At the National Center for Genome Resources in Santa Fe, New Mexico, scientists run a world-renowned sequencing service facility that’s heavy on long reads and bioinformatics expertise. It also supports a wide range of microbial, plant, and animal projects.

It’s not every core facility that can tout a Human Genome Project pedigree. The sequencing service lab at the National Center for Genome Resources (NCGR), however, has its roots in the earliest days of that game-changing project: in 1994, the center was spun out of Los Alamos National Laboratory, where staffers provided the bioinformatics support for some of the earliest human genome sequence data ever generated.

Today, NCGR continues to be a bioinformatics powerhouse, but now it produces plenty of sequencing data as well – for its own scientific staff and for customers of its service lab. The team is adept at delivering advanced, tailored bioinformatics analysis for clients’ projects, setting the service facility apart from other providers that might use a single best-practices pipeline for all genomic data. The center also specializes in its own scientific research (with 200+ publications to date) as well as in bioinformatics education, creating extensive training resources for the scientific community.

The NCGR core facility is a member of NIH’s IDeA Networks of Biomedical Research Excellence (INBRE), which helps it connect to other research institutions in its home state of New Mexico and other regional and national INBRE centers. But clients come from all over the world, and most of them request SMRT Sequencing: about 70 percent of projects are run on the lab’s PacBio RS II Sequencing System.

### The Sequencing Business

After its emergence from Los Alamos National Laboratory, birthplace of GenBank, the NCGR team kept its sole focus on bioinformatics for more than a decade. Their foray into service-based sequencing began in 2007, when NCGR bought its first next-gen sequencing instrument. They put their analysis background to work immediately, developing a tool called Alpheus to clean up the error-prone NGS data being generated at the time, says Faye Schilkey, NCGR’s director of strategic projects. In 2009, Alpheus won a Computerworld Laureate Award.

A year later, the lab leapfrogged most of its competitors by becoming one of the first to acquire a PacBio System. “We realized it was the future,” says Schilkey, noting that its value was expected to be complementary to the short-read data the facility was already producing. Now, however, many customers prefer results from just the long-read data. “As SMRT Cell yield is getting higher, people are opting for PacBio-only assemblies, even with genomes in the gigabase range,” she says. “Researchers see the value of very long reads which we’ve seen reach 75 kb in length.”

NCGR sequencing customers submit a broad array of projects – “from the antimicrobial properties of the American cockroach to large agricultural genomes to biofuels, biofilms, West Nile virus, and beyond,” Schilkey says. For the PacBio System, primary interests range from de novo genome assemblies to epigenome and transcriptome studies for plants, bacteria, fungi, insects, animals, and humans. The Iso-Seq method for elucidating the transcriptome landscape within a sample has become very popular; Schilkey says in one recent project, the Iso-Seq data outperformed a short-read transcriptome assembly so convincingly that the customer abandoned the short-read data set.

### Bioinformatics Value

With roots in the team that created the first relational human genome sequence database and developed a series of innovative software tools and practices to extract value from ‘omics data, it’s no wonder the NCGR sequencing staff prides itself on delivering a better bioinformatics-based scientific impact to its customers. Housed in a modern 32,000-square-foot facility, “we’re a one-stop destination for sequencing and bioinformatics,” Schilkey says. “We have enviable computational resources and petabyte-scale storage in a secure local infrastructure, which for some customers is an important factor.”

The NCGR team offers any number of analyses, from difficult de novo polyploidy plant genome assemblies to transcriptome assemblies, differential gene expression, pathway, small RNA, and custom analysis. They alter bioinformatics pipelines based on what the client needs, and even run similar algorithms to see where results overlap or diverge. “This is where we spend the extra
time,” Schilkey says. “We know how
to turn the knobs and work the
bioinformatics tools to produce
optimal results.”

Much of this expertise comes from
NCGR’s internal research efforts –
their scientists use these platforms
for their own work, and can provide
helpful suggestions to customers for
how best to conduct an experiment,
including sample prep protocols,
new R&D techniques, sequencing
approaches, and analysis workflows.
“We also have check points where
we review intermediate results and
make recommendations. We’re
more engaged with the client and
the outcome than other service
providers that might just take your
sample, sequence it, and what you
get is what you get,” Schilkey adds.
“At the end of an analysis project,
we have a webinar to ensure the
customer understands their results
and molecular insights.”

For PacBio projects, this attention
to detail means the bioinformatics
team dives into the SMRT Analysis
Portal and other analysis programs
to customize parameters for each
client. “For example, we change
settings and do iterative optimization
of assembly pipelines,” Schilkey
says. “Customers benefit because
we can provide them superior
results by improving key assembly
characteristics and metrics.”

She anticipates that demand for
SMRT Sequencing will continue to
keep NCGR’s scientists on their toes
going forward. “PacBio has a strong
hold now,” she says, “and has really
taken off as a preeminent technology
in molecular research.”

PAPERS OF INTEREST


