

Share Tweet

AFFILIATED REPORTS COLLECTION
Global Bioinformatics Partnering, Terms & Agreements
 ORDER TODAY

AFFILIATED REPORTS COLLECTION
Cloud Testing Market - Research Report Forecast
 ORDER TODAY

AFFILIATED REPORTS COLLECTION
Clinical NGS Process Report
 ORDER TODAY

ONE STOP SHOP
 for over 2,000 Life Science Research Reports spanning over 80 Research Areas
 INSIGHT PHARMA REPORTS

By Melissa Kruse

September 11, 2003 | Two years ago, 2,792 lives were lost in the collapse of the World Trade Center. While rescuers labored night and day to recover the bodies, a small Michigan software company set about salvaging their identities.

Here, a team of eight software engineers would work around the clock. Several times a day, the tangle of electrical wires and cords would accidentally be kicked out of the power strip, crashing the network until one of the programmers crawled underneath the table and restored power.

This was the unlikely setting of a revolution in mass fatality identification science.

Staffers at Gene Codes, the Ann Arbor, Mich., bioinformatics company, had witnessed the carnage of Sept. 11, 2001, on a television that Howard Cash, the company's founder and CEO, hastily purchased that fateful day. They saw what the world saw, but were about to be charged with describing it — in numbers.

The tools for the task were as bizarre as the environment in which it was carried out. Toothbrushes, razors, and clothing provided reference samples that might yield a billionth of a gram of DNA — invaluable clues **Men in black:** Gene Codes engineer Dave Relyea (left) and CEO Howard Cash.



For forensic biologists, the Sept. 11 attacks on the World Trade Center (WTC) marked the largest mass fatality identification project in history. Families and friends of the victims demanded answers — and closure. While the relentless recovery effort at Ground Zero produced many heroes, others were to be found far from the spotlight.

Heading the 9/11 Call

Four weeks after the disaster, on October 8, Robert Shaler, director of the Department of Forensic Biology at the Office of the Chief Medical Examiner (OCME) of New York City, asked Cash to create catharsis out of code.

Cash flew to New York expecting to donate some existing software to the recovery effort. But Shaler needed new software that would inventory and match the victims' remains, as well as catalog the reference samples required to name the dead — and reunite them with their families.

The deadline for this unprecedented task: yesterday.

Cash warned his staff that identifying the victims of 9/11 would be a 24/7 marathon. And there could be absolutely no mistakes. His 16 colleagues readily agreed, and after a few additional hires, they set about reinventing the science of DNA mass identification. "We'd do it again in a heartbeat," Cash says. "Every employee, every shareholder was completely behind it," even though it would mean enduring arduous 12-hour programming shifts (see "The Right Decision").

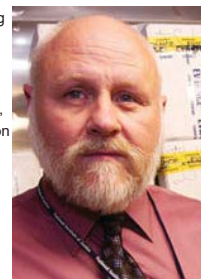
Whether out of patriotism or professionalism, staffers routinely arrived for work at 7 a.m. and left at midnight. Engineers like Dave Relyea just wanted to help. "We thought about the victims, the families, and the people at the Office of the Chief Medical Examiner working around the clock. What they were going through made us feel like we could never work hard enough."

Occasional relief came in the form of some old toys lying around the office. A boxing nun became their "integration token." Staff could submit source code changes, but only if they had the nun. A boxing Godzilla was also on hand, although there was little use of the Nerf guns.

Every Friday, Cash flew to New York to deliver the latest software release to Shaler, returning the following Monday. His colleagues are convinced that Cash subsisted on coffee and airline pretzels.

On Dec. 13, 2001, Gene Codes' Mass Fatality Identification System "M-FISys" (pronounced "emphasis") was born. M-FISys combines DNA profiles from three sources: victims' personal effects (toothbrushes, razors, combs, etc.); kinship references (relatives' cheek swabs); and the remains themselves. The software is able to crossmatch thousands of DNA profiles in minutes, a task that previously might have taken two weeks.

That first day alone, the OCME, normally about the business of solving homicides and sexual assaults, made 55 victim identifications from Ground Zero. However, while M-FISys can present almost all the knowledge necessary for making identifications, only the chief medical examiner, Charles Hirsch, can legally determine the identity of each set of remains.



Tall order: Robert Shaler, head of forensic biology at the OCME, presented Gene Codes with its mission.

INSIGHT PHARMA REPORTS
 Expert Intelligence for Better Decisions

Over 1000 Scientific Reports Focusing on:

- Biomarker & Diagnostics
- Bioprocess & Manufacturing
- Cancer Research & Therapeutics
- Chemistry
- Clinical Trials
- Drug Discovery & Development
- Drug & Device Safety
- Genomics
- Healthcare
- IT & Informatics
- Translational Medicine
- And many other research areas

InsightPharmaReports.com

CSI Manhattan

The collapse of the twin towers did not merely kill thousands of people; many literally disappeared. Although most victims left behind traces of DNA (albeit often highly damaged), those who were vaporized or pulverized may never be identified, because they left nothing behind.

The WTC foundation was a six-story, 70-foot-deep, watertight shell designed to keep the New York Harbor at bay. Following the attacks, 1.6 million tons of debris, water, bodies, and burning jet fuel collected in the well, as if putting everything in a pot and cooking it. Fires burned for three months, stewing some of the remains for as long as nine months.

Conventional means of identification such as dental records became virtually irrelevant. "If someone's been lying in burning jet fuel for three months, it's harder to figure out," Cash says. "DNA is relatively fragile stuff."

Only 287 nearly intact bodies were recovered from Ground Zero. The remains of one individual were found scattered in nearly 200 places. These remains were collected in 16 refrigerated trailers backed under a tent dubbed "Memorial Park."

The identification effort required new technologies to salvage readable sequence from the remnants of DNA. To identify individuals, forensic biologists usually measure the length of microsatellite, or short tandem repeat (STR), patterns in nuclear DNA. These are naturally occurring variations in the length of 13 to 15 stretches of repetitive DNA strewn across the human genome (see "Forensic Profiling," right).

As many of the DNA samples were badly degraded, M-FISys also included the option of matching the genetic patterns of mitochondrial DNA (mtDNA) and single nucleotide polymorphisms (SNPs).

Each human cell contains 1,000 mitochondria, each carrying a loop of mtDNA some 16,500 bases long. Two highly variable regions of about 1,000 bases within the mtDNA are checked. Unrelated individuals might have a handful of single-base differences within those regions. But about 7 percent of the Caucasian population has the identical sequence.

SNP technology is still under development at Orchid Cellmark, a Dallas-based business unit of Orchid BioSciences dedicated to forensic DNA testing services, but after hearing about Orchid's SNP panel for parental testing, the New York medical examiner asked if the technology could be used on WTC specimens. Orchid developed a technique to examine mere 100-base DNA fragments, rather than the standard 100- to 400-base lengths, which helped with such degraded samples.

According to Orchid BioSciences' Bob Giles, a panel of 71 SNPs provides more powerful identification on average than a full profile of 13 STR loci. "Many samples recovered at the WTC gave a partial STR profile, but typically four to five markers are not enough to make an identification. If you couple a partial STR profile with 20 to 30 SNP markers, identification is feasible. It's not an either-or; in some situations, it's both."

So far, Orchid has tested 2,500 tissue specimens and continues to study the remaining samples. The SNP data are sent to Gene Codes to be incorporated into M-FISys so OCME analysts can plug gaps in the STR profiles if necessary.

Bodes Well

Another key contributor to M-FISys and the identification project was The Bode Technology Group in Virginia. Mitchell Holland, Bode's laboratory director, also received an urgent call from Shaler in September 2001 asking "if we could develop a system for processing a very large number of bone samples."

Within a month, Bode was processing 1,000 bone samples per week. "Historically, we'd be lucky to do 100 to 200 per week," Holland says. "We increased that by a factor of 10 by developing a method that rapidly prepared bones for DNA extraction."

Extracting pure DNA from bones is hampered by contaminating material (e.g., soft tissue, dirt) that blocks amplification. Only when the surface is absolutely clean is a core sample taken. Bode reduced the processing time from 20 minutes to about four, boosting efficiency five- to tenfold.

Bode also enhanced the quality of results by developing a polymerase chain reaction (PCR) STR system for highly degraded DNA samples, re-engineering the PCR primers to halve the size of the target. "If you reduce the size of the target you increase the amount of DNA available for amplification." So far, Bode has performed more than 18,000 analyses and sent more than 30,000 results to the OCME.

Several other groups were also on the scene when Gene Codes arrived. Myriad Genetics, for example, had a pre-existing contract with the State Police lab in Albany to process rape kits. The Salt Lake City firm analyzed nearly 20,000 DNA samples, using a dozen high-speed sequencing machines. DNA technicians from Celera Genomics in Rockville, Md., also sequenced mtDNA samples from victims and relatives.

An IT Nightmare

Gene Codes spent October of 2001 looking at databases in New York with the medical examiner. At the Family Assistance Center, New York Police Department personnel interviewed friends and family members of victims, taking buccal swabs from family members. All swabs and personal effects were then sent to the Forensic Investigation Center (FIC) at the New York State Police headquarters in Albany. All of these "exemplars" were recorded and either tested within the FIC or shipped to collaborating commercial labs for STR analysis.

But the collection of family reference samples and personal effects bordered on chaos. Volunteers at Pier 94, a makeshift outreach center for victims' families, sometimes accepted toothbrushes with no names on them. Other family members brought personal effects or donated cheek swabs separately, resulting in some people entering the system twice. Handwritten interviews with mourning family members also slowed down data entry.

Researchers navigated through databases from FileMaker Pro to Oracle, collapsing data that were once held in 22 different laboratory databases across five states into neatly compiled aggregated profiles in M-FISys. With more than 164,000 lines of code, M-FISys links all the information in the identification project: 11,641 cheek swab samples from 7,166 family members; 7,681 personal effects and the results of the three types of DNA test; and nearly 20,000 human remains.

The Right Decision

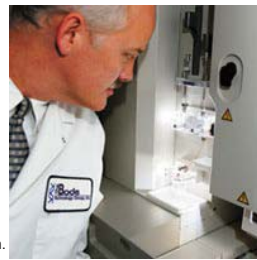
To create M-FISys, Gene Codes had to put its flagship software product Sequencher on hold for about a year and a half.

[Read More](#)



Painstaking inspection:

Workers at a landfill in Staten Island sift through debris from the World Trade Center, searching for evidence, property, and the remains of victims.



Quick work: Mitchell Holland, lab director at The Bode Technology Group, observes the Applied Biosystems 3100 Genetic Analyzer, which was used to process WTC samples.

Prior to M-FISys, the medical examiner's laboratory tried to use the Combined DNA Index System (CODIS), a tool used by the FBI to allow government labs worldwide to match criminals' DNA profiles, for victims' identification. CODIS generates a report every time two DNA patterns match. But because so many bodies were fragmented into multiple pieces, most of the reports generated by CODIS were redundant, slowing the work. And with no way to shut it off, analysts had to verify every find against multiple reports.

"We were literally running from computer terminal to terminal to verify information," recalls Elaine Mar, a criminalist and lead supervisor of the WTC DNA Identification Unit at the OCME. "It was a logistical nightmare because we had to comb through so many databases to answer a question."

Mar says M-FISys solved most of her team's problems. Each victim sample was given a number according to when it was found. DM0100001 was the first sample found in "Disaster Manhattan 2001." Numbers with similar profiles collapse into aggregates. For every profile, M-FISys calculates the probability of another person having the same one.

Initially, M-FISys was programmed not to match remains to personal effects unless the likelihood was 1010 or better. As more remains were identified, this threshold was lowered and more pieces fell into the same aggregates. "CODIS couldn't give us a red flag, but M-FISys does," Mar says. "That's one of the amazing things about M-FISys — it does the searching for us."

Extreme Programming

Because the software for M-FISys had to be developed so quickly, Gene Codes couldn't write specifications. So it adopted extreme programming (XP) to safeguard quick development against errors. In XP, programmers work in pairs, constantly checking and testing each other's code every step of the way (see "XP Philosophy," right).

At the end of each week's iteration, the staff holds a retrospective — a ritual since November 2001. They list things that worked well, and what needs improvement, on fluorescent pink, green, and yellow Post-it Notes, transforming an entire wall into a case of art imitating life. Under "Worked Well," a note says "Figured out how to use debug form on a wrapped test class." One square under the "Needs Improvement" category simply reads, "I'm tired."

Amy Sutton, Gene Codes' manager of software quality assurance, says the early weeks were tough. Her team was tasked with safeguarding the integrity of M-FISys

by automating tests to "break" it. "We have tests for crazy things they never even thought of. It's been a stressful project for most of us," Sutton says, though she seemingly thrives on the pressure.

Sutton was the first person to see the names and data together, working alone one night. "Everyone is always very aware of what the data represent," she says. "You can think you're prepared for dealing with the reality of these data, and then it hits you."

Sutton enjoys watching forensic biologists use M-FISys because it helps her to fine-tune the user experience. "Usability is our watchword here. When you're digging a ditch, any shovel will do, but one with a padded handle makes a big difference. I like to pad the handle."

Cash takes pride that his team has never made a software mistake that could result in a misidentification. "If we write a bug that destroys a computer, we buy a new one," Cash says. "If all the lights go out on the Eastern Seaboard, they'll eventually come back on. But none of these [errors] are as serious as misidentifying a person. How do you tell a family they have to give the body back, that their funeral didn't count?"

The chance of a false match based on the coincidental sharing of loci is less than 1 in 3.58 million.

The Future for M-FISys

The Gene Codes team has come a long way since M-FISys 0.1, which simply imported STR data and grouped them on screen. The latest iteration — Version 6.03 — is the 68th release of the program. "Only people who have written complex code understand how difficult it is and what an extraordinary job Gene Codes has done in providing it to us as quickly as they did," Shaler wrote in an August 2002 letter to Cash.

Once a positive ID has been made, the relatives may have a funeral director collect the remains from the OCME and request notification if additional remains are found. Others prefer not to be notified, in which case remains pass into the common memorial, individually tagged in case the family changes its mind. Families can also ask the OCME to retain the remains until all identification efforts are completed, and have the funeral director collect everything at once. Meanwhile, some families still missing any trace of their relative can request not to be informed, even if remains are discovered. Authorities have had to deal with 80 attempts at fraudulent insurance claims for supposed victims. One woman is alleged to have fled to Peru after collecting \$70,000 in compensation for falsely reporting a death of a relative. But according to the chief medical examiner, the forensic effort will continue until every victim is identified. For now, unidentified remains are being saved for testing under future technologies. M-FISys recently added "virtual profiles," which are made up of several attempts to extract DNA from the same sample.

While investigators wrap up as much as they can in New York, M-FISys is being recruited for other natural and manmade tragedies. Gene Codes has offered to donate the finished software to nonprofit forensic organizations like the International Commission on Missing Persons, which is working to identify remains from the war in Kosovo.

Gene Codes may also issue licenses to countries that would like to have the software as a national resource. "Several countries have expressed interest in M-FISys, and we will continue to do development in order to support

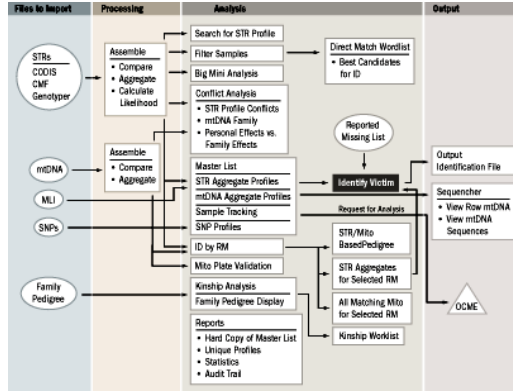


Noteworthy: Gene Codes staff keeps track of what works well and what needs improvement by using a collage of Post-it Notes.

those requests," Cash says. "But for now, the Office of the Chief Medical Examiner has absolute priority."

Approaching the second anniversary of Sept. 11, 1,521 of the 2,792 people who perished in the WTC disaster have been identified. ●

Melissa Kruse is a reporter for The Grand Rapids Press. She can be reached at melissakruse@earthlink.net.



Profiles in courage: M-FISys collects disparate DNA data for analysis, leading to a provisional victim identification. (CMF=Common Message Format. MLI=medicolegal investigator. RM=reported missing.)

PHOTO CREDITS: SPENCER PLATT / GETTY IMAGES; ROBERT SHALER BY DANIEL ACKER/BLOOMBERG NEWS; DAVE RELYEA BY BLAKE DISCHER; GENE CODES STAFF BY BLAKE DISCHER

Cambridge Healthtech Institute's Inaugural
Data Science, Precision Medicine and Machine Learning
Bringing Biomedicine and Computer Science Together
March 14-15, 2019 San Francisco, CA MOSCONE SOUTH CONVENTION CENTER

Bio-IT World WEST
CONFERENCE & EXPO
 Molecular Med
 Part of **TRI-CON**

SUBMIT A STORY
 TO **Bio-IT World**



A division of Cambridge Innovation Institute (CII)

250 First Avenue, Suite 300
 Needham, MA 02494

P: 781.972.5400
 F: 781.972.5425
 E: chi@healthtech.com



Life Science Portals

- | | |
|--|-------------------------------------|
| Biological Therapeutic Products | Drug Discovery & Development |
| Biomarkers & Diagnostics | Drug Targets |
| Biopharma Strategy | Genomics |
| Bioprocess & Manufacturing | Healthcare |
| Chemistry | IT & Informatics |
| Clinical Trials & Translational Medicine | Technology & Tools For Life Science |
| Drug & Device Safety | Therapeutic Indications |

CHI Divisions

- Conferences
- Reports & Market Research
- Barnett Educational Services
- News & Advertising
- Professional Services

Corporate Information

- Cambridge Innovation Institute
- Executive Team
- Testimonials
- Mailing List
- Careers
- Request Information
- Privacy Policy