Big Data and Bacteria: Mapping the New York Subway's DNA

Scientists in 18-Month Project Gather DNA Throughout Transit System to Identify Germs, Study Urban Microbiology

Researcher Christopher Mason from Weill Cornell Medical College collecting DNA samples at the 68th Street subway station in New York City last summer. The scientists identified hundreds of types of bacteria, most harmless, in the transit system as a way to study the microbiology of urban environments. Photo: Katie Orlinsky for The Wall Street Journal

By
Robert Lee Hotz
Feb. 5, 2015 1:07 p.m. ET

Aboard a No. 6 local train in Manhattan, Weill Cornell researcher Christopher Mason patiently rubbed a nylon swab back and forth along a metal handrail, collecting DNA in an effort to identify the bacteria in the New York City subway.

In 18 months of scouring the entire system, he has found germs that can cause bubonic plague uptown, meningitis in midtown, stomach trouble in the financial district and antibiotic-resistant infections throughout the boroughs.

Frequently, he and his team also found bacteria that keep the city livable, by sopping up hazardous chemicals or digesting toxic waste. They could even track the trail of bacteria created by the city's taste for pizza—identifying microbes associated with cheese and sausage at scores of subway stops.

The big-data project, the first genetic profile of a metropolitan transit system, is in many ways "a mirror of the people themselves who ride the subway," said Dr. Mason, a geneticist at the Weill Cornell Medical College.

It is also a revealing glimpse into the future of public health.

Across the country, researchers are combining microbiology, genomics and population genetics on a massive scale to identify the micro-organisms in the buildings and confined spaces of entire cities.

[Interactive: Mapping the Bacteria in New York's Subways]

Click on the map for a station-by-station look at the bacteria found by the scientists. Above, stations where bacteria related to sepsis were identified.

By documenting the miniature wildlife, microbiologists hope to discover new ways to track disease outbreaks—including contagious diseases like Ebola or measles—detect bioterrorism attacks and combat the growing antibiotic resistance among microbes, which causes about 1.7 million hospital infections every year.

"We know next to nothing about the ecology of urban environments," said evolutionary biologist Jonathan Eisen at the University of California at Davis. "How will we know if there is something abnormal if we don't know what normal is?"

Dr. Mason and his research team gathered DNA from turnstiles, ticket kiosks, railings and benches in a transit system shared by 5.5 million riders every day. They sequenced the genetic material they found at the subway's 466 open stations—more than 10 billion fragments of biochemical code—and sorted it by supercomputer. They compared the results to genetic databases of known bacteria, viruses and other life-forms to identify these all-but-invisible fellow travelers.

In the process, they uncovered how commuters seed the city subways every day with bacteria from the food they eat, the pets or plants they keep, and their shoes, trash, sneezes and unwashed hands. The team detected signs of 15,152 types of life-forms. Almost half of the DNA belonged to bacteria—most of them harmless; the scientists said the levels of bacteria they detected pose no public-health problem. Data from the PathoMap Project, as Dr. Mason calls it, was published online in the journal Cell Systems on Thursday.

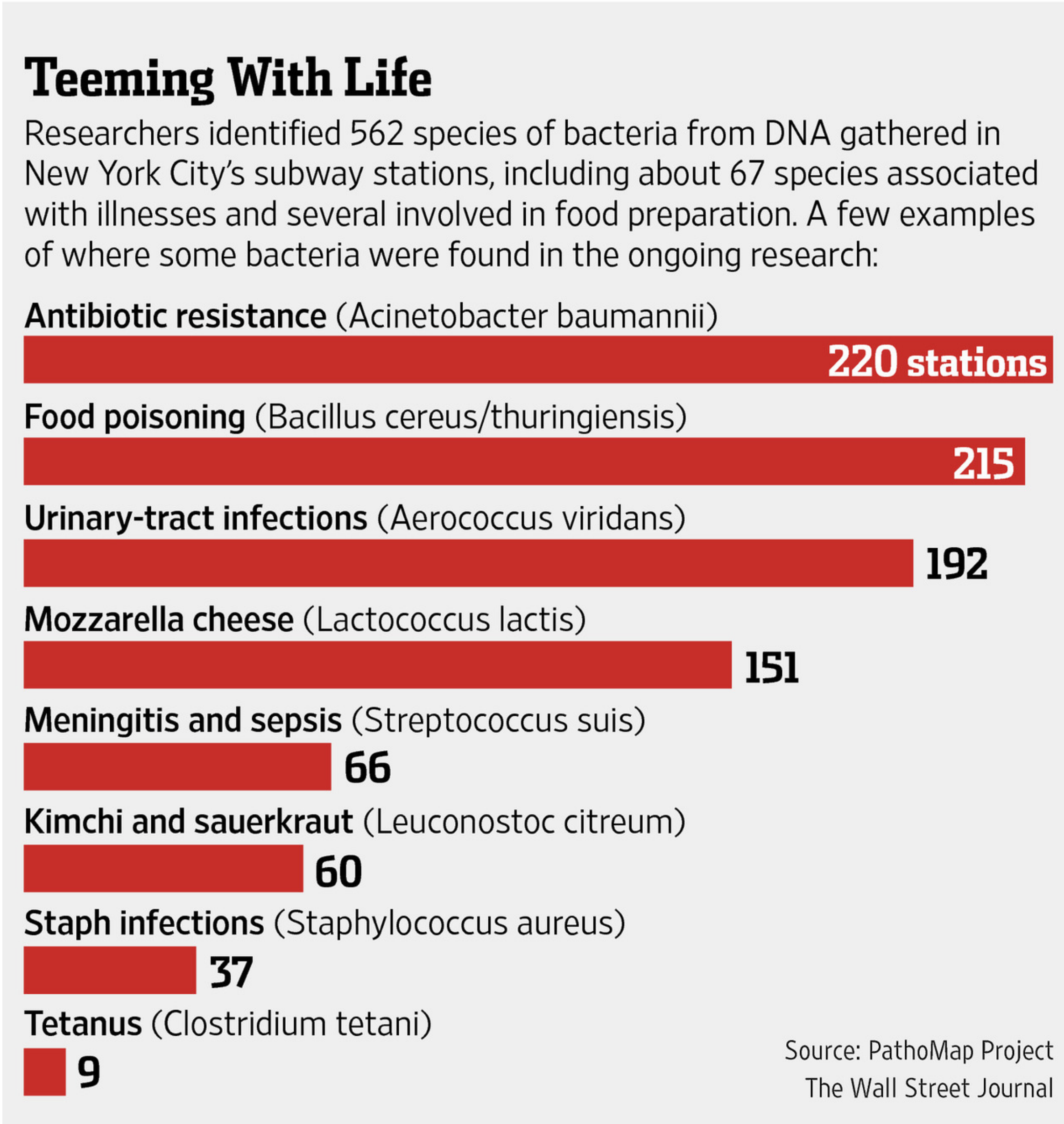
As more and more scientists probe urban microbiology, they are also hoping to find ways to foster beneficial bacteria through building design and to learn how to eliminate construction practices that create living conditions for the germs that make people ill.

This emerging field reflects the growing awareness that the human body swarms with bacteria. Typically, every person is home to about a hundred trillion microbial cells bearing five million different genes, totaling about 5 pounds of micro-organisms per person. Indeed, microbes in and on the body outnumber human cells about 10 to one.

“You are a minority party in the democracy of the body,” Dr. Mason said.

The body's collection of microbes, called the microbiome, influences health in ways that researchers are only beginning to understand. They may be key to proper digestion, vitamin synthesis and brain function, new research suggests. Changes among the millions of microbes living in the human stomach also may promote obesity, trigger ulcers or affect how well a flu vaccine works.

Broadly speaking, city living leaves its mark on people. That includes the sorts of microbes that collect inside them. A recent comparison of urban and rural residents in Russia found that city dwellers had different sets of stomach microbes than people in the countryside.



Every person trails a distinctive collection of microbes, by shedding about 1.5 million microscopic skin cells every hour. Bacteria from a person's body can colonize a hotel room in less than six hours, scientists at the U.S. Department of Energy's Argonne National Laboratory in Illinois recently discovered.

"A city is like an organism," said IBM Corp. computational biologist Robert Prill, who is among those at the company investigating ways to better collect and analyze these immense new public-health genome databases. "It has a circulating system consisting of the movement of people."

In New York City, the Cornell scientists and student volunteers gamely dodged rats and gingerly worked around discarded pregnancy tests, used condoms, puddles of vomit and rotting food to swab surfaces in every subway station. On more than one occasion, suspicious police stopped them and escorted them to the street.

The subway findings might unsettle some people, Dr. Mason acknowledged, but he said they illustrate the remarkable microbial diversity of a healthy city. “I don’t want people to be terrified,” he said. “I want them to be intrigued.”

Large-scale microbe studies, called metagenomics, are made possible by recent advances in low-cost, high-speed gene sequencing machines that for the first time allow researchers to study millions of micro-organisms in the wild that normally can’t be grown in a laboratory. The Sloan Foundation in New York jump-started the field by funding 50 building-genome projects since 2005, although not the PathoMap survey.

In Oregon, researchers are mapping the ebb and flow of bacteria that inhabit the rooms of a busy classroom building. In Virginia, biologists are analyzing the microbes that live inside a building’s plumbing and drinking water pipes. In Chicago, scientists are documenting the microbial life of a new hospital, to see how bacteria in its offices, operating rooms and patient recovery areas changed as the health-care facility became operational.

Researchers are learning that quirks of building materials, ventilation systems, humidity and interior design affect the kinds of bacteria people encounter indoors.

Depending on the material involved, some surfaces can have thousands of different types of bacteria while others may have only a few hundred, researchers monitoring the new Chicago hospital found. Pathogens responsible for common infections, such as the strep germs that cause an estimated 700 million infections world-wide every year, can survive for months on a dry surface, researchers in Germany reported in September in the journal BMC Infectious Diseases.

Upholstery fabric can make a difference. Drug-resistant staph germs can live for up to a week on materials used for airplane seat pockets, while E. coli can last 96 hours on the covering used for an airliner armrest, researchers at Auburn University said at a meeting of the American Society of Microbiology last year.

Air conditioning matters too. Studies of indoor air quality at shopping centers in Singapore and homes in the United Arab Emirates revealed up to 300 distinct species of airborne bacteria, fungi and viruses carried through ventilation systems.

“Unintentionally, architects and engineers are creating ecosystems without much thought at all as to whether they are healthy or harmful to humans,” said biologist Jessica Green, director of the University of Oregon’s Biology and The Built Environment Center. “Different urban conditions might promote the growth of different microbial ecosystems.”

Researchers at the University of Chicago and the Argonne National Laboratory started installing a network of sensors throughout Chicago’s downtown that, among other things, will sample the air periodically for microbes. They hope to have 500 in place by next year and as many as 5,000 sensors when the monitoring system is completed.

“We are now in a position where we can go into a city and characterize its microbial wildlife,” said Jack Gilbert, an environmental microbiologist who is involved in the project. “We can see the genomes of thousands of different species, millions of different species.”

On a quiet morning in New York this past spring, commuters waiting for the No. 6 train at the 68th Street subway station paid little attention to three student interns from Cornell intently stroking the platform benches and railings with small nylon-tipped swabs. A panhandler played the accordion.

“As the train goes through, it kicks up the air and it coats the entire subway, like snow, with DNA,” said Dr. Mason, who was diligently gathering genetic material from a stairway handrail nearby.

The team boarded the Lexington Avenue local train. They spread out through the car and started swabbing overhead handholds and passenger seats. Each sample took about three minutes to collect. They sealed each one in a plastic bag, then photographed and logged the location.

A woman and a boy in a Yankees cap sat down without a glance at the scientist swabbing the seat next to them.

A researcher at Weill Cornell Medical College collected DNA samples on a New York City subway train last summer.
Katie Orlinsky for The Wall Street Journal

Michael Rivera, 16, in an orange T-shirt and camo pants, was taking the subway uptown to Harlem. He gawked at them. At 96th Street, he finally asked: “What’s going on?”

As Dr. Mason explained the work, Mr. Rivera nodded appreciatively. On other occasions, commuters have been suspicious. “We’ve been accused of spreading HIV or collecting alien DNA,” Dr. Mason said. “People get very fearful and say they don’t want to know what we find.”

The researchers recently began sampling DNA in other subways around the world, including Shanghai, Tokyo, Paris and São Paulo.

In New York, the Metropolitan Transportation Authority, which runs the subway system, declined to comment on the Weill Cornell study or discuss its subway sanitation procedures.

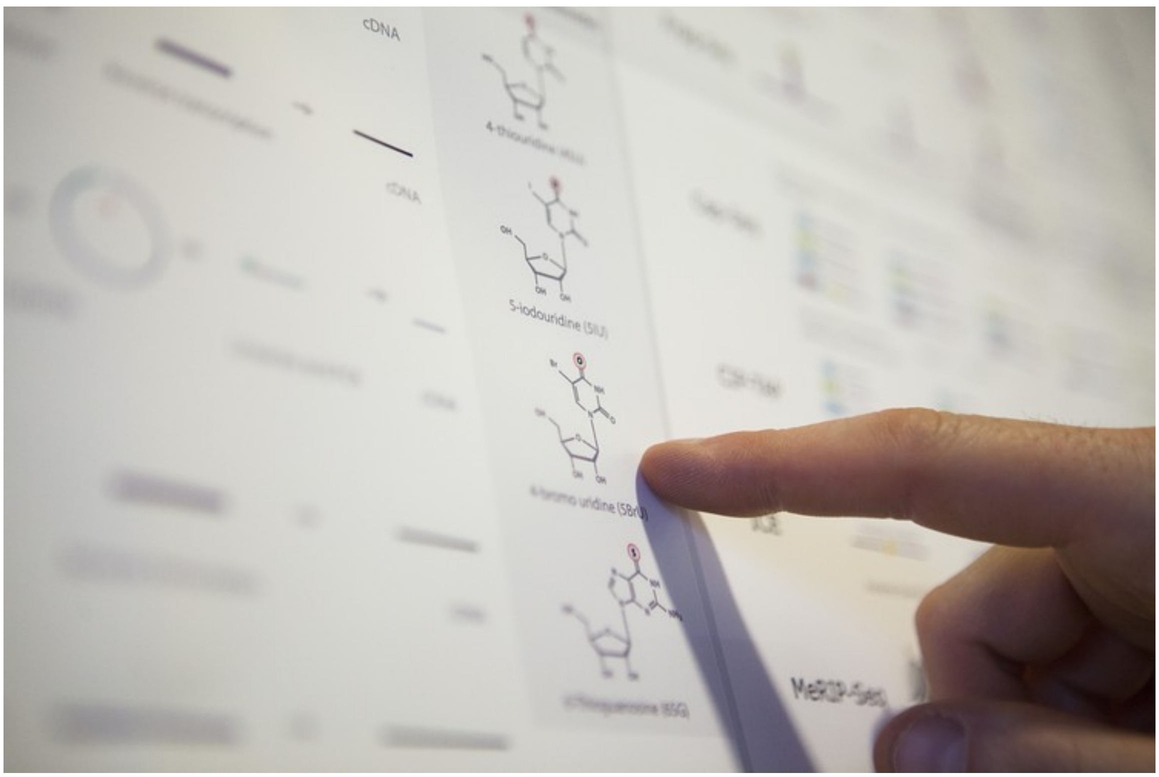
The subway project has been a proving ground for the new technology of public-health genomics. “If you can sequence a city, you can sequence anything and everything,” said James Kaufman, who runs the public-health research project at IBM’s Almaden Research Center in California.

The New York subway study quickly hit the current limits of science. Most microbes have never been isolated or studied. Only a few thousand creatures of any sort have ever had their entire set of genes analyzed, so identifications of DNA sequences through online computer comparisons can be inaccurate.

Initial database searches with subway DNA, for instance, turned up false matches to the Tasmanian devil, the Himalayan yak and the Mediterranean fruit fly—all creatures highly unlikely to be found in a New York transit system.

All told, the biodiversity of the subway isn't as rich as normal soil. The dirt in Central Park contains 167,000 types of micro-organisms—about 11 times the number of species in the transit system—soil ecologists at Colorado State University and the University of Colorado Boulder reported in October.

But a deep breath of subway air contains about as many free-floating bacteria as fresh air at street level, researchers at the University of Colorado Boulder reported last year in the journal Applied and Environmental Microbiology.



Dr. Mason working on the project in the Weill Cornell Medical College lab. Illustration: Katie Orlinsky for The Wall Street Journal

No two subway stations were exactly the same, said Weill Cornell project leader Ebrahim Afshinnekoo, who helped analyze the data.

The greatest subway biodiversity was found at the Myrtle-Willoughby Avenue stop for the G train in Bedford-Stuyvesant, Brooklyn, where 95 unique bacteria groups were detected.

The most unusual bacteria inhabit the South Ferry Station, which has been closed since it was flooded during superstorm Sandy in 2012. "We saw bacteria there that previously were only seen in Antarctica," Dr. Mason said.

Among the DNA of higher organisms, the researchers found across the system that genetic material from beetles and flies was the most prevalent—the cockroach genome hasn't been sequenced yet so that DNA wasn't identified. Cucumber DNA ranked third—possibly from lunch leftovers, or from the computer grouping partial DNA from other plants into the nearest known species.

Human DNA ranked fourth. The genetic leavings of mice, fish and lice were commonplace. (The fish DNA is likely swept in on the 14 million gallons of water that city crews pump out of the subways every day.) In some stations, about 15% of that higher order DNA belonged to rats.

So far, scientists have identified 562 species of bacteria, most of them benign or low risk. At least 67 of those species can make people sick. Even these infectious bacteria were all detected at such low levels that they were unlikely to cause illness in a healthy person.

Among the pathogenic and infectious bacteria, the Cornell researchers identified DNA related to strep infections at 66 stations and urinary tract infections at 192 stations. They found E. coli at 56 stations and other bacteria related to food poisoning at 215 stations.

A multidrug resistant bacterium called Stenotrophomonas maltophilia, associated with respiratory ailments and hospital infections, turned up at 409 stations. Another antibiotic resistant infectious microbe, called Acinetobacter baumannii, turned up at 220 stations.

At spots in three stations—on a garbage can, a MetroCard vending machine and a stairway railing—they also turned up traces of the bacteria that cause bubonic plague. While common among rodents in the western U.S., plague infections are extremely rare along the Eastern Seaboard. It has been 12 years since a human case has been diagnosed in New York City, according to the U.S. Centers for Disease Control and Prevention.

"We think the rats are the likely carrier [of the plague bacteria], since we see plenty of rat and mouse DNA," said Dr. Mason.

They also found a trace of anthrax DNA on a railing at one station and on a handhold in a subway car. "The results do not suggest that the plague or anthrax is prevalent, nor do they suggest that NYC residents are at risk," the researchers reported.

The New York City Department of Health and Mental Hygiene "strongly" disputed that the bacteria were correctly identified. "The interpretation of the results are flawed, and the researchers failed to offer alternative, much more plausible explanations for their findings," a department spokeswoman said in a written statement. "The NYC subway system is not a source of plague or anthrax disease, and the bacteria that cause these diseases do not occur naturally in this part of North America."

The subway DNA was also a measure of urban appetites.

The scientists detected DNA from bacteria associated with the production of mozzarella cheese at 151 stations. DNA from chickpeas, a key ingredient in hummus and falafel, was detected on many subway platforms and benches.

The researchers also found bacteria that readily dine on arsenic, sup on oil spills and digest sulfates commonly found in the subways. Some species in the subterranean system are unusually resistant to extremes of acidity, aridity, temperature and radiation.

"They are like New Yorkers," Dr. Mason said. "They can survive anywhere."

Write to Robert Lee Hotz at sciencejournal@wsj.com

The subway project has been a proving ground for the new technology of public-health genomics. Weill Cornell Medical College scientists sequenced the genetic material they found in the subway—more than 10 billion fragments of biochemical code—and sorted it by supercomputer. They compared the results to genetic databases of known bacteria to identify species.
Katie Orlinsky for The Wall Street Journal

How Scientists Wrangled 10 Billion Fragments of Genetic Code

Researchers Collected DNA From Hundreds of New York City Subway Stations and Used High-Speed Supercomputers to Analyze and Identify Species

By
Robert Lee Hotz
Feb. 5, 2015 2:11 p.m. ET

When is a cucumber not a cucumber?

When so much of its DNA appears in the New York City subway system, say researchers at Weill Cornell Medical College, who analyzed 10.4 billion fragments of genetic material from the transit system.

The scientists collected DNA samples from 466 subway stations, building the first genetic profile of a metropolitan transit system. They hope the database can be used to discover new ways to track disease outbreaks, detect bioterrorism attacks and combat the growing antibiotic resistance among microbes.

In this new field of metagenomics, researchers use faster, cheaper gene sequencers and supercomputer software to study and identify DNA from millions of micro-organisms that live on every surface people touch.

"This study would have been impossible even five years ago," said Weill Cornell geneticist Christopher Mason, who conducted the subway project.

To sort the huge quantity of intermingled DNA, the researchers used a method called shotgun metagenomic sequencing, which allows them to amplify and label the genetic fragments and then reassemble them into a collection of plausible sequences. Those genetic sequences were then analyzed by computer to see whether or not they lined up with the genomes of known organisms.

The Weill Cornell researchers compared the sequences to species already on file in four independent genetic databases. They identified 15,152 types of life-forms overall, and 562 species of bacteria.

Researchers can determine genetic material belongs to bacteria though a genetic marker called the 16s ribosomal RNA gene, which turns up only in microbes. But relatively few species have been mapped sufficiently to be identified by name.

In all, 46 scientists from 20 research centers were involved in the Weill Cornell project, which is still ongoing. Illumina Inc. in San Diego, which makes high-speed genetic sequencing machines for large-scale genome analysis, and Qiagen NV in the Netherlands donated the chemicals needed to process the samples. The IBM Almaden Research Center in San Jose, Calif., contributed supercomputer resources.

The researchers found that among the higher-order DNA in the subway system, cucumbers, in a surprise, were the third most common life-form. Humans ranked fourth.

Some of the cucumber DNA likely came from salads discarded in subway garbage cans, where the Weill Cornell researchers collected samples of DNA, but much of it might be a measure of the technical limitations found in these early days of metagenomics.

The current genetic databases are incomplete and often unreliable, microbiologists said. Since many plants share basic DNA, for example, cucumbers may have been the closest known plant match that turned up in the database.

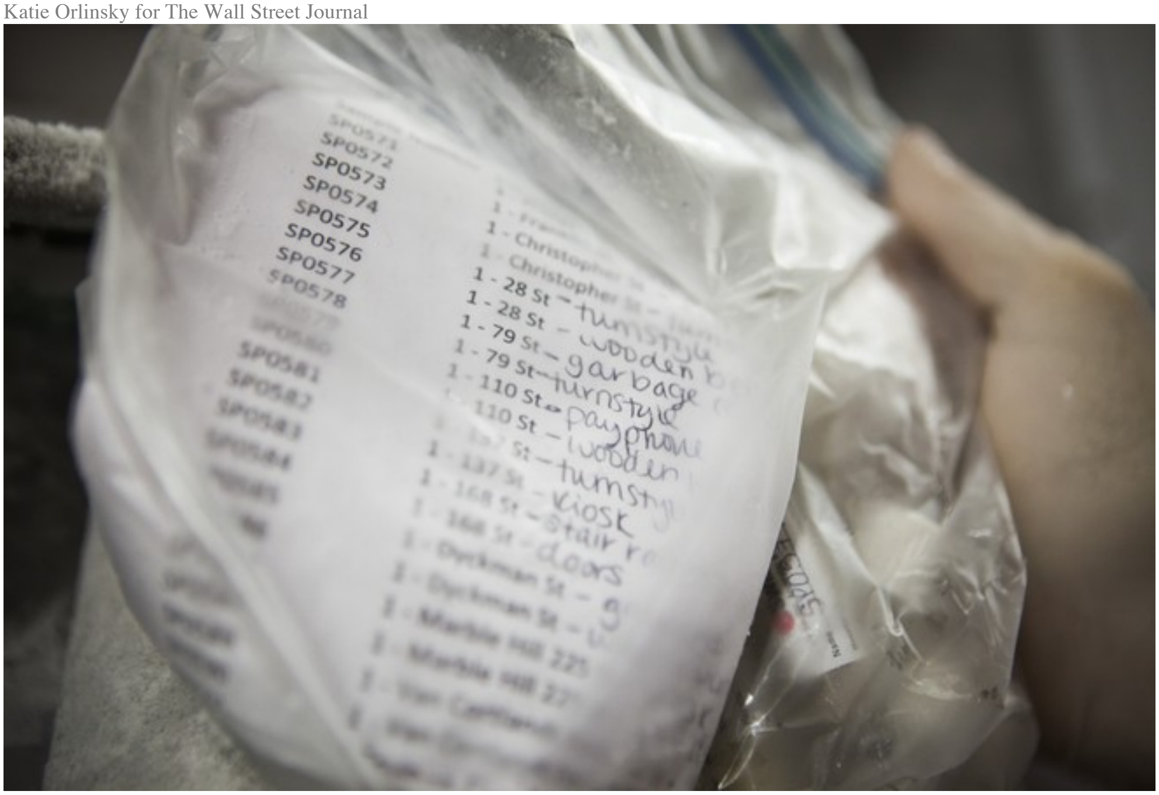
Many studies also are flawed by contamination or distorted by various error-correcting measures in the analytic software, experts said.

"We are limited by the knowledge of what has been sequenced, which we know is woefully incomplete," said Dr. Mason. "We are seeing things that no one has cultured or seen or studied. We know nothing about them."

NOTE TO READERS (Aug. 3, 2015): Researchers at Weill Cornell Medical College have revised their study of the microbiome of the New York City subway system to indicate that there is not enough evidence to say that they found infectious traces of anthrax and bubonic plague in the system. The revision was published in the journal Cell Systems. Geneticist Christopher Mason calls the mistake "an error of interpretation" that does not apply to the broader findings.

Dr. Mason and his team detected signs of 15,152 types of life-forms from the DNA in New York's subways. Almost half the DNA belonged to bacteria, most of them harmless.
Katie Orlinsky for The Wall Street Journal

The team spent 18 months swabbing surfaces in 466 subway stations—platform benches, turnstiles, ticket-machine keypads—to collect genetic material.
Katie Orlinsky for The Wall Street Journal



DNA samples from the subway. Scientists have identified 562 species of bacteria. At least 67 of those species can make people sick, but they were found in such low levels they don’t pose a risk to healthy people.
Katie Orlinsky for The Wall Street Journal