

T2T-CHM13 – adds nearly 200 million base pairs to the 2013 version of the human genome sequence.

This time, instead of taking DNA from a living person, the researchers used a cell line derived from a complete hydatidiform mole, a type of tissue that forms in humans when a sperm inseminates an egg with no nucleus. The resulting cell contains chromosomes only from the father, so the researchers don't have to distinguish between two sets of chromosomes from different people.

Miga says the feat probably wouldn't have been possible without new sequencing technology from Pacific Biosciences in Menlo Park, California, which uses lasers to scan long stretches of DNA isolated from cells – up to 20,000 base pairs at a time. Conventional sequencing methods read DNA in chunks of only a few hundred base pairs at a time, and researchers reassemble these stretches like puzzle pieces. The larger pieces are much easier to put together, because they are more likely to contain sequences that overlap.

T2T-CHM13 is not the last word on the human genome, however. The team had trouble resolving a few regions and estimates that about 0.3% of the genome might contain errors. There are no gaps, but Miga says quality-control checks have proved difficult in those areas. And the sperm cell that formed the hydatidiform mole carried an X chromosome, so the researchers have not yet sequenced a Y chromosome, which typically triggers male biological development.

T2T-CHM13 represents only one person's genome. But the T2T Consortium has teamed up with a group called the Human Pangenome Reference Consortium, which aims over the next 3 years to sequence more than 300 genomes from people all over the world. Miga says that the teams will be able to use T2T-CHM13 as a reference to understand which parts of the genome tend to differ between individuals. They also plan to sequence an entire genome that contains chromosomes from both parents, and Miga's group has been working on sequencing the Y chromosome, using the same new methods to help fill in gaps.

Miga expects that genetics researchers will quickly find out the whether newly sequenced areas and possible genes are associated with human diseases. "When the human genome came out, we didn't have the tools poised and ready to go," she says, but information about the function of the newly sequenced genes should come much faster now, because "we've built up a ton of resources".

She hopes that future human genome sequences will cover everything, including the newly sequenced sections – not just the parts that are easy to read. This should be easier now that the reference genome has been completed. "We need to reach a new standard in genomics where this isn't special, but routine," she says.



A university investigation found that Geoffrey Marcy had infringed sexual-harassment policies.

US SCIENCE ACADEMY EXPELS MEMBER OVER HARASSMENT COMPLAINT

Astronomer Geoffrey Marcy is first to be kicked out for violating elite group's amended code of conduct.

By Nidhi Subbaraman

The US National Academy of Sciences (NAS) has terminated astronomer Geoffrey Marcy's membership, in light of sexual-harassment complaints – the first time the respected group has expelled a member.

The action comes two years after the NAS introduced a code of conduct that would allow the organization to expel members "for the most egregious violations ... including for proven cases of sexual harassment".

The 158-year-old academy changed its by-laws following pressure from the scientific community and after the #MeToo movement, which highlighted pervasive workplace harassment and institutional failures to prevent it. In 2018, a report by the US National Academies of Sciences, Engineering, and Medicine warned that sexual harassment is widespread in academic science.

Marcy, an exoplanet researcher, resigned from his tenured position at the University of California, Berkeley, in 2015, after BuzzFeed News reported that a university investigation had found that he had infringed sexual-harassment policies in several cases over nearly a decade, until 2010.

Responding to his NAS expulsion, Marcy

told *ScienceInsider* that he has been "completely out of organized academia for over 5 years", and that he "always supported equal opportunity and success for women in academia and science".

"My engaging and empathic style could surely be misinterpreted, which is my fault for poor communication," he said. "I would never intentionally hurt anyone nor cause distress." (Marcy did not respond to *Nature's* request for comment.)

Last September, *Nature* reported that, despite the NAS's by-law amendment, the organization was yet to expel any harassers, even though there were public reports of investigations and findings involving current members. NAS president Marcia McNutt told *Nature* at the time that the organization requires that a complaint be filed before it can adjudicate on a member's status, and that no complaints had been filed.

That report prompted François-Xavier Coudert, a chemist at the French national research agency CNRS in Paris, to e-mail the organization and file complaints about four scientists, including Marcy, citing findings of inappropriate behaviour that had been reported in the press. "I found it was ridiculous as an argument for the academy to say, 'We have a policy, but no one is filing a complaint,'"

News in focus

says Coudert. He had previously posted about the *Nature* story on Twitter; McNutt replied, urging him to take action.

“The NAS has chosen a policy that is very weak and that protects them in a way,” says Coudert.

The academy has said in the past that it does not have the resources for formal investigations, apart from for internal NAS business. The group relies on publicly documented investigations carried out by other organizations to begin inquiries into its members.

The NAS informed Coudert of Marcy’s termination last month; the chemist says it is a preliminary step in the right direction.

Membership of the highly selective NAS is regarded as a top honour in US science, burnishing the profile of elected members. It also confers a degree of influence – the group is regularly tapped by US agencies to offer scientific views on national affairs.

Seyda Ipek, a theoretical particle physicist at the University of California, Irvine, also submitted a complaint last September, including public details of harassment investigations and findings concerning Marcy. “It’s really important to not allow these people in these prestigious communities, because they are doing bad things for science,” says Ipek. She

says she was surprised and angry to learn that scientists continued to collaborate with the astronomer, pointing out that manuscripts posted on the arXiv preprint server in the past six months still listed Marcy as a co-author. “Where is the justice for women pushed out of the field if people continue to work with him?”

Some of those papers point to Berkeley as Marcy’s affiliation. A Berkeley spokesperson

“Where is the justice for women pushed out of the field if people continue to work with him?”

says that Marcy is currently a retired professor at the university, and that University of California policy allows retirees to refer to themselves as emeritus faculty members at those institutions. They added that the university’s 2015 announcement of Marcy’s resignation was accurate at that time.

A spokesperson for the NAS confirmed that Marcy’s membership had been rescinded as of 24 May. They did not say how many other members were under review as a result of sexual-harassment complaints.

of Toulouse, France, came up with a new idea: searching for key grammatical phrases characteristic of SCIGen’s output. Last May, he and Cabanac searched for such phrases in millions of papers indexed in the Dimensions database.

After manually inspecting every hit, the researchers identified 243 nonsense articles created entirely or partly by SCIGen, they report in a study published on 26 May (G. Cabanac and C. Labbé *J. Assoc. Inf. Sci. Technol.* <https://doi.org/gj7b8h>; 2021). These articles, published between 2008 and 2020, appeared in various journals, conference proceedings and preprint sites, and were mostly in the computer-science field. Forty-six of them had already been retracted or deleted from the websites on which they were first published.

Since last year, the researchers have added another 20 papers to their list, including gibberish articles created by MATHgen (software that generates mathematics papers) and the SBIR proposal generator (which creates nonsense grant proposals).

CV padding

Most of the latest batch of SCIGen papers were authored by researchers from China (64%) or India (22%), although Labbé notes that the manuscripts could have been submitted in anyone’s name without their knowledge. One author of several of the papers told Labbé and Cabanac that he’d submitted them as hoaxes. But other manuscripts seem to have been edited with genuine reference lists, suggesting that they might have been generated to inflate scientists’ citation counts. “I think the vast majority are created to pad CVs in order to fulfil a need to publish papers,” says Labbé.

The researchers found only two SCIGen papers that hadn’t been retracted at the IEEE – which is evaluating both of them – and one Springer paper that included a fragment of MATHgen text. But other publishers were caught out more badly. IOP Publishing, a subsidiary of the London-based Institute of Physics, says it retracted ten papers “as there was clear evidence they had been computer-generated”, and is investigating why they weren’t identified during peer review at the conference for which they were accepted. “We have reasonable evidence to suggest that the peer-review process for some of these papers was compromised,” says Kim Eggleton, the publisher’s integrity and inclusion manager.

The publishers that posted the most SCIGen content were Trans Tech Publications based in Bäch, Switzerland, which published 57 SCIGen papers; Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP), based in Bhopal, India, which had 54; and Atlantis Press, a Paris-based publisher that was acquired by Springer Nature this March, with 39. Both Trans Tech Publications and Atlantis told *Nature* that they were investigating and were in the process of retracting the articles, but a spokesperson

HUNDREDS OF GIBBERISH PAPERS STILL LURK IN THE SCIENTIFIC LITERATURE

Nonsensical articles, spotted years after the problem was first seen, could lead to a wave of retractions.

By Richard Van Noorden

Nonsensical research papers generated by a computer program are still popping up in the scientific literature many years after the problem was first seen, a study has revealed. Some publishers have told *Nature* they will take down the papers, which could result in more than 200 retractions.

The issue began in 2005, when three PhD students created paper-generating software called SCIGen for “maximum amusement”, and to show that some conferences would accept meaningless papers. The program cobbles together words to generate research articles with random titles, text and charts, easily spotted as gibberish by a human reader. It is free to download, and anyone can use it.

By 2012, computer scientist Cyril Labbé had found 85 fake SCIGen papers in conference

proceedings published by the Institute of Electrical and Electronic Engineers (IEEE); he went on to find more than 120 fake SCIGen papers published by the IEEE and by Springer (C. Labbé and D. Labbé *Scientometrics* **94**, 379–396; 2013). It was unclear who had generated the papers or why. The articles were subsequently retracted – or sometimes deleted – and Labbé released a website allowing anyone to upload a manuscript and check whether it seems to be a SCIGen invention. Springer also sponsored a PhD project to help spot SCIGen papers, which resulted in free software called SciDetect. (Springer is now part of Springer Nature; *Nature*’s news team is editorially independent of its publisher.)

Labbé, who works at the University of Grenoble Alpes in France, originally searched manuscripts for words typical of SCIGen’s vocabulary. But he and another computer scientist, Guillaume Cabanac at the University