Science & Society

EMBO reports

The Frankenthesis

This stitched-together monster is alive and coming to a campus near you

David Roy Smith

When my father passed away, I climbed up the steep steps to his attic, dug out his old, dusty PhD thesis, and dragged it halfway across the country to put it in a prime spot on my office shelves at the University of Western Ontario. As the afternoon sun peeks through the blinds, it is easy to make out the gold-colored words on the green leather binding: Frank Smith, PhD 1971, Chemistry, University of New South Wales. The title—Metal complexes of 2-substituted pyridines and 1, 10-phenanthrolines—is enough to intimidate even the most prodigious students that come to my office. And, being a biologist, most of the thesis is lost on me as well.

> *"All my peers in the department were building their own Frankentheses in the twilight hours of dimly lit laboratories, and most of us were advised to do so by our mentors."*

Still, sometimes I gently pull it down from the shelf and skim through the chapters, glance at the figures, and reread the acknowledgments. What strikes me whenever I do this is how different my father's dissertation is from my own—not because his is in chemistry and mine genetics, but because his has a single overarching question and hypothesis that runs throughout the chapters, whereas mine is a hodgepodge of ideas and arguments. I'm the academic product of the DNA sequencing generation and, as such, my thesis is a conglomerate of loosely strung together genome papers, all of which were published before I even arrived at a defense date. Of course, I tried to develop a coherent theme unifying the different chapters, but I did this as an afterthought, and it was obvious to the examiners that I had created—as one of them creatively phrased it—a Frankenthesis.

I wasn't alone in practicing the dark arts of dissertation writing. All my peers in the department were building their own Frankentheses in the twilight hours of dimly lit laboratories, and most of us were advised to do so by our mentors. As my old silverback of a supervisor would say: "No one is going to read your thesis, Smitty. It is only the publications that matter, and the more the better". And, so, my grad school cohort became paper-writing fanatics, aided in large part by the exploding field of genomics and the new high-throughput tools in molecular biology.

Next-generation sequencing

Since finishing my doctorate, next-generation (and now third-generation) sequencing technologies have been delivered to the scientific masses, which, among many things, has breathed even more life into the Frankenthesis. "It's alive!" shouts the student from the PhD examination room. "What, in Darwin's name, is it?" ask the committee members, recoiling in fear. It may be published, it may be high impact, and it may have taken years of hard work, but, in many instances, it is not a doctoral thesis, at least not in the traditional sense. But is that necessarily a bad thing?

Being one of the few people in my department who specializes in genomics, I'm regularly asked to read and examine graduate dissertations that employ next-generation sequencing data. Although some are superbly written and address fundamental questions in biology, recently I have found that more and more are big on data but small on scientific substance. For instance, during a defense, I usually ask: "What do you think is the most significant scientific contribution to come from this PhD?" After a long pause, I'm apt to get a reply like: "Well, these five transcriptomes will help other researchers study important aspects of fish biology". Indeed, I think to myself—the "big data" baton gets passed along and along, as does the responsibility for actually solving a major scientific problem.

> *"It may be published, it may be high impact, and it may have taken years of hard work, but, in many instances, it is not a doctoral thesis, at least not in the traditional sense."*

Some of these issues might arise because an increasing number of my colleagues from diverse disciplines are incorporating high-throughput sequencing into their research programs, and consequently are recruiting students with expertise in bioinformatics and computer science but who typically have little or no experience on the biological system they will be studying. Although supposedly focusing on their graduate projects, these students often become the bioinformatics technicians for the entire laboratory, and, for better or worse, their theses reflect this fact.

Not long ago, I reviewed a doctoral dissertation in which an entire chapter was on how the candidate set up a public online search repository for a eukaryotic genome sequence. Other chapters centered on the development of a genome annotation pipeline and the deposition of assembly data into GenBank. Some might sneer at the thought

Department of Biology, University of Western Ontario, London, ON, Canada. E-mail: dsmit242@uwo.ca
DOI 10.15252/embr.201745446

of such seemingly mundane bioinformatics tasks forming the brunt of a PhD thesis, but anyone who has ever participated on a major genome project will know that this kind of work can take hundreds of hours and is far from trivial. Whatever your stance, the student who wrote the thesis in question went on to get a full-time, high-profile bioinformatics position at a hospital shortly after the defense.

Overproduction of PhDs

Maybe we need to change how we think about and define the structure and content of PhD theses [1,2]. Doctoral (and masters) research can expand the bounds of knowledge, develop new hypotheses, and solve unanswered problems. But it can also be about acquiring the skills needed to find a fulfilling and well-paying job, within or outside of academia. Obviously, the requirements, length of training, quality control, and academic culture of a PhD program can differ drastically between different universities and countries, and there is no denying that some institutes are producing low-quality graduates [3]. Much has been made, moreover, about the apparent over-production of doctorates, with some going so far as to suggest that universities have become mere PhD factories, and asking: "Is it time to stop?" [3]. Not to mention the declining interest in an academic career among graduate students [4].

"Perhaps, the Frankenthesis is a reflection of the natural selection of bad science, and possibly fast, cheap, and near-unlimited access to genetic sequencing data is adding fuel to the process."

Even more disconcerting is the fact that institutional incentives for high research output—in the form of publications and grants, for example—is leading to bad science. Studies have shown that "the most powerful incentives in contemporary science actively encourage, reward and propagate poor research methods" [5]. Some have dubbed this process the natural selection of bad science because "it requires no conscious strategizing nor cheating on the part of researchers. Instead, it arises from the positive selection of methods and habits that lead to publication" [5]. This, in turn, is contributing to the increase in the number of fatal errors in and retractions of research articles, particularly in prominent journals [6,7], as well as the apparent reproducibility crisis [8]. Perhaps, the Frankenthesis is a reflection of the natural selection of bad science, and possibly fast, cheap, and near-unlimited access to genetic sequencing data is adding fuel to the process.

Lessons from history

Back in my office, I'm preparing for yet another thesis defense. I have chugged through six chapters, covering three distinct species, five different next-generation sequencing datasets, and four publications, and not one clear hypothesis is discernible from the document. My sense is that the student has never seen or worked with the organisms described in the dissertation, a crime that I have been guilty for as well [9]. I look up at my dad's thesis on the shelf and remember how he described to me the burgeoning popularity in the 1960s and 1970s of X-ray crystallography for solving the structures of proteins and other biological macromolecules.

"It sounds a lot like your genome papers, sport", he said. "When I was a graduate student, lots of chemists specialized in crystallography and would pump out one structure paper after another with no clear reason for doing so aside from getting more publications". He described how, like today's genome paper junkies, many of his classmates built their entire PhDs on the backs of unrelated crystallography papers. In fact, the field of crystallography has many parallels with that of genomics, and its history can act, in certain ways, as an example and forewarning for genomicists.

In a special issue of *Nature* celebrating the 100th anniversary of the technique, Laura Cassidy describes how "until recent decades, only specialists with years of training and expensive equipment could perform X-ray crystallography. But in the 1990s, the technique became much more accessible … improvements in methods for solving structures and a boost in computing power greatly sped up the process, giving researchers time for other scientific pursuits" [10]. Cassidy notes how increased competition for research grants forced crystallography laboratories to become more well-rounded. "Instead of just solving one structure after another, researchers must now link the structure of a molecule to its function through biochemistry and cell-biology experiments" [10]. Most importantly, she stresses how "crystallography work increasingly requires a good scientific question rather than just solving structures". One could similarly argue that nucleotide sequencing experiments should be used to tackle important scientific topics rather than to just assemble an endless array of genomes and transcriptomes.

"Most would agree that the best scientists ask the right questions and that the best PhD theses address such questions."

Most would agree that the best scientists ask the right questions and that the best PhD theses address such questions. One question that I cannot shake from my mind is what will become of the Frankenthesis? Will it grow even more malformed over time or will it fade away? Will scientists come to accept it as the standard model for a dissertation or will they revolt against it? Personally, I hope that the denouement of the Frankenthesis mirrors that of Mary Shelly's gothic classic: "Thou didst seek my extinction, that I might not cause greater wretchedness … My spirit will sleep in peace; or, if it thinks, it will not surely think thus. Farewell".

Acknowledgement

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

1. Fuhrmann C, Halme D, O'Sullivan P, Lindstaedt B (2011) Improving graduate education to support a branching career pipeline: recommendations based on a survey of doctoral students in the basic biomedical sciences. *CBE Life Sci Educ* 10: 239 – 249
2. Taylor M (2011) Reform the PhD system or close it down. *Nature* 472: 261
3. Cyranoski D, Gilbert N, Ledford H, Nayar A, Yahia M (2011) The PhD factory. *Nature* 472: 276 – 2799

4. Roach M, Sauermann H (2017) The declining interest in an academic career. *PLoS One* 12: e0184130
5. Smaldino PE, McElreath R (2016) The natural selection of bad science. *R Soc Open Sci* 3: 160384
6. Ioannidis JPA (2012) Why science is not necessarily self-correcting. *Perspect Psychol Sci* 7: 645 – 654
7. Steen RG, Casadevall A, Fang FC (2013) Why has the number of scientific retractions increased? *PLoS One* 8: e68397
8. Baker M (2016) 1,500 scientists lift the lid on reproducibility. *Nature* 533: 452
9. Smith DR (2013) Are we divorced from the species we study? *F1000Res* 2: 254
10. Cassiday L (2014) More than a crystallographer. *Nature* 505: 711 – 713