

# Introduction to computational biology

Daniel Rice  
Chris Poras  
Lauren Blake  
Eric Friedlander

University of Chicago  
20 May 2019

# Computation has four main uses in biology today

- Data analysis
- Visualization
- Simulation
- Communication

# An RNA sequence in 1967

TABLE 5  
*Summary of the evidence for the base sequence of the  
larger nucleotides from ribonuclease  $T_1$  digests*

Spot no. (Plates I & II)	Sequence†
13	<p style="text-align: center;">S ↓</p> <p>→      →      →      →      →</p> <p>— A A C U C A G —</p> <p>→      →      →      →      →</p> <p>— S      S —</p> <p>→      →      →      →      →</p> <p>— P      P —</p>
14	<p style="text-align: center;">S ↓</p> <p>→      →      →      →      →</p> <p>— A C C C C A U G —</p> <p>→      →      →      →      →</p> <p>— S      S —</p> <p>→      →      →      →      →</p> <p>— P      P —</p>
15	<p style="text-align: center;">S ↓</p> <p>→      →      →      →      →</p> <p>— U C C C A C C U G —</p> <p>→      →      →      →      →</p> <p>— S      S —</p> <p>→      →      →      →      →</p> <p>— Sp      —</p> <p>→      →      →      →      →</p> <p>— P      —</p>
16	<p style="text-align: center;">S ↓</p> <p>→      →      →      →      →</p> <p>— U C U C C C C A U G —</p> <p>→      →      →      →      →</p> <p>— S      S —</p> <p>→      →      →      →      →</p> <p>— P      —</p>
17	<p style="text-align: center;">S ↓</p> <p>→      →      →      →      →</p> <p>— C C U U A G —</p> <p>→      →      →      →      →</p> <p>— S      S —</p>

# DNA sequences today are longer

```
CCACACCACACCCACACACCCACACACCACACCACACACCACACCCACACACACA  
CATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTT  
ACCCTCCATTACCCTGCCTCCACTCGTTACCCTGTCCCATTCAACCATAACCACTCCGAAC  
CACCATCCATCCCTCTACTTACTACCACTCACCCACCGTTACCCTCCAATTACCCATATC  
CAACCCACTGCCACTTACCCTACCATTACCCTACCATCCACCATGACCTACTCACCATAC  
TGTTCTTCTACCCACCATATTGAAACGCTAACAAATGATCGTAAATAACACACACGTGCT  
TACCCTACCACTTTATACCACCACCACATGCCATACTCACCTCACTTGTATACTGATTT  
TACGTACGCACACGGATGCTACAGTATATACCATCTCAAACCTTACCCTACTCTCAGATTC  
CACTTCACTCCATGGCCCATCTCTCACTGAATCAGTACCAAATGCACTCACATCATTATG  
CACGGCACTTGCCTCAGCGGTCTATACCCTGTGCCATTTACCCATAACGCCCATCATTAT  
CCACATTTTGATATCTATATCTCATTTCGGCGGTCCCAAATATTGTATAACTGCCCTTAAT  
ACATACGTTATACCACTTTTGCACCATATACTTACCACTCCATTTATATACACTTATGTC  
AATATTACAGAAAAATCCCCACAAAAATCACCTAAACATAAAAAATATTCTACTTTTCAAC  
AATAATACATAAACATATTGGCTTGTGGTAGCAACACTATCATGGTATCACTAACGTAAA  
AGTTCCTCAATATTGCAATTTGCTTGAACGGATGCTATTTTCAGAATATTTTCGTACTTACA  
CAGGCCATACATTAGAATAATATGTCACATCACTGTCGTAACACTCTTTATTCACCGAGC  
AATAATACGGTAGTGGCTCAAACCTCATGCGGGTGCTATGATACAATTATATCTTATTTCC  
ATTCCCATATGCTAACCGCAATATCCTAAAAGCATAACTGATGCATCTTTAATCTTGTAT  
GTGACACTACTCATACGAAGGGACTATATCTAGTCAAGACGATACTGTGATAGGTACGTT  
ATTTAATAGGATCTATAACGAAATGTCAAATAATTTTACGGTAATATAACTTATCAGCGG  
CGTATACTAAAACGGACGTTACGATATTGTCTCACTTCATCTTACCACCCTCTATCTTAT  
TGCTGATAGAACTAACCCCTCAGCTTTATTTCTAGTTACAGTTACACAAAAAACTATG  
CCAACCCAGAAATCTTGATATTTTACGTGTCAAAAAATGAGGGTCTCTAAATGAGAGTTT  
GGTACCATGACTTGTAACCTCGCACTGCCCTGATCTGCAATCTTGTTCTTAGAAGTGACGC  
ATATTCTATACGGCCCGACGCGACGCGCAAAAAATGAAAAACGAAGCAGCGACTCATTT  
TTATTTAAGGACAAAGGTTGCGAAGCCGCACATTTCCAATTTCAATTGTTGTTTATTGGAC  
:
```

DNA sequences today  
are **much** longer



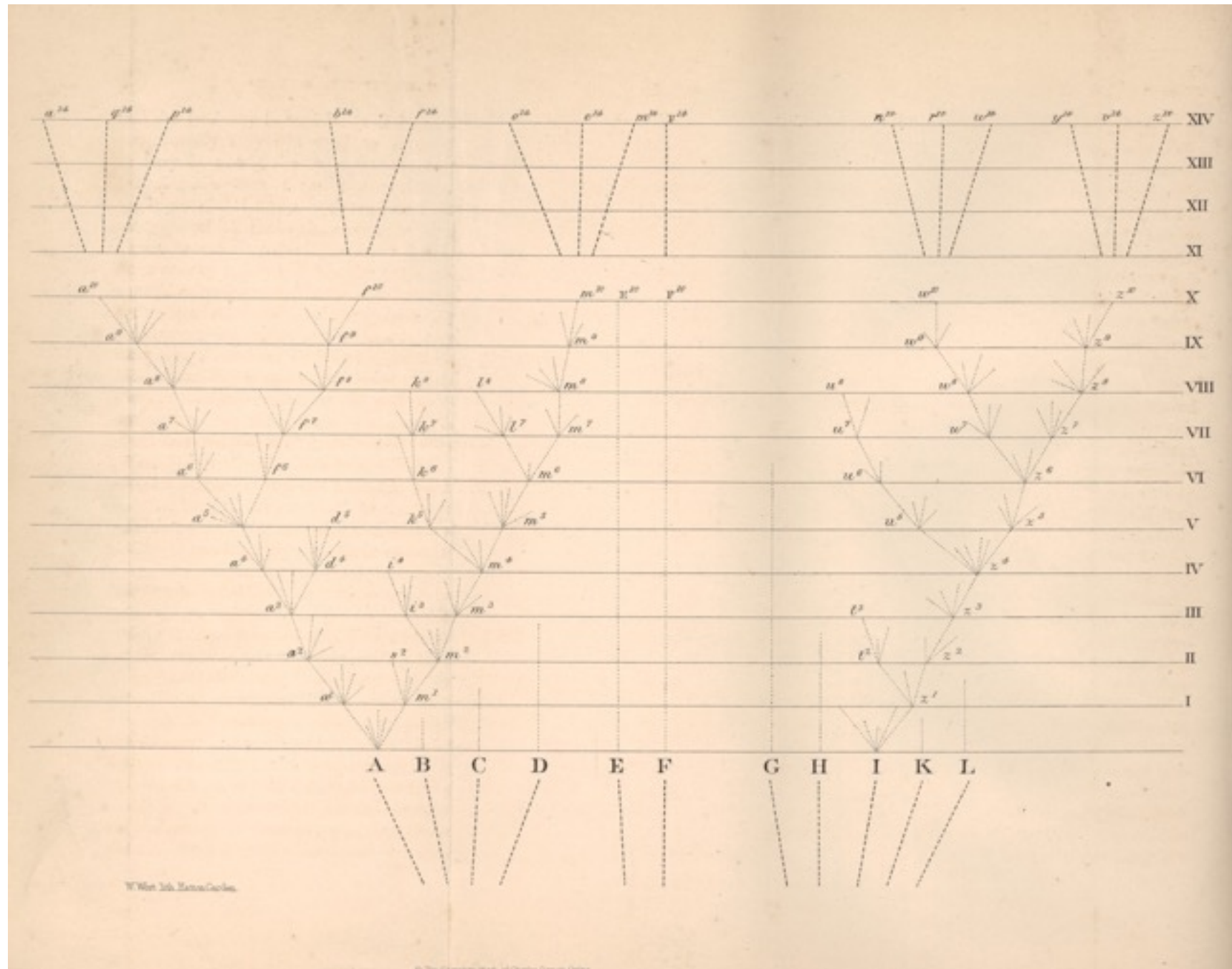
DNA sequences today  
are **much** longer

A large, dark blue rectangular area with a fine, horizontal-line texture. In the center of this area is a white rectangular box containing text.

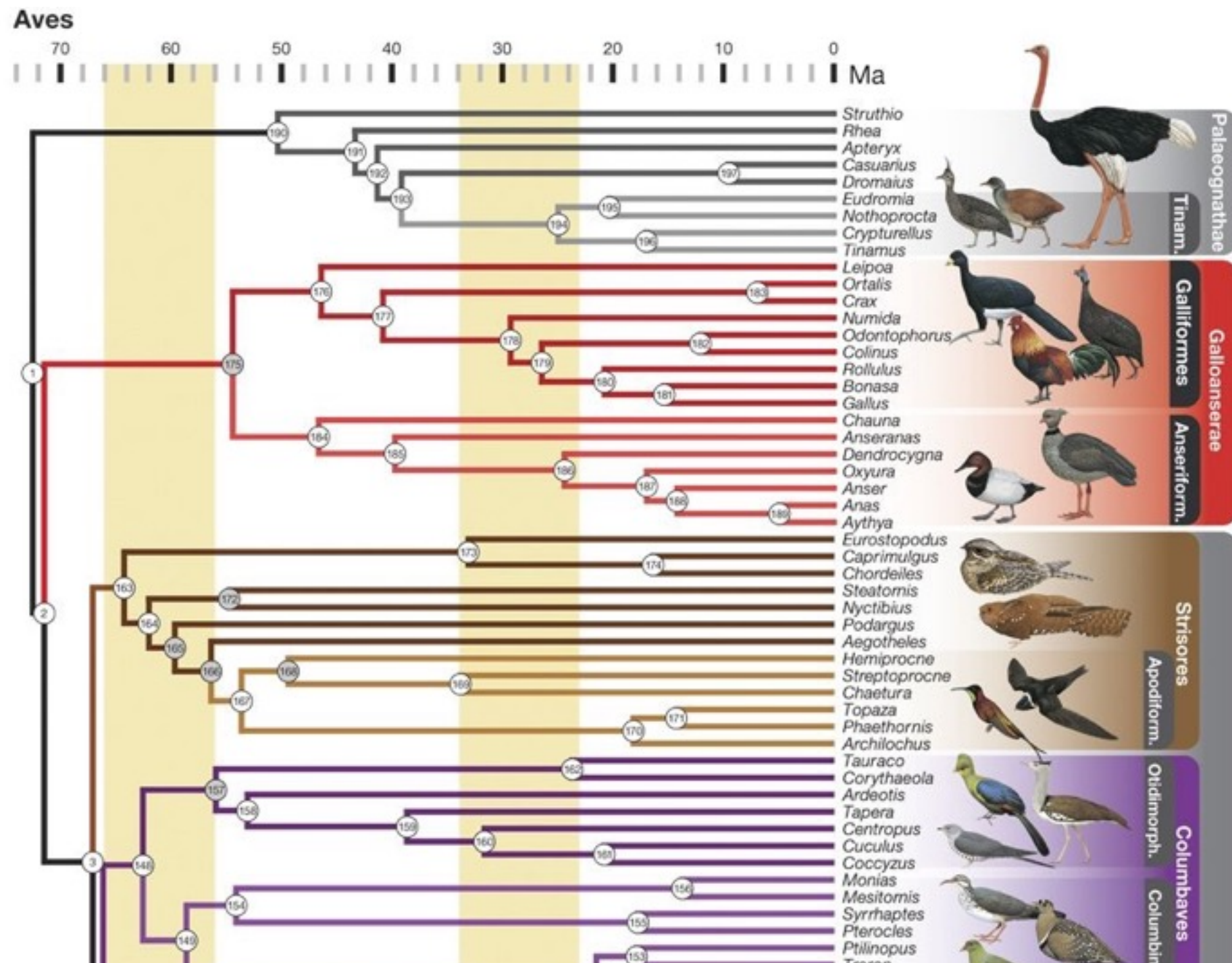
0.000375% of the human genome!



Without computers,  
figures were ... rudimentary

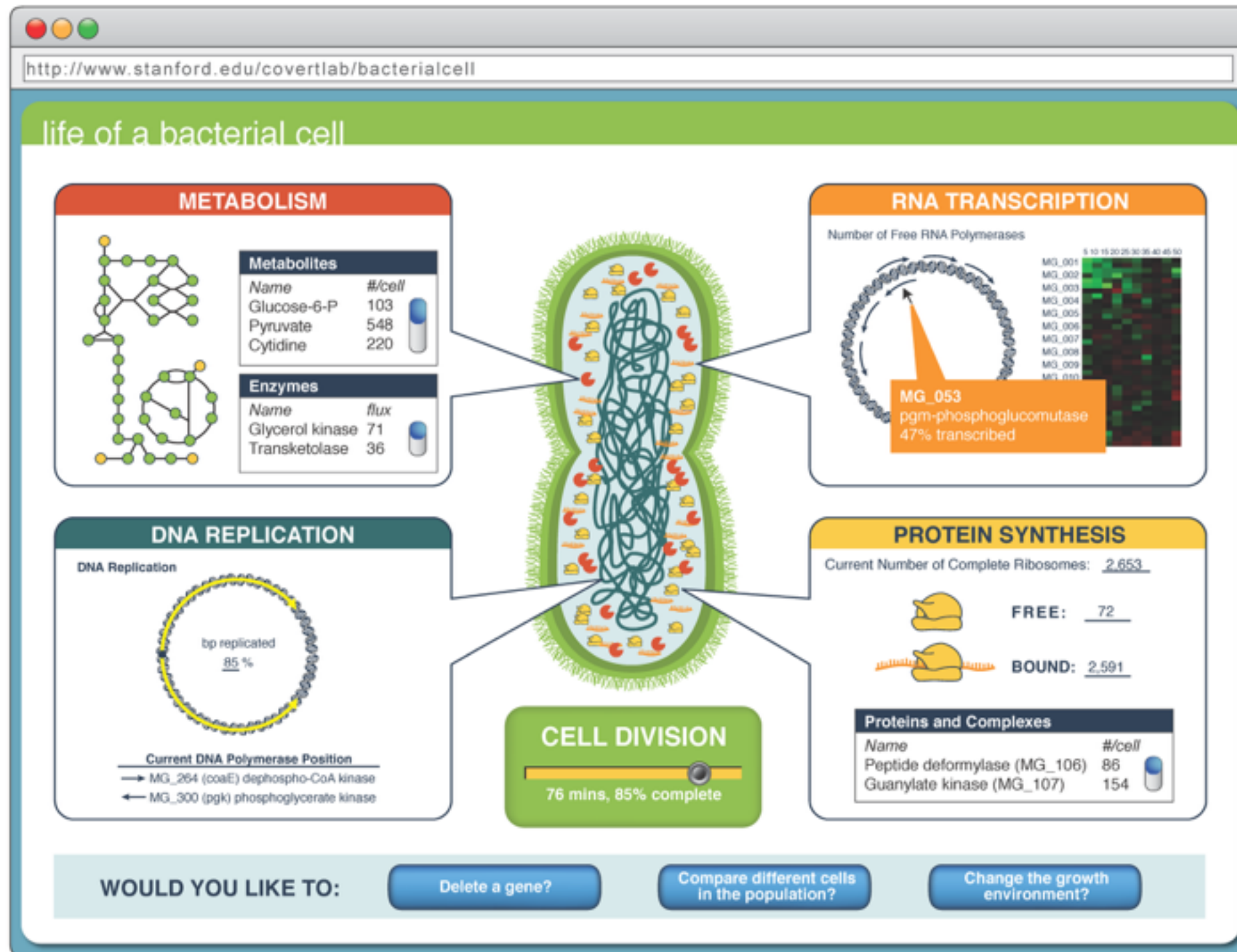


# With computers, we can make sophisticated graphics

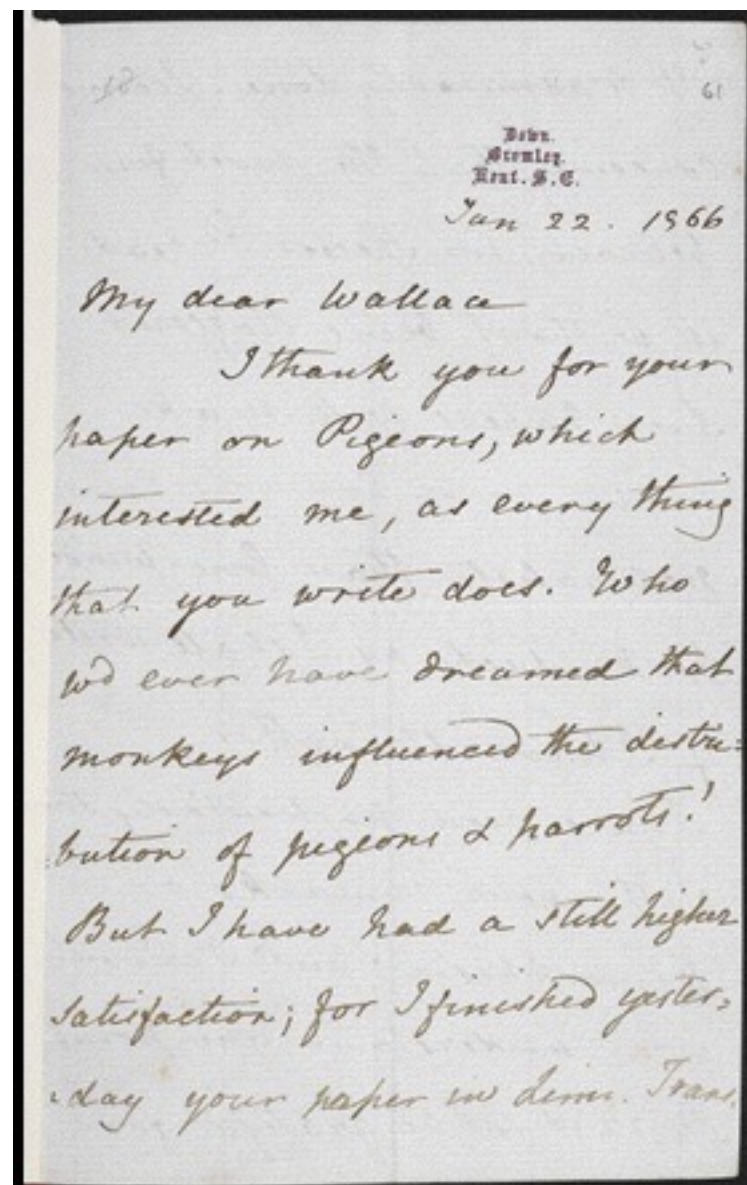




# With computers, we can simulate complex processes




Scientific communication  
used to be difficult





# Scientific communication is much easier now

 **Leo Speidel** @leo\_speidel · Feb 14  
Our preprint is up!  
A method for genome-wide genealogy estimation for thousands of samples: [biorxiv.org/content/10.1101/17](https://biorxiv.org/content/10.1101/17) 1/7



**A method for genome-wide genealogy estimation f...**  
Knowledge of genome-wide genealogies for thousands of individuals would simplify most evolutionary analyses for humans and other species, but has remained com...  
[biorxiv.org](https://biorxiv.org)

1 54 93  
[Show this thread](#)

 John Novembre Retweeted


 **Diego Ortega** @dortega\_delv · Feb 13  
The Computational Population Genetics Group, or CpG group, has now a website! [ligh.unam.mx/dortega/](http://ligh.unam.mx/dortega/) If you are interested in our research, please get in touch. I am currently looking to recruit a PhD student to join our group.


4 59 116  
[Show this thread](#)

 John Novembre Retweeted

 **Paul Appelbaum** @appelbap · Feb 12  
Selecting smarter embryos based on polygenic scores for IQ is looming. My colleagues Erik Parens and @WendyKChung and I share our thoughts on the ethical problems and the need for regulation: [statnews.com/2019/02/12/emb...](https://statnews.com/2019/02/12/emb...) via @statnews




 **Google Scholar**

 **Articles** About 3,160,000 results (0.03 sec)

**Any time**  
Since 2019  
Since 2018  
Since 2015  
Custom range...

**Sort by relevance**  
Sort by date

☒ include patents  
☒ include citations

 Create alert

**Next-generation DNA sequencing methods**  
[ER Mardis](#) - Annu. Rev. Genomics Hum. Genet., 2008 - [annualreviews.org](https://annualreviews.org)  
Recent scientific discoveries that resulted from the application of next-generation **DNA sequencing** technologies highlight the striking impact of these massively parallel platforms on genetics. These new **methods** have expanded previously focused readouts from...  
☆ 99 Cited by 2333 Related articles All 28 versions Web of Science: 1091

**New DNA sequencing methods**  
[A Marziali](#), [M Akesson](#) - Annual review of biomedical engineering, 2001 - [annualreviews.org](https://annualreviews.org)  
Abstract The Human Genome Project and other major genomic **sequencing** projects pushed the development of **sequencing** technology. In the past six years alone, instantaneous throughput has increased 15-fold. New technologies are now on the horizon that could...  
☆ 99 Cited by 171 Related articles All 4 versions Web of Science: 78

**[PDF] Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence**  
[H Swerdlow](#), [JZ Zhang](#), [DY Chen](#), [HR Harke](#)... - Analytical ..., 1991 - ACS Publications  
Capillary gel electrophoresis is demonstrated for the four-spectral-channel **sequencing** technique of Smith, the two-spectral-channel **sequencing** technique of Prober, and the spectral-channel **sequencing** technique of Richardson and Tabor. **Sequencing** rates...  
☆ 99 Cited by 317 Related articles All 4 versions Web of Science: 214

**Methods and apparatus for dna sequencing**  
[KM Ulmer](#) - US Patent 5,674,743, 1997 - [Google Patents](https://google.com/patents)  
The present invention provides a method and apparatus for automated **DNA sequencing**. The method of the invention includes the steps of: a) using a processive exonuclease to cleave from a single **DNA** strand the next available single nucleotide on the strand; b...  
☆ 99 Cited by 470 Related articles All 2 versions

# Our goals for this workshop:

1. **Identify** uses of computers in biology
2. **Write** and **execute** basic python code.
3. **Apply** these programming skills to: data analysis and simulations

Now, Chris is going to  
get us started with  
python