

Addgene achieves faster processing and lower plasmid analysis costs with Nextflow

CUSTOMER



<https://addgene.org>

LOCATION

Watertown, MA

INDUSTRY

Biomedical research

OBJECTIVES

To support expanded on-premises sequencing of plasmids for research, Addgene needed to scale analysis and QC capacity rapidly while leveraging existing in-house developed scripts and tools

CHALLENGES

- Increasing volume of incoming materials
- Growing plasmid library
- Analysis requirements outgrowing capacity
- Need for automated end-to-end processing

SOLUTION

- Illumina MiSeq™ Next Generation Sequencer
- Nextflow
- nf-core community pipelines

RESULTS

- Faster runtimes, lower costs
- Shorter cycle times
- Increased productivity
- Reuse existing scripts and tools
- Seamless access to cloud resources
- Automate multi-step processes

Summary

Addgene is a global, non-profit repository created to help scientists share plasmids and other DNA-based tools. Plasmids are small, circular pieces of DNA found in bacteria and other single-cell organisms. Biomedical laboratories routinely use them to modify existing genes or introduce new ones into an organism, making them a powerful research tool. Applications for plasmids include advancing genetic research, creating disease-resistant crops, and developing new treatments for various diseases and conditions.

By using Nextflow to automate sequence assembly and QC workflow, Addgene was able to scale its internal sequencing and analysis to handle hundreds of plasmid genomes per week. This has helped them improve research productivity while minimizing costs, enabling them to better serve the research community.

The business

Addgene manages an extensive repository of DNA-based research reagents commonly used in life sciences. When researchers publish papers, they typically deposit plasmids with Addgene. Once deposited, these plasmids are accessible to other researchers for future experiments. Addgene also facilitates MTA implementation, manages quality control and shipping, and maintains detailed and accurate records.

The company serves organizations seeking to share DNA-based tools and supplies plasmids, antibodies, and ready-to-use viral preparations to research labs that request them. With more than 5,500 laboratories contributing materials, Addgene has accumulated more than 300,000 nucleotide sequences to date.

The challenge

In the early days of its evolution, Addgene would simply spot-check critical regions of deposited plasmids using Sanger Sequencing. Since obtaining an Illumina MiSeq™ sequencer in-house, Addgene has evolved to the point where it now performs full plasmid sequencing in-house as part of its extensive QC process.

As Addgene began to sequence plasmids themselves, they faced the challenge of automating the bioinformatic steps needed to transfer raw sequence data into complete plasmid sequences that were ready for QC teams to analyze. The volume of data quickly overwhelmed existing processes.

Addgene's software was designed to work with 96-well microplates. They needed to perform multiple sequential computational steps on each sample. Steps included retrieving data from instruments, converting raw reads to FASTQ format, adapter trimming, quality scoring, k-mer counting, and sequence assembly. They needed a software solution that could scale to facilitate high-throughput sequencing while also helping control costs.

The solution

After experimenting with several bioinformatic pipeline solutions, Addgene selected open-source Nextflow, a project maintained by Seqera Labs, to automate their plasmid sequencing and analysis pipelines.

Nextflow provided multiple advantages over other pipeline orchestration software, including:

- Mixing scripting languages within pipelines to maximize the reuse of existing scripts and code.
- Mature container support, making pipelines portable across environments.
- The ability to easily switch between local and cloud-based environments without code changes.
- An extensive developer community and curated nf-core pipelines.

Results

FASTER RUNTIMES, LOWER COSTS

Deploying Nextflow enabled Addgene to sequence, analyze, and ultimately share hundreds of QC-verified plasmids per week by tapping resources in the cloud. Jason Niehaus, Director of Data Science at Addgene, observed, "**By scaling up Nextflow pipelines, we were able to achieve a 9-fold reduction in manual effort by simply consolidating and automating our workflow. The time and cost savings extended to our platform engineering as well, since we no longer needed to purchase and maintain our own infrastructure.**"

SHORTER CYCLE TIMES, BETTER PRODUCTIVITY

With effortless switching between on-premises and cloud computing resources enabled by Nextflow and containerized pipeline processes, Addgene could quickly jump back and forth between environments without changes to pipeline code. This enabled them to iterate on development and testing more quickly, improving overall productivity.

REUSE EXISTING SCRIPTS AND TOOLS

Addgene had already developed custom in-house scripts to automate various process steps. With Nextflow supporting multiple pipeline scripting languages, including Python, Perl, Java, and R, Addgene was able to reuse its existing scripts and leverage off-the-shelf components and modules to simplify pipeline development and maintenance.

AUTOMATE MULTI-STEP PROCESSES

Before adopting Nextflow, process steps such as retrieving instrument data, converting file formats, and assembling reads into plasmid sequences were performed manually or guided by custom scripts. Nextflow enabled Addgene to automate these processes, avoiding the potential for human error and making them more efficient and reliable.

LEARN MORE

To learn more, visit www.seqera.io