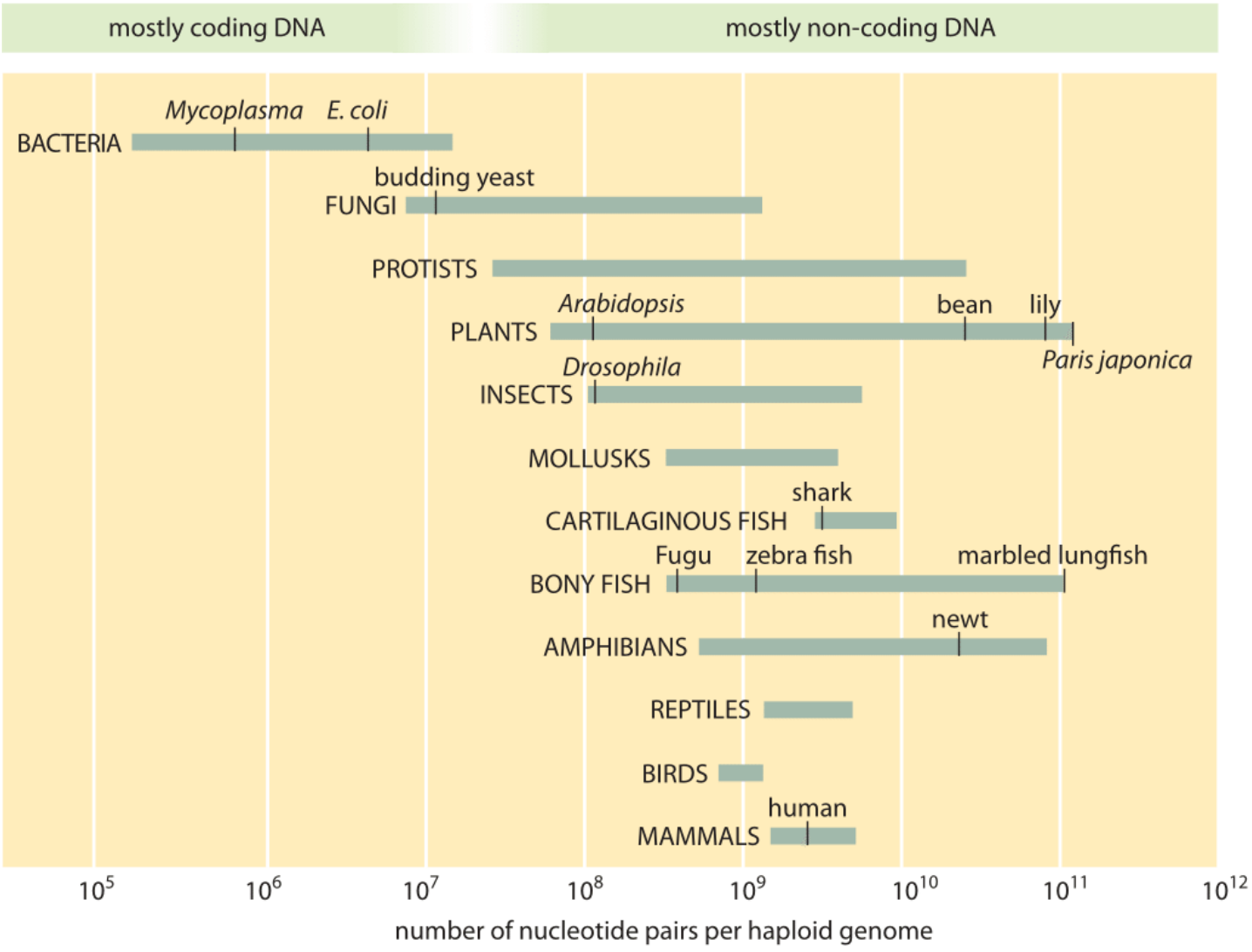
Twitter Thread by Michael Nielsen

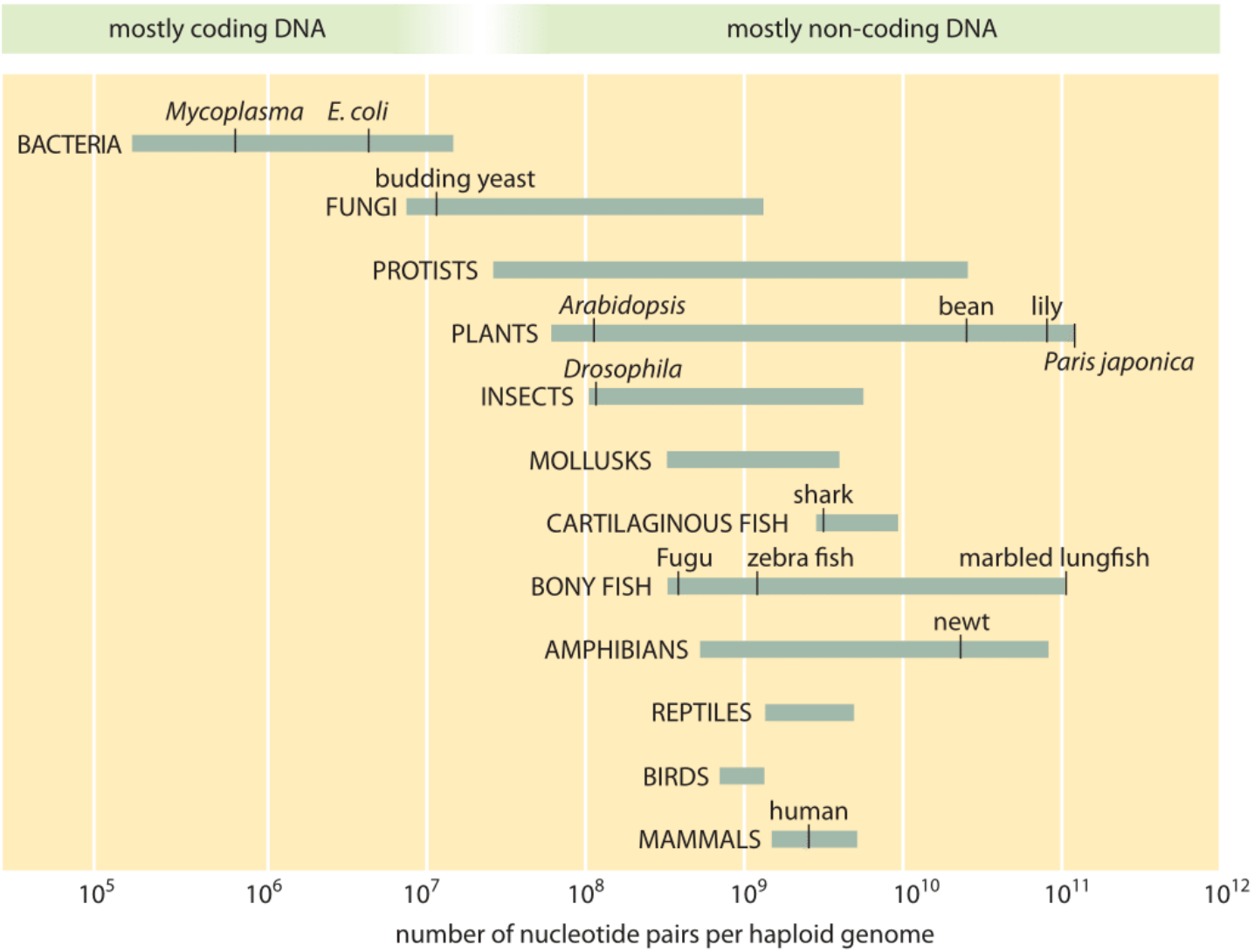
Michael Nielsen
@michael_nielsen

I've been reading the wonderful book "Cell Biology by the Numbers" (https://t.co/rLzAiv32Dq). Here's a thread of striking things I learned from the book (or related reading)

1. In terms of raw genomic information, human beings are unimpressive. Many plants & amphibians, for instance, have larger genomes. We're not so much the genomic crown of creation as a cul-de-sac of creation.



I haven't checked, but this graph suggests it may be possible to have a pet snail more genetically complex than you. Or keep (be kept by?) a houseplant that's more genetically complex



2. In fact, the herb Paris japonica has the largest known genome, ~150 billion base pairs, about 50x the human genome. The lungfish has more than 100 billion.

3. What are Paris japonica and the lungfish doing with all that genomic information? No idea! But what a great question!

4. With some caveats, a test substrate molecule, "collides with each and every protein in the cell on average about once per second".

Amazingly rapid mixing!

> A striking quantitative insight into the possibilities and rate of interactions at the molecular level can be gleaned from a clever interpretation (D. S. Goodsell, “The machinery of life”, Springer, 2010) of the diffusion limit. Say we drop a test substrate molecule into a cytoplasm with a volume equal to that of a bacterial cell. If everything is well mixed and there is no binding, how long will it take for the substrate molecule to collide with one specific protein in the cell? The rate of enzyme substrate collisions is dictated by the diffusion limit which as shown above is equal to ≈109 $s^{-1}M^{-1}$ times the concentrations. We make use of one of our tricks of the trade which states that in E. coli a single molecule per cell (say our substrate) has an effective concentration of about 1nM (i.e. 10-9 M). The rate of collisions is thus 109 $s^{-1}M^{-1}$ x 10-9 M ≈ 1 s^{-1}, i.e. they will meet within a second on average. This allows us to estimate that every substrate molecule collides with each and every protein in the cell on average about once per second. As a concrete example, think of a sugar molecule transported into the cell. Within a second it will have an opportunity to bump into all the different protein molecules in the cell. The high frequency of such molecular encounters is a mental picture worth carrying around when trying to have a grasp of the microscopic world of the cell.

5. I don't really know what it means for two molecules to collide, or what's required for a reaction to take place (how binary is it? how close do they need to be? does rotation matter etc?) Relatedly: I have no detailed picture of how enzymes speed up reactions, either.

6. A single strand of human DNA would, if stretched out linearly, be about a meter long. It's all bundled up tight inside the nucleus, and a whole lot of fascinating and complex machinery is needed to make use of that rather tight bundle.

7. Related: the volume of a DNA base is ~1 nm cubed, and that of an E. coli is ~1 micron cubed. These numbers are surprisingly useful to know.

(For a physicist, a bit like knowing that visible light has a wavelength roughly 500 nm, that a Hydrogen atom is roughly an angstrom in size, or that light travels one foot per nanosecond. These turn out to be just incredibly useful, over and over.)

8. The smallest virus genome is that of porcine circovirus, which contains 1759 bases, or just over 400 bytes of information(!!).

If you wanted a really viral tweet, you could just put the porcine circovirus genome in your tweet!

More later, just having some fun.

9. Upon reflection, I don't really know what cells are _for_. (I know, I know, that may be the wrong framing entirely.) Should I think of them as little factories, taking simple inputs (sugars etc) and pumping out complex proteins? Not sure, exactly.

10. If you look naively at a lot of diagrams, it often seems the cell wall is a relatively small part of the cell. In yeast, at least, it makes up 10-25% of the dry mass of the cell(!) I don't know the % of the volume, unfortunately.

11. There's something like 6 orders of magnitude between small-genome life and large-genome life! In terms of linear dimension it's something like the difference between the height of a large ant hill and the height of Mt. Everest.

12. Related: One thing the book does well is just put in-your-face over and over and over the staggering diversity of life. Protons, electrons, and neutrons are nifty things.

13. Also related: the book rubs in your face the extent to which the biological world is a repository of extraordinary nanomachines which we humans can go discover. It's just this incredible extant resource of ideas and principles and

machinery.

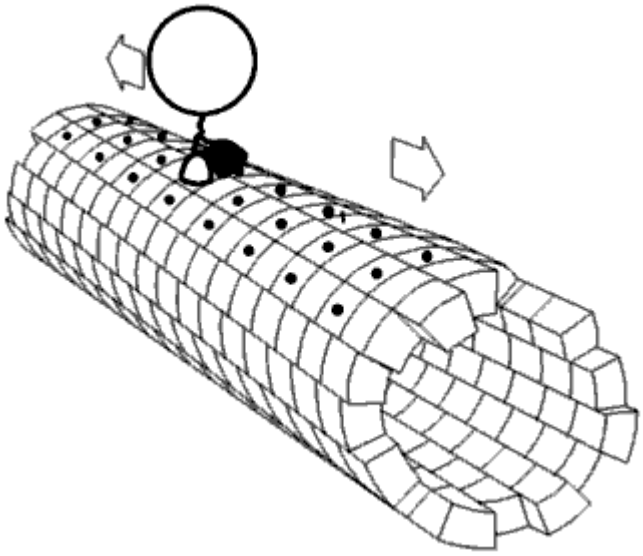
And we understand so little about it still. Fun to realize, for instance, that we only pretty recently understood the basic structure of the ribosome - the nanomachine that turns messenger RNA into proteins.

(I know, I know, the broad point here certainly isn't news. Still, the book is fun to read for the onslaught of lovely examples.)

14. Question: Has anyone understood in detail how Coase's "Nature of the Firm" (and the modern followons) relate to multi-cellular life? Lots of very similar problems...

15. Wikipedia's "Molecular Machines" article is as much fun as you should be allowed to have on the Internet: https://t.co/GOliGurSB8

(It's possible I need to get out more. I blame 2020!)



16. Charles Darwin on the evolutionary origin of the eye. I find this quite moving:

evolution could have given rise to such specialized organs. Chapter VI of “The Origin of Species” is entitled “Difficulties on Theory” and is used by Darwin as a forum to explain what he referred to as a “crowd of difficulties” that “will have occurred to the reader”. He notes that some of these difficulties are “so grave that to this day I can never reflect on them without being staggered; but, to the best of my judgment, the greater number are only apparent, and those that are real are not, I think, fatal to my theory.” One of the most significant of those difficulties was what Darwin thought of as “organs of extreme perfection” such as our eye. He goes on to say that “To suppose that the eye, with all its inimitable contrivances for adjusting the focus to different distances, for admitting different amounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree. Yet reason tells me, that if numerous gradations from a perfect and complex eye to one very imperfect and simple, each grade being useful to its possessor, can be shown to exist; if further, the eye does vary ever so slightly, and the variations be inherited, which is certainly the case; and if any variation or modification in the organ be ever useful to an animal under changing conditions of life, then the difficulty of believing that a perfect and complex eye could be formed by natural selection, though insuperable by our imagination, can hardly be considered real.” Our understanding of the long evolutionary history of eyes continues to evolve itself and a current snapshot can be attained by reading a recent review (such as Lamb et al, Nat. Rev. Neuro. 8:960, 2007).

17. @rob_carlson's excellent book "Biology is Technology" has a great discussion of the necessity of predictive quantitative models for design and engineering. Here's an excerpt, which repays thought IMO:

There are many unspoken assumptions built into this representation of engineering. Among the most important are that (1) all the parts are the result of a careful design process, (2) the parts can be constructed to function according to the design, and (3) when assembled from those parts, the resulting whole actually behaves as predicted by the design.

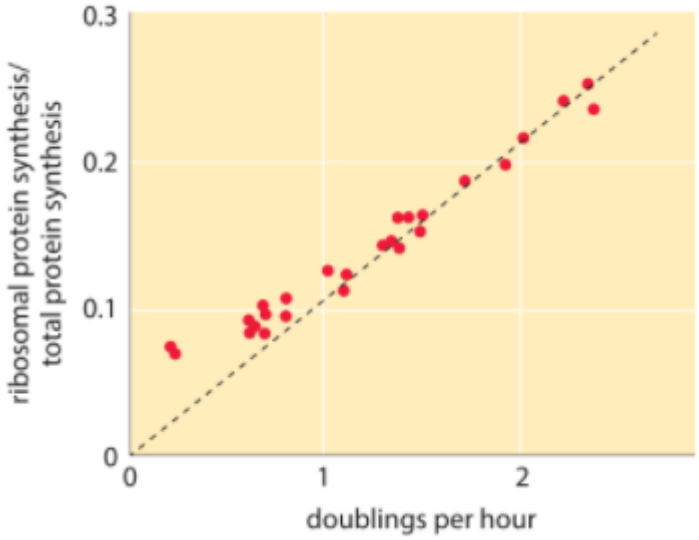
The experience of everyone who watches the commercial contributes to the communication of these unspoken assumptions. Not only do we the viewers have considerable exposure to other products of this engineering process, but, given the number of Honda cars and trucks on the road, many of us demonstrate confidence in the engineering and manufacturing prowess of Honda in particular.

Just as the broad public understanding of LEGOS can be used to imbue a sense of careful design and manufacturing into a new Honda, the notion that the products of modern engineering are safe and predictable can be used to sell other technologies.

Unfortunately, in comparison, current genetic "engineering" techniques are quite primitive, akin to swapping random parts between cars to produce a better car. Biological engineering in general does not yet exist in the same way that electrical, mechanical, and aeronautical engineering do. Mature engineering fields rely on computer-aided design tools—software packages like SolidWorks for

One of the most striking things about "Cell Biology by the Numbers" is that such predictive quantitative models are _lacking_. This isn't a criticism. Rather, it's an opportunity. The book is chock-full of wonderful observations that could, plausibly, help lead to such models.

18. A fun set of examples of inchoate models comes from this vignette about the relationship b/w how many ribosomes a bacteria produces, & how often the bacteria reproduces. Very tight relationship! Why, exactly? What else impacts it? Many nascent ideas! https://t.co/ij531Rt2bX



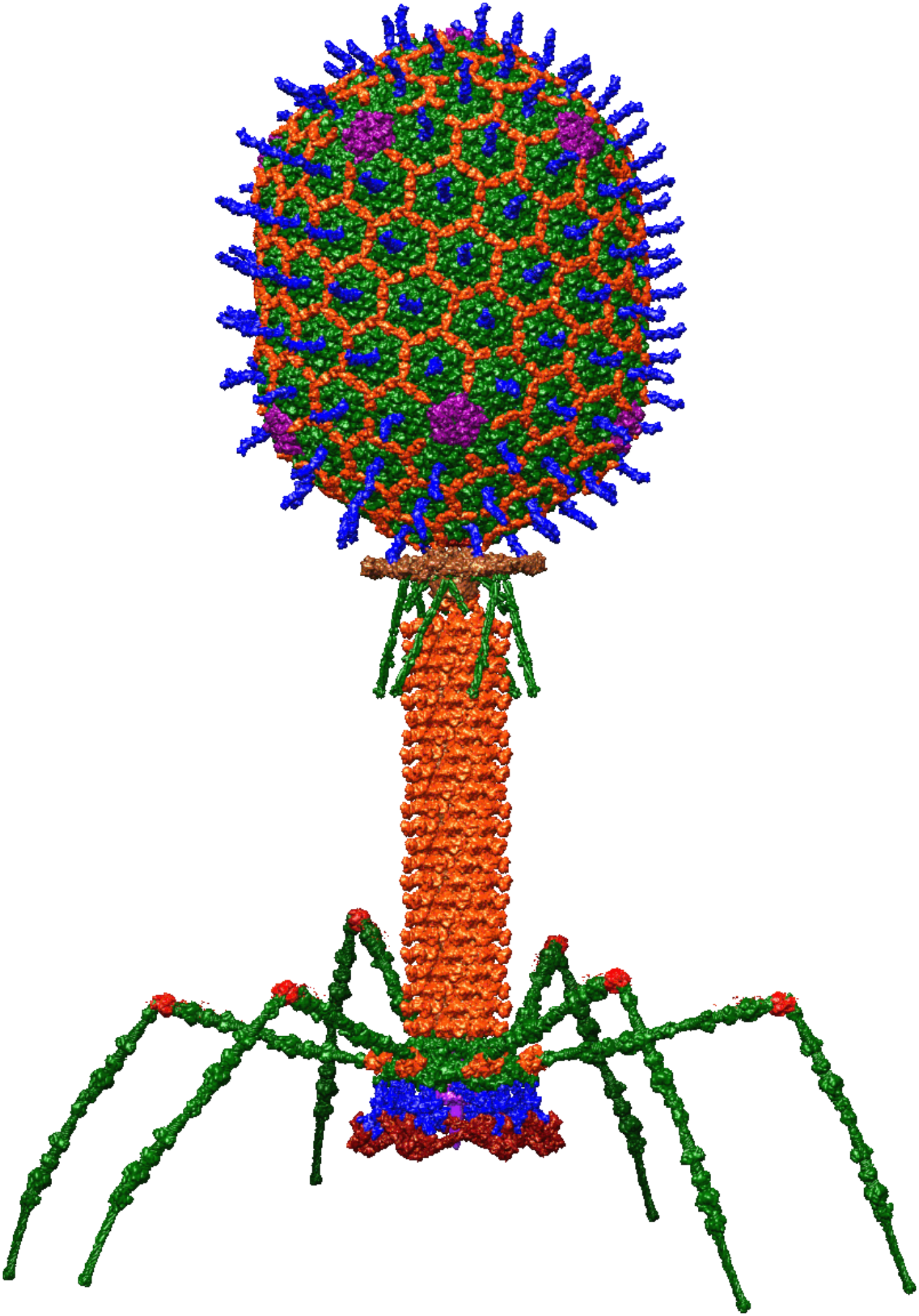
19. Very very roughly, and with much variation, there's about 10,000 ribosomes per femtoliter of cell. The way these numbers are determined are interesting - a trend away from clever weighing techniques, and toward direct microscopy / counting techniques.

Traditionally, measuring the number of ribosomes per cell was based on separating the ribosomes from the rest of the cell constituents, measuring what fraction of the total mass comes from these ribosomes and then with conversion factors based on estimations of cell size and mass, ribosomal molecular weight etc. inferring the abundance per cell. Recently a more direct approach is becoming available based on explicitly counting individual ribosomes. In cryo-electron microscopy, rapidly frozen cells are visualized from many angles to create what is known as a tomographic 3D map of the cell. The known structure of the ribosome is then used as a template that can be searched in the complete cell tomogram. This technique was applied to the small, spiral-shaped prokaryote *Spiroplasma melliferum.* As shown in Figure 3, in this tiny cell, 10-100 times smaller than *E. coli* by volume (BNID 108949, 108951) and slower in growth, researchers counted on average 1000 ribosomes per cell (BNID 108945). Similar direct counting efforts have been made using the super-resolution techniques that have impacted fluorescence microscopy as shown in Figure 4 where a count was made of the ribosomes in *E. coli*. A comparison of the results from these two methods is made in Figure 5 where a simple estimate of the ribosomal density is made from the cryo-electron microscopy images and this density is then scaled up to a full *E. coli* volume, demonstrating an encouraging consistency between the different methods.

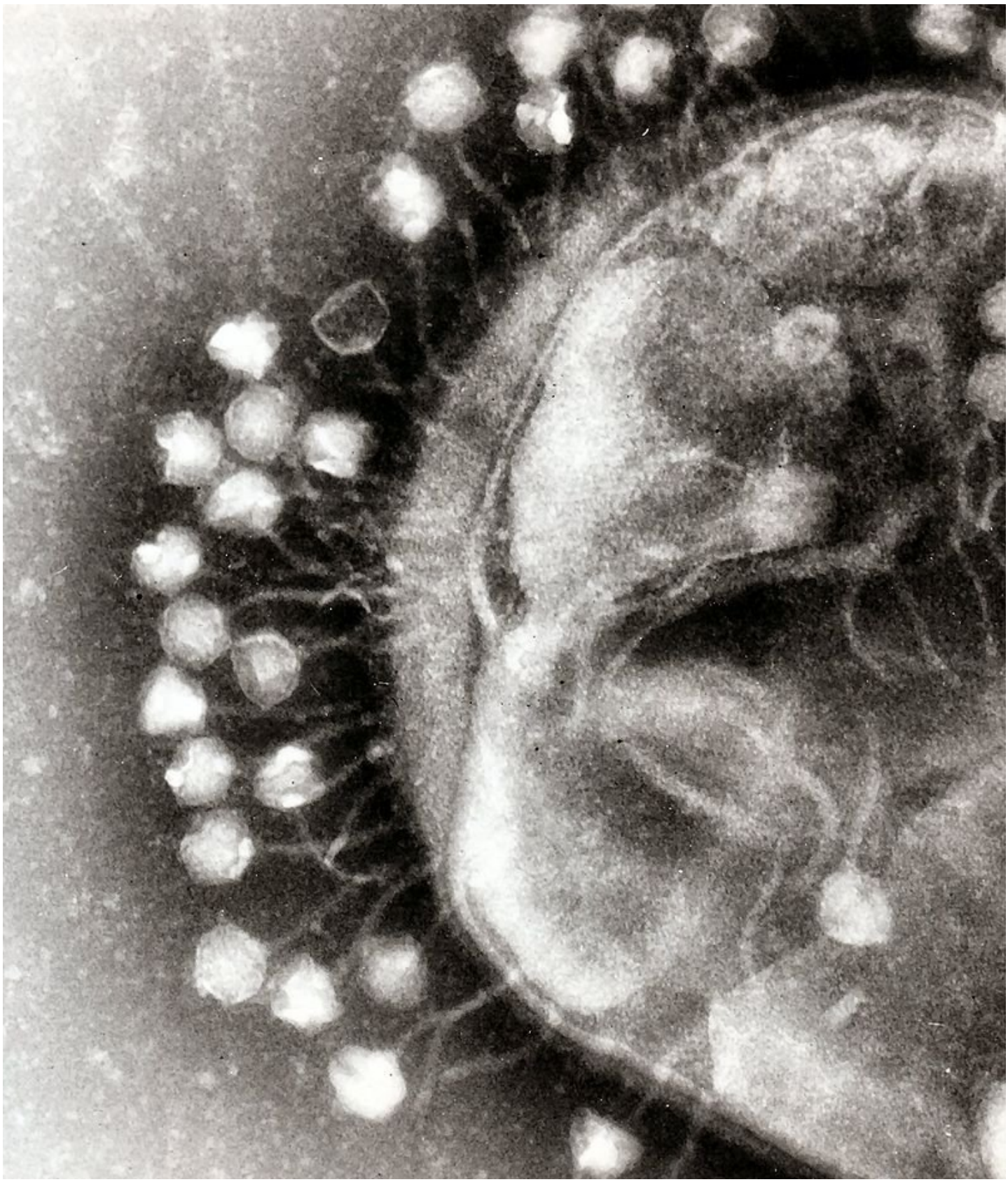
20. Roughly, an E. coli is about a femtoliter in size. Common cells are often in the range of femtoliters to picoliters (and may go larger).

A Hewlett-Packard researcher once told me they think about controlling printer fluid with femtoliter precision.

21. In general, viruses are incredibly cool looking. Here's the t4 phage.



And here's a bunch of phages attacking a bacterium. (Both images from Wikipedia.)



22. Utterly amazing range of variation in sizes of human cells - a factor of more than 100,000!

It's incredible that all these things are called "cells". If I met a person 100,000 times larger than me, I'd think a new category was called for!

cell type	average volume (μm³)	BNID
sperm cell	30	109891, 109892
red blood cell	100	107600
lymphocyte	130	111439
neutrophil	300	108241
beta cell	1,000	109227
enterocyte	1,400	111216
fibroblast	2,000	108244
HeLa, cervix	3,000	103725, 105879
hair cell (ear)	4,000	108242
osteoblast	4,000	108088
alveolar macrophage	5,000	103566
cardiomyocyte	15,000	108243
megakaryocyte	30,000	110129
fat cell	600,000	107668
oocyte	4,000,000	101664

In general, one of the main things I'm getting from the book is that many categories of biological thing have far more variation than I'd previously appreciated.

23. "The fact.. all organisms are built of basic units, namely cells, is one of the great revelations of biology.. [O]ften now taken as a triviality, it is one of the deepest insights in the history of biology...

... & serves as a unifying principle in a field where diversity is the rule rather than the exception."

[Loved this. One of those things so easy to overlook until someone really knowledgeable points it out. And then can keep you thinking for decades.]

Reminded of the famous (& very beautiful) way Feynman began the Feynman Lectures:

1–2 Matter is made of atoms

SU

If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generations of creatures, what statement would contain the most information in the fewest words? I believe it is the *atomic hypothesis* (or the atomic *fact*, or whatever you wish to call it) that *all things are made of atoms—little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another*. In that one sentence, you will see, there is an *enormous* amount of information about the world, if just a little imagination and thinking are applied.