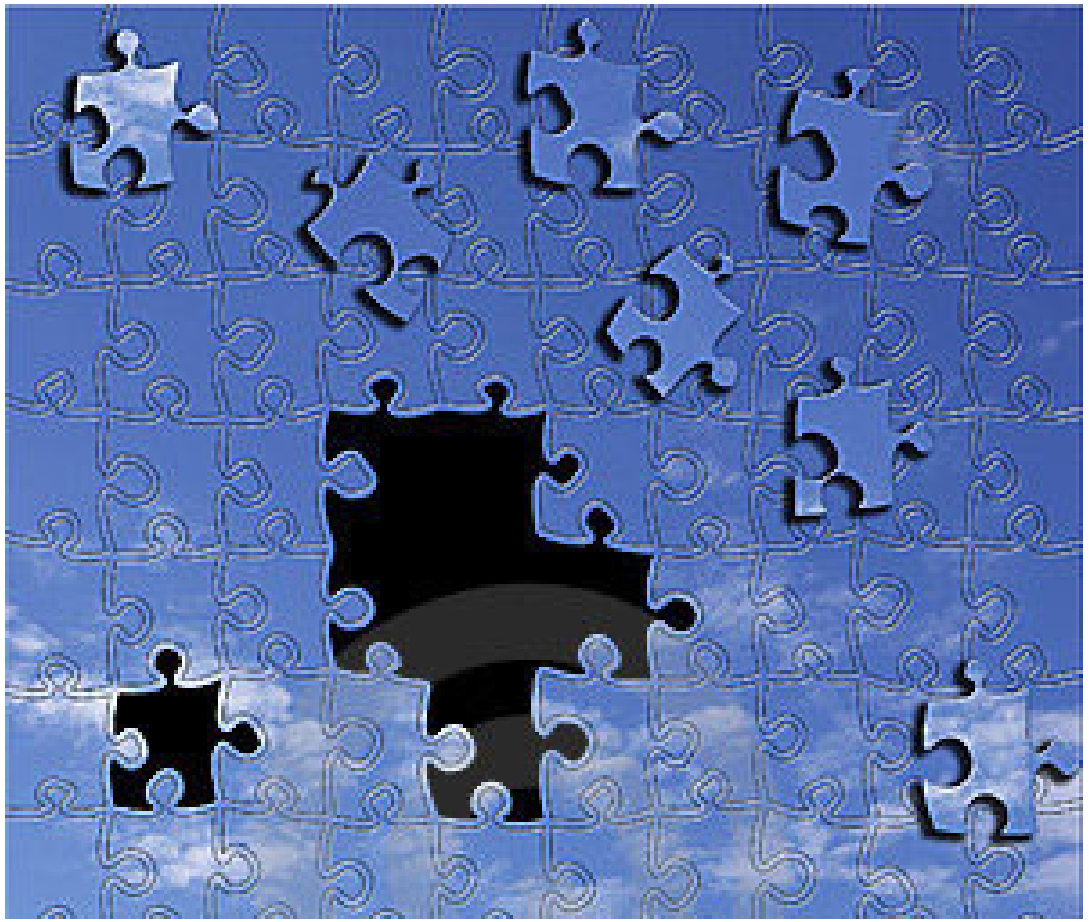
These correspond to two edge-disjoint directed cycles joined by node AB

AB is a repeat: ZABCDABEFABY

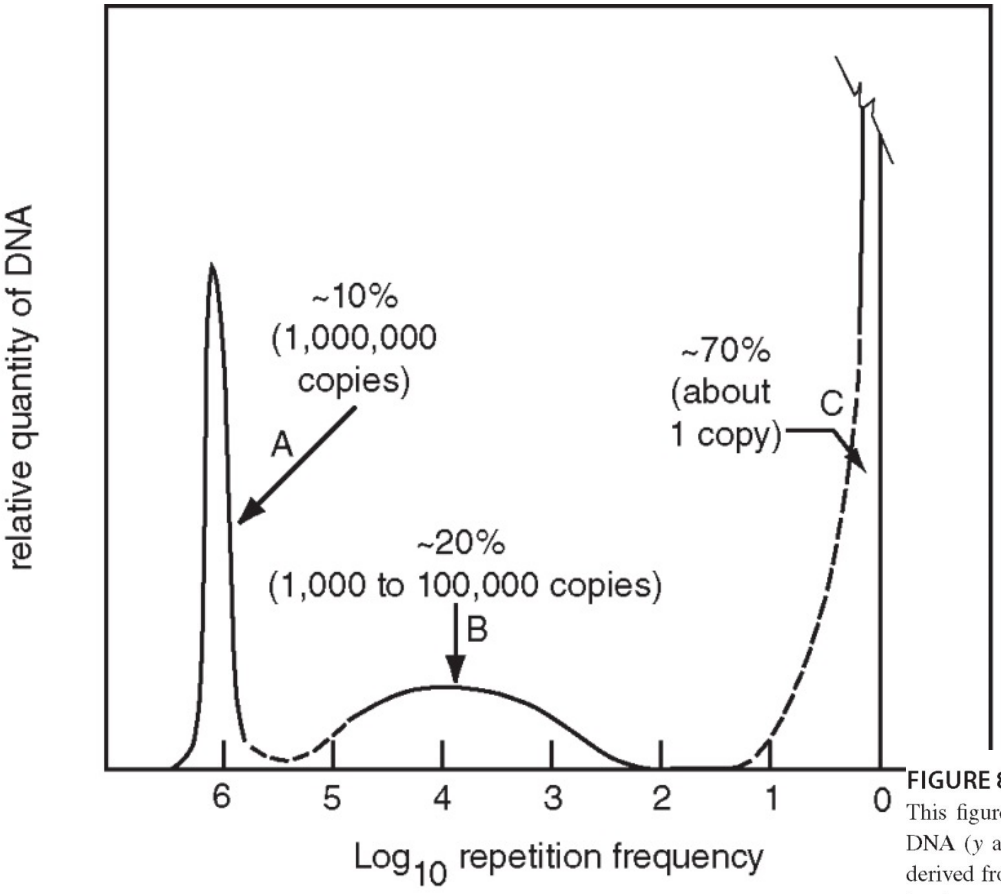
k=3 is to short to resolve. But k=4 would be enough



Repeats are like sky puzzle pieces



How many repeats are in eukaryotic genomes?



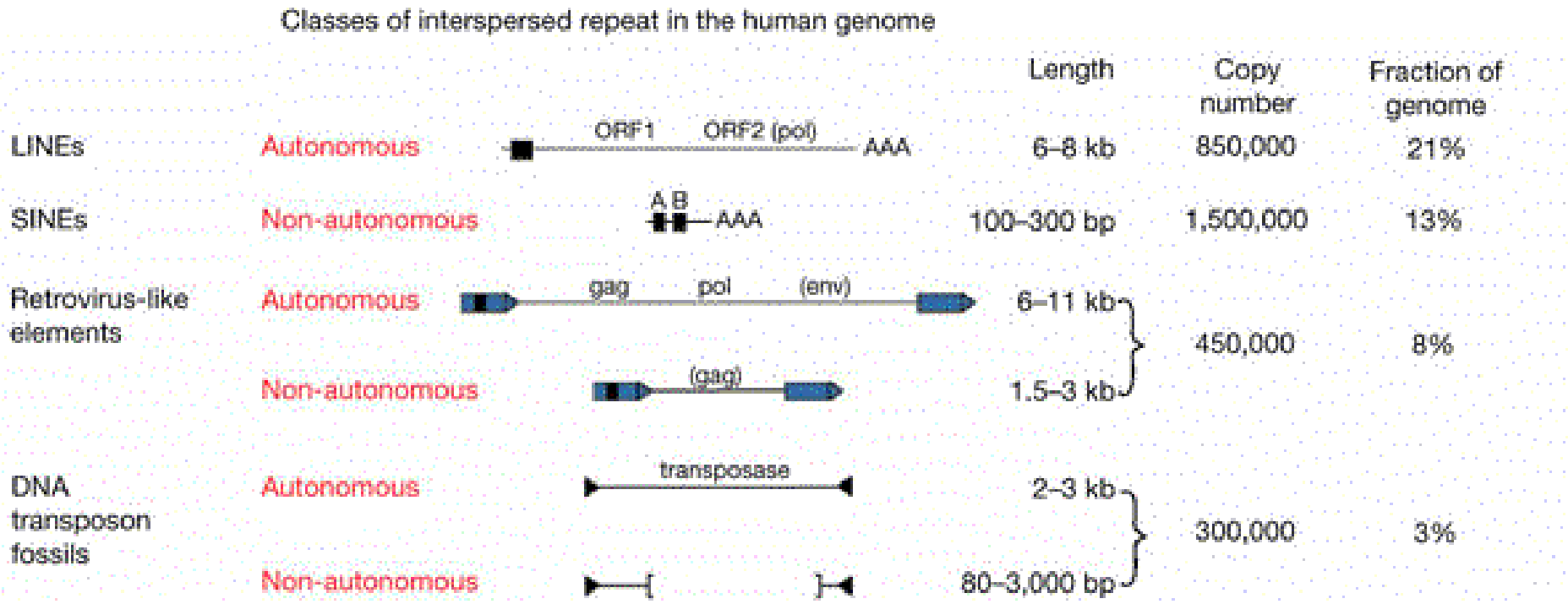
Data for mouse genome obtained in 1961 (sic!) using DNA denaturation and renaturation curves

FIGURE 8.6 The complexity of genomic DNA can be estimated by denaturing then renaturing DNA. This figure (redrawn from Britten and Kohne, 1968) depicts the relative quantity of mouse genomic DNA (y axis) versus the logarithm of the frequency with which the DNA is repeated. The data are derived from a $C_0 t_{1/2}$ curve, which describes the percent of genomic DNA that reassociates at particular times and DNA concentrations. A large $C_0 t_{1/2}$ value implies a slower reassociation reaction. Three classes are apparent. The fast component accounts for 10% of mouse genomic DNA (arrow A), and represents highly repetitive satellite DNA. An intermediate component accounts for about 20% of mouse genomic DNA and contains repeats having from 1000 to 100,000 copies. The slowly reassociating component, comprising 70% of the mouse genome, corresponds to unique, single-copy DNA. Britten and Kohne (1968) obtained similar profiles from other eukaryotes, although distinct differences were evident between species. Used with permission.

I showed my masterpiece to the grown-ups and asked them if my drawing frightened them.

The Little Prince, Antoine de Saint-Exupéry, 1943

Almost all transposable elements in humans fall into one of four classes



The most abundant SINE lineage, Alu, has about 50,000 active copies.
It is non-autonomous and depends on LINE with which it co-evolved

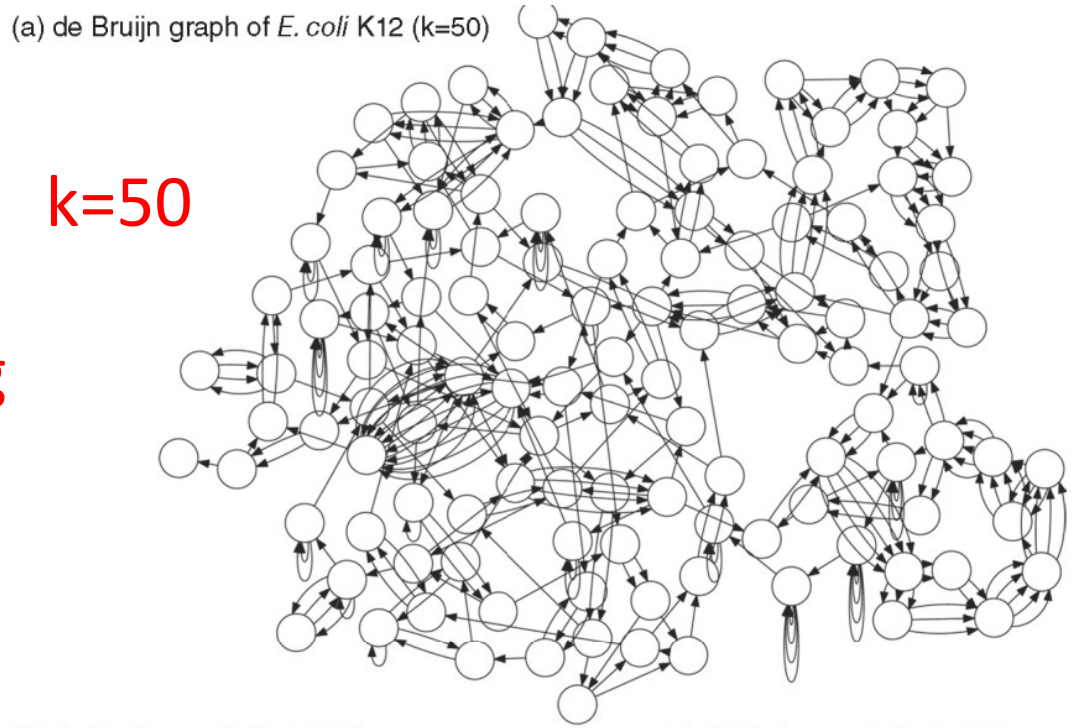
The only active LINE in human, LINE-1, has about 100 active copies per genome (the number varies between people). Sleep hormone melatonin has been shown to reduce LINE-1-induced genome instability.

How to assemble a genome with repeats?

- Answer:
  longer reads
- But:
  cheap sequencing
  =
  short reads

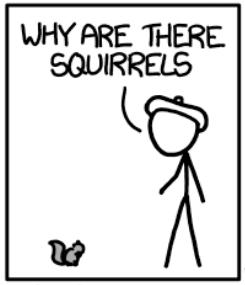
Technology	Read length (bp)
Roche 454	700
Illumina	50–250
SOLiD	50
Ion Torrent	200
Pacific Biosciences	2900
Sanger	400–900

~10 cents per 10^6 bases



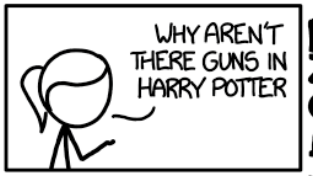
Credit: XKCD comics

QUESTIONS
FOUND IN GOOGLE AUTOCOMPLETE

WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY ARE THERE MALE AND FEMALE BIKES
WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE ARTERIES
WHY ARE OLD KLINGONS DIFFERENT
WHY ARE THERE SQUIRRELS

WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL
WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY
WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD
WHY IS SEX SO IMPORTANT

WHY ARE THERE SLAVES IN THE BIBLE
WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT
WHY AREN'T MY ARMS GROWING

WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY
WHY IS YKK ON ALL ZIPPERS
WHY ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS
WHY IS THERE LAVA
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE
WHY ARE THERE ANTS IN MY LAPTOP
WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN
WHY ARE THERE GHOSTS

WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47S SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS
WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA
WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE
WHY AREN'T THERE GUNS IN HARRY POTTER

WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG
WHY ARE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND
WHY IS LIFE SO BORING
WHY ARE THERE FEMALE MR MIMES
WHY IS GPS FREE
WHY IS THERE HELL IF GOD FORGIVES
WHY DO IGUANAS DIE
WHY AREN'T THERE DINOSAUR GHOSTS

A gallery of useful discrete probability distributions

Geometric Distribution

- A series of Bernoulli trials with probability of success =*p*. continued **until the first success**. X is the number of trials.
- Compare to: Binomial distribution has:
  - Fixed number of trials =n.
  - Random number of successes = x.

$$P(X=x)=C_x^n p^x (1-p)^{n-x}$$

- Geometric distribution has reversed roles:
  - Random number of trials, *x*
  - Fixed number of successes, in this case 1.
  - Success always comes in the end: so no combinatorial factor C_x^n
  - $P(X=x) = p(1-p)^{x-1}$ where:

    x-1 = 0, 1, 2, … , the number of failures until the 1st success.
- NOTE OF CAUTION: Matlab, Mathematica, and many other sources use x to denote the number of failures until the first success. We stick with Montgomery-Runger notation

Geometric Mean & Variance

- If *X* is a geometric random variable (according to Montgomery-Bulmer) with parameter *p*,

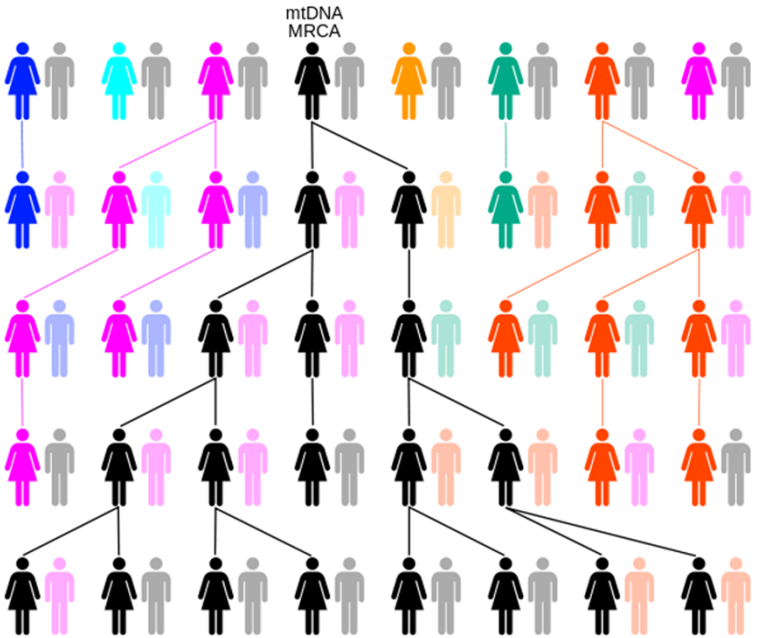
$$\mu = E(X) = \frac{1}{p} \quad \text{and} \quad \sigma^2 = V(X) = \frac{(1-p)}{p^2} \qquad (3\text{-}10)$$

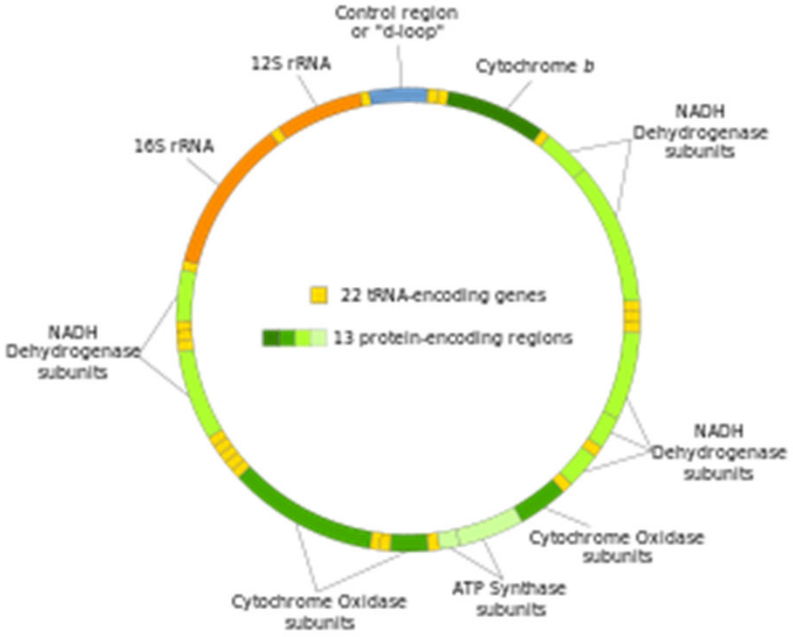
- For small *p* the standard deviation ~= mean

- Very different from Poisson, where it is

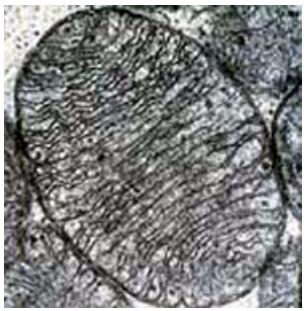
variance = mean and standard deviation = mean$^{1/2}$

Geometric distribution in biology



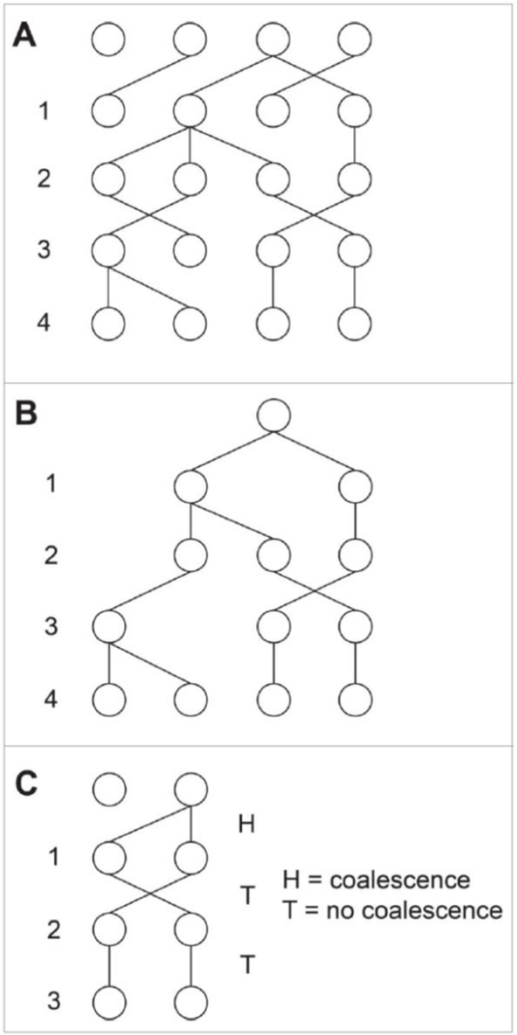


- Each of our cells has mitochondria with 16.5kb of mtDNA **inherited only from our mother**
- Human mtDNA has 37 genes encoding 13 proteins, 22+2 tRNA & rRNA
- Mitochondria appeared 1.5-2 billion years ago as a symbiosis between an alpha-proteobacterium (1000s of genes) and an archaeaon (of UIUC's Carl R. Woese fame)
- Since then most mitochondrial genes were transferred to the nucleus
- Plants also have plastids with genomes related to cyanobacteria



Time to the last common (maternal) ancestor follows geometric distribution

- Constant population of N women
- Random number of (female) offsprings. Average is 1 (but can be 0 or 2)
- Randomly pick two women. Question: how many generations T since their last maternal ancestor?
- T is a random variable What is its PMF: P(T=t)? Answer: P(T=t) follows a geometric distribution
- Do these two women have the same mother? Yes: "success" in finding their last common ancestor (p=1/N). P(T=1)=1/N.
- No? "failure" (1-p=1-1/N). Go to their mothers and repeat the same question.
- $P(T=t)=(1-1/N)^{t-1}(1/N) \approx (1/N)\exp(-(t-1)/N)$
- T can be inferred from the density of differences on mtDNA $=2\mu t$



Maddamsetti R, MOBILE GENETIC ELEMENTS, 6, e1137380(2016)

- There are about $N=3.5x10^9$ women living today
- For a random pair of women the average number of generations to the last common maternal ancestor is:

$$E(T)=\sum_{T=1}^{\infty} T \cdot exp(-T/N)=1/p=N$$

- Most Recent maternal Common Ancestor (MRCA) of all people living today lived $T_{MRCA}=2N$ generations ago
- $T_{MRCA} = 2 \cdot 3.5x10^9$ generations
- If the generation time 20 years it is 140 billion years > 10 times the time since the Big Bang.
- Something is wrong here!

- Population is not constant and for a long time was very low
- Change N to "effective" size N_e
- Current thinking is that for all of us including people of African ancestry N_e~7500 people
- For humans of European + Asian ancestry N_e~ 3100 people

- Mito Eve lived ~

2*Ne*20 years=
=2*7500*20 years=
300,000 years ago

Recent human effective population size estimated from linkage disequilibrium

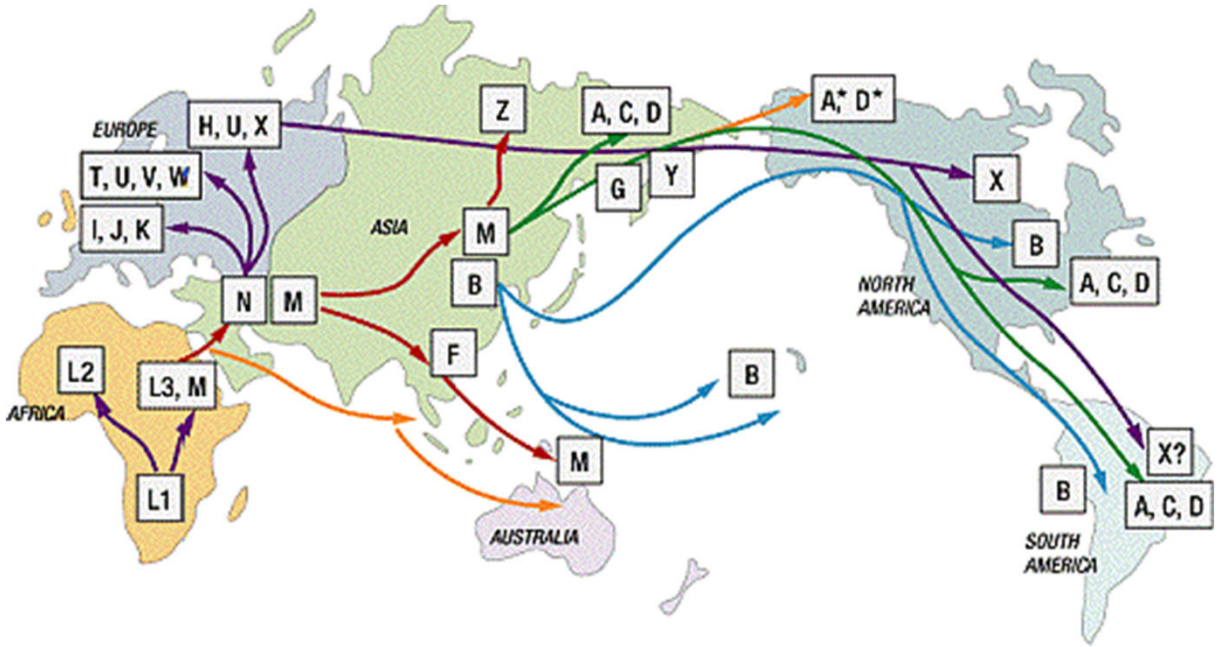
Albert Tenesa,[1,2,3] Pau Navarro,[3] Ben J. Hayes,[4] David L. Duffy,[5] Geraldine M. Clarke,[6] Mike E. Goddard,[4,7] and Peter M. Visscher[3,5,8]

[1]Colon Cancer Genetics Group, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom; [2]MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, United Kingdom; [3]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom; [4]Victorian Institute of Animal Science, DPI, Attwood 3049, Australia; [5]Queensland Institute of Medical Research, Royal Brisbane Hospital, Brisbane 4006, Australia; [6]The Wellcome Trust Centre for Human Genetics, The University of Oxford, Oxford OX3 7BN, United Kingdom; [7]Institute of Land and Food Resources, University of Melbourne, Parkville 3010, Australia

Effective population size (N_e) determines the amount of genetic variation, genetic drift, and linkage disequilibrium (LD) in populations. Here, we present the first genome-wide estimates of human effective population size from LD data. Chromosome-specific effective population size was estimated for all autosomes and the X chromosome from estimated LD between SNP pairs <100 kb apart. We account for variation in recombination rate by using coalescent-based estimates of fine-scale recombination rate from one sample and correlating these with LD in an independent sample. Phase I of the HapMap project produced between 18 and 22 million SNP pairs in samples from four populations: Yoruba from Ibadan (YRI), Nigeria; Japanese from Tokyo (JPT); Han Chinese from Beijing (HCB); and residents from Utah with ancestry from northern and western Europe (CEU). For CEU, JPT, and HCB, the estimate of effective population size, adjusted for SNP ascertainment bias, was ~3100, whereas the estimate for the YRI was ~7500, consistent with the out-of-Africa theory of ancestral human population expansion and concurrent bottlenecks. We show that the decay in LD over distance between SNPs is consistent with recent population growth. The estimates of N_e are lower than previously published estimates based on heterozygosity, possibly because they represent one or more bottlenecks in human population size that occurred ~10,000 to 200,000 years ago.

Tenesa, Albert, et al. *Genome research* 17.4 (2007): 520-526.

“Mitochondrial Eve” lived in Africa

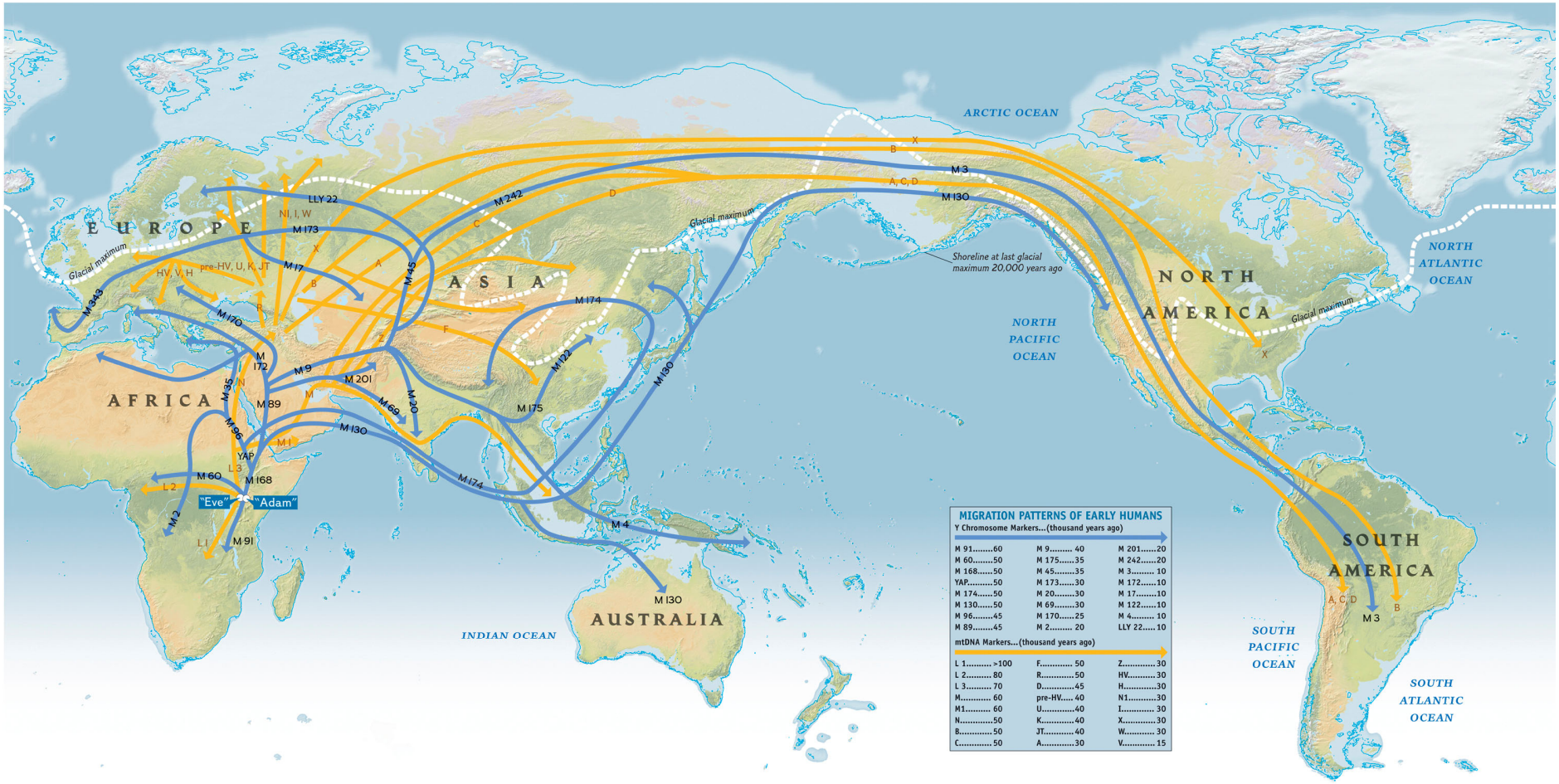


EXPANSION TIMES *(years ago)*	
Africa	120,000 - 150,000
Out of Africa	55,000 - 75,000
Asia	40,000 - 70,000
Australia/PNG	40,000 - 60,000
Europe	35,000 - 50,000
Americas	15,000 - 35,000
Na-Dene/Esk/Aleuts	8,000 - 10,000

FamilyTreeDNA
mtDNA Migrations Map

- “Mitochondrial Eve” lived in Africa between 100,000 and 150,000 years ago (or 240,000?)
- *Poznik GD, et al (Carlos Bustamante lab in Stanford), Science **341**: 562 (August 2013).*

“Adam” and “Eve” are both out of Africa



- “Mitochondrial Eve” lived in Africa between 100,000 and 150,000 years ago (or 240,000?)
- “Y-chromosome Adam” also lived in Africa between 120,000 and 160,000 years ago
- *Poznik GD, et al (Carlos Bustamante lab in Stanford), Science **341**: 562 (August 2013).*

Mitochondrial Eve (maternally transmitted ancestry)
Y-chromosome Adam (paternally transmitted ancestry)
lived ~200,000 years ago.

When lived the latest common ancestor shared by all of us based on nuclear DNA?

A. 1 million years ago
B. 200,000 years ago
C. 3400 years ago
D. 660 years ago
E. Yesterday, I really have no clue

Get your i-clickers

Mitochondrial Eve (maternally transmitted ancestry)
Y-chromosome Adam (paternally transmitted ancestry)
lived ~200,000 years ago.

When lived the latest common ancestor shared by all of us based on nuclear DNA?

A. 1 million years ago
B. 200,000 years ago
C. 3400 years ago
D. 660 years ago
E. Yesterday, I really have no clue

Get your i-clickers

Last common ancestor in nuclear (non Y-chr) DNA is another matter

- Nuclear DNA gets mixed with every generation
  - Each of us gets 50% of nuclear DNA from father & 50% from mother
  - Each has 2 parents, 4 grandparents, 8 great-grand parents ….
- If one assumes:
  - Well-mixed marriages (not true: mostly local until recently)
  - Constant size population (not true: much smaller)
  - In 33 generations the number of ancestors: 2^{33} =8 billion > 7 billion people living today
- Every pair of us living today should have at least one shared ancestor who lived
  - 33 generations * 20 years/generation=660 years ago ~1300 AD

Corrected for mostly local marriages rare migrations

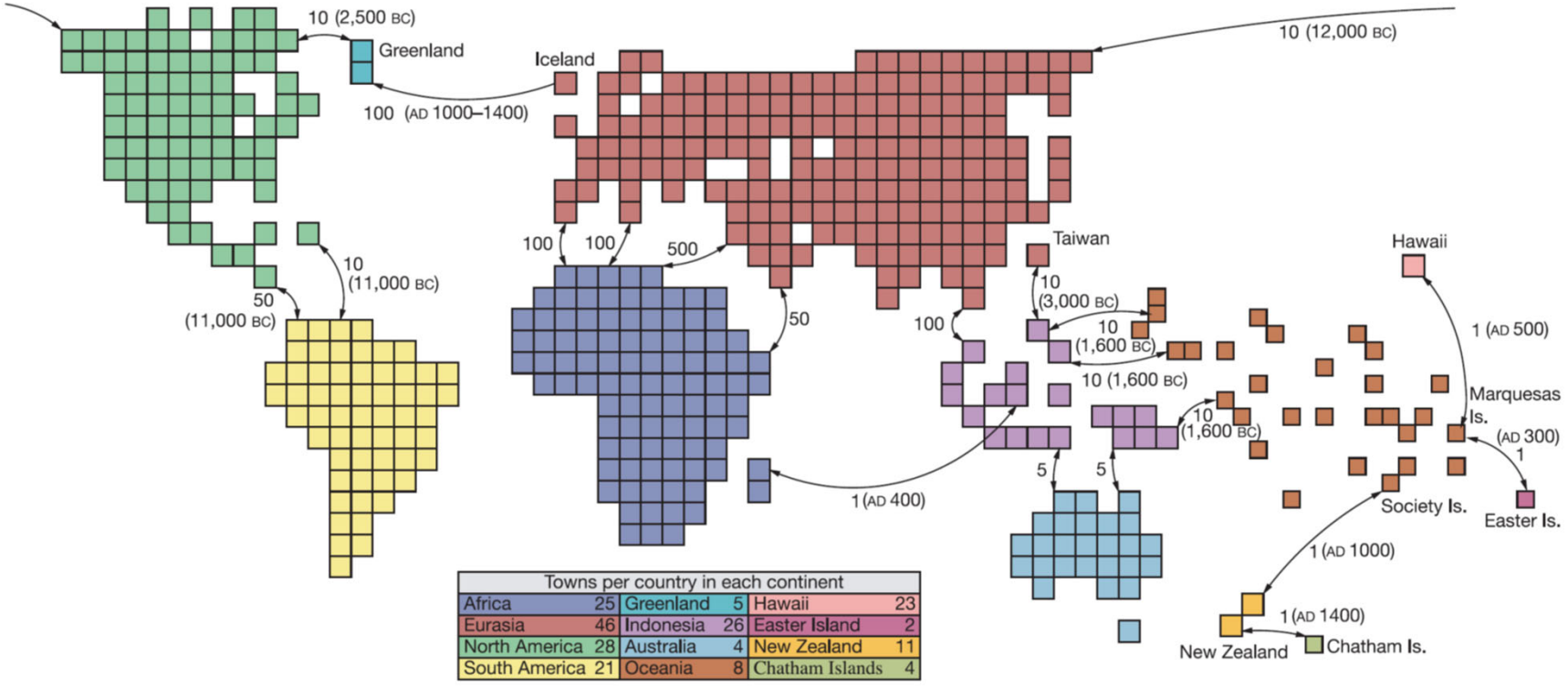
562 NATURE | VOL 431 | 30 SEPTEMBER 2004 |

Modelling the recent common ancestry of all living humans

Douglas L. T. Rohde[1], Steve Olson[2] & Joseph T. Chang[3]

[1]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*7609 Sebago Road, Bethesda, Maryland 20817, USA*
[3]*Department of Statistics, Yale University, New Haven, Connecticut 06520, USA*

With 5% of individuals migrating out of their home town, 0.05% migrating out of their home country, and 95% of port users born in the country from which the port emanates, the simulations produce a mean MRCA date of 1,415 BC and a mean IA date of 5,353 BC.

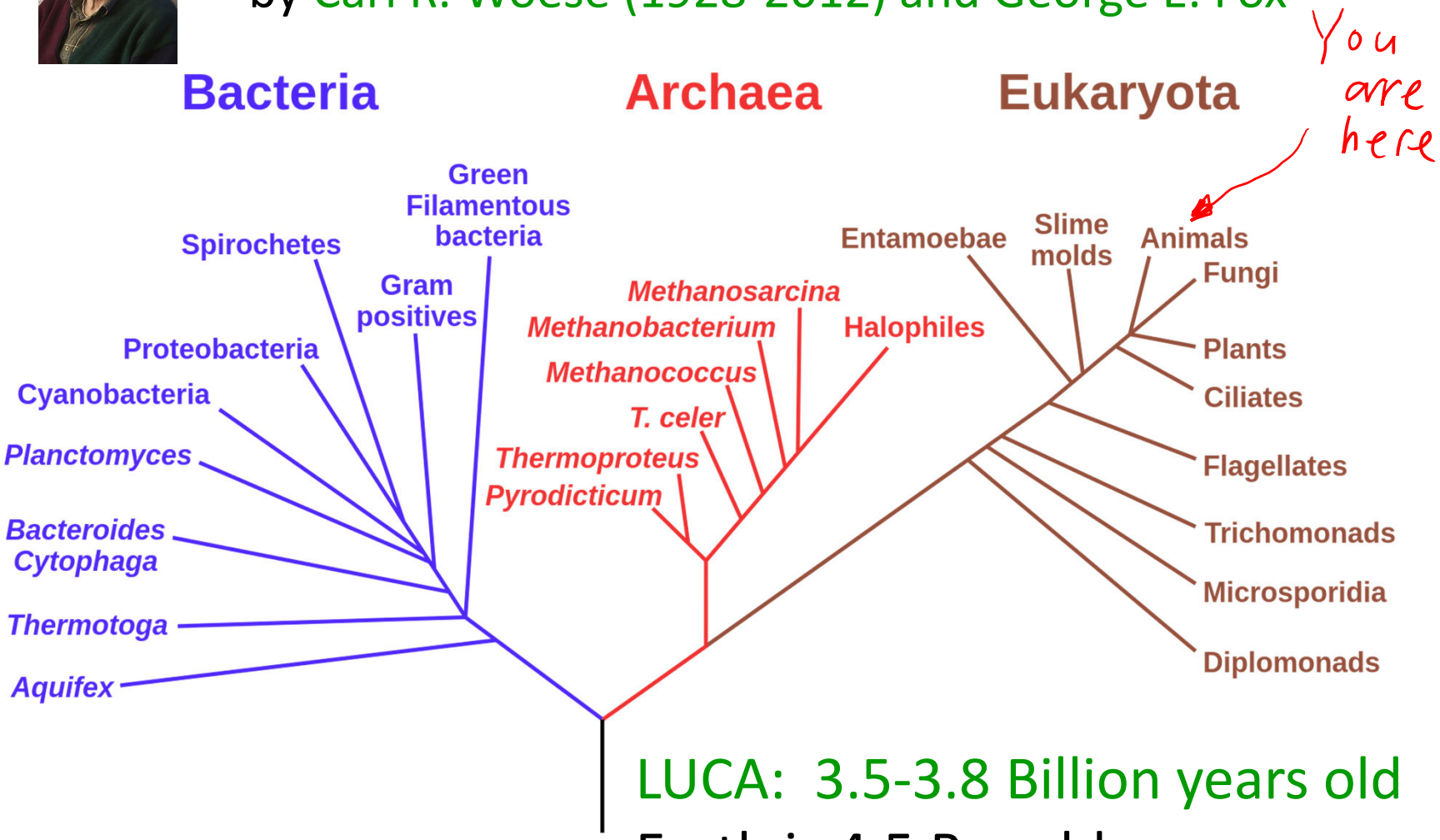


Towns per country in each continent					
Africa	25	Greenland	5	Hawaii	23
Eurasia	46	Indonesia	26	Easter Island	2
North America	28	Australia	4	New Zealand	11
South America	21	Oceania	8	Chatham Islands	4

Figure 2 Geography and migration routes of the simulated model. Arrows denote ports and the adjacent numbers are their steady migration rates, in individuals per generation. If given, the date in parentheses indicates when the port opens. Upon opening, there is usually a first-wave migration burst at a higher rate, lasting one generation.

Last Universal Common Ancestor (LUCA)

Archaea were discovered here at UIUC in 1977
by Carl R. Woese (1928-2012) and George E. Fox



LUCA: 3.5-3.8 Billion years old

Earth is 4.5 Byr old