GENOMICS

Genome giant offers data service

Chinese sequencing institute launches remote computing networks to crunch DNA data.

BY EWEN CALLAWAY

The world's largest DNA-sequencing outfit is looking to the clouds. The BGI (formerly the Beijing Genomics Institute) this month announced plans to roll out its cloud computing capabilities, which it hopes will help it to dominate bioinformatics in the same way it does the world of sequencing (see *Nature* 464, 22–24; 2010).

Research labs often lack the storage, computing power and technical know-how to cope with the current deluge of genomic data. Services such as the BGI's offer a solution, says Cliff Reid, chief executive of Complete Genomics in Mountain View, California, a competitor of the BGI that sequences and analyses human genomes. "The cloud is going to be central in the entire world of DNA sequencing."

Cloud computing marshals the power of a network of computers that can be accessed remotely to store and analyse data. Creating

a bespoke cloud network to boost the data-crunching power of the Shenzhen-based BGI was a logical move, says Magic Fang, director of the institute's bioinformatics centre. "Competition in sequencing service is becoming more and more intense," he notes. Meanwhile,

"The cloud is going to be central in the entire world of DNA sequencing."

the flood of genome data is outstripping scientists' capacity to handle it, as next-generation sequencing techniques drive down the price of reading genomes even faster than the cost of data storage is falling (see 'DNA and chips').

Increasingly, Fang says, scientists who turn to the BGI for its sequencing muscle are also asking for help processing and analysing the data that its next-generation sequencers — which number more than 150 — churn out. By developing its own cloud service, the institute hopes to lure high-profile projects and collaborators,

even if they never set foot in China.

Xueping Quan, a bioinformatician at Imperial College London, made the trip to Shenzhen with raw sequencing data from a plant species whose DNA sequence was too large and complex for her team to assemble into a complete genome. The BGI used its cloud to stitch the data together within a month, and now Quan plans to work with the institute to produce additional genomic data and create a more thorough genome sequence.

At present, the principal use of the BGI's cloud is in the assembly of genomes such as Quan's plant. However, Sifei He, who manages its cloud service, says it will be able to run other bioinformatics software, such as programmes to scour genomes for single-nucleotide variants, or to find places where large chunks of the genome are duplicated or missing.

The BGI is not the only sequencing centre pursuing this strategy. Complete Genomics has been storing data for researchers on ►

► Amazon's well established Elastic Compute Cloud for nearly a year, says Reid. He expects that scientists will eventually analyse and store data entirely in such clouds, and only download the final results. Meanwhile, companies such as GenomeQuest of Westborough, Massachusetts — which inked a deal last year to provide bioinformatics to biotechnology giant Syngenta of Basel, Switzerland — and DNAnexus in Palo Alto, California, already run genome-analysis software on clouds.

But it is the BGI's combination of sheer sequencing muscle and in-house cloud computing that makes it stand out as a 'one-stop shop', says David Dooling, a bioinformatician at

DNA AND CHIPS

The price of DNA sequencing is falling faster than computer storage costs, making cloud computing an increasingly important tool in genomics.

the Genome Institute at Washington University in St Louis, Missouri. "It certainly makes sense," he says. "As more and more people are sequencing and those people are less and less experienced with sequence analysis, you're going to see the proliferation of these vertically integrated solutions."

Paul Flicek, a bioinformatician at the European Bioinformatics Institute near Cambridge, UK, says that it's too early to tell whether the BGI's cloud will address the bioinformatics needs of naive users. But "if somebody can produce a cloud service that's ideal for bioinformatics and costs less than Amazon, there's a niche market there they could really capture", he says. ■

SOURCE: L. D. STEIN GENOMEBIOL. 11, 207 (2010)

MEASUREMENT

Physicists count on updated constants

The latest revision of fundamental quantities bodes well for the proposed overhaul of the international system of units.

BY EUGENIE SAMUEL REICH

As cheat sheets go, it is a long one. Every four years, a task force of the Committee on Data for Science and Technology (CODATA), headquartered in Paris, releases a short-hand description of the physical world: the latest, painstaking determination of the values and uncertainties of more than 300 physical constants, from the familiar speed of light, c , to the more obscure mass of the tau lepton, τ .

The most recent figures, from 2010, were first placed online in June and announced on 19 July by the US National Institute for Standards and Technology (NIST) in Gaithersburg, Maryland. The numbers include reduced uncertainties for several key constants, which physicists say is encouraging because these will allow for better tests of theory. The more precise figures will also aid plans to redefine familiar units of measurement, such as the kelvin and the kilogram, in terms of unchanging fundamental constants rather than relying on a material object that might not be stable (as for the kilogram standard).

"There's progress in the right direction," says Peter Mohr, a member of the CODATA task

force at NIST. "It gives you more confidence doing science."

The task force revised the numbers that it had released in 2006 by reviewing the latest experiments and theory constraining fundamental quantities. Uncertainties decreased for the fine-structure constant, α , which is used in atomic physics; Planck's constant, h , which defines the size of energy quanta; Avogadro's constant, N_A , the number of atoms in a sample of an element, the mass of which in grams equals its atomic mass; and Boltzmann's constant, k , which relates energy and temperature.

Much of the progress arises from advances in measurement science. For example, a 2010 measurement of Avogadro's constant made by counting the atoms in a lattice of highly enriched silicon produced a revised number (see *Nature* 467, 892; 2010). This brought the figure closer to another value obtained using Planck's constant (which is related to Avogadro's constant) by measuring the electric current needed to support a kilogram test mass suspended in a magnetic field against the force of gravity. Combining the two methods reduced the relative uncertainty in CODATA's Avogadro's constant from $\pm 5.0 \times 10^{-8}$ to $\pm 4.4 \times 10^{-8}$. For Boltzmann's constant, five new measurements lowered its relative uncertainty to $\pm 9.1 \times 10^{-7}$ from $\pm 1.7 \times 10^{-6}$ in 2006, when it was based on just two measurements.

"That's a piece of good news," says Mohr.

The improved precision, particularly in the values of Avogadro's constant and Boltzmann's constant, is a boost for a proposal to rationalize the international system of standard units (known as SI, *Système International d'Unités*) by linking four of them — the ampere, mole, kilogram and kelvin — to physical constants, says Terry Quinn, emeritus director of the International Bureau of Weights and Measures (BIPM) in Sèvres, France. The proposal is due to be voted on at the General Conference on Weights and Measures in Paris in October, and would put the SI on a path to be revamped by 2015.

Plans for the ampere's revision have long been in good shape because measurements of the electron's charge are very precise. Using the latest values for Boltzmann's constant, the kelvin could also be redefined with ease, says Mohr. The sticking point has been the kilogram, which

"There's progress in the right direction. It gives you more confidence doing science."

is currently defined by a prototype mass held at the BIPM that many physicists consider an embarrassment because it is thought to vary over time. The new proposal would define the kilogram in terms of Planck's constant, which has units of kilogram metres

squared per second. Before that can happen, some experts would like to see the uncertainty in that constant's value decrease to $\pm 2 \times 10^{-8}$.

Quinn says that given recent progress, he expects that to be achieved before 2015. But Ian Mills, a chemist and physicist at the University of Reading, UK, who is president of the BIPM's consultative committee on units, says that the uncertainty of Planck's constant is already low enough to redefine the kilogram reliably. "We could go ahead and make the change to SI today, but everyone is very anxious that we get everything tickety-boo." ■

➔ **NATURE.COM**
For more on how to improve the accuracy of SI units:
go.nature.com/oujgwf