DELL Technologies

LIFE SCIENCES
UNITED KINGDOM

wellcome sanger institute

Scaling genome sequencing to preserve biodiversity

The Wellcome Sanger Institute utilizes the Dell AI Factory with NVIDIA on-premises solutions, improving our understanding of life on earth.

Business needs

The Wellcome Sanger Institute is a world leader in genome research that delivers insights into human, evolutionary and pathogen biology. It aims to sequence 70,000 species in the British and Irish Isles as part of the Darwin Tree of Life project. It has also contributed over 70 percent of the genomes to the global Earth BioGenome Project, which seeks to decode the DNA of all eukaryotic life on Earth to support scientific research. As genome sequencing becomes more complex and data requirements continue to grow, on-premises computational power and AI systems are essential for scalability and efficiency.

Business results

- Produces one genome every seven hours, up from three genomes in ten years.
- Contributes approximately 70 percent of global data for Earth BioGenome Project.
- Processes vast genomes such as mistletoe at nearly 100 gigabase pairs—30x the size of a human's.
- Maintains over 100 petabytes of curated genetic sequencing data on-premises.

Solutions at a glance

Dell AI Factory with NVIDIA

- Dell PowerEdge XE-Series servers with NVIDIA accelerated computing

NVIDIA

From 3 genomes in 10 years to 1 every 7 hours.

Biodiversity research meets technical scalability

Understanding biodiversity is critical to the health of the planet. The Wellcome Sanger Institute's Tree of Life program is part of a revolutionary effort, the global Earth BioGenome Project, to sequence and catalog the genetic blueprints of 1.5 million species by 2032. Within this effort, the Institute aims to decode the genomes of 70,000 British and Irish Isles species as a first step toward a broader global understanding of ecological networks and conservation challenges.

As Dr. Kerstin Howe, head of production genomics at Sanger Institute, explains, "We are discovering the fundamental building blocks of what makes a species a species. By decoding life, we are revealing new insights into our ecosystems, agriculture, biomedicine and conservation." However, such an ambitious genetic project comes with enormous computational and operational challenges.

The transformative effect of accelerated computing

The sheer complexity of genomic sequencing necessitated transformative advancements in computational infrastructure. The project relies on over 50,000 high-performance computing cores managed by the Institute's Informatics Support Group Team, led by Dr. Peter Clapham, to integrate the latest hardware innovations. Dell PowerEdge XE-Series servers powered by NVIDIA accelerated computing became the foundation of their genomic engine, designed to sequence and assemble genomes with unparalleled precision and speed. "The Dell AI Factory with NVIDIA ensures every investment is architected for success. It optimizes our platforms to scale effortlessly and accelerate the science required for species mapping," Clapham adds.

In practical terms, this infrastructure enables Sanger Institute to handle species with vast genomic scales, such as mistletoe—a plant with a genome 30 times larger than the human genome. Sequencing what seemed computationally impossible became possible. "Mistletoe's genome demonstrates the breakthroughs Dell Technologies and NVIDIA help us achieve, opening the door to possibilities in research we once deemed unfeasible," he explains.

With recent advancements in technology and AI, the pace of production has skyrocketed. The Institute now sequences and assembles a genome every seven hours, a dramatic leap from earlier methods, which took years for a single genome draft. Howe notes the contrast: "We went from creating three genomes in ten years to one genome every seven hours in specific projects. The transformation in speed and quality redefines what's possible."

Agile innovation with on-premises infrastructure

Considering the colossal amount of data processed daily, their choice for on-premises infrastructure proved critical. With over 100 petabytes of curated genetic sequence data generated through their projects, local solutions provided unparalleled efficiency. "Having data on-prem speeds up iteration cycles and eliminates latency, allowing us to fail fast and iterate frequently," says Clapham. This agile approach enables the Institute to meet current demands while paving the way for collaboration with external national systems such as the UK's AI infrastructure when scale demands arise.

Dell Technologies titanium partner, Boxxe, plays a crucial role as a systems integrator. Their partnership ensures high-quality standards in system performance, helping Sanger Institute scale and evolve with confidence. "Boxxe have been a fantastic enabler for us," says Clapham. "They make sure that systems are deployed in an effective, consistent manner that allow us to actually grow, build and develop again as the platforms need to."

> *We are discovering the fundamental building blocks of what makes a species a species. By decoding life, we are revealing new insights into our ecosystems, agriculture, biomedicine and conservation.*
>
> **Dr. Kerstin Howe**
> Head of Production Genomics, The Wellcome Sanger Institute

> *The Dell AI Factory with NVIDIA ensures every investment is architected for success. It optimizes our platforms to scale effortlessly and accelerate the science required for species mapping.*
>
> **Dr. Peter Clapham**
> Informatics Support Group Team Leader,
> The Wellcome Sanger Institute

Saving biodiversity before it's too late

Though locally focused, the insights from the genome data will ripple globally. The data produced is published openly, making this library of knowledge accessible to researchers for a better understanding of life on our shared planet. By sequencing biodiversity hotspots and rare species before their potential extinction, the Institute contributes to vital conservation strategies. For example, genomic insights aid breeding efforts for at-risk species or help combat invasive species that threaten ecosystems. As Howe remarks, "Our UK-centric sequencing approach creates a foundation for global comparison. Even without all species present in this region, we find up to 50 percent representation by close relatives, allowing a broader understanding of ecological relationships worldwide."

Accelerating the next chapters in genomics

By collaborating with Dell Technologies, Sanger Institute ensures its customized solutions are future-proof. Regular roadmap discussions aid long-term operational efficiency and sustainability across systems, crucial for maintaining AI momentum. "Every computational task considers its CO^2 cost. Our projects aim to execute world-changing science within a framework of sustainability," adds Clapham.

AI is set to revolutionize how Sanger Institute creates and deploys genomic insights. Employing the Dell AI Factory with NVIDIA, the Institute explores generative genomic models to refine protocols and sanitize sequencing data in almost real-time. Beyond enhanced results, the shift democratizes genomic research globally. Howe emphasizes the importance of equitable access, stating, "We hope this technology becomes widely accessible, empowering biodiversity-rich regions to define their ecosystems and reap tailored benefits."

As the Institute scales such vital work, its partnership with Dell Technologies and NVIDIA exemplifies how innovation and operational excellence drive outcomes that resonate far beyond laboratory walls.

> *Mistletoe's genome demonstrates the breakthroughs Dell Technologies and NVIDIA help us achieve, opening the door to possibilities in research we once deemed unfeasible.*
>
> **Dr. Peter Clapham**
> Informatics Support Group Team Leader,
> The Wellcome Sanger Institute

Learn More About Dell Technologies AI Solutions.

Connect on Social.

DELL Technologies

NVIDIA