# How do we use data science in EEB?

## (and the sciences at large)?

Dr. Tomomi Parins-Fukuchi
tomo.fukuchi@utoronto.ca

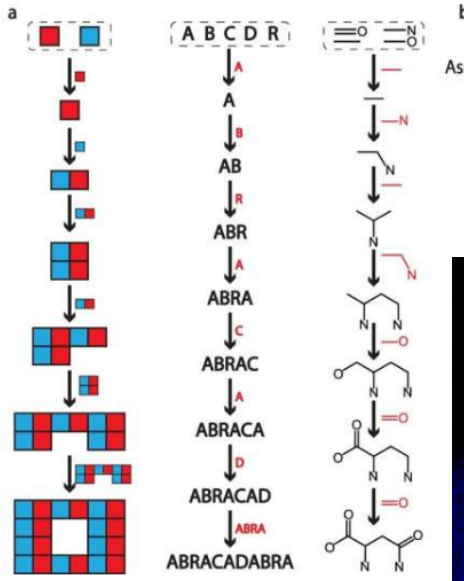
# **Can we better understand** the potential for life on other planets?

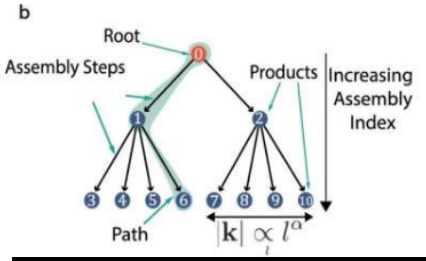
Home / Astronomy & Space / Astrobiology

MAY 24, 2021

## Complex molecules could hold the secret to identifying alien life

by University of Glasgow



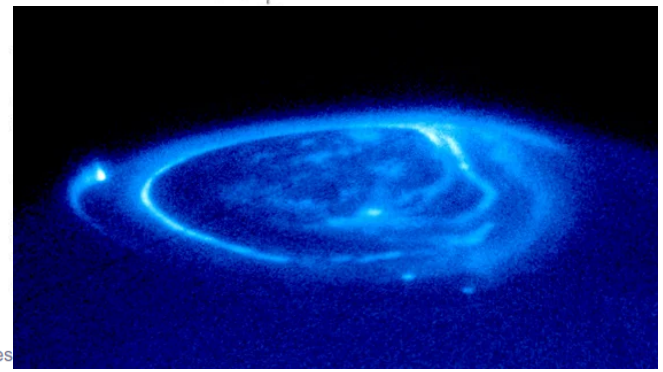



Fig. 1: Assembly pathways. From: Identifying molecules

"...the team used their method to assign MA numbers to a database containing about **2.5 million molecules**..."

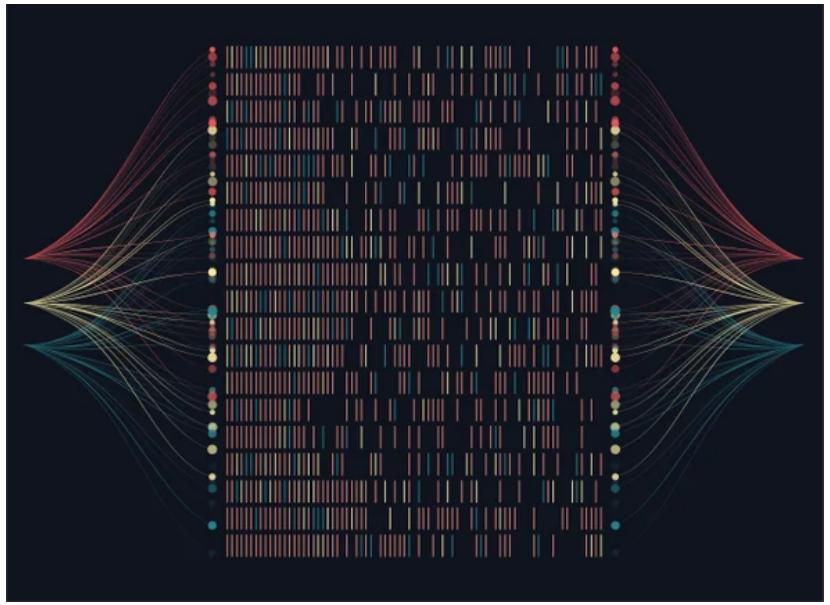
## **Can we better understand** human neurological diversity?

NEUROLOGY | OPINION

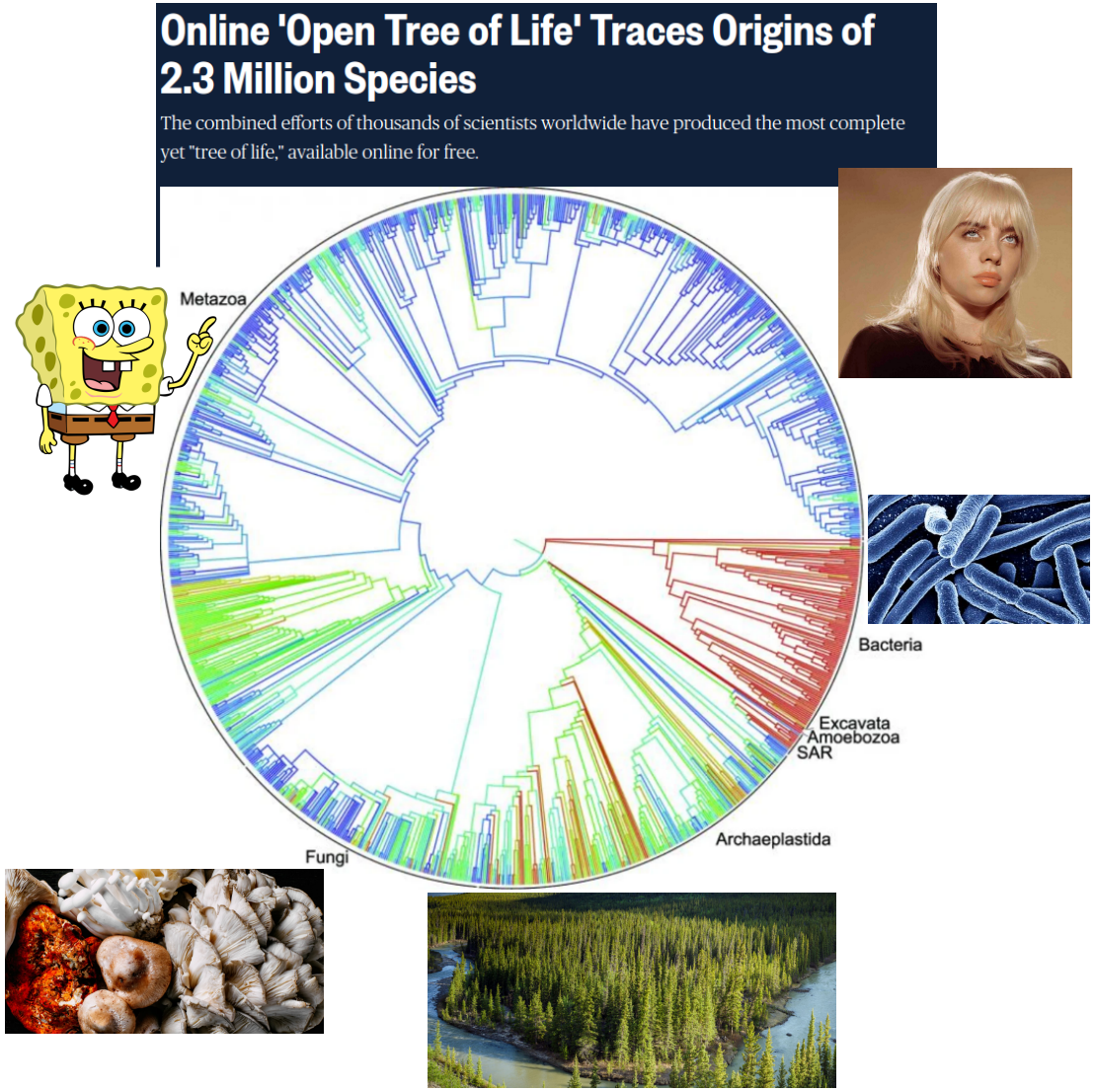
# How Big Data Are Unlocking the Mysteries of Autism

Better genetic insights can help support people across the spectrum

By Wendy Chung on April 30, 2021



# **Can we better understand** the shared history across all life?



# **Can we better understand** emotional expressiveness in music?

Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv.org > physics > arXiv:2103.16737

Search... All fields Search

Help | Advanced Search

**Physics > Popular Physics**

*[Submitted on 31 Mar 2021]*

## I Knew You Were Trouble: Emotional Trends in the Repertoire of Taylor Swift

Megan Mansfield, Darryl Seligman

As a modern musician and cultural icon, Taylor Swift has earned worldwide acclaim via pieces which predominantly draw upon the complex dynamics of personal and interpersonal experiences. Here we show, for the first time, how Swift's lyrical and melodic structure have evolved in their representation of emotio volume of the relevant discography, and that uniquely identifying a song that optimally describes a hy song, we separate the criteria into the level of optimism ($H$) and the strength of commitment to a rel the entire repertoire. We find an overall trend toward positive emotions in stronger relationships, with happiness ($H$) within individual albums over time. The mean relationship score ($R$) shows trends wh with blue eyes and/or bad reputations may lead to overall less positive emotions, while those with gr that these trends are based on small sample sizes, and more data are necessary to validate them. F specific mood.

Comments: 11 pages, 8 figures. Submitted to Acta Prima Aprila. taylorswift code available at this http URL
Subjects: **Popular Physics (physics.pop-ph)**; Earth and Planetary Astrophysics (astro-ph.EP)
Cite as: arXiv:2103.16737 **[physics.pop-ph]**
(or arXiv:2103.16737v1 **[physics.pop-ph]** for this version)

**Submission history**

From: Megan Mansfield [view email]
**[v1]** Wed, 31 Mar 2021 00:21:15 UTC (286 KB)

**Download:**

- PDF
- Other formats



Current browse context:
**physics.pop-ph**

< prev | next >

The Chicago Maroon

GREY CITY / July 24, 2021 / 2:27 p.m.

## *"Sad, Beautiful, Tragic": UChicago Researchers Analyze the Emotional Range of Taylor Swift's Music*

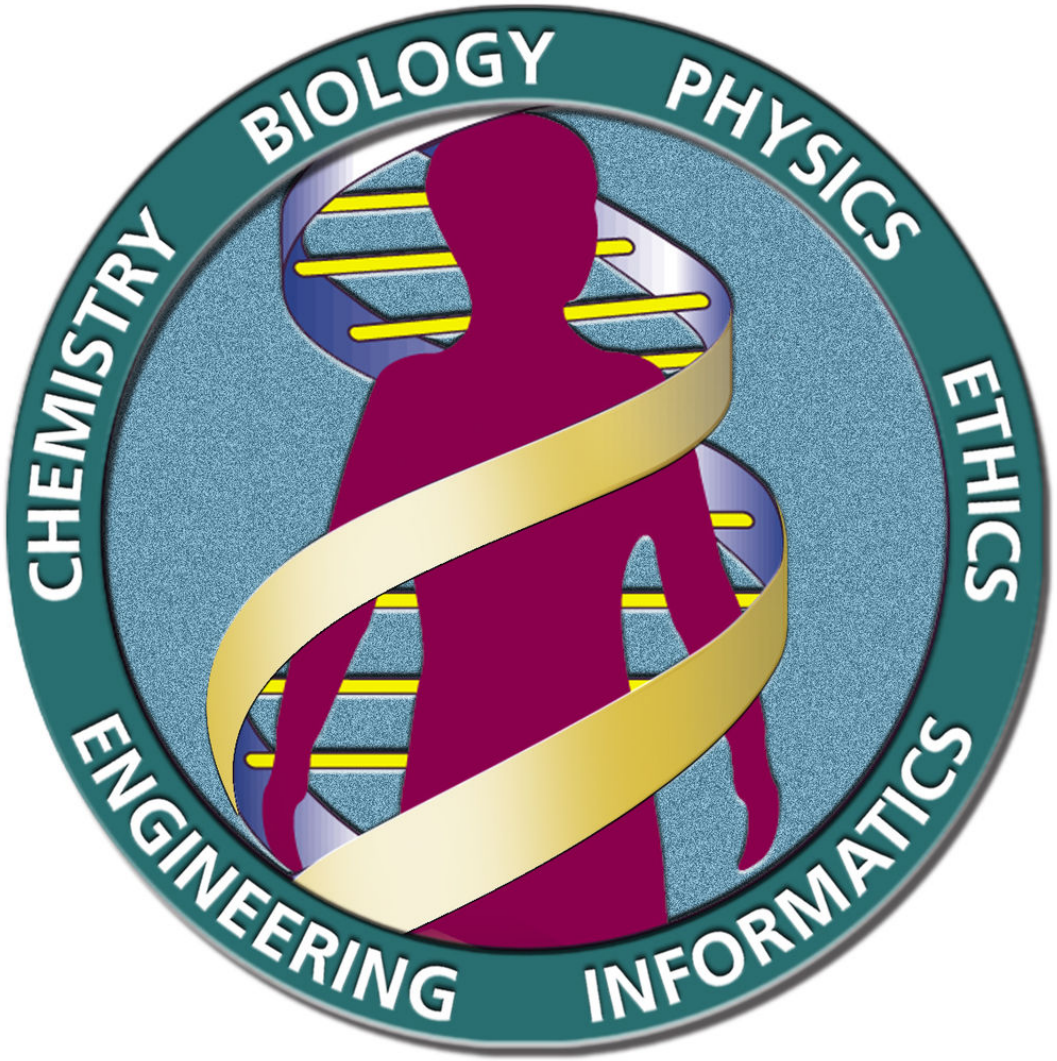
These are all **data science** questions

These are all **data science** questions

Data science has *revolutionized* biology

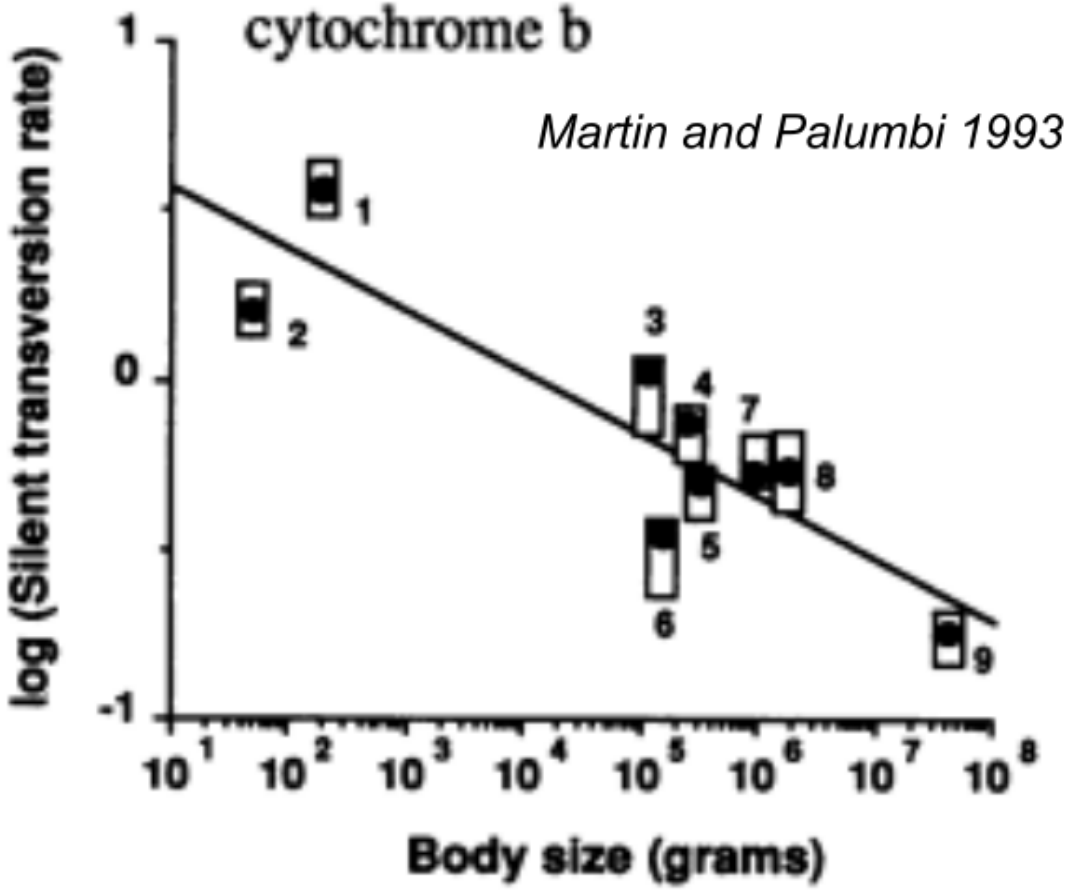
## The Human Genome Project (1990-2003)

- Reshaped everything:
  - biology
  - medicine
  - psychology
  - computation
  - statistics



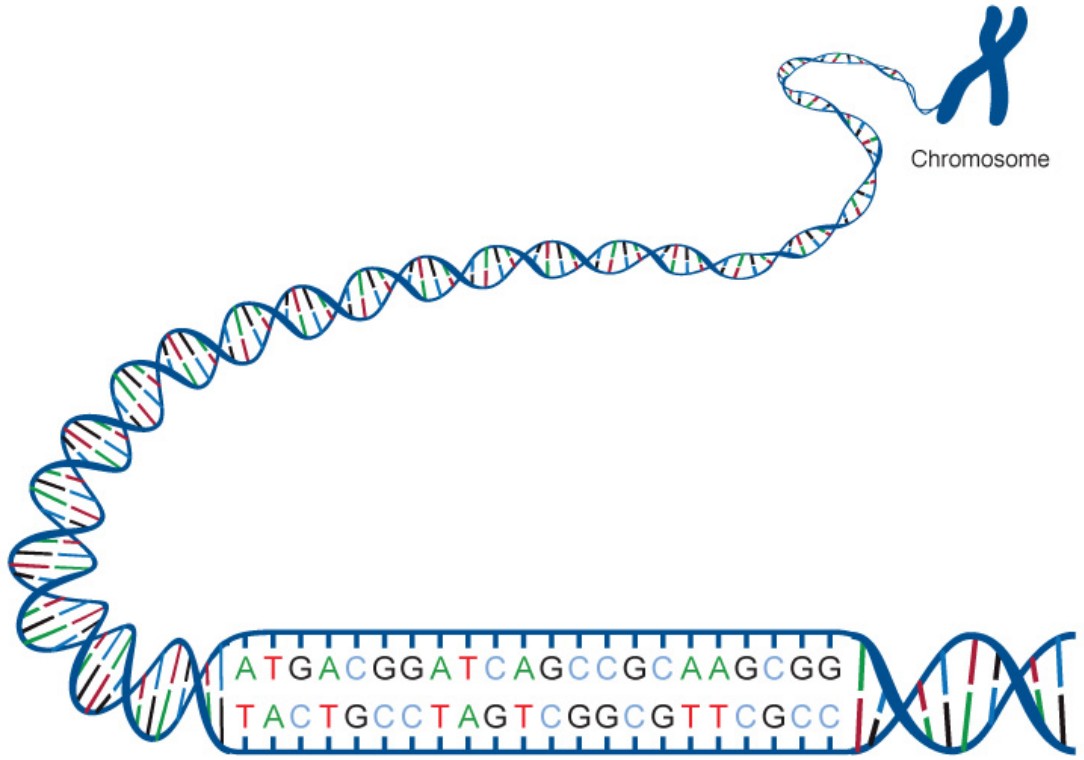
## Scientific datasets have become massive

The field used to focus mostly on targetted questions using small datasets



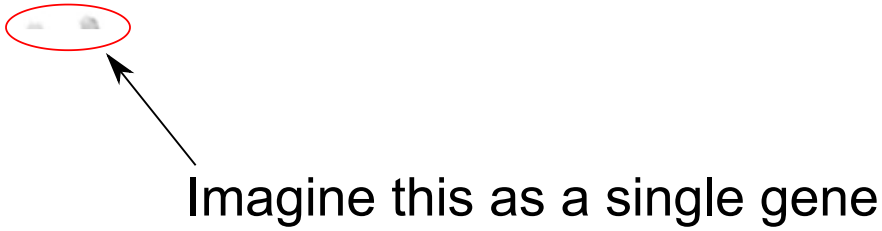
## 'Genetics' -> 'Genomics'

- 'Genes' are comprised of nucleotides
- **Genome**: All nucleotides possessed by an individual



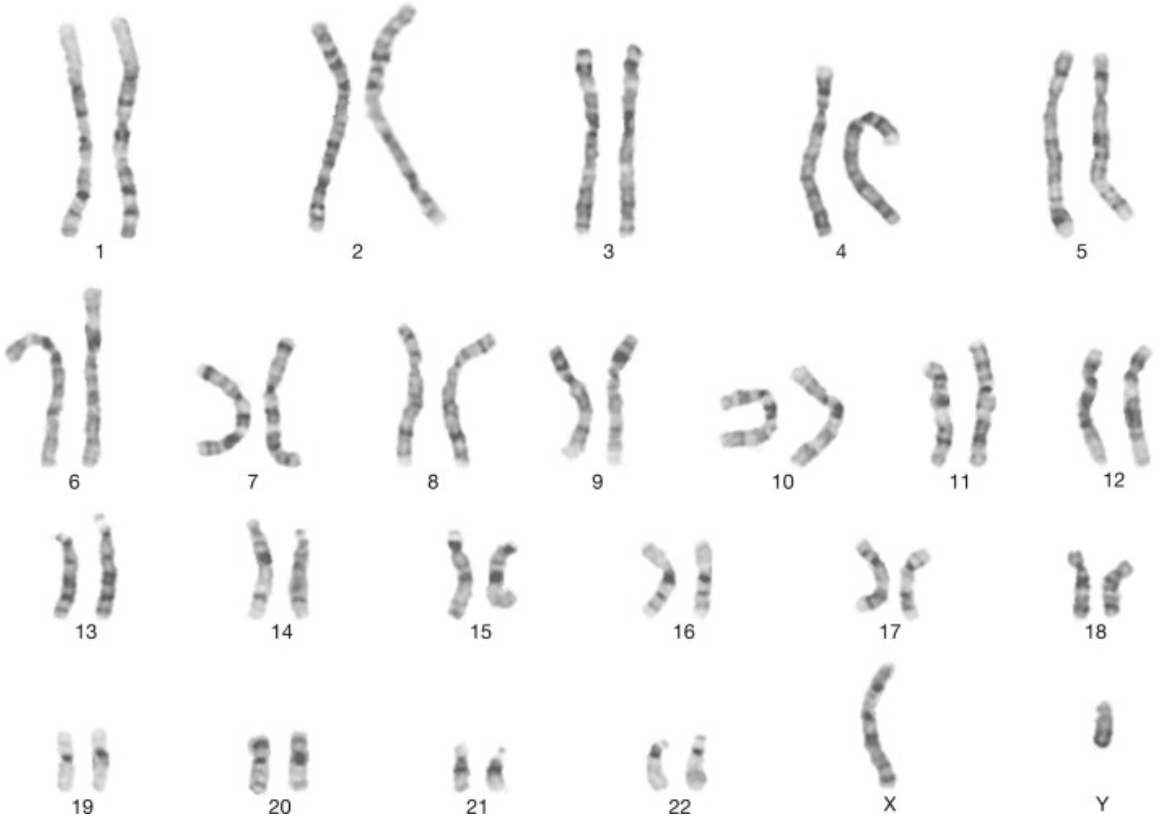
**Pre-2003:**

- Analyze ~3 **thousand** nucleotides at once



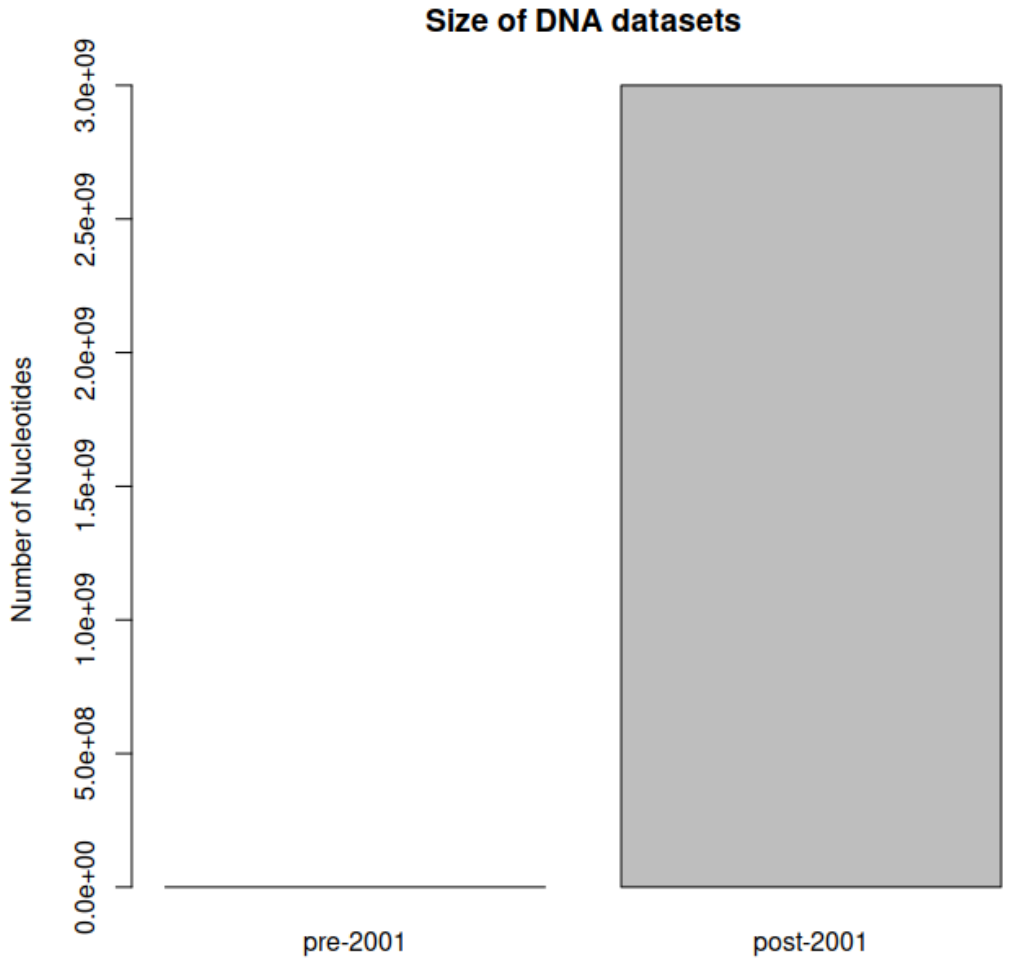
## Post-2003:

- Analyze ~3 **billion** nucleotides at once



That is a **million times** bigger.

That is a **million times** bigger.



These discoveries depend on new approaches for dealing with data

These discoveries depend on new approaches for dealing with data

We often call this mix of new approaches "data science"

## How do we practice data analysis?

- Develop new statistical approaches

Statistics/Math

Computation

## How do we practice data analysis?

- Develop new statistical approaches
- Write computer code to analyze data using stats

Statistics/Math

Computation

## How do we practice data analysis?

- Develop new statistical approaches
- Write computer code to analyze data using stats
- Apply code to address scientific questions

Statistics/Math

Computation

**Most of the science you encounter in your daily life is DataSci!**

## DataSci breakthroughs

We have learned a lot about human biology using DataSci

NEWS | 22 August 2018 | Correction 23 August 2018

# Mum's a Neanderthal, Dad's a Denisovan: First discovery of an ancient-human hybrid

Genetic analysis uncovers a direct descendant of two different groups of early humans.

Matthew Warren

Denny inherited one set of chromosomes from her Neanderthal ancestors, depicted in this model. Credit: Christopher Rynn/University of Dundee

A female who died around 90,000 years ago was half Neanderthal and half Denisovan, according to genome analysis of a bone discovered in a Siberian cave. This is the first time scientists have identified an ancient individual whose parents belonged to distinct human groups. The findings were published on 22 August in *Nature*[1].

## DataSci breakthroughs

We have learned a lot about human biology using DataSci

# Our Neanderthal genes linked to risk of depression and addiction

LIFE 11 February 2016

By Colin Barras

**Having sex with Neanderthals meant some of us still carry their DNA, and with it, a higher risk of depression and nicotine addiction**
Nikola Solic/Reuters

## Dealing with data is hard

Neanderthal genes are *risk-inducing* for severe COVID?

Article | Published: 30 September 2020

# The major genetic risk factor for severe COVID-19 is inherited from Neanderthals

Hugo Zeberg ✉ & Svante Pääbo ✉

*Nature* **587**, 610–612 (2020) | Cite this article

## Dealing with data is hard

Neanderthal genes are *protective* against severe COVID?

RESEARCH ARTICLE

# A genomic region associated with protection against severe COVID-19 is inherited from Neandertals

Hugo Zeberg and Svante Pääbo

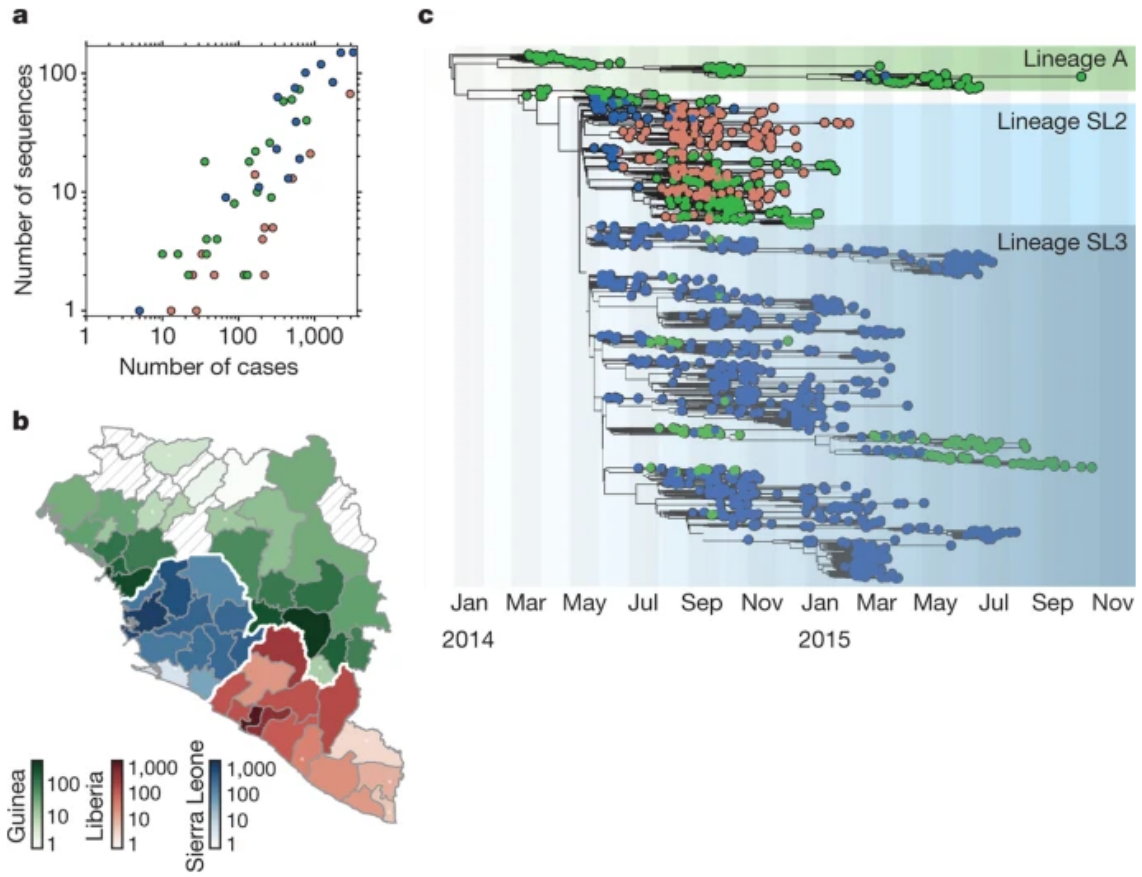
+ See all authors and affiliations

PNAS March 2, 2021 118 (9) e2026309118; https://doi.org/10.1073/pnas.2026309118

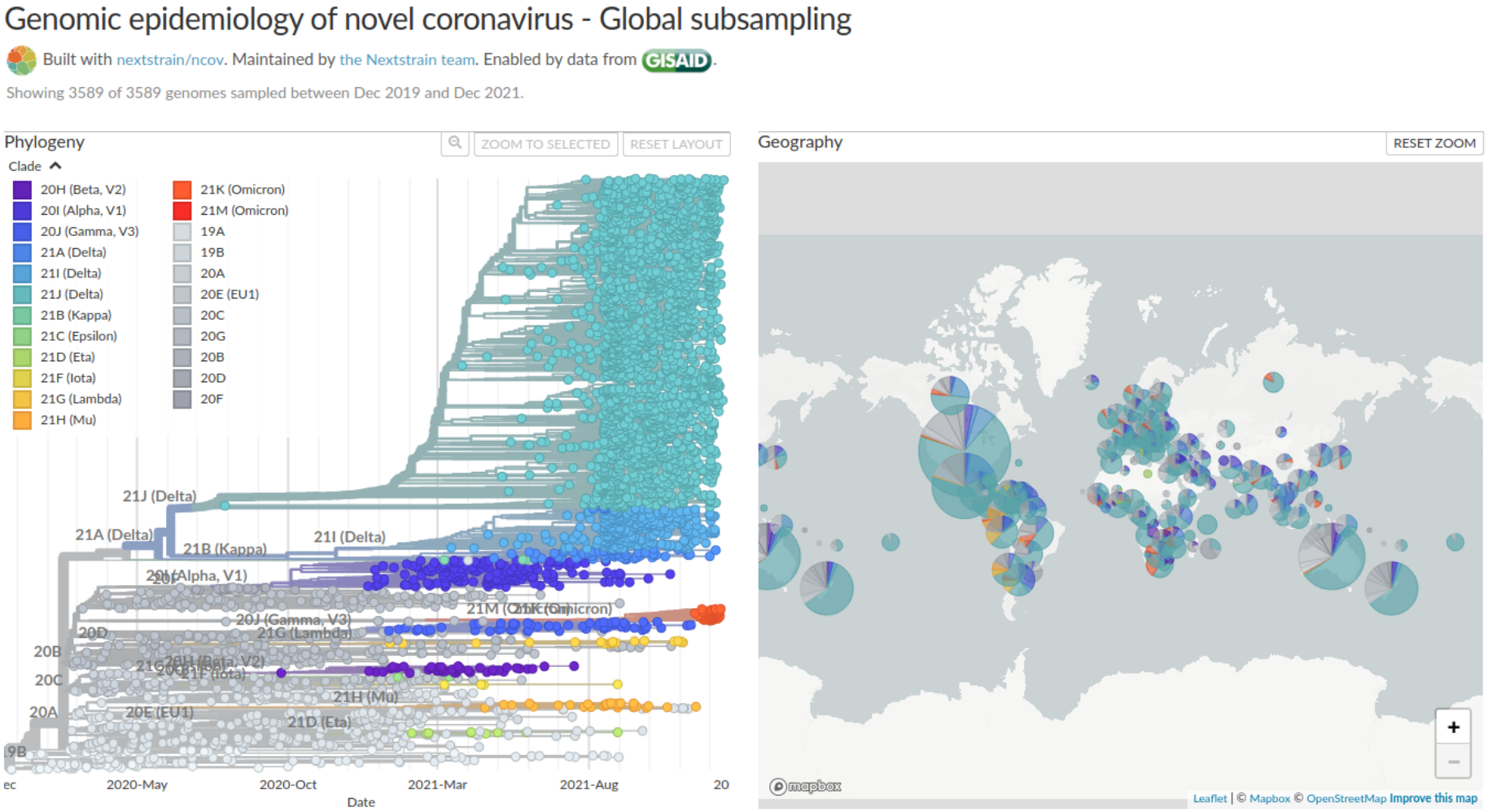
Contributed by Svante Pääbo, January 22, 2021 (sent for review December 21, 2020; reviewed by Tobias L. Lenz and Lluis Quintana-Murci)

# DataSci helps us understand human health

**Figure 1: Evolution of EBOV during the 2013–2016 outbreak showing the extent and location of virus sampling.**



# DataSci helps us understand human health



# DataSci tools are also marketed to the public

23andMe®

ANCESTRY BREAKDOWN

## Dig deeper into your ancestry.

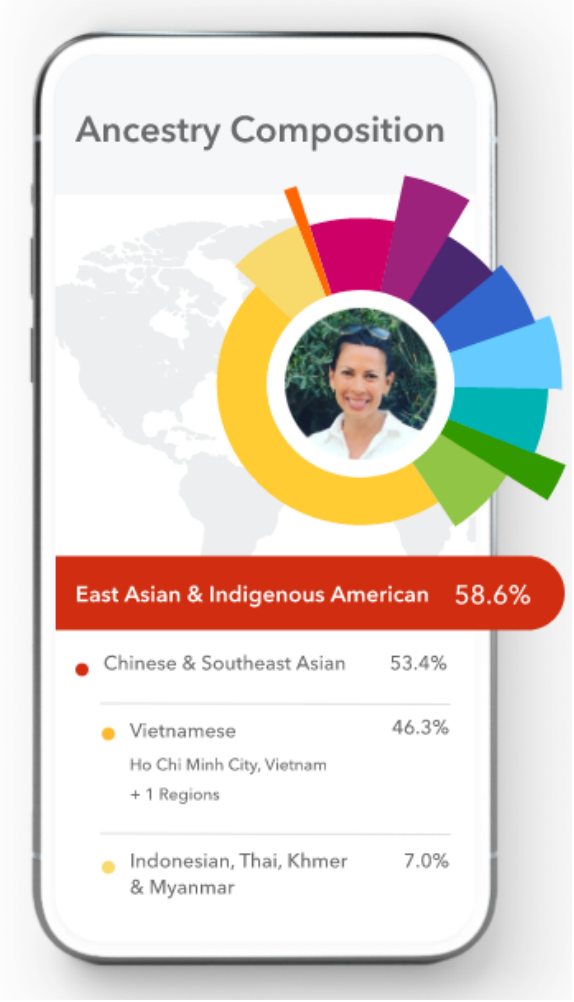
It's the most complete genetic breakdown on the market, and the most comprehensive portrait of you yet.

- **Ancestry Composition**

  Discover where in the world your DNA is from across 2000+ regions — in some cases, down to the county level.

- **Ancestry Detail Report**

See all regions



## How does DataSci fit into the life sciences?

- The influx of new data has changed the landscape of research

## How does DataSci fit into the life sciences?

- The influx of new data has changed the landscape of research
- Understanding these tools is *essential* to understanding modern science

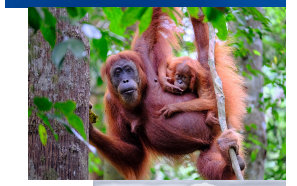
## Who am I?

- Evolutionary biologist
- I develop new computational approaches to ask 'big picture' questions about evolution
- For example...

# A population of small, burrowing mammals gave rise to all of this ecological diversity

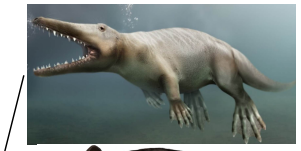
KPg Boundary (66mya)

Present



# ...in just a few million years

KPg Boundary (66mya)



Aquatic
Massive size
Swimming locomotion

Grasping
Enhanced vision
Long lifespan

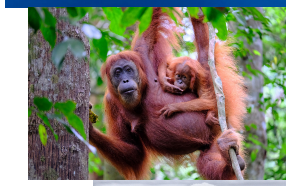
Carnivory traits
Large size
Locomotor

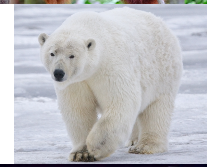
Echolocation
Flight

Hooves
Long legs
Grazing traits

Carnivory
Large size
Walk on toes

Present





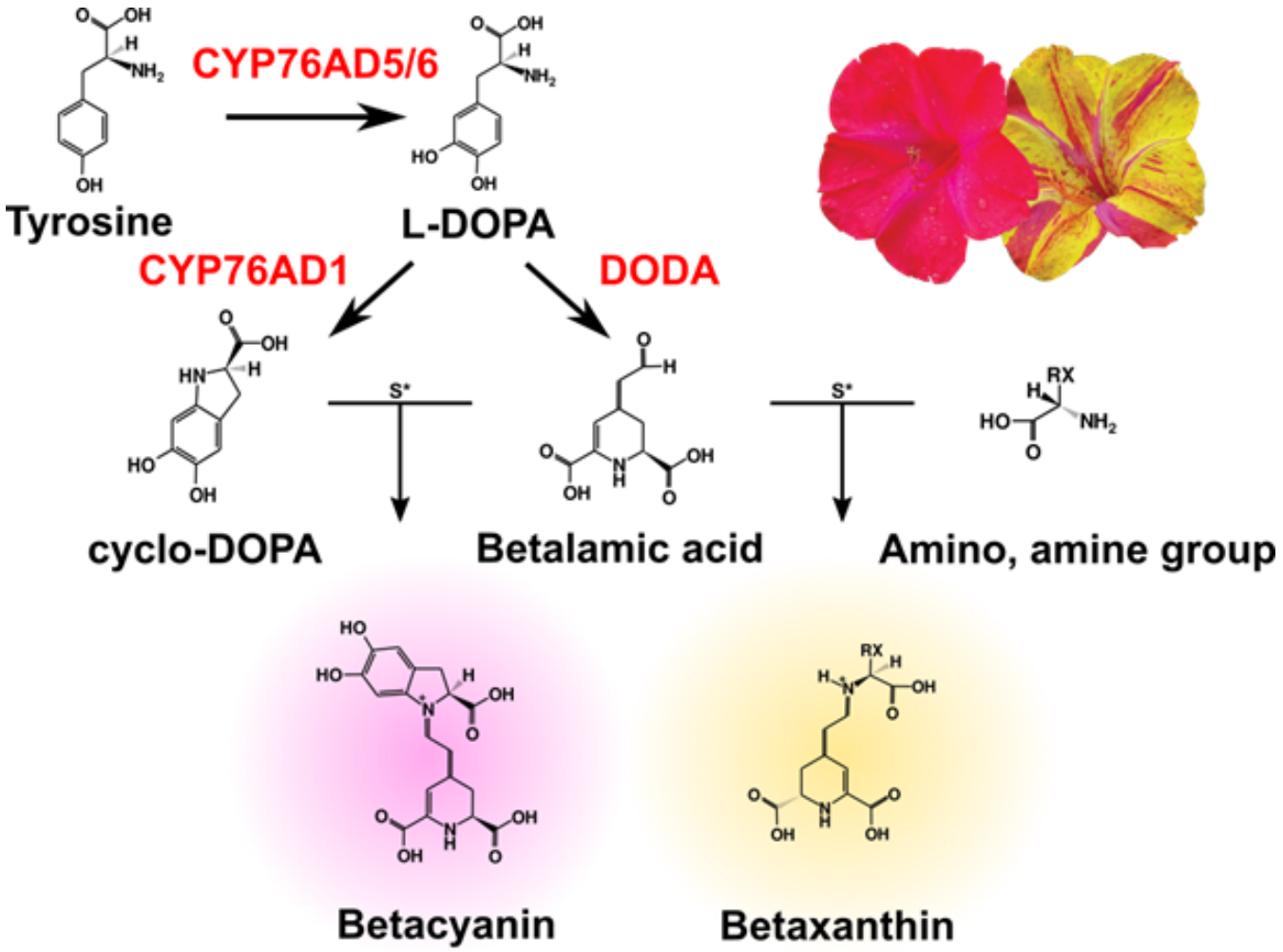
# Most biologists are interested in the origin of cool features

# Most biologists are interested in the origin of cool features

- flowers

# Most biologists are interested in the origin of cool features

- novel biochemical pathways (C4, betalain, etc)



## Most biologists are interested in the origin of cool features

- skeletal features associated with novel locomotor modes

# Apes

Ape skeletons display many locomotor innovations

# Skeletal features display complex functional and genetic relationships

## Computational Evo Bio

- This requires data science approaches to reconstruct complex patterns in:
  - Morphological traits
  - Genomes
  - Ecological information

## Computational Evo Bio

- Computational skills empower us as scientists
- Allow us to reconstruct complex patterns and test hypotheses in creative and novel ways