CRIMEFIGHTING WITH FAMILY TREES

Parabon Nanolabs shot to fame in a controversial field, using DNA and genealogy analysis to catch criminals. Then it hit a setback. **By Carrie Arnold**

It was May 2019 when Parabon Nanolabs ran into a major controversy. At the time, it was the most famous forensic-genetics company on the planet. From its headquarters in Reston, Virginia, Parabon was helping police to crack cold-crime cases almost weekly, such as the murder of a Canadian couple in 1987 and the case of a young woman who was sexually assaulted and killed in the 1960s.

The company had made its name by comparing suspects' DNA to profiles on genealogy databases and piecing together family trees to track down alleged offenders.

But then controversy erupted over a case Parabon helped to solve, in which a teenage boy had violently assaulted a septuagenarian in a Mormon meeting house in Utah. The Utah case generated public backlash because of concerns over privacy.

Genealogists at Parabon had been generating leads by sifting through a database of DNA tests called GEDMatch, a free-to-use website that allows users to upload test results in the hope of finding long-lost relatives. At the time, GEDMatch allowed law-enforcement agencies access to the profiles to help solve murders and sexual assaults, unless users specifically opted out. The police, aided by Parabon and companies like it, made new arrests weekly.

But the Utah case was not a murder or a sexual assault – and so was not covered by the website's disclaimer. The assailant had left

traces of blood at the scene, and the detective in charge of the case, Mark Taggart, made a personal plea to GEDMatch's founder, Curtis Rogers, for access to the database. When it was granted, Parabon, which had initially refused the case, signed on. The company traced several partial DNA matches to individuals living in the area, and narrowed in on a suspect, a teenaged boy who was a relative of one of them. Taggart made an arrest.

That triggered an immediate backlash from genealogists, privacy experts and the wider public at the violation of GEDMatch's agreement with its users. In response, Rogers required the site's millions of users to specifically opt in to law-enforcement use. Overnight, Parabon lost its major source of DNA data.

That proved to be a challenge for the company, and for forensic genetic genealogy. In the year since then, the restrictions on GEDMatch's data have forced Parabon to forge ahead while navigating new controls limiting access to genealogy data. It has also continued work on another strategy: attempting to use DNA to reconstruct faces. At the same time, it is facing competition from forensic-geneaology companies that are trying to stake their own claims in the field.

Parabon acknowledges that the rule change at GEDMatch substantially restricted its main source of DNA data, but it says this was a temporary setback. It adds that it has continued to solve cases using data from another genealogy

company and with GEDMatch data from people who have opted to allow its use in criminal cases.

Just as the prominence of forensic genetic profiling has grown, so has its notoriety. Ethicists have raised concerns over China's use of genetic profiling to target the Uyghurs, a predominantly Muslim minority population in the country's northwestern provinces. In the past year, the US government has launched two programmes that have begun taking DNA samples from immigrant detainees and some asylum seekers. The US Department of Justice issued guidelines last November that tried to set boundaries on the use of forensic genetic genealogy, but concerns about police brutality and systemic racism against Black Americans have raised questions as to whether these guidelines provide enough protection to people of colour, who are disproportionately stopped by police and overrepresented in criminal DNA databases. These legal, ethical and social concerns have left industry experts wondering what's next for forensic genomics.

"Because DNA is so powerful, we tend to see

Traces of crime-scene DNA have been matched to suspects using genealogy databases.

it as a silver bullet," says Yves Moreau, a biologist and engineer at the Catholic University of Leuven in Belgium. But law-enforcement agencies are using databases and techniques not designed for solving crimes or generating leads, he says. "It's like a knife — people underestimate just how sharp they can be."

Family ties

In December 2017, genetic genealogist Barbara Rae-Venter got the call that would propel family-tree forensics into the public eye. She was running a business that used GEDMatch to find clients' long-lost relatives when she heard from a California detective who had found some old DNA evidence and was trying to reopen the case of the Golden State Killer, a serial rapist and murderer who committed a string of crimes in the 1970s and 1980s.

Combining DNA samples with family trees is the core of forensic genetic genealogy. The process rests on the simple statistical rules of genetics. A parent and child, or two siblings, share 50% of their DNA. Grandparents and grandchildren share 25%. Even distant relatives share small portions of DNA. This allows consumer genetic-testing companies such as Ancestry in Lehi, Utah, and 23andMe in Sunnyvale, California, to estimate relationships between two individuals who have submitted samples, as far out as fourth cousins (who share a pair of great-great-great grandparents). Anyone can upload the results of their own DNA test to databases such as GEDMatch.

Rae-Venter found two GEDMatch profiles that looked to be distant cousins of the suspect, and used that information to work backwards and find their great-grandparents.

"

WE'RE STILL ASKING WHETHER THESE TECHNIQUES ARE SCIENTIFICALLY VALID."

Then, she moved forward in time to trace their descendants, focusing on California during the time the crimes were committed. After two months, Rae-Venter handed the detective the names of three brothers. DNA from a cigarette discarded by one brother matched the sample, and on 24 April 2018, police arrested Joseph DeAngelo — in the first criminal case to be solved using the technique. (DeAngelo pleaded guilty to multiple counts of rape and murder and was sentenced to life in prison last month.)

Following DeAngelo's arrest, forensic genetic genealogists such as Rae-Venter and CeCe Moore (who joined Parabon in May 2018) helped to solve similar rape and murder cases at a rapid clip. Although a few ethicists raised concerns about privacy, media coverage of the cases was overwhelmingly positive. "I was actually surprised there wasn't more criticism," says geneticist Ellen McRae Greytak, bioinformatics chief at Parabon.

And then the Utah case hit the media, and the criticism came crashing in.

Active case

Late on Saturday 17 November 2018, 71-year-old Margaret Orlando dialled 911 from a Mormon meeting house in Centerville, Utah. Someone had thrown a rock through a window, climbed in, and attacked her as she was practising the organ, strangling her until she passed out. Taggart was called to the scene, where he found three drops of blood, presumably from her attacker having cut himself on the broken glass. The DNA profile didn't match anyone in state and federal databases, but a chance conversation with a genealogist friend gave Taggart hope: if police couldn't identify the suspect, perhaps they could track down a relative. He reached out to GEDMatch and got permission to use the site.

In the same way that Rae-Venter helped to identify the Golden State Killer, Parabon provided Taggart with three possible names, one of which he recognized right away. The man, who lived near the meeting house, had had several run-ins with the police, and Taggart discovered that he had a 17-year-old nephew living with him — a nephew who matched the description the organist had given.

The next day, Taggart managed to get a DNA sample from a milk carton the suspect had thrown in the rubbish at school. It matched. So did a follow-up swab. Taggart arrested the suspect (whose name was not disclosed as he was a minor) on 24 April 2019 — one year to the day after the arrest of the Golden State Killer. "It was like a puzzle coming together," he says.

With the relief, however, came the publicity. "We were a little surprised at how positive the response was to the Golden State Killer and how negative the response was to this," Greytak says. She points to a study in *PLoS Biology*[1] that found 90% of Americans supported police use

Migrants detained at the US border. The government takes DNA from some asylum seekers.

of forensic genetic genealogy, and says that a small but vocal group led the outcry against the Utah case.

Ethicist Matthias Wienroth at Northumbria University in Newcastle, UK, sees it differently. Wienroth raised privacy concerns about this type of search almost as soon as news of the arrest broke. It's your right to relinquish some of your own privacy by uploading your DNA profile to sites such as GEDMatch, Wienroth says, but these sites also reduce the privacy of some of your distant relatives. Indeed, the proliferation of at-home DNA tests has made some genetic genealogy databases so large that a 2018 *Science* paper[2] estimated that the troves could identify 60% of North Americans of European descent, even if they had never themselves taken one of these tests. Greytak and Armentrout say that they have uploaded their own results to GEDMatch and are untroubled by the idea that they might incriminate a distant relative.

"We're still asking whether these techniques are scientifically valid. No one's talking about failures — all I ever hear about are the successes," Wienroth says. He points to the fact that the California police first chased leads from a different branch of the family tree before they realized their mistake and focused on DeAngelo.

But Greytak doesn't see that as a failure. She says that investigative genetic genealogy was never intended to serve as the final answer in a case. Instead, she sees it as a tool to help law enforcement to generate leads.

Face value

Steven Armentrout started Parabon in his basement to provide supercomputing services. Parabon's first big breakthrough was in 2011, when the fledgling company applied for a US Department of Defense (DoD) grant to try to reconstruct a person's appearance from their DNA — a technique called DNA phenotyping. The DoD wanted to develop the technology to identify makers of improvised explosive devices from the tiny amounts of DNA left on bombs, but they also knew that law enforcement would be interested. Most labs studying DNA phenotyping look for relationships between changes to individual letters of a person's genetic code, known as

> ## "YOU HAVE A RIGHT TO PRIVACY. YOU ALSO HAVE THE RIGHT NOT TO BE MURDERED OR RAPED."

single-nucleotide polymorphisms (SNPs), and physical characteristics such as eye or hair colour. But Parabon framed the challenge as a machine-learning exercise. Its plan was to collect a large number of DNA samples and face photographs, and train algorithms to pick out relationships. Parabon got the grant.

Its approach worked well with large amounts of high-quality DNA from blood samples and cheek swabs. But forensic samples are often small and degraded. When Armentrout hired Greytak in 2014, the company's first goal was to see whether commercial genotyping arrays could get information from forensic samples. When Parabon sent out its first sample, the lab manager phoned and said it would never work. The chips needed 200 nanograms of DNA.

"In the forensics world, 200 nanograms is a truckload," Armentrout says. Parabon had sent a sample with just 1 nanogram. Everyone involved — including Armentrout and Greytak — was surprised to find that it worked. Parabon says it can now sequence enough SNPs to trace family history and build a face with less than 1 nanogram of DNA. Greytak says that the sequencing runs that use such scant quantities of DNA often leave parts of the genetic code blank because the sample is too degraded or too dilute to read. The company's response was to build proprietary algorithms to anticipate such blank spots in its mathematical models. Greytak says that lower-quality DNA can sometimes mean that predictions are made with less confidence — but that problems are rare.

Parabon's goal was ambitious: rather than just telling police that a suspect had fair hair and green eyes, it wanted to provide a comprehensive analysis of someone's ancestry and a composite facial sketch from a DNA sample. The procedure, dubbed Snapshot, was released in December 2014. Parabon says that since 2018 the police have solved more than 120 cases with the help of their genetic genealogy and phenotyping methods (the company declined to disclose the total number of cases for which they were used, citing ongoing investigations).

Other companies have also developed DNA-phenotyping strategies, including the now-defunct Identitas, which specialized in predicting physical appearance using SNPs, and Illumina, the DNA-sequencing giant in San Diego, California, that spun off its forensics branch into a new company, Verogen, also in San Diego, in 2017.

Several academic labs are also researching DNA phenotyping. At Erasmus University Medical Center in Rotterdam, the Netherlands, Manfred Kayser (once an adviser to Identitas) developed IrisPlex in 2011 to predict eye colour from DNA[3]. Since then, his team has added more SNPs to capture more genetic variation and to add other identifiable characteristics, such as hair colour and texture. The Netherlands police began using Kayser's techniques once they were vetted in the scientific literature. The most famous example was in 2012 when they showed that the rape and murder of 16-year-old Marianne Vaatstra was probably not committed by a member of a refugee settlement located close to where her body was discovered.

Unlike Parabon, Kayser does not attempt to weave together different features to try to recreate a person's face. Instead, he uses the individual traits (say, auburn hair and hazel eyes) as law-enforcement leads. He finds Snapshot to be problematic because the technology hasn't been evaluated in the peer-reviewed literature. "It's very limited, what we know about the face, and this particular company says they can predict it from DNA. It's pretty bad that they don't publish how they do this and how they validated this," Kayser says. Scientists have published hundreds of papers about the relationship between specific genetic

variants and physical features, Kayser says, but researchers still don't know how these individual traits become a unique human face.

Mark Shriver, a geneticist who researches DNA phenotyping at Pennsylvania State University (Penn State) in University Park, says that because the effects of ancestry on facial appearance are so strong, he suspects that Parabon's data are creating a set of average, generic faces that the company then tweaks to fill in the blanks. Without seeing the data and algorithms the company uses in its machine-learning system, Shriver says, "we don't know whether their ability to estimate a face's appearance is better than chance, or if it's an approximation based on what we know about ancestry".
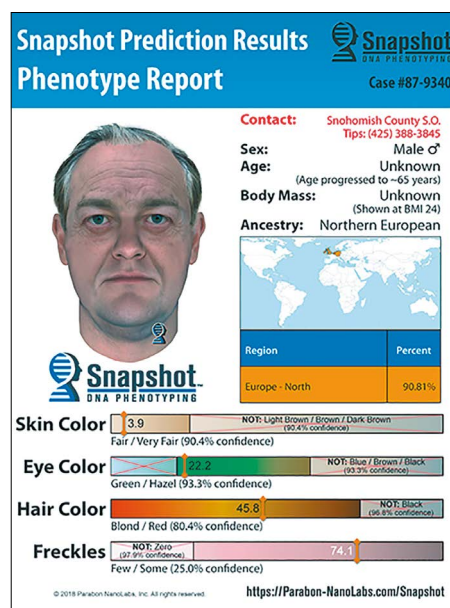
Armentrout says that Parabon doesn't need to know how each gene contributes to appearance in order to create the image of a face; he says the associations between SNPs and faces in the company's database is good enough for its mathematical models, and that police-department satisfaction is all the proof he needs. Just because the firm doesn't publish doesn't mean its method is flawed, Armentrout says. "We're not in business to write papers," he says. "The results speak for themselves." But Shriver says that making an arrest doesn't mean that Snapshot works as Parabon claims. Nor do the police have a rigorous way to show that the Snapshot profile matches their suspect, he says.

Forensic future

While Parabon was adding DNA phenotyping to its portfolio, other companies, including Verogen and commercial DNA-testing company FamilyTreeDNA in Houston, Texas, began testing the waters with forensic genetic genealogy. Last December, Verogen announced it had bought GEDMatch, which now has 280,000 of its 1.45 million DNA profiles opted in to police searches. Chief executive Brett Williams says that Verogen recognized GEDMatch as the linchpin to forensic genetic genealogy, and wanted to safeguard the company's access. What this means for Parabon and the millions of private GEDMatch users remains to be seen, but Williams says he's committed to striking a balance between privacy and safety. "You have a right to privacy. You also have the right not to be murdered or raped," Williams says. This July, however, GEDMatch was hacked and users' opt-out settings were overridden for a few hours, potentially exposing their data to law-enforcement searches without their consent. In a statement, Verogen said that it had taken down GEDMatch "until such time that we could be absolutely sure that user data was protected against potential attacks".

There have been attempts to gain access to users' profiles through official channels, too. A detective in Orlando, Florida, announced last October that he had obtained a search warrant to use all GEDMatch profiles to try to find relatives from DNA left by a suspect. Genealogy company Ancestry successfully fought against a Pennsylvania search warrant this February. Williams says he will fight against any warrants Verogen receives in the future. In the meantime, the US Department of Justice has issued interim guidelines to help police with their use of forensic genetic genealogy, permitting use of the technology only for serious violent crimes such as rape and murder, and only after other leads have been exhausted. Notably, the document specifies that suspects

Snapshot Prediction Results Phenotype Report — Snapshot DNA PHENOTYPING — Case #87-9340

Contact: Snohomish County S.O. Tips: (425) 388-3845
Sex: Male ♂
Age: Unknown (Age progressed to ~65 years)
Body Mass: Unknown (Shown at BMI 24)
Ancestry: Northern European

Region	Percent
Europe - North	90.81%

Skin Color 3.9 — Fair / Very Fair (90.4% confidence) — NOT: Light Brown / Brown / Dark Brown (90.4% confidence)
Eye Color 22.2 — Green / Hazel (93.3% confidence) — NOT: Blue / Brown / Black (93.3% confidence)
Hair Color 45.8 — Blond / Red (80.4% confidence) — NOT: Black (96.8% confidence)
Freckles 74.1 — Few / Some (25.0% confidence) — NOT: Zero (97.9% confidence)

© 2018 Parabon NanoLabs, Inc. All rights reserved. https://Parabon-NanoLabs.com/Snapshot

Parabon's Snapshot tool uses DNA to reconstruct faces. This suspect was later convicted of a 1987 murder.

cannot be arrested on genealogy alone — conventional forensic genetics must be used to provide a conclusive match.

Sociologist Helena Machado at the University of Minho in Braga, Portugal, isn't against law-enforcement use of genetic genealogy or DNA phenotyping, but says she's concerned that work linking genealogy and crime might lead to biases against certain families or ethnic groups. "It might reinforce the idea that there is a higher prevalence of criminality in certain families," she says. An overemphasis on the links between genetics and crime means that researchers could be less likely to focus on the social and economic factors that lead to lawbreaking.

Both Armentrout and Kayser say that DNA technologies could help to reduce police bias by providing concrete evidence to bolster eyewitness accounts, and that DNA phenotyping could decrease racial profiling by providing more details on a potential suspect's appearance to police.

But sociologist Amade M'charek at the University of Amsterdam says this thinking is naive, especially given the incidence of police brutality against people from racial minorities.

"If we don't know the individual, often all we see is race," she says.

M'charek's concerns are not unfounded: these technologies are already being used to target and discriminate against people from minority groups, Moreau says. The US Department of Homeland Security announced in January that its Immigration and Customs Enforcement (ICE) division had launched a pilot programme to collect DNA from immigrant detainees and upload the resulting sequences to the Federal Bureau of Investigation's official forensic DNA database, the Combined DNA Index System (CODIS). The initiative joined last year's announcement that homeland security would be using 'rapid DNA technology' to test whether families applying for asylum were relatives. (ICE did not respond to requests for comment.)

In China's northwest, officials are using genetic ancestry to identify members of the Uyghur minority group. In July 2017, as part of China's Physicals for All programme, the government began collecting iris scans, fingerprints and DNA of everyone between the ages of 12 and 65 in the Xinjiang Uyghur Autonomous Region. The programme has been criticized by human-rights groups. Dispatches from Xinjiang from the non-governmental organization Human Rights Watch in New York City, reported that more than one million Uyghurs have so far been placed in detention camps. "When you give any authority such important information and such strong leverage against individuals, you start to worry very, very much about the shape society's going to take," Moreau says. "You put people in a database because you want to control them." Some Chinese scientists, says Moreau, are also working to turn Uyghur DNA into facial portraits, just as Snapshot does. Parabon says it is not involved in the Chinese research.

Despite the controversy over the Utah case — or perhaps because of it — Rogers is bullish about the future of genetic techniques in forensics. "I think that in time — and probably not very long — people will accept that law-enforcement use of genetic genealogy is there and not to be feared," he says.

For his part, Taggart doesn't regret using GEDMatch. The suspect he narrowed in on pleaded guilty and is still in detention, and Taggart is confident that his community is safer that way. "I believe that Curtis Rogers doing this for us saved a life."

Carrie Arnold is a science journalist based near Richmond, Virginia.

1. Guerrini, C. J., Robinson, J. O., Petersen, D. & McGuire, A. L. *PLoS Biol.* **16**, e2006906 (2018).
2. Erlich, Y., Shor, T., Pe'er, I. & Carmi, S. *Science* **362**, 690–694 (2018).
3. Walsh, S. *et al. Forensic Sci. Int. Genet.* **7**, 98–115 (2011).

Correction
This Feature erred in saying the Golden State Killer case was solved using data from two second cousins; they were distant cousins. It also mistakenly said that the assault of Margaret Orlando would have been a crowning achievement for the company, as the first active case that it had helped to solve. In fact, Parabon had helped to solve other active cases — this was the first to contravene the GEDMatch terms of use. Finally, the story stated that the restrictions on GEDMatch data forced Parabon to return to an earlier business strategy of using DNA to reconstruct faces. However, Parabon says that the restrictions were a temporary setback and didn't significantly affect its business.