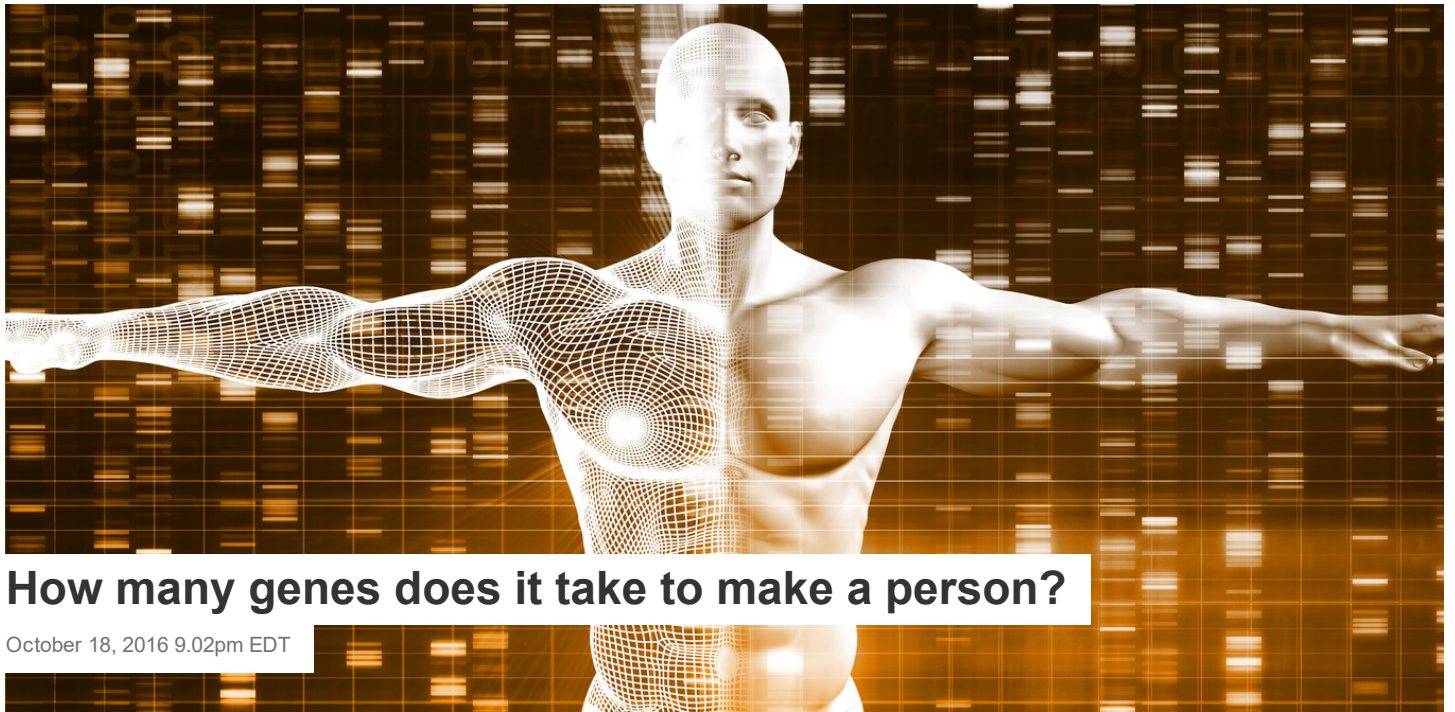


THE CONVERSATION

Academic rigor, journalistic flair



How many genes does it take to make a person?

October 18, 2016 9.02pm EDT

Do we contain the most elaborate set of instructions? Genome image via www.shutterstock.com.

We humans like to think of ourselves as on the top of the heap compared to all the other living things on our planet. Life has evolved over three billion years from simple one-celled creatures through to multicellular plants and animals coming in all shapes and sizes and abilities. In addition to growing ecological complexity, over the history of life we've also seen the evolution of intelligence, complex societies and technological invention, until we arrive today at people flying around the world at 35,000 feet discussing the in-flight movie.

It's natural to think of the history of life as progressing from the simple to the complex, and to expect this to be reflected in increasing gene numbers. We fancy ourselves leading the way with our superior intellect and global domination; the expectation was that since we're the most complex creature, we'd have the most elaborate set of genes.

This presumption seems logical, but the more researchers figure out about various genomes, the more flawed it seems. About a half-century ago the estimated number of human genes was in the millions. Today we're down to about 20,000. We now know, for example, that bananas, with their 30,000 genes, have 50 percent more genes than we do.

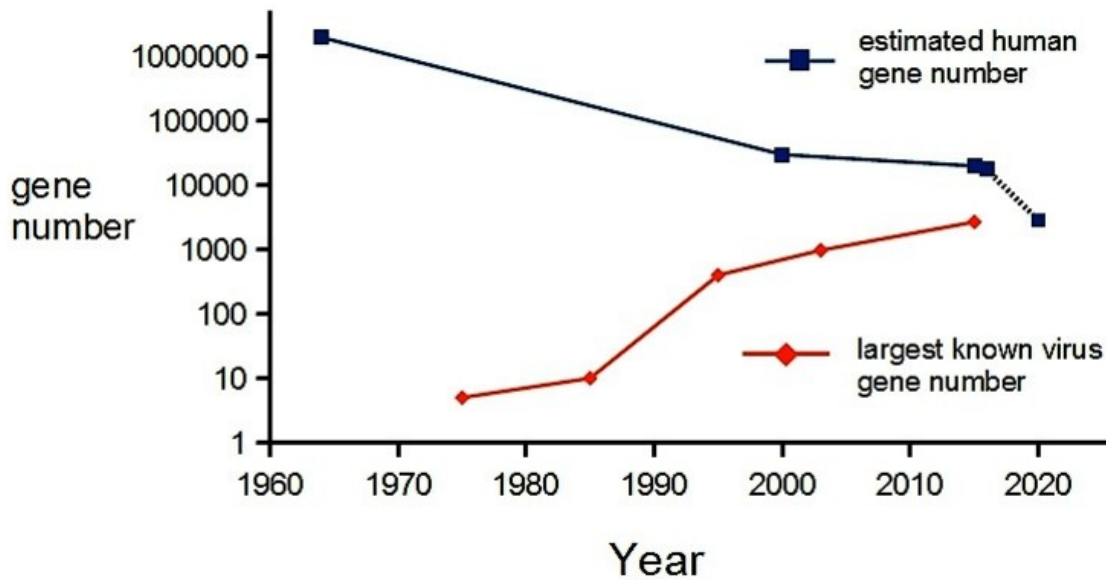
As researchers devise new ways to count not just the genes an organism has, but also the ones it has that are superfluous, there's a clear convergence between the number of genes in what we've always thought of as the simplest lifeforms – viruses – and the most complex – us. It's time to rethink the question of how the complexity of an organism is reflected in its genome.

Author



Sean Nee

Research Professor of Ecosystem Science and Management, Pennsylvania State University



The converging estimated number of genes in a person versus a giant virus. Human line shows average estimate with dashed line representing estimated number of genes needed. Numbers shown for viruses are for MS2 (1976), HIV (1985), giant viruses from 2004 and average T4 number in the 1990s. Sean Nee, CC BY

Counting up the genes

We can think of all our genes together as the recipes in a cookbook for us. They're written in the letters of the bases of DNA – abbreviated as ACGT. The genes provide instructions on how and when to assemble the proteins that you're made of and that carry out all the functions of life within your body. A **typical** gene requires about 1000 letters. Together with the environment and experience, genes are responsible for what and who we are – so it's interesting to know how many genes add up to a whole organism.

When we're talking about numbers of genes, we can display the actual count for viruses, but only the estimates for human beings for an important reason. One **challenge** counting genes in eukaryotes – which include us, bananas and yeast like *Candida* – is that our genes are not lined up like ducks in a row.

Our genetic recipes are arranged as if the cookbook's pages have all been ripped out and mixed up with three billion other letters, about **50 percent** of which actually describe inactivated, dead viruses. So in eukaryotes it's hard to count up the genes that have vital functions and separate them from what's extraneous.

In contrast, counting genes in viruses – and bacteria, which can have **10,000** genes – is relatively easy. This is because the raw material of genes – nucleic acids – is relatively expensive for tiny creatures, so there is strong selection to delete unnecessary sequences. In fact, the real challenge for viruses is discovering them in the first place. It is startling that all **major virus discoveries**, including HIV, have not been made by sequencing at all, but by old methods such as magnifying them visually and looking at their morphology. Continuing advances in molecular technology have taught us the remarkable diversity of the virosphere, but can only help us count the genes of something we already know exists.

Flourishing with even fewer

The number of genes we actually need for a healthy life is probably even lower than the current estimate of 20,000 in our entire genome. One author of a recent study has reasonably extrapolated that the count for essential genes for human beings may be much lower.

These researchers looked at thousands of healthy adults, looking for naturally occurring “knockouts,” in which the functions of particular genes are absent. All our genes come in two copies – one from each parent. Usually, one active copy can compensate if the other is inactive, and it is difficult to find people with *both* copies inactivated because inactivated genes are naturally rare.

Knockout genes are fairly easy to study with lab rats, using modern genetic engineering techniques to inactivate both copies of particular genes of our choice, or even remove them altogether, and see what happens. But human studies require populations of people living in communities with 21st century medical technologies and known pedigrees suited to the genetic and statistical analyses required. Icelanders are one useful population, and the British-Pakistani people of this study are another.

This research found over 700 genes which can be knocked out with no obvious health consequences. For instance, one surprising discovery was that the PRDM9 gene – which plays a crucial role in the fertility of mice – can also be knocked out in people with no ill effects.

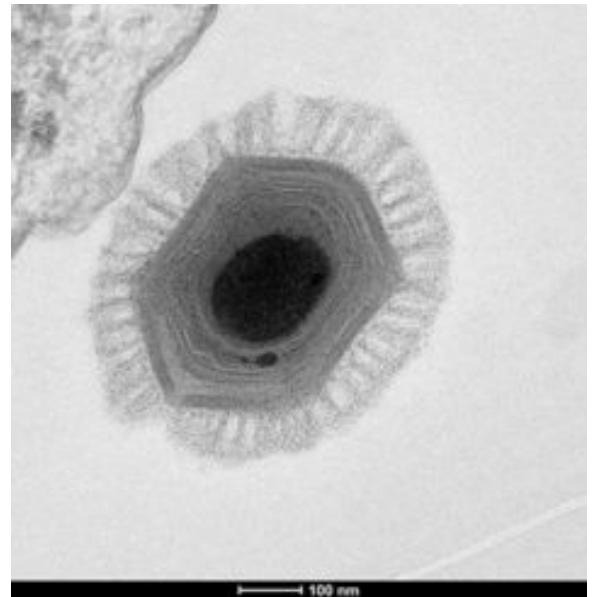
Extrapolating the analysis beyond the human knockouts study leads to an estimate that only 3,000 human genes are actually needed to build a healthy human. This is in the same ballpark as the number of genes in “giant viruses.” Pandoravirus, recovered from 30,000-year-old Siberian ice in 2014, is the largest virus known to date and has 2,500 genes.

So what genes do we need? We don't even know what a quarter of human genes actually do, and this is advanced compared to our knowledge of other species.

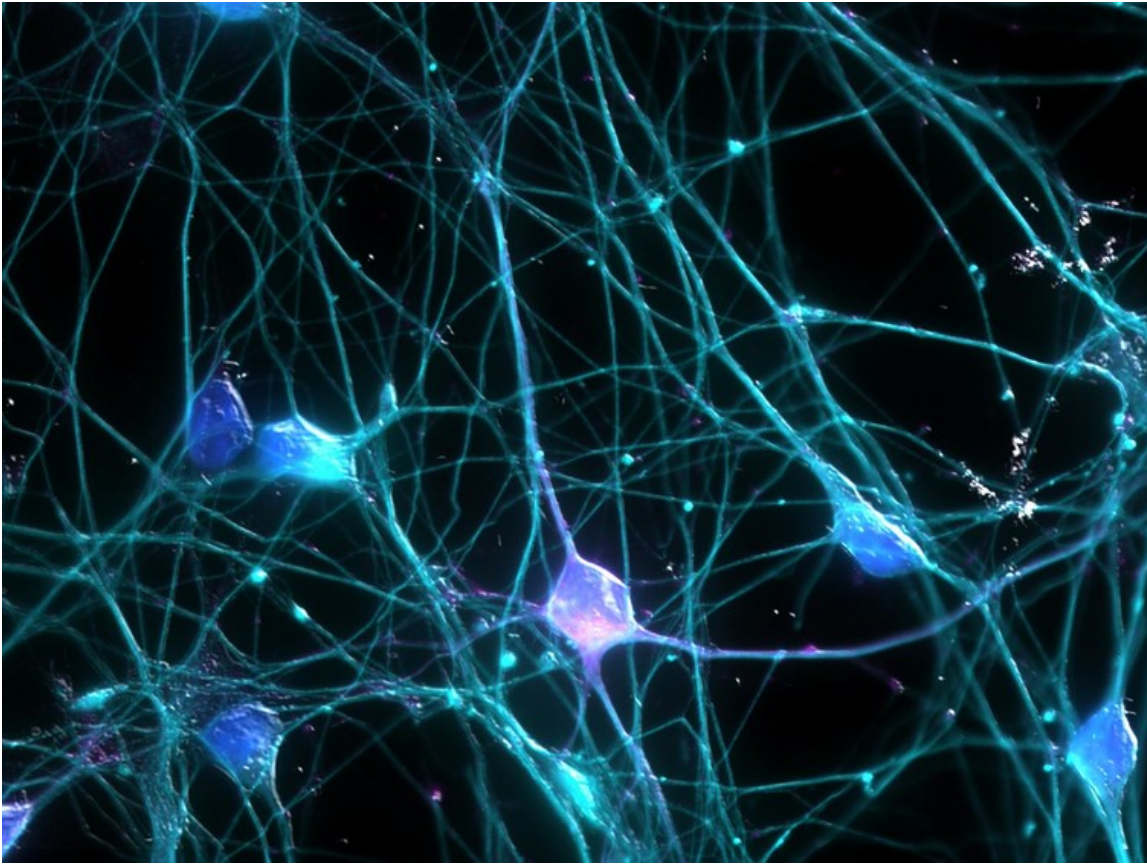
Complexity arises from the very simple

But whether the final number of human genes is 20,000 or 3,000 or something else, the point is that when it comes to understanding complexity, size really does not matter. We've known this for a long time in at least two contexts, and are just beginning to understand the third.

Alan Turing, the mathematician and WWII code breaker established the theory of multicellular development. He studied simple mathematical models, now called “reaction-diffusion” processes, in which a small number of chemicals – just two in Turing's model – diffuse and react with each other. With simple rules governing their reactions, these models can reliably generate very complex, yet coherent structures that are easily seen. So the biological structures of plants and animals do not require complex programming.



Megavirus has over a thousand genes, Pandoravirus has even more. Chantal Abergel, CC BY-SA



The simple building blocks of neurons together generate immense complexity. UCI Research/Arady Rahman, CC BY-NC

Similarly, it is obvious that the 100 trillion connections in the human brain, which are what really make us who we are, cannot possibly be genetically programmed individually. The recent breakthroughs in artificial intelligence are based on neural networks; these are computer models of the brain in which simple elements – corresponding to neurons – establish their own connections through interacting with the world. The results have been spectacular in applied areas such as handwriting recognition and medical diagnosis, and Google has invited the public to play games with and observe the dreams of its AIs.

Microbes go beyond basic

So it's clear that a single cell does not need to be very complicated for large numbers of them to produce very complex outcomes. Hence, it shouldn't come as a great surprise that human gene numbers may be of the same size as those of single-celled microbes like viruses and bacteria.

What is coming as a surprise is the converse – that tiny microbes can have rich, complex lives. There is a growing field of study – dubbed “sociomicrobiology” – that examines the extraordinarily complex social lives of microbes, which stand up in comparison with our own. My own contributions to these areas concern giving viruses their rightful place in this invisible soap opera.

We have become aware in the last decade that microbes spend over 90 percent of their lives as biofilms, which may best be thought of as biological tissue. Indeed, many biofilms have systems of electrical communication between cells, like brain tissue, making them a model for studying brain disorders such as migraine and epilepsy.

Biofilms can also be thought of as “cities of microbes,” and the integration of sociomicrobiology and medical research is making rapid progress in many areas, such as the treatment of cystic fibrosis. The social lives of microbes in these cities – complete with cooperation, conflict, truth, lies and even suicide – is fast becoming the major study area in evolutionary biology in the 21st century.

Just as the biology of humans becomes starkly less outstanding than we had thought, the world of microbes gets far more interesting. And the number of genes doesn't seem to have anything to do with it.

[Bacteria](#)[Genome](#)[Virus](#)[Microbiology](#)[Genes](#)[Human genome](#)[Biofilm](#)[Bacterial biofilm](#)