Opinion

EMBO reports

The genome paper is dead, long live the genome paper!

David R Smith*

For now they kill me with a living death.

Richard III, Shakespeare

Nearly ten years ago, I wrote a lighthearted essay for a genetics journal lamenting the downfall of my favorite genre of research article—the genome paper (Smith, 2013). I argued that this once lauded form of presenting genomic data had become outdated, unoriginal and, the worst fate for any paper, low impact. As I recall, the reviewer comments of the piece were encouraging, and publishing it was straightforward. What happened next, however, was anything but ordinary.

Within a week of being online, the essay accrued thousands of reads and hundreds of shares through social media, quickly making it the most popular article I had ever written. Admittedly, the provocative title, *Death of the genome paper*, helped spur interest and was the equivalent of scientific clickbait.

Soon, I started receiving emails from readers. At first, they were mostly positive: "Thanks for the fun read, David. Couldn't agree with you more". "Shared the article at our lab meeting . . . lots of fruitful discussion ensued". Then, more critical feedback began to trickle in: "Interesting points but you should really give more thought to your arguments before condemning a field of research and format of publication that's been around for longer than you". Some emails were downright angry: "The genome paper is alive and well and you are too naïve and short-sighted to see this!"

Years later, I came across someone at a conference who glanced at my nametag and said, "Wait a minute, aren't you that person who wrote off the genome paper?", as if my entire contribution to science could be distilled into a short satirical reflection article. Even now, the odd humorous or insensitive online comment comes my way through Twitter or ResearchGate: "Hypocritomics: can you believe that the author of this gem is still publishing genome sequences?" (I can and I am.).

Like it or not, my essay touched a nerve. Worst of all, time has proven that I was oh so wrong. Today, if you skim through a major research journal in the life sciences, you will almost certainly come across a genome paper. Earlier this year, for example, *Nature* published a full-length article describing the genome of the Australian lungfish *Neoceratodus forsteri* (Meyer *et al*, 2021), and at > 35 gigabases what an extraordinary genome it is. Similarly, in a recent issue of *Science* one can find the "genome of a ~ 34,000-year-old hominid skull cap discovered in the Salkhit Valley in northeastern Mongolia" (Massilani *et al*, 2020). Even in my own area of research, protistology, genome papers abound. *Current Biology*, for instance, recently showcased the genome of an Antarctic green alga (*Chlamydomonas* sp. ICE-L) that lives on the bottom of sea ice (Zhang *et al*, 2020).

What is it about genome papers that is so enduring? And why was I so fast to put the nail in their coffin? The answer to the latter question is easy. I began my research career by publishing organelle genomes, which had their start in *Nature* in 1981 with the sequencing of the human mitochondrial DNA (mtDNA) (Anderson *et al*, 1981). At the beginning of my PhD, in the mid-2000s, it was not uncommon to get organelle genome data into good journals, but by the time I became a postdoc in 2010 their impact had plummeted. Burgeoning next-generation sequencing technologies made it fast and easy to generate organelle genome sequences en masse, especially animal mtDNAs, and consequently their novelty declined.

Today, most newly sequenced mtDNAs appear as short, rapid-communications—often called "genome announcements" or "genome reports"—which are typically < 1,000 words, have few if any figures or tables, and do not necessarily undergo formal peer review (Smith, 2017). Despite their brevity, it can cost hundreds of dollars to publish a genome report, and new journals or subsections within journals specializing in these kinds of papers are popping up (and disappearing) all the time. Genome reports have also become a popular way for presenting viral and bacterial genomes.

What's the point of these bite-sized articles, apart from being an expensive way of letting researchers know that a genome has been sequenced and where to find it? The cynic in me says, they are primarily an easy source of revenue for journals and a quick route to "peer-reviewed" papers for scientists. (I write this having published my fair share of mitochondrial genome reports.) I would argue that the field of genomics has outgrown genome reports and that what is most important when presenting new nucleotide sequences is not the publication but the quality of the submission to a public databank, like NCBI. There is nothing worse than downloading a genome from GenBank and discovering that the annotations are missing or of poor quality. The same can be said of the raw sequencing reads and other associated datasets used for genome analyses. Moreover, there are now many high-quality, user-friendly bioinformatics software suites allowing users to download and view GenBank data. The figures and images generated by some of these programs are so

University of Western Ontario, London, ON, Canada
*Corresponding author. E-mail: dsmit242@uwo.ca
David R Smith is a regular columnist for *EMBO reports*.
DOI 10.15252/embr.202154434 | EMBO Reports (2021) e54434

interactive, intuitive, and beautiful that they make the standard chromosome maps from genome reports look like old-school Polaroid pictures, which is even more reason to dispense with these types of articles.

I can hear the readers' complaints already ("Fool me once, shame on you. Fool me twice …"), which is why this time around I won't be so quick to sign the death certificate of the genome paper or genome report. Indeed, my collaborators and I have spent the past four years over our heads in data wrestling with the nuclear genome sequence of a chlamydomonadalean green alga (Zhang *et al*, 2021). I can attest that despite the massive leaps forward in genomic technologies it can be bloody hard assembling and annotating *de novo* a decently sized nuclear genome. Hence the reason why these sequences continue to merit publication in a top-tier journal. And I will not be running out to assemble another one anytime soon. Still, the fact that a small team—there were 5 authors on our paper—from a small university can complete such a project suggests that nuclear genomes may soon be going the way of the mitogenome—to be packaged in tiny, figureless notes called "nuclear DNA announcements".

Of course, there is also the multi-genome paper. If publishing one genome is old hat, then why not pack 2, 10, or 100 genomes into a single article? I know this game well. Not long ago, my collaborators and I crammed 72 yeast mtDNAs into an 800-word paper, all in the hopes of getting the data into a leading journal. It worked, but is the research community well-served by a paper that has ~10 words per genome?

Whatever the fate of the genome paper and its various offshoots, no one can deny that it gave us some of the most exciting science of the past 35 years. When the true obituary is eventually written, it may go something like this: "Born 1974. Father: Frederik Sanger. Closest friend: tree of life. Life-long partner: the transcriptome. Hailed as a prodigy early on but fell into disrepute in its later years. No stranger to scandal and tabloid press. Helped usher many into the upper echelons of science and was often seen dining with the elites of Silicon Valley. Outlived many of its detractors, including the little-known scientist David R. Smith who once incorrectly declared it dead in 2013".

Acknowledgements

DRS is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Conflict of interest

The author declares that he has no conflict of interest.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F *et al* (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290: 457 – 565

Massilani D, Skov L, Hajdinjak M, Gunchinsuren B, Tseveendorj D, Yi S, Lee J, Nagel S, Nickel B, Devièse T *et al* (2020) Denisovan ancestry and population history of early East Asians. *Science* 370: 579 – 583

Meyer A, Schloissnig S, Franchini P, Du K, Woltering JM, Irisarri I, Wong WY, Nowoshilow S, Kneitz S, Kawaguchi A *et al* (2021) Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* 590: 284 – 289

Smith DR (2013) Death of the genome paper. *Front Genet* 4: 72

Smith DR (2017) Goodbye genome paper, hello genome report: the increasing popularity of 'genome announcements' and their impact on science. *Brief Funct Genomics* 16: 156 – 162

Zhang X, Cvetkovska M, Morgan-Kiss R, Hüner NPA, Smith DR (2021) Draft genome sequence of the Antarctic green alga *Chlamydomonas* sp. UWO241. *iScience* 24: 102084

Zhang Z, Qu C, Zhang K, He Y, Zhao X, Yang L, Zheng Z, Ma X, Wang X, Wang W *et al* (2020) Adaptation to extreme Antarctic environments revealed by the genome of a sea ice green alga. *Curr Bio* 30: 3330 – 3341

