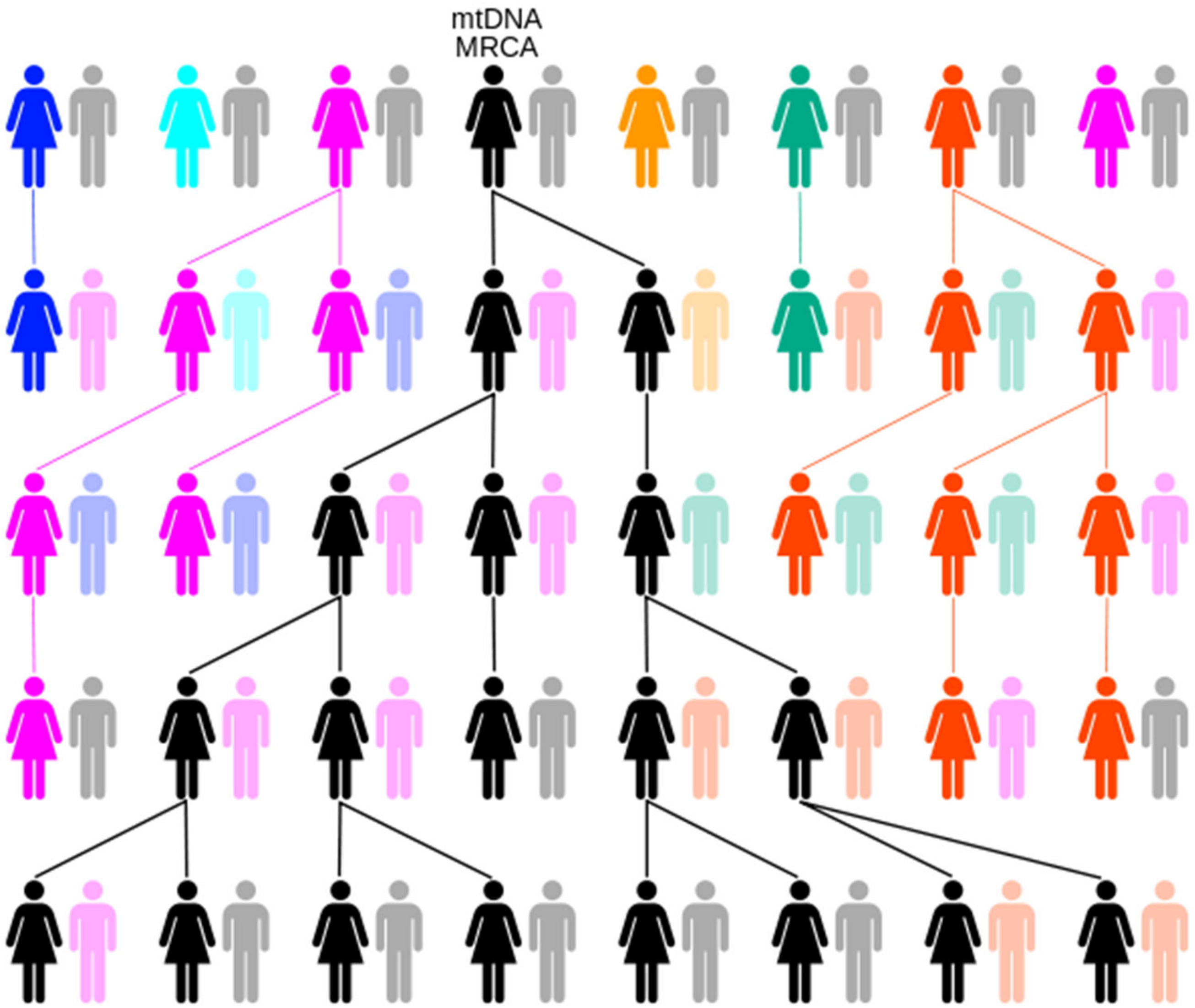


Assume that *x* – the # of daughters per each mother follows a Poisson distribution

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

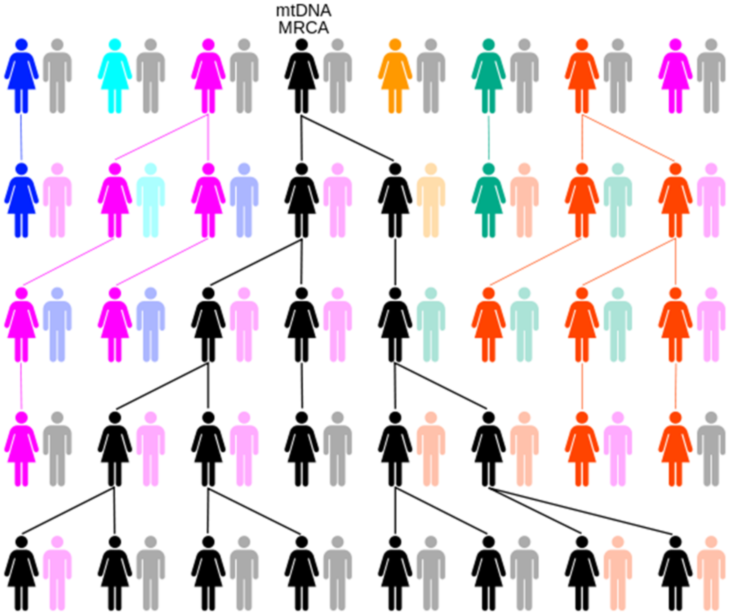
Population does not grow → $\lambda=1$

Prob(merge)= $=E[x(x-1)]/N= = \lambda^2/N=1/N$

$$P(T=t)=(1-1/N)^{t-1}(1/N) \approx (1/N)\exp(-(t-1)/N)$$

Most Recent Common Ancestor (MRCA)

- Start with N individuals. Time for one pair to merge is $E(T) = \sum_{t=1}^{\infty} t \cdot (1/N)\, exp(-t/N) = N$
- Any of $\frac{N(N-1)}{2}$ pairs can merge first. The average time for the first pair to merge is $\frac{2}{N(N-1)} N$
- After merger $N \to N-1$,
- So, the time until the next merger is $\frac{2}{(N-1)(N-2)}$



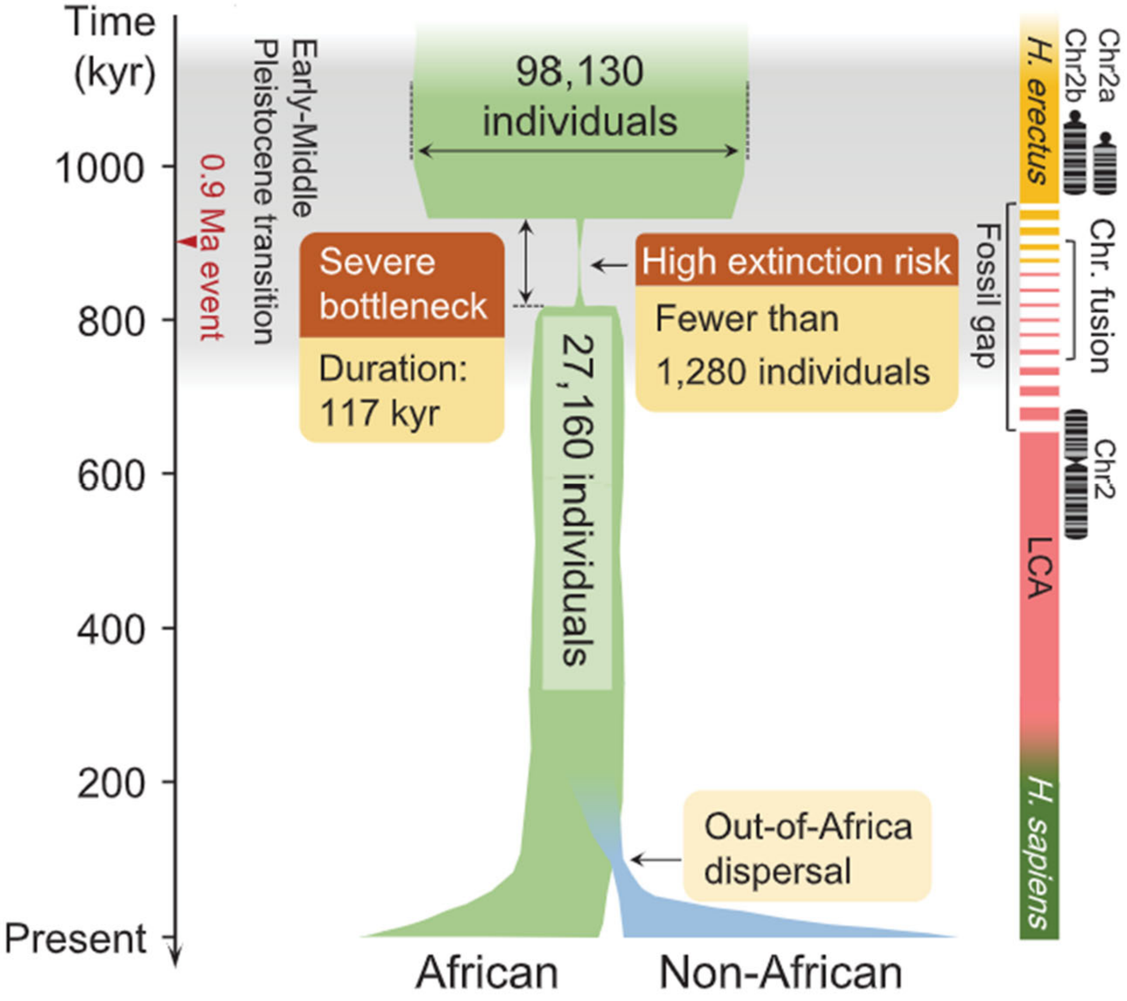
Most Recent Common Ancestor (MRCA)

Total time until the MRCA

$$T_{MRCA} = N \cdot \sum_{k=2}^{N} \frac{2}{k(k-1)}$$

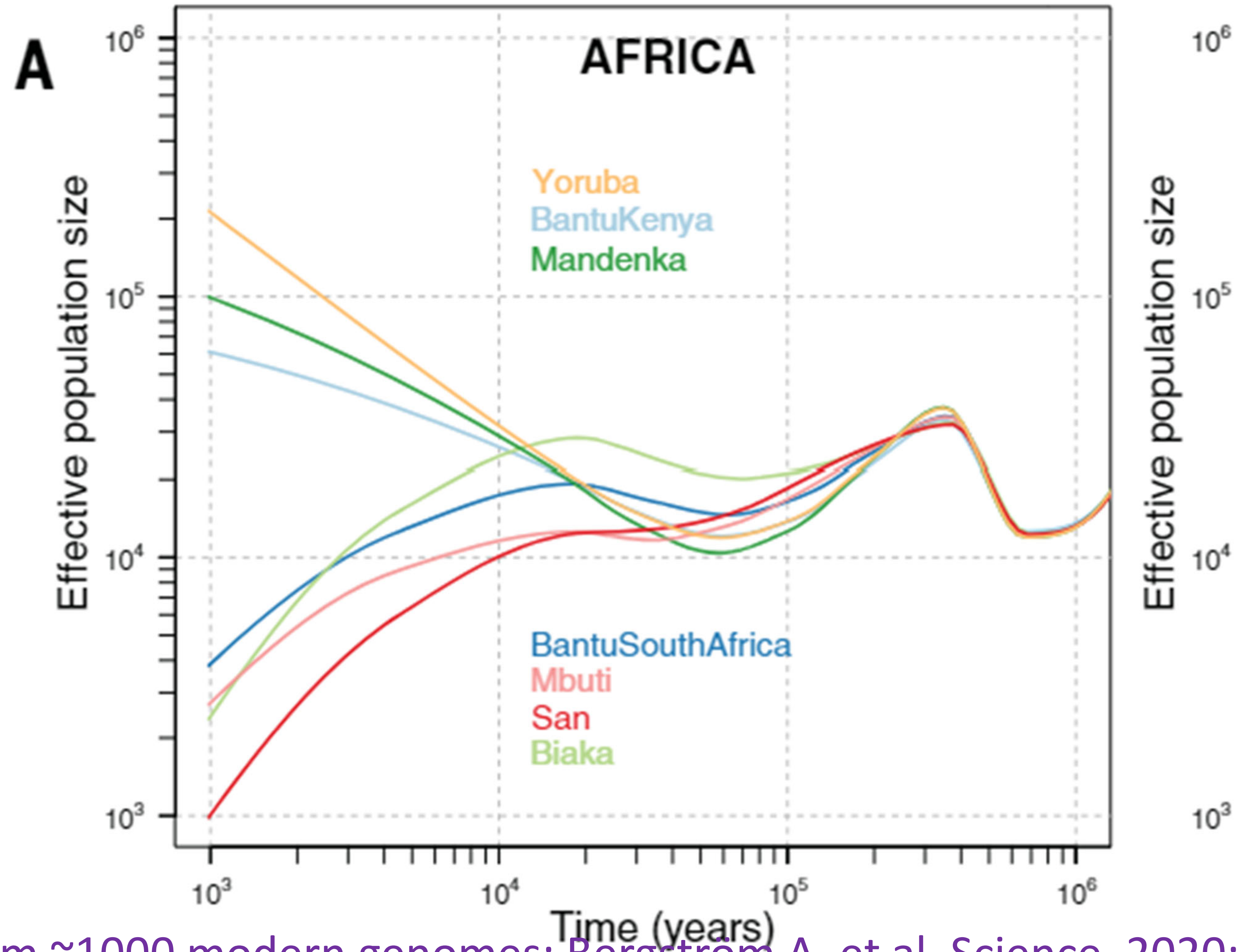
$$= 2N \sum_{k=2}^{N} \left(\frac{1}{k-1} - \frac{1}{k}\right) = 2N \left(1 - \frac{1}{N}\right) \approx 2N$$

Hot off the press: human ancestors almost got extinct about 1M years ago



Hu W, et al. Science. 2023;381: 979–984

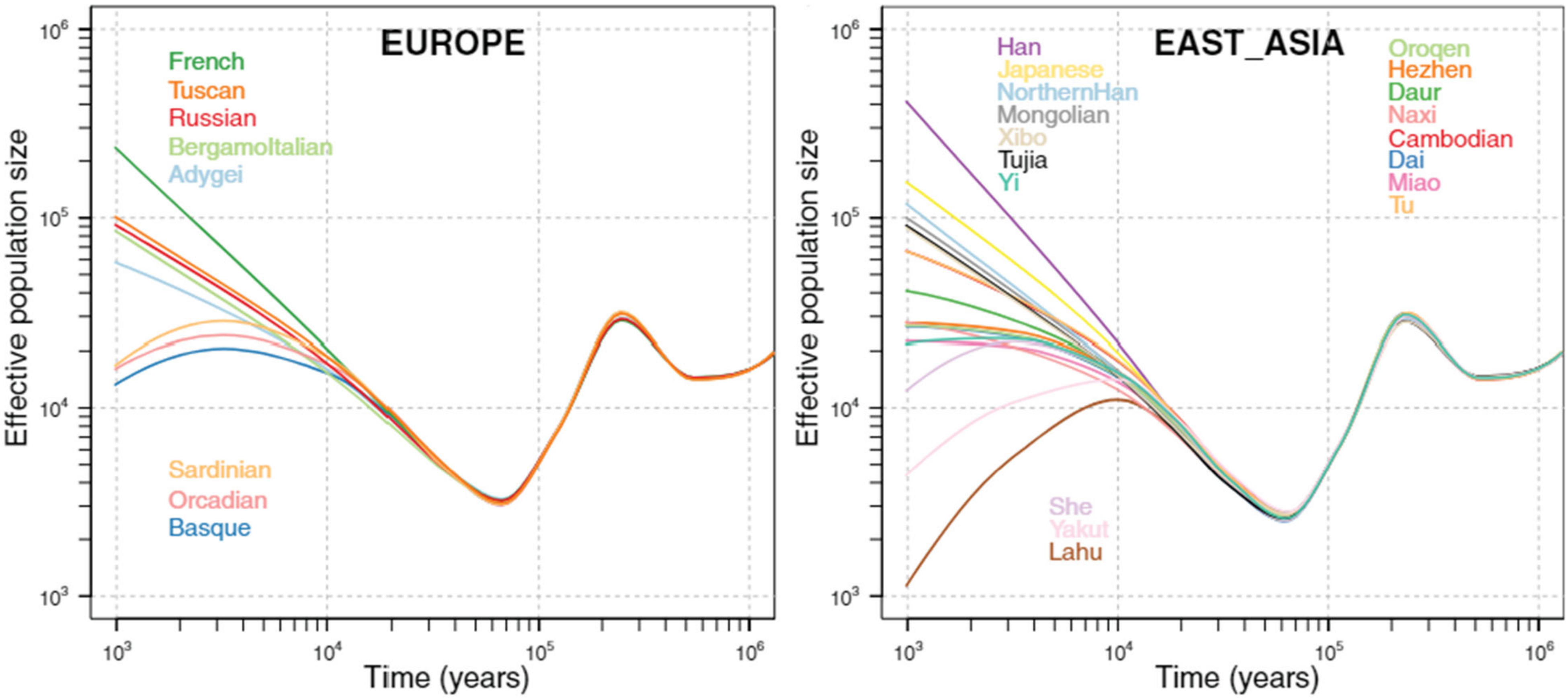
Effective human population size ~10,000



From ~1000 modern genomes: Bergström A, et al. Science. 2020;367

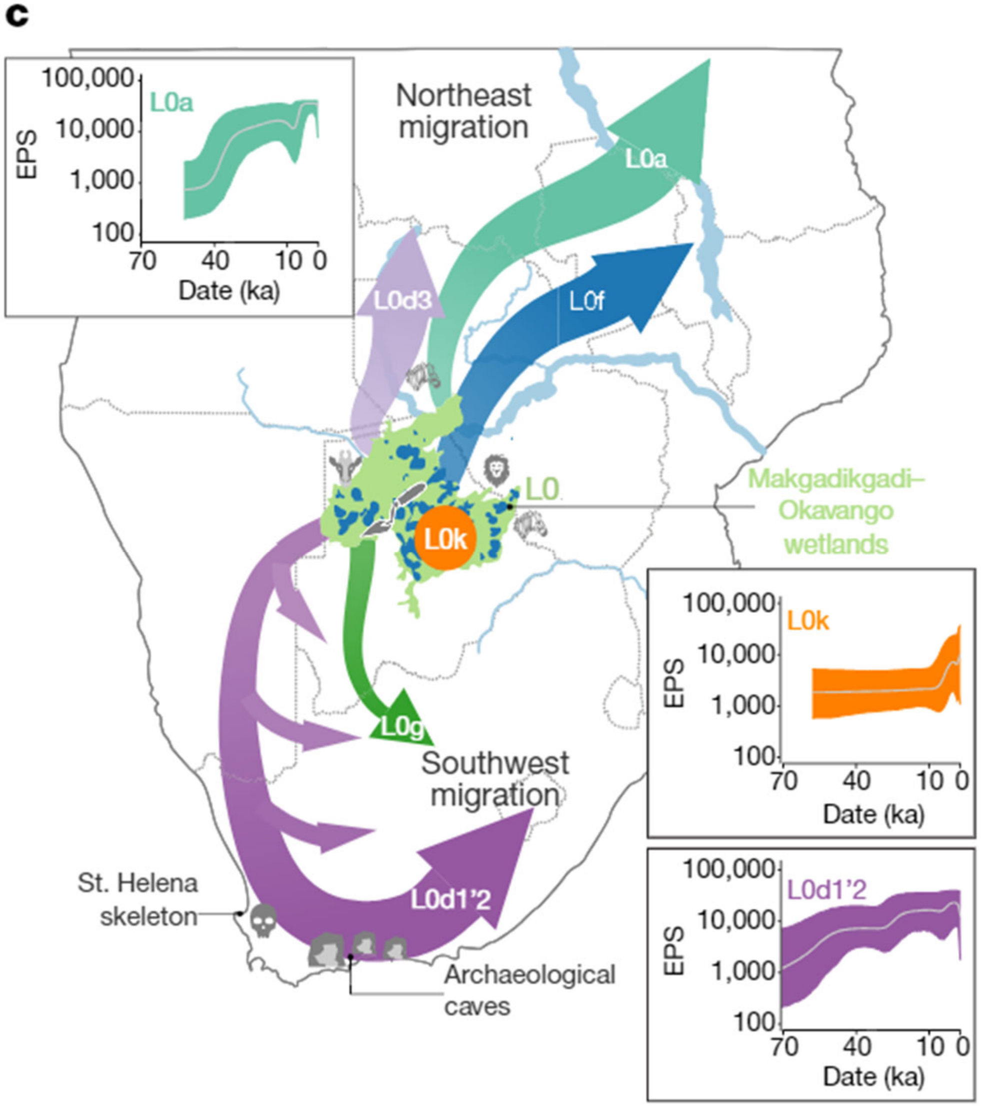
- Population is not constant and for a long time was very low
- Change N to the “effective” size N_e
- Current thinking is that for all of us including people of African ancestry N_e~10,000 people
- For humans of European + Asian ancestry N_e~ 3000 people
- Mito Eve lived in Africa ~2*(Ne/2)*20 years=10,000*20 years= 200,000 years ago

Effective human population size in Europe and Asia ~3000 people ~60,000 years ago



From ~1000 modern genomes: Bergström A, et al. Science. 2020;367

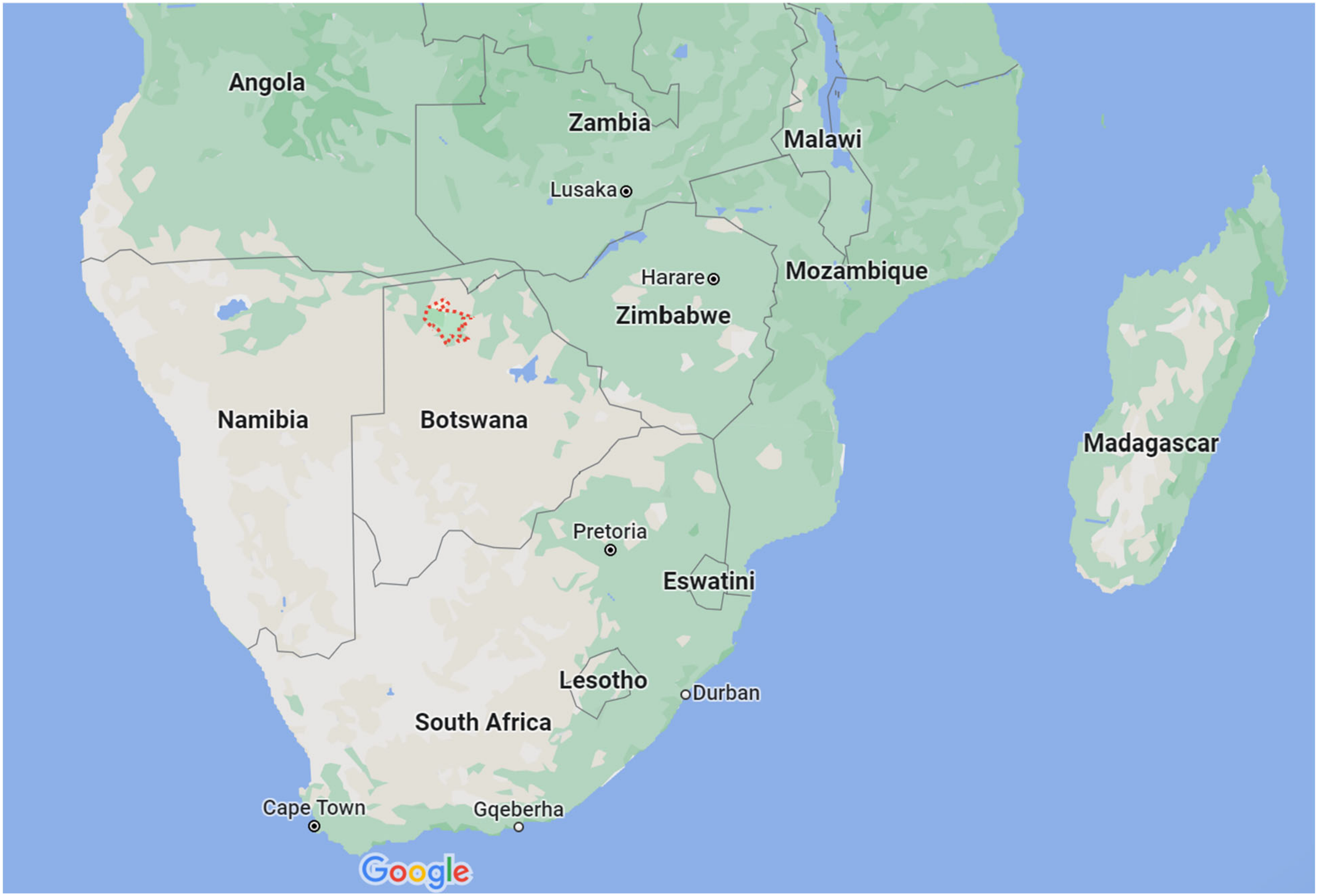
“Mitochondrial Eve” lived in Africa



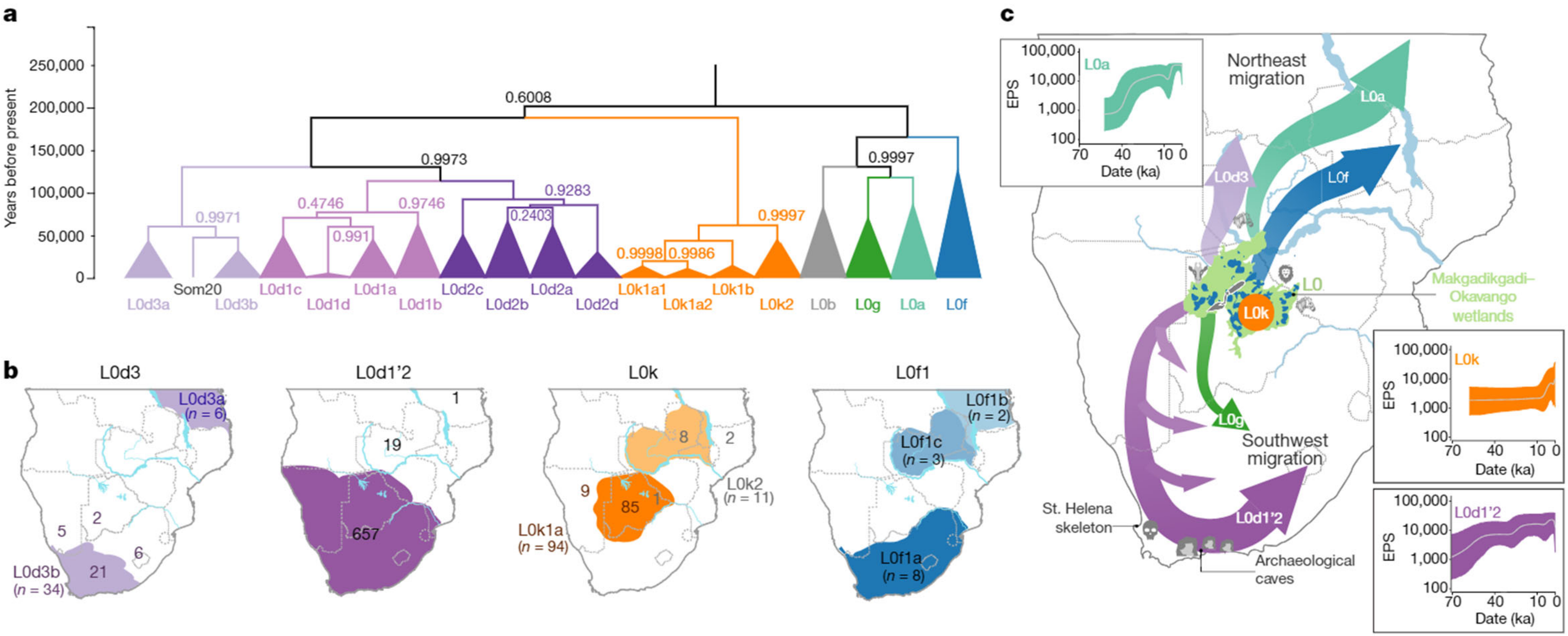
“Mitochondrial Eve” lived in Makgadikgadi–Okavango paleo-wetland of southern Africa ~200,000 years ago (between 165,000 and 240,000 years ago)

Chan EKF, et al. Nature. 2019; 575: 185–189.

Okavango Delta now



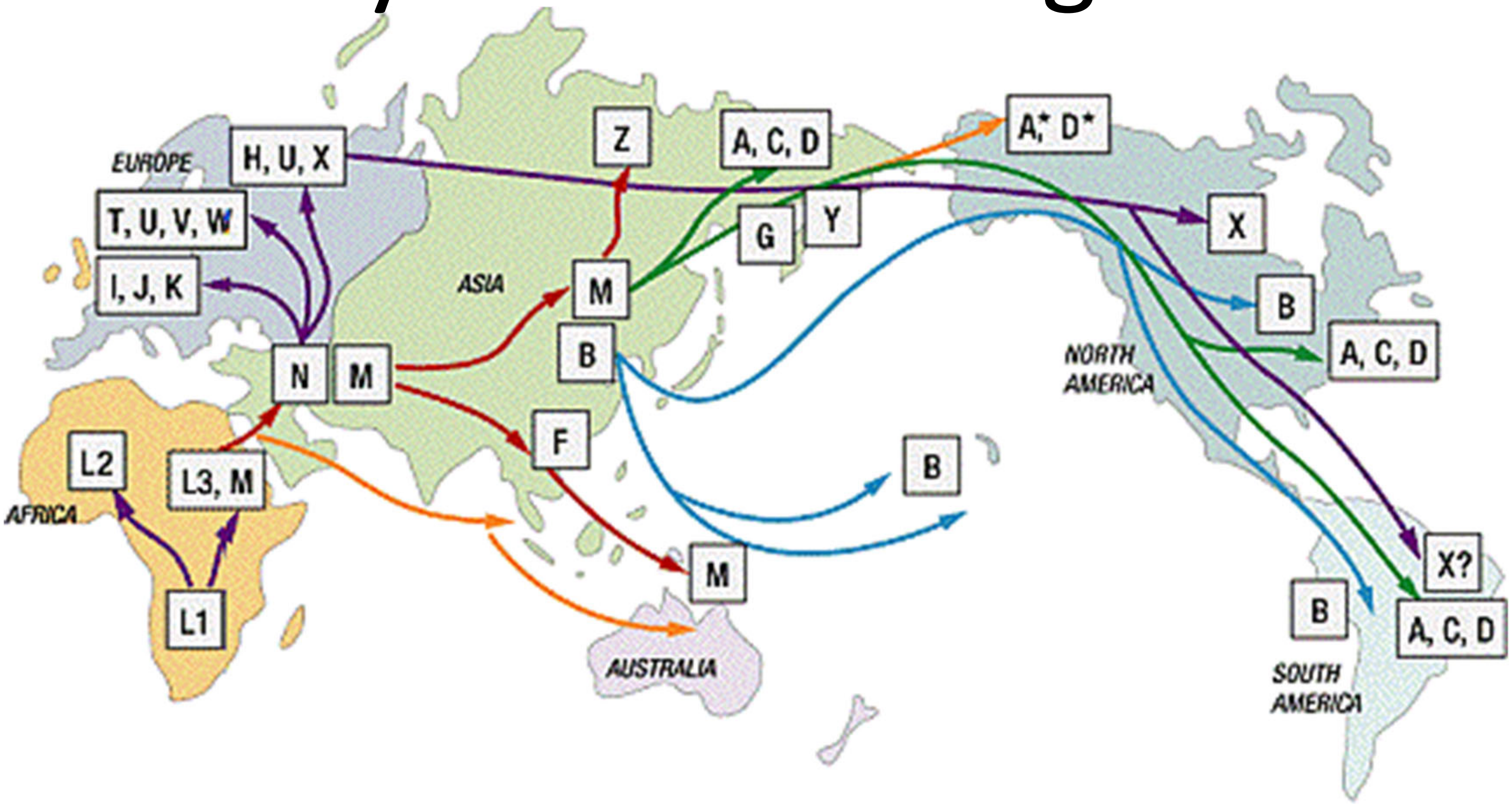
“Mitochondrial Eve” lived in Africa

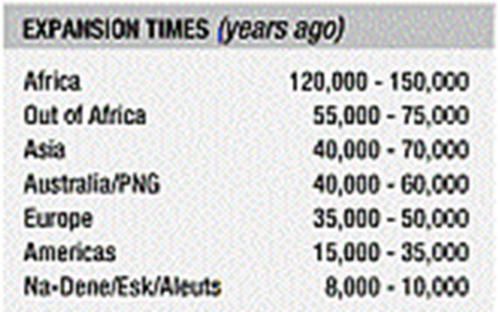


“Mitochondrial Eve” lived in Makgadikgadi–Okavango paleo-wetland of southern Africa ~200,000 years ago (between 165,000 and 240,000 years ago)

Chan EKF, et al. Nature. 2019; 575: 185–189.

Modern mitochondrial DNA contains history of human migrations





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

*Poznik GD, et al (Carlos Bustamante lab in Stanford), Science **341**: 562 (August 2013).*

What about men?

- Y-chromosome is transferred from father to son
- Like mitochondria it can be used to trace ancestry of all men to the “Y-chromosome Adam”
- Where did “Adam’” live? Did he meet the “mitochondrial Eve”?

Y-chromosomal Adam also lived in Africa

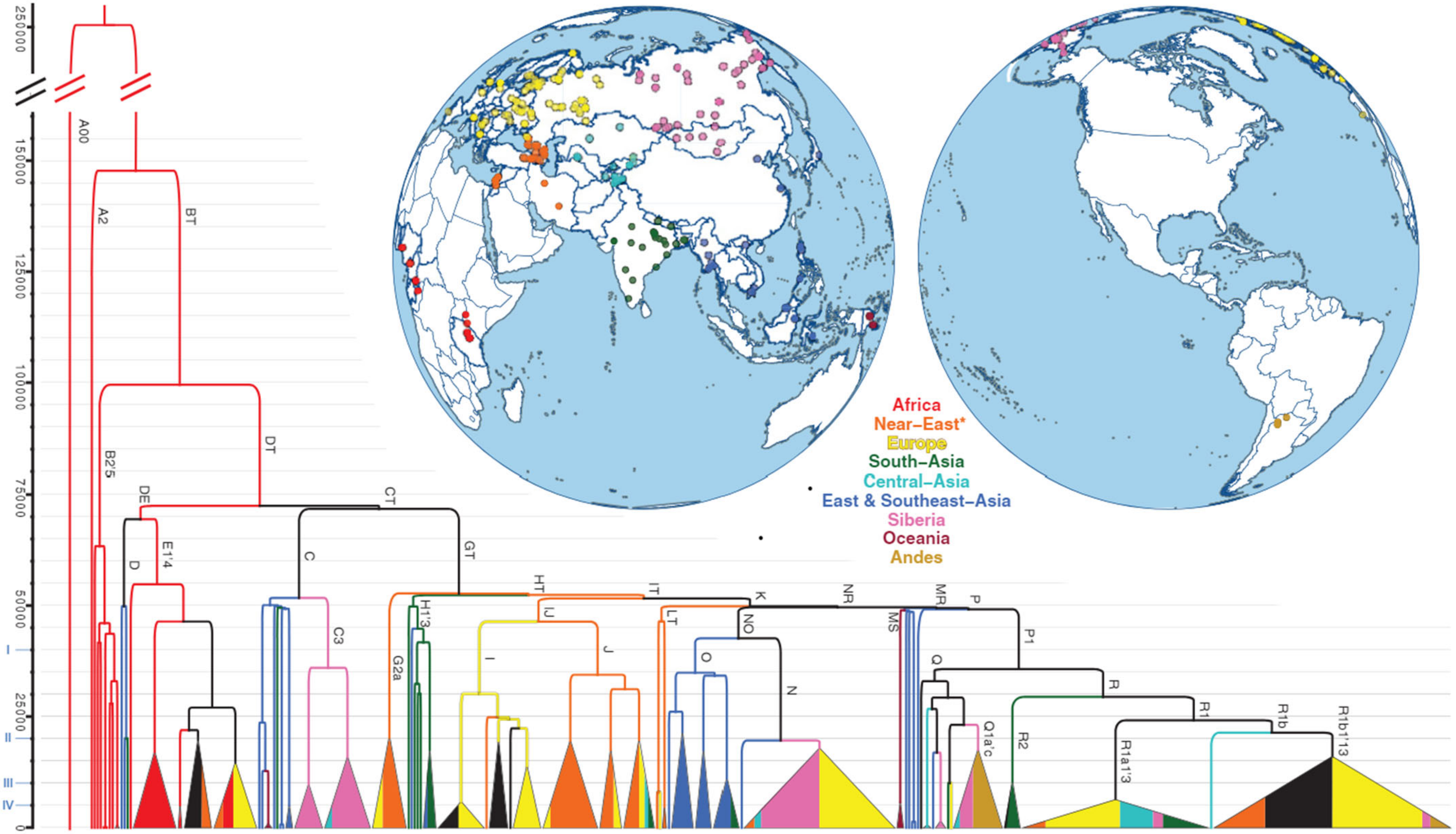
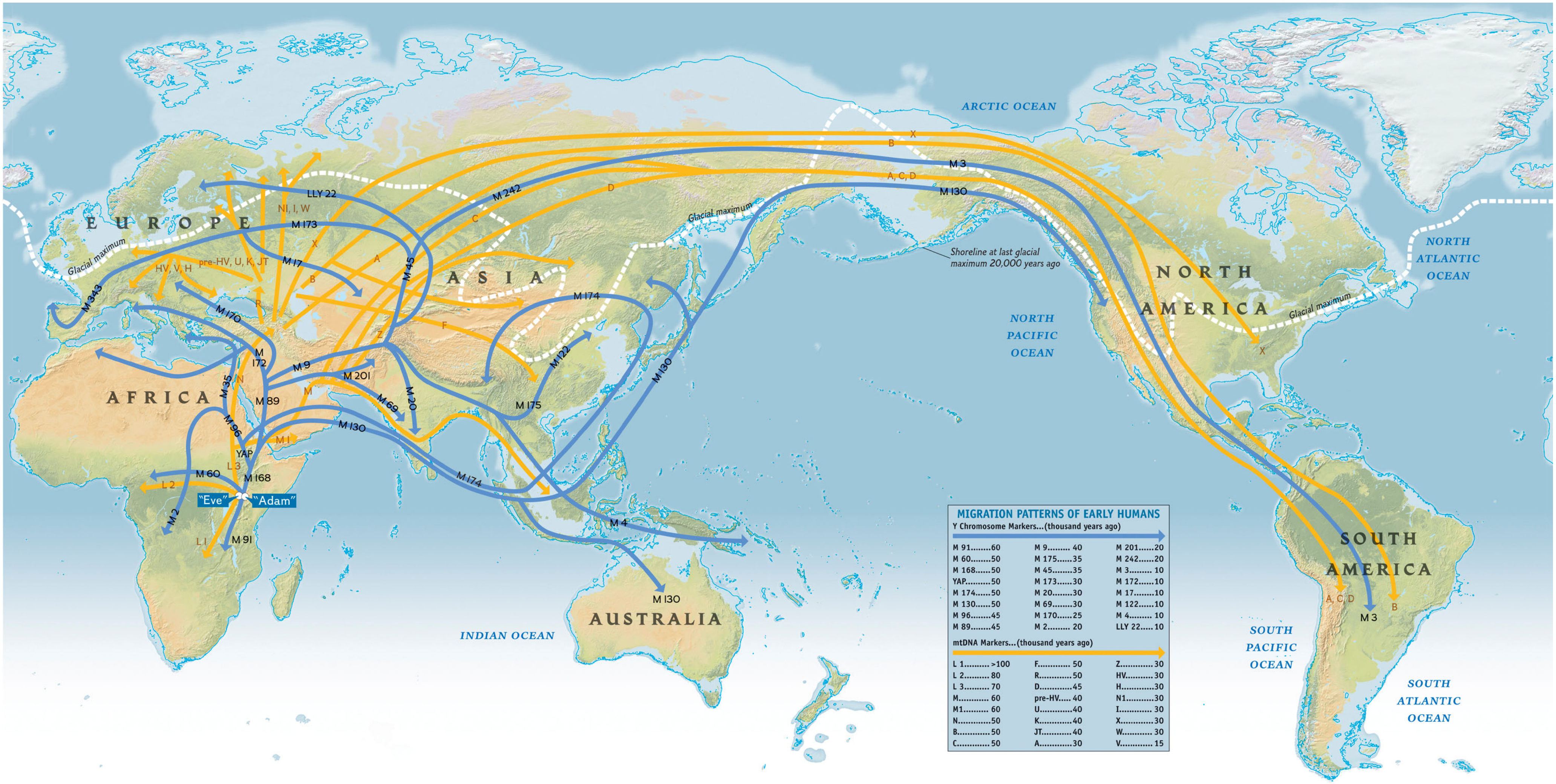


Figure 1. The phylogenetic tree of 456 whole Y chromosome sequences and a map of sampling locations. The phylogenetic tree is reconstructed using BEAST. Clades coalescing within 10% of the overall depth of the tree have been collapsed. Only main haplogroup labels are shown (details are provided in Supplemental Information 6). Colors indicate geographic origin of samples (Supplemental Table S1), and fill proportions of the collapsed clades represent the proportion of samples from a given region. Asterisk (*) marks the inclusion of samples from Caucasus area. Personal Genomes Project (http://www.personalgenomes.org) samples of unknown and mixed geographic/ethnic origin are shown in black. The proposed structure of Y chromosome haplogroup naming (Supplemental Table S5) is given in Roman numbers on the *y*-axis.

Karmin M, Saag L, Vicente M, Sayres MAW, Järve M, Talas UG, et al. Genome Res. 2015;25: 459–466.

“Adam” and “Eve” both lived in Africa



MIGRATION PATTERNS OF EARLY HUMANS

Y Chromosome Markers...(thousand years ago)

M 91........60	M 9......... 40	M 201......20
M 60........50	M 175......35	M 242......20
M 168......50	M 45........35	M 3......... 10
YAP..........50	M 173......30	M 172......10
M 174......50	M 20........30	M 17........10
M 130......50	M 69........30	M 122......10
M 96........45	M 170......25	M 4......... 10
M 89........45	M 2......... 20	LLY 22..... 10

mtDNA Markers...(thousand years ago)

L 1.......... >100	F............ 50	Z............30
L 2.......... 80	R............50	HV..........30
L 3.......... 70	D............45	H............30
M............ 60	pre-HV..... 40	N1..........30
M1.......... 60	U............40	I............ 30
N............50	K............40	X............30
B............50	JT...........40	W........... 30
C............50	A............30	V............ 15

- “Mitochondrial Eve” lived in Africa between 100,000 and 240,000 years ago
- “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

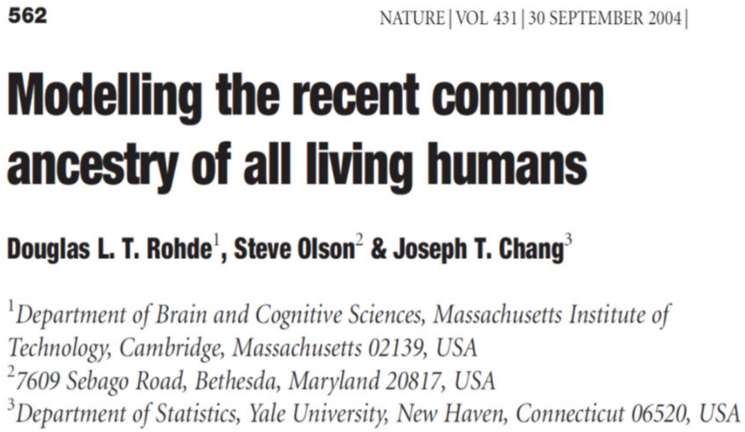
E. Yesterday, I really have no clue

Get your i-clickers

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

- Unlike Mito or Y-chromosome, **nuclear DNA gets mixed with every generation**
  - Each of us gets 50% of nuclear DNA from the father & 50% from the mother
  - Each of us has 2 parents, 4 grandparents, 8 great-grand parents …
- If one assumes:
  - Well-mixed marriages (not true: mostly local marriages until recently)
  - Constant size population (not true: much smaller in the past)
  - 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 and rare migrations

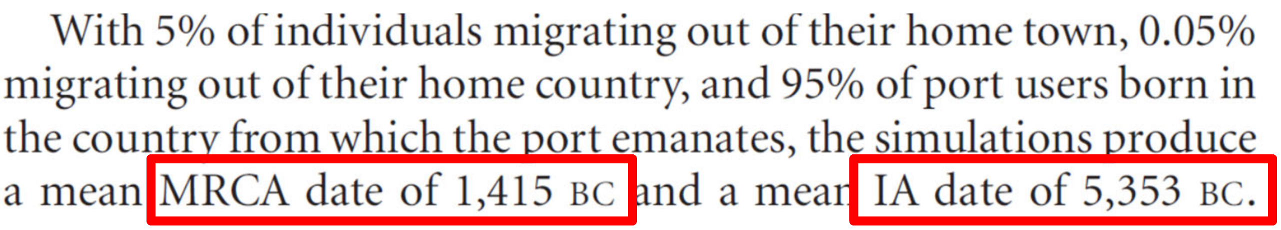


562 NATURE | VOL 431 | 30 SEPTEMBER 2004 |

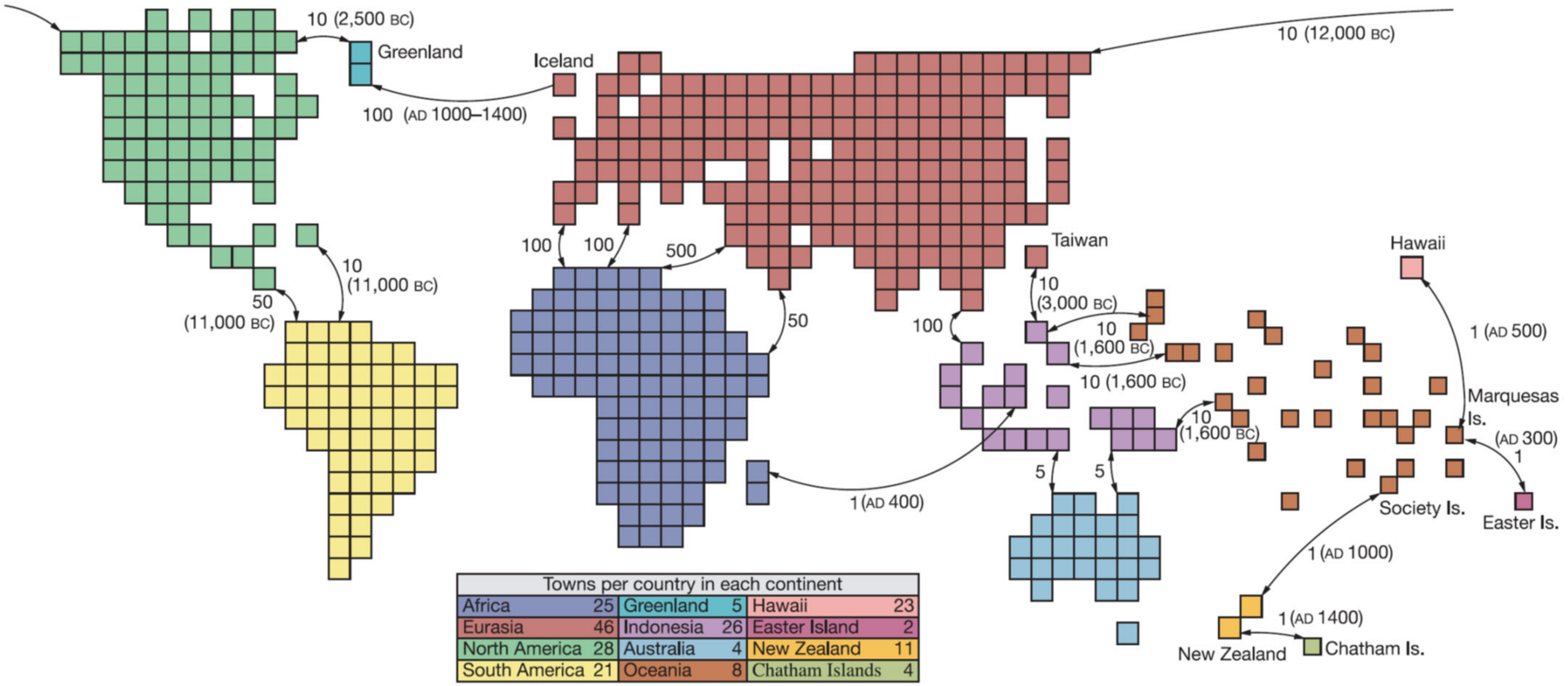
Modelling the recent common ancestry of all living humans

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*



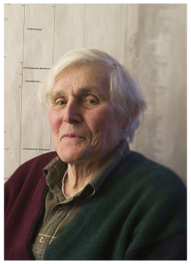
With 5% of individuals migrating out of their home town, 0.05% migrating out of their home country, and 95% of port users born in the country from which the port emanates, the simulations produce a mean MRCA date of 1,415 BC and a mean IA date of 5,353 BC.



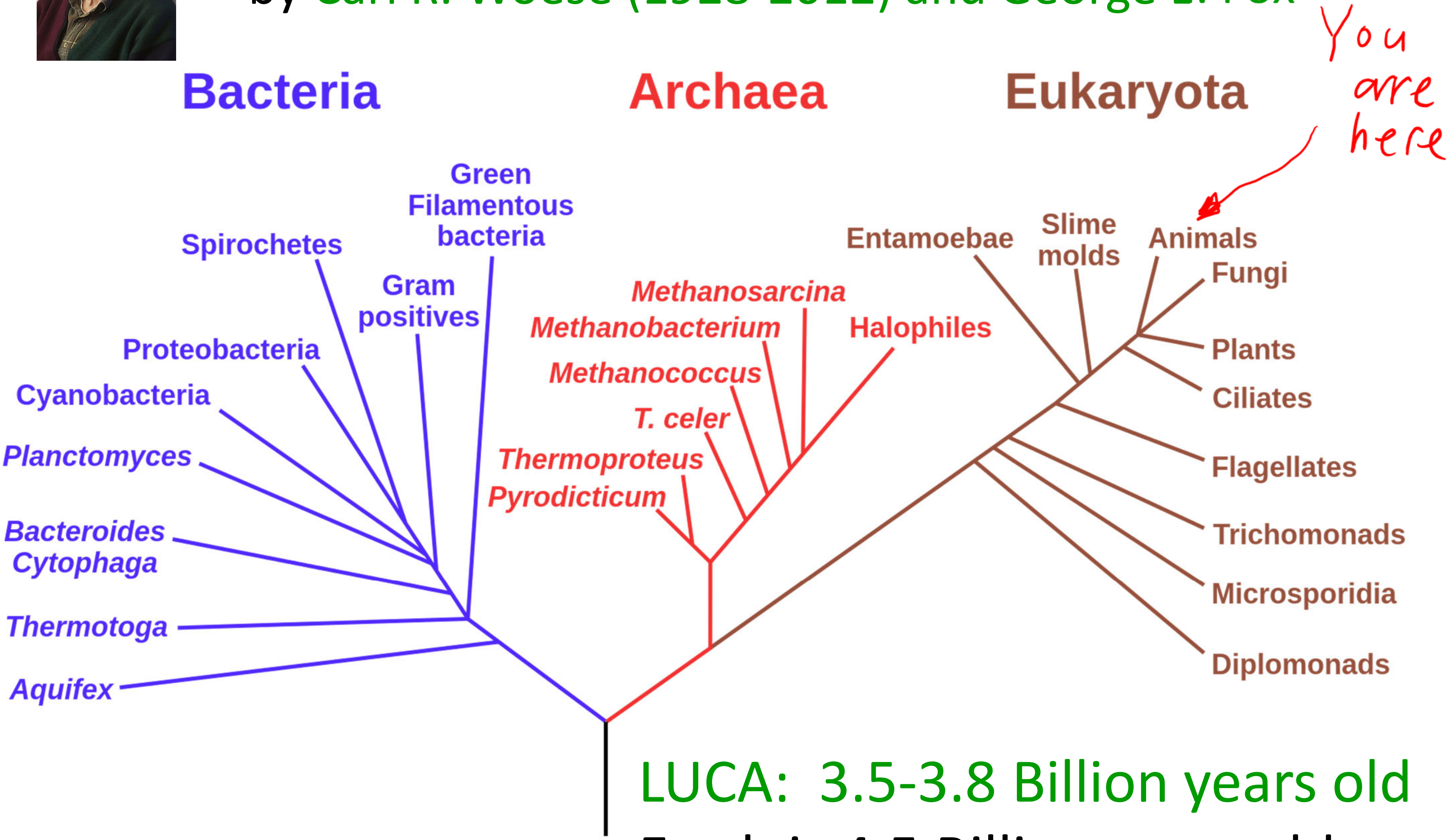
Towns per country in each continent					
Africa	25	Greenland	5	Hawaii	23
Eurasia	46	Indonesia	26	Easter Island	2
North America	28	Australia	4	New Zealand	11
South America	21	Oceania	8	Chatham Islands	4

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)

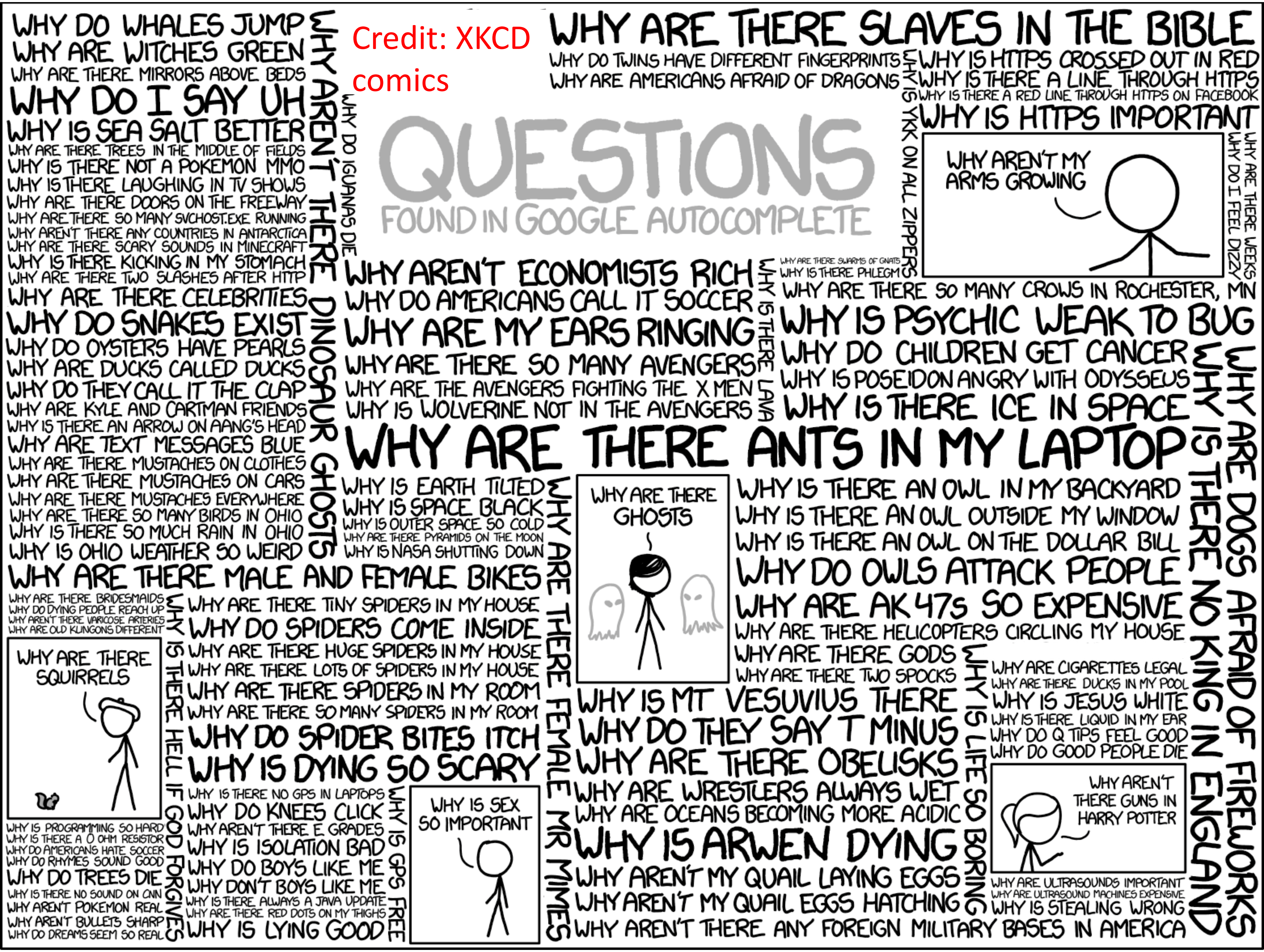


Archaea were discovered here at UIUC in 1977 by Carl R. Woese (1928-2012) and George E. Fox



LUCA: 3.5-3.8 Billion years old

Earth is 4.5 Billion years old


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 AREN'T THERE DINOSAUR GHOSTS
WHY DO IGUANAS DIE
WHY ARE THERE SLAVES IN THE BIBLE
WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS YKK ON ALL ZIPPERS
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 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 ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
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 ARE THERE FEMALE MR MIMES
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 THERE NO KING IN ENGLAND
WHY ARE DOGS AFRAID OF FIREWORKS
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 IS THE HELL IF GOD FORGIVES
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 GPS FREE
WHY IS SEX SO IMPORTANT
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 IS LIFE SO BORING
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

Credit: XKCD comics

Negative Binomial Definition

- In a series of independent trials with constant probability of success, p, let the random variable X denote the number of trials until r successes occur. Then X is a negative binomial random variable with parameters:
  $0 < p < 1$ and r = 1, 2, 3, ...
- The probability mass function is:

$$f(x) = C_{r-1}^{x-1} p^r (1-p)^{x-r} \quad \text{for } x = r, r+1, r+2 \ldots \qquad (3\text{-}11)$$

- Compare it to binomial

$$f(x) = C_x^n p^x (1-p)^{n-x} \quad \text{for } x = 1, 2, \ldots \text{ n}$$

NOTE OF CAUTION: Matlab, Mathematica, and many other sources use x to denote the number of failures until one gets r successes.
We stick with Montgomery-Runger.

Negative Binomial Mean & Variance

- If *X* is a negative binomial random variable with parameters *p* and *r*,

$$\mu = E\left(X\right) = \frac{r}{p} \quad \text{and} \quad \sigma^2 = V\left(X\right) = \frac{r\left(1-p\right)}{p^2} \qquad (3\text{-}12)$$

- Compare to geometric distribution:

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

Matlab exercise

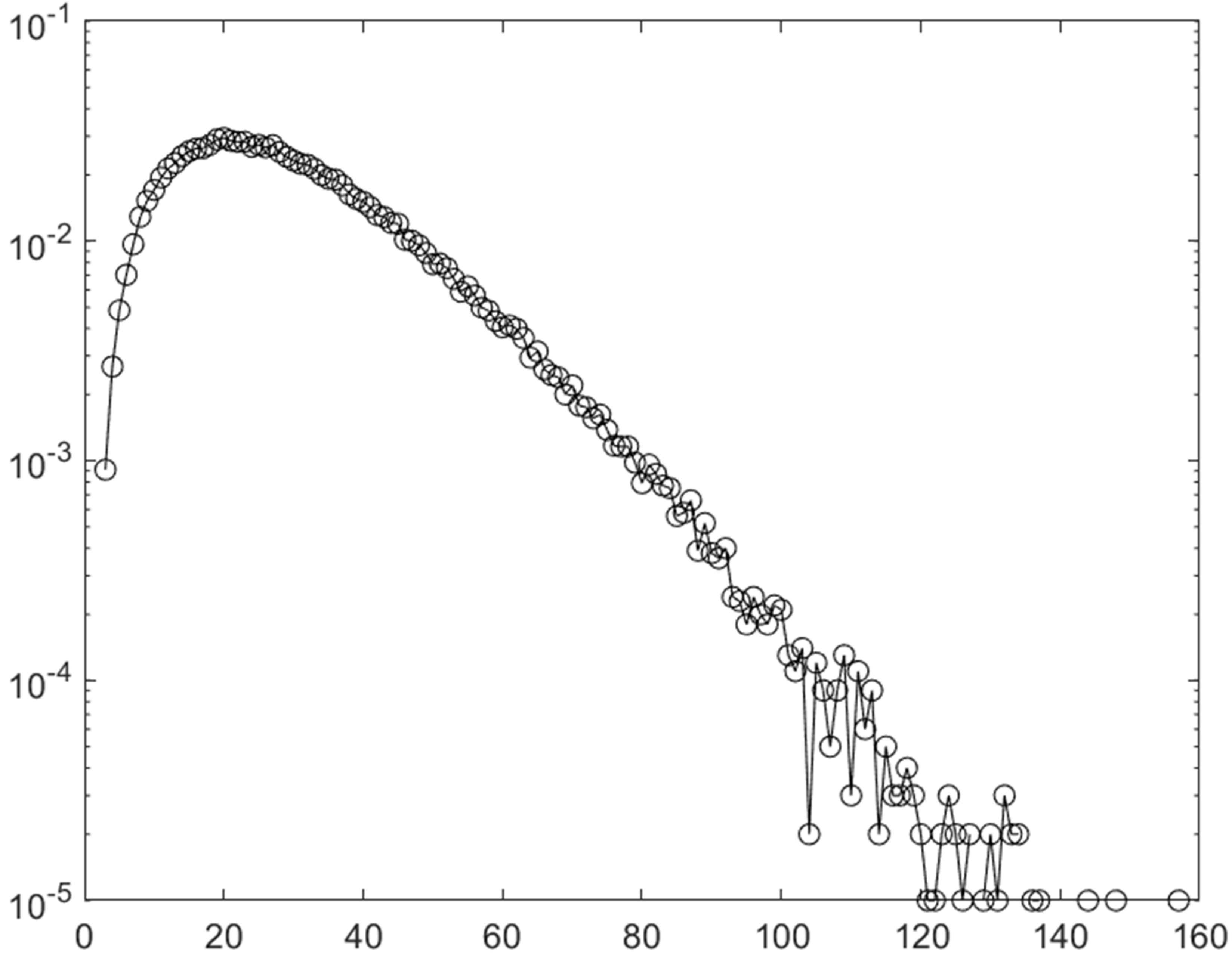
- Estimate mean, variance, and PMF based on 100,000 random variables drawn from a negative binomial distribution
with p=0.1, r=3

- Repeat with negative binomial distribution with p=0.1, r=100

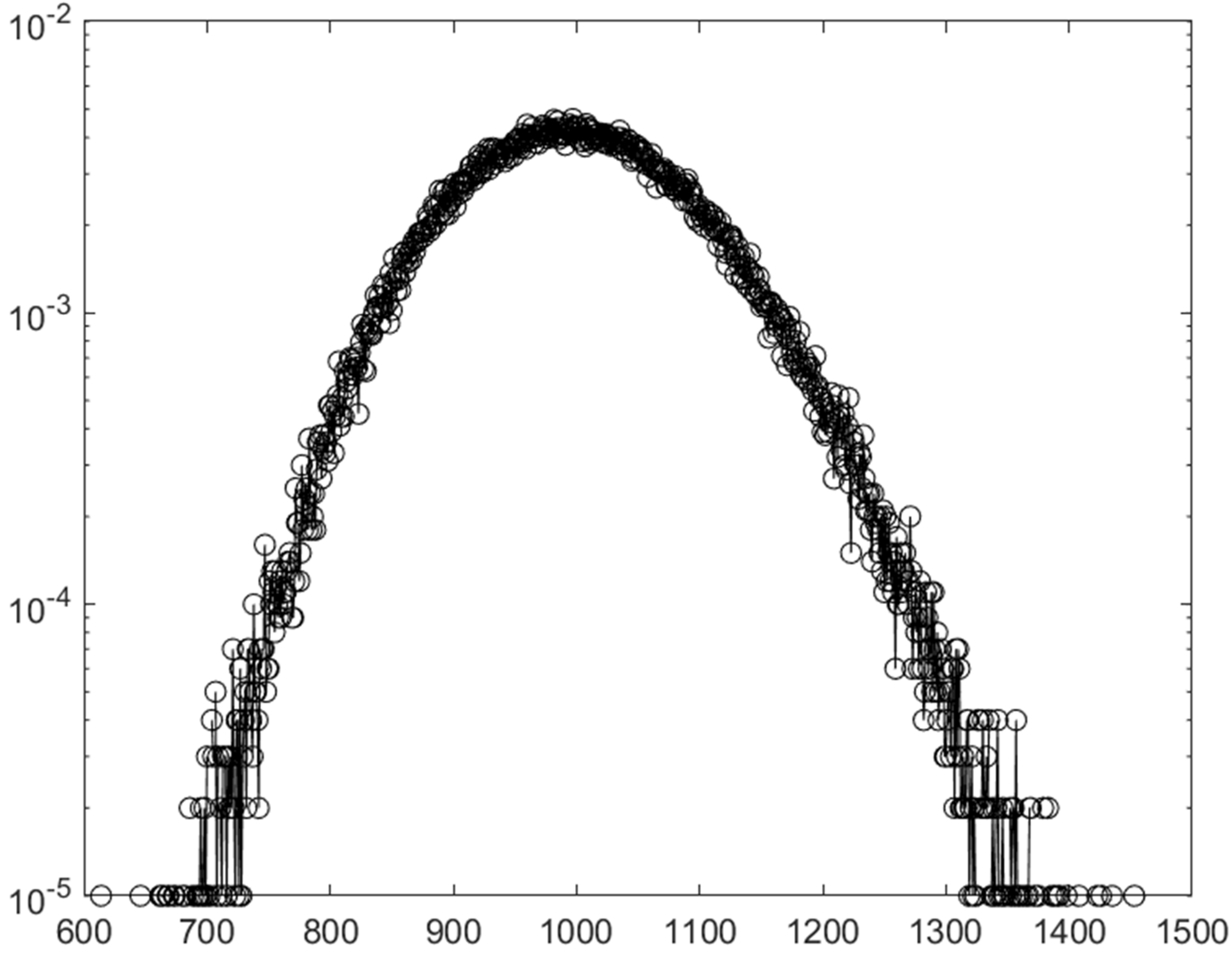
Matlab: Negative binomial distribution

```
Stats=100000;
r=3; p=0.1;
r2=zeros(Stats,1);
for k=1:Stats
    n_trials=0;
    n_successes=0;
    while n_successes<r
        if rand<p
            n_successes=n_successes+1;
        end;
        n_trials=n_trials+1;
    end;
    r2(k)=n_trials;
end;
disp('Observed average value’); disp(sum(r2)./Stats);
disp('Expected average value’); disp(r./p);
disp('Observed variance’); disp(sum(r2.^2)./Stats-(sum(r2)./Stats).^2);
disp('Expected variance’); disp(r.*(1-p)./p^2);
[a,b]=hist(r2, 1:max(r2));
p_nb=a./sum(a);
figure; semilogy(b,p_nb,'ko-');
```

Negative binomial PMF, p=0,1 r=3



Negative binomial PMF, p=0,1 r=100



Cancer is scary!

- Approximately 40% of men and women will be diagnosed with cancer at some point during their lifetimes (source: NCI website)

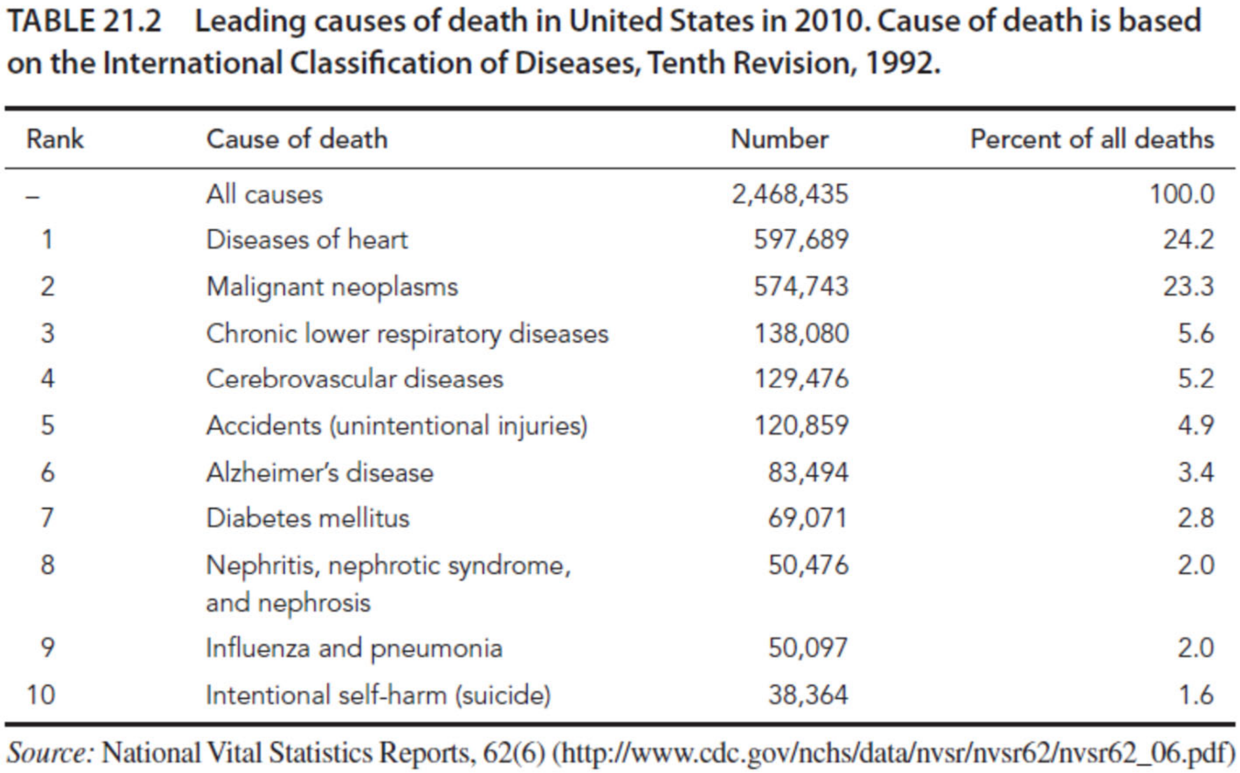


TABLE 21.2 **Leading causes of death in United States in 2010. Cause of death is based on the International Classification of Diseases, Tenth Revision, 1992.**

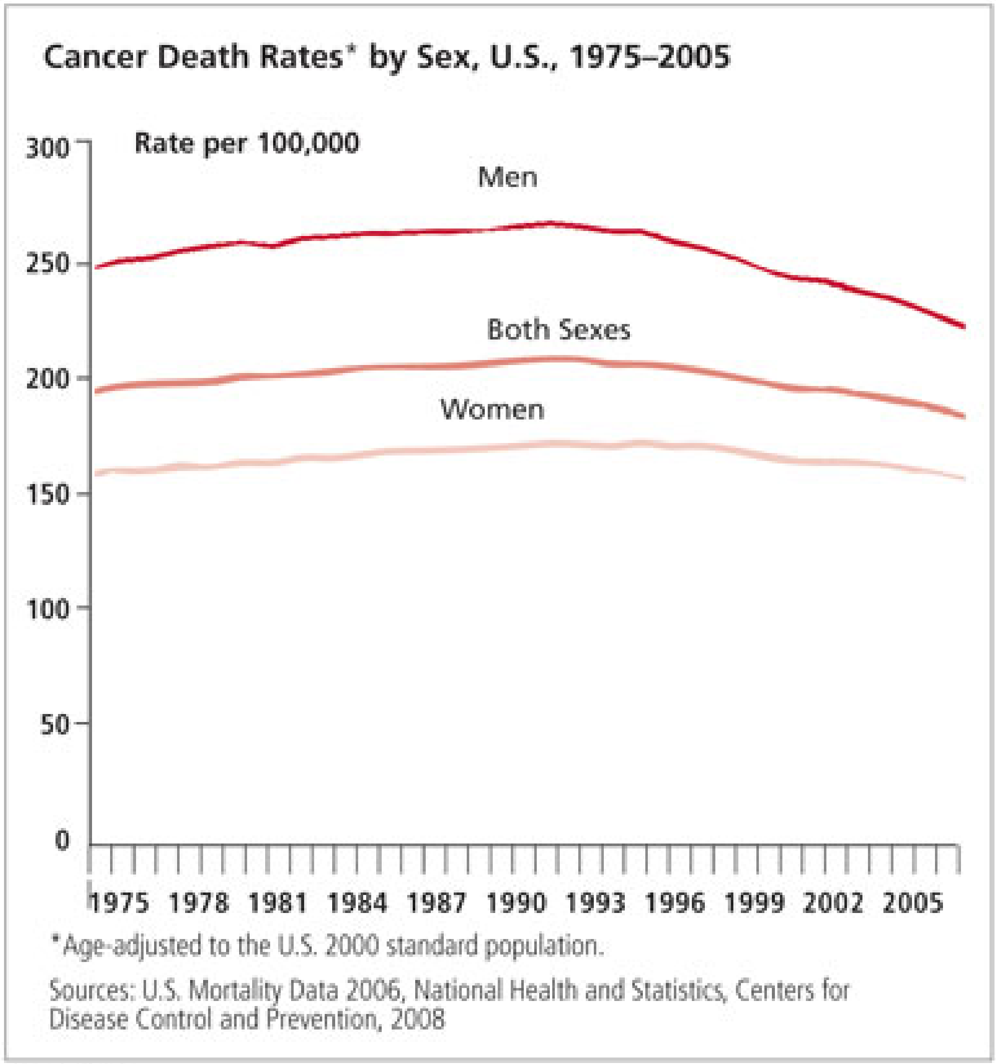
Rank	Cause of death	Number	Percent of all deaths
–	All causes	2,468,435	100.0
1	Diseases of heart	597,689	24.2
2	Malignant neoplasms	574,743	23.3
3	Chronic lower respiratory diseases	138,080	5.6
4	Cerebrovascular diseases	129,476	5.2
5	Accidents (unintentional injuries)	120,859	4.9
6	Alzheimer's disease	83,494	3.4
7	Diabetes mellitus	69,071	2.8
8	Nephritis, nephrotic syndrome, and nephrosis	50,476	2.0
9	Influenza and pneumonia	50,097	2.0
10	Intentional self-harm (suicide)	38,364	1.6

Source: National Vital Statistics Reports, 62(6) (http://www.cdc.gov/nchs/data/nvsr/nvsr62/nvsr62_06.pdf)

Table from J. Pevsner 3rd edition

- "War on Cancer" – president Nixon 1971.
"Moonshot to Cure Cancer" – vice-president Joe Biden 2016

“War on Cancer” progress report

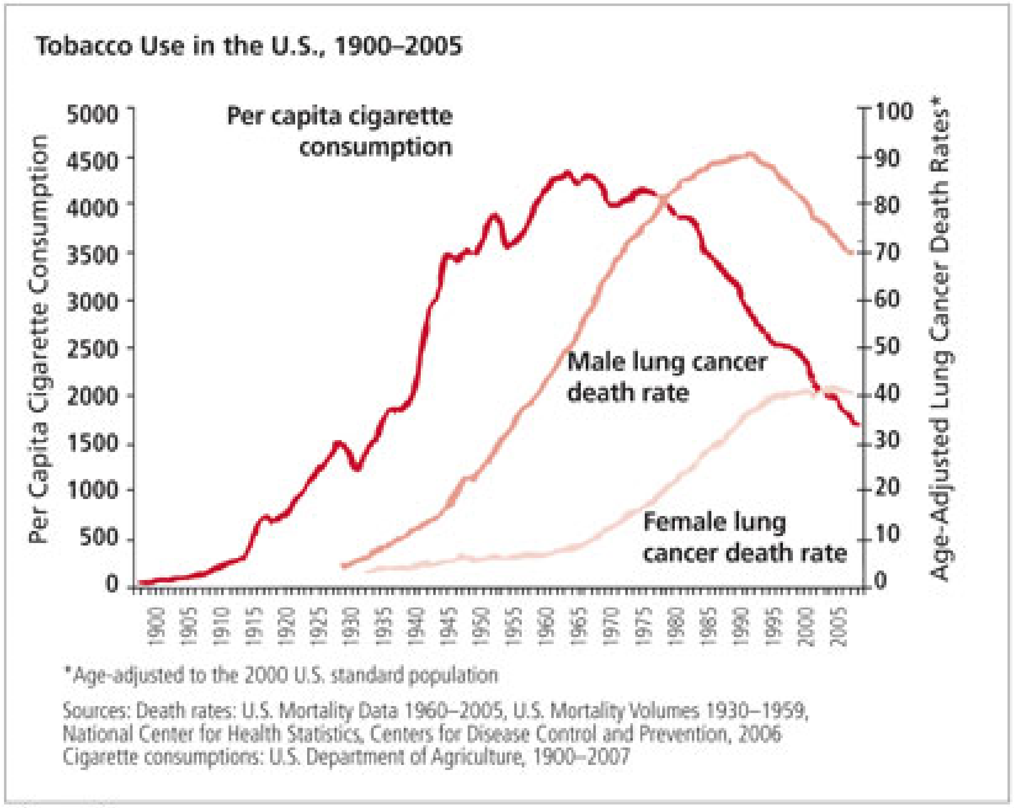


Cancer Death Rates* by Sex, U.S., 1975–2005

*Age-adjusted to the U.S. 2000 standard population.

Sources: U.S. Mortality Data 2006, National Health and Statistics, Centers for Disease Control and Prevention, 2008

Figure 2



Tobacco Use in the U.S., 1900–2005

*Age-adjusted to the 2000 U.S. standard population

Sources: Death rates: U.S. Mortality Data 1960–2005, U.S. Mortality Volumes 1930–1959, National Center for Health Statistics, Centers for Disease Control and Prevention, 2006
Cigarette consumptions: U.S. Department of Agriculture, 1900–2007

Figure 3

Probability theory and statistics

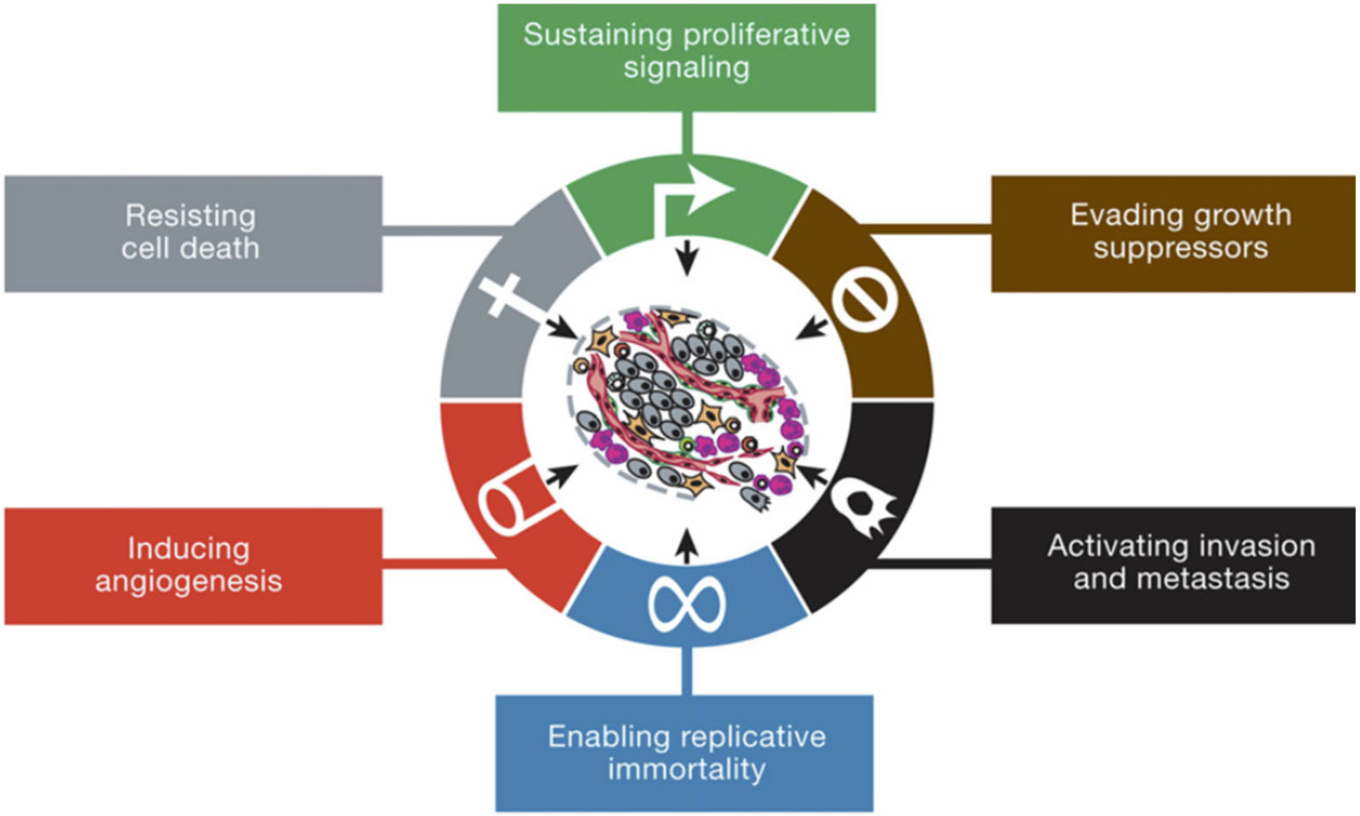
is a powerful tool to

learn new cancer biology

“Driver genes” theory

- Progression of cancer is caused by accumulation of mutations in a handful of “driver” genes
- Mutations in driver genes boost the growth of a tumor
- Oncogenes: expression needs to be elevated for cancer
- Tumor suppressors (e.g. p53) need to be turned off in cancer

Douglas Hanahan and Robert A. Weinberg
Hallmarks of Cancer: The Next Generation
Cell 144, 2011


Sustaining proliferative signaling
Evading growth suppressors
Activating invasion and metastasis
Enabling replicative immortality
Inducing angiogenesis
Resisting cell death

Statistics of cancer incidence vs age

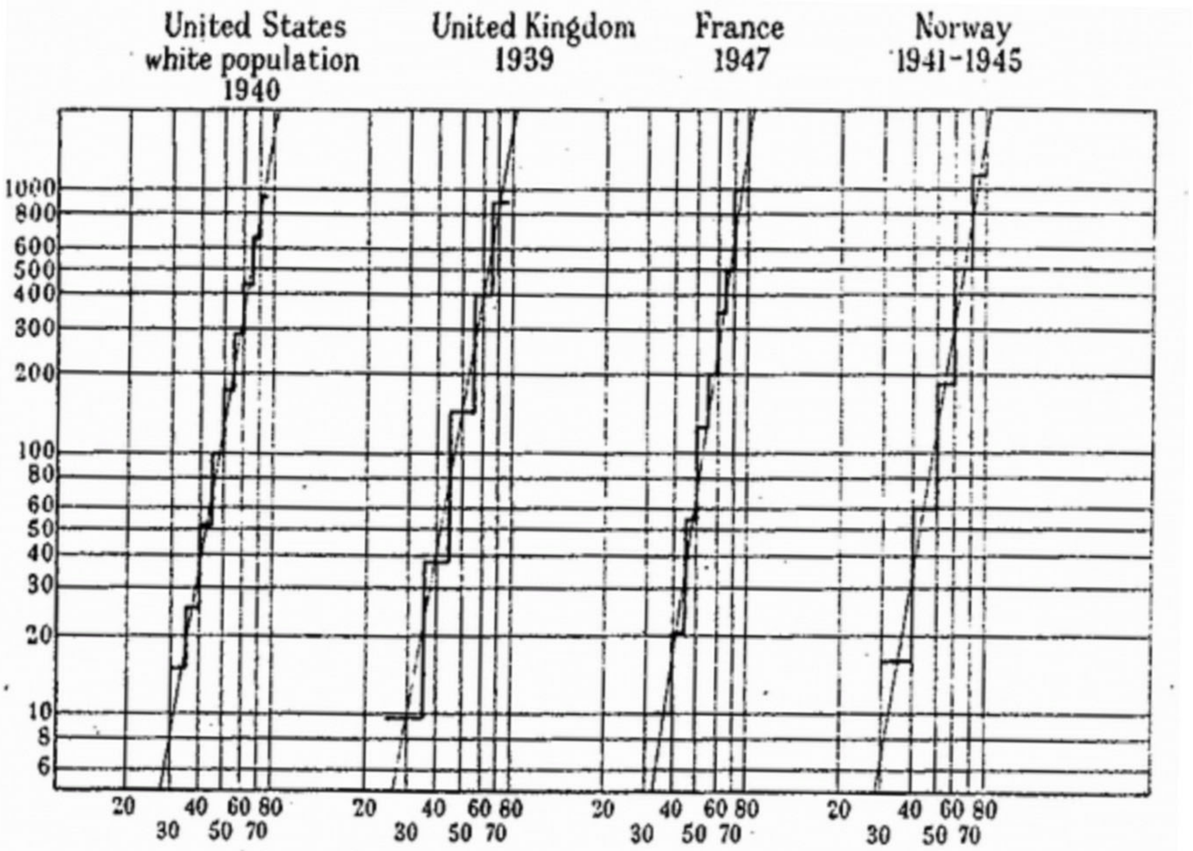


FIG. 1.—Diagram drawn to double logarithmic (log/log) scale showing the cancer death-rate (in the case of the United Kingdom, the carcinoma death-rate) in males at different ages. Deaths per 100,000 males are shown on the vertical scale, age figures on the horizontal scale.

Multi-mutation theory of cancer:
Carl O. Nordling (British J. of Cancer, March 1953):

Cancer death rate
~ (patient age)^6

It suggests the existence of
k=7 driver genes

$$P(T_{cancer} \le t) \sim (u_1 t)(u_2 t)..(u_k t) \sim u_1 u_2..u_k\, t^k$$

$$P(T_{cancer} = t) \sim \frac{d}{dt}(u_1 t)(u_2 t)..(u_k t) \sim k\, u_1 u_2..u_k\, t^{k-1}$$