

Powering groundbreaking plant genomics research.

How the **University of Delhi Department of Genetics** uses a Lenovo high-performance computing (HPC) system, featuring 2nd Gen Intel® Xeon® Scalable processors, to reduce time to insight.

Lenovo Infrastructure Solutions
for The Data-Centered



Lenovo

1

Background

The University of Delhi is one of India's largest and most renowned higher education institutions, with a community of over 62,000 students. Since it was established in 1984, the university's Department of Genetics has emerged as a major hub for teaching and research in genetics, genomics, and molecular biology.

The department is recognized internationally for its scientists' work on plant and human genetics, plant breeding, crop improvement, and crop health. The human geneticists at the University of Delhi perform Omics (e.g. genomics, transcriptomics) analytics to understand the genetic basis of several human diseases, including schizophrenia, Parkinson's, and rheumatoid arthritis. The Centre for Genetic Manipulation of Crop Plants (CGMCP) was established in 1996 at the South Campus of the University of Delhi with funding from the National Dairy Development Board (NDDB).

Since its inception, a major mandate of CGMCP has been to improve the productivity of oilseed mustard – *Brassica juncea*. This crop is grown extensively in the dryland areas of the Indian subcontinent as a source of edible oil for human consumption and seed meal for livestock feeding.

Recently, scientists at CGMCP started using Omics approaches to understand plant-pathogen interactions and to understand factors underlying in high yield of the crop for different varieties of the species.

2

Challenge

With dozens of pioneering new research projects starting up at the University of Delhi's Department of Genetics each year, demand for high-performance computing (HPC) resources is extremely high. The department operates a central HPC cluster that all members of the faculty can access—but there was often a long wait for essential compute resources.

To accelerate vital research on oilseed brassicas, CGMCP decided to invest in its own dedicated HPC infrastructure. CGMCP sought an affordable solution that would deliver the high performance and high throughput capacity required to support large-scale genomics analytics.



“The latest genome sequencing techniques produce very large amounts of data, which requires enormous compute and memory capacity to handle large-scale processing and analysis.”

Dr. Kumar Paritosh

Scientist, Centre for Genetic Manipulation of Crop Plants,
University of Delhi South Campus

Cutting-edge HPC.

CGMPC engaged Lenovo to design and deploy an HPC solution that met its demanding performance and memory requirements.

Lenovo proposed the Genomics Optimization and Scalability Tool (GOAST) Plus, an HPC architecture specifically optimized for genomics analytics. The pre-configured hardware and software bundle, based on Lenovo ThinkSystem SR950 servers with 2nd Gen Intel® Xeon® Scalable processors, is extremely fast, affordable, and easy to use.

Dr. Paritosh recalls: “We were very impressed by the technical specifications of the proposed solution. The ThinkSystem SR950 is an eight-socket server, and memory capacity is an important consideration for genomics analyses. We are impressed by the performance of the system for *de novo* genome assembly with data generated from the third-generation sequencing approaches.”

Why Lenovo? All-in-one HPC for genomics.

Working closely with Lenovo, CGMPC installed two Lenovo ThinkSystem SR950 servers, each equipped with eight Intel® Xeon® Platinum 8260 Processors, and a Network File System (NFS) based on a Lenovo ThinkSystem DE2000H Hybrid Storage Array.

The GOAST system includes the CentOS operating system, the Genome Analysis Toolkit (four different types of variant-calling workflows in the GATK) and the software suite from the Broad Institute, as well as all other software dependencies required to run GATK.

The Lenovo team configured and optimized the GOAST system to deliver optimized performance, tuning the hardware so it can run at full capacity when performing genomics analytics. Lenovo also created easy-to-run scripts that wrap around the complex and often difficult-to-set-up GATK workflows to simplify submission, monitoring, and management of genome samples thus greatly improving usability for researchers.



“The Lenovo team went the extra mile to deliver an HPC system that meets researchers’ needs both in terms of performance and usability.”

Dr. Kumar Paritosh

Scientist, Centre for Genetic Manipulation of Crop Plants, University of Delhi South Campus

3

Results

With the Lenovo GOAST system in place, scientists at CGMCP have access to powerful HPC resources that are extremely fast, affordable, and easy to use. Crucially, high throughput (more samples analyzed at any one time) and short execution times are accelerating time to insight.

One key area of research at CGMCP is a genetically modified variety of mustard that has demonstrated increased yields over existing varieties. To reduce India's need for edible oil imports, scientists at CGMCP are working to further improve this mustard variety, using insights from the Lenovo HPC solution. For example, while identifying and manipulating genes related to drought and disease resistance to breed hardier crops, Dr. Kumar could also map, tag, and introgress traits in *Brassica juncea* for resistance to *Albugo candida* – a type of white rust fungus that can cause serious damage to the plants.

Dr. Paritosh confirms: “I can now process more genomes concurrently and get results faster. Previously, it took 48 hours to process a ~110X *Brassica juncea* genome. Now it takes just six hours, thanks to GOAST’s 8x boost in performance for ultra-deep genomes.”¹



दिल्ली विश्वविद्यालय
University of Delhi



36x and 8x increase in performance for human and plant genomics analytics, respectively²



1.3 hr. to process a 30X human whole genome; 6 hours to process a 110X *Brassica juncea* whole genome³



Faster time to insight for researchers

^{1, 2, 3} Data provided by the University of Delhi.



“With the Lenovo GOAST system, I can analyze more data and uncover new insights faster. Our new HPC environment is powering cutting-edge research that will help us to breed more nutritious, more drought and disease-tolerant, high-yield plants to feed the world.”

Dr. Kumar Paritosh

Scientist, Centre for Genetic Manipulation of Crop Plants,
University of Delhi South Campus

What will you do with Lenovo HPC solutions?

The Data-Centered accelerate vital plant genomics research with
Lenovo smarter infrastructure solutions, powered by Intel® Xeon® Scalable processors.

[Explore HPC Solutions](#)



Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo.

Intel, the Intel logo, Xeon, and Xeon Inside are trademarks of Intel Corporation or its subsidiaries in the U.S.
and/or other countries.

Other company, product and service names may be trademarks or service marks of others.

© Lenovo 2021. All rights reserved.