environmental fluctuations can perturb the experimental system.

Two key technological developments enabled the success of Storz and colleagues' Bell experiment. By achieving a single-qubit readout of around 50 nanoseconds, much faster than the few hundred nanoseconds that define the multi-qubit state-of-the-art systems[12,13], the authors were able to reduce the required qubit separation to around 30 metres. Then, they developed a low-loss cryogenic waveguide of this size and integrated it with the qubits to reach a high-fidelity connected system.

Photon-based implementations typically violate Bell's inequality by a small margin, but with a data-production rate high enough to show a statistically significant violation in a relatively short collection time. Matter-based implementations usually violate the inequality by a larger margin, but have low data-acquisition rates, making it difficult, or at least time consuming, to reach high statistical certainty. Storz and colleagues' set-up violates Bell's inequality by a higher margin than previous photon-based experiments, with a higher rate of data production than that obtained in previous matter-based experiments[8–11].

This Bell experiment sets a record for the longest separation between two entangled superconducting qubits, and is impressive because of its physical size and precision. Although the 50-nanosecond readout demonstrated here cannot readily be applied to multi-qubit quantum computers, it pushes this qubit technology to new limits. Similarly, although the superconducting-waveguide approach does not scale to arbitrary distances, it represents a path towards quantum-information transfer between superconducting-qubit chips, a technology that will be needed in a large-scale quantum computer. With the achievement of this foundational quantum milestone, and the technological advancements that enabled it, Storz *et al.* have expanded the superconducting-qubit toolbox and given further credibility to this promising platform.

Marissa Giustina is at Google Quantum AI, Santa Barbara, California 93111, USA.
e-mail: mgiustina@google.com

1. Storz, S. *et al. Nature* **617**, 265–270 (2023).
2. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **47**, 777–780 (1935).
3. Bell, J. S. *Physics* **1**, 195–290 (1964).
4. Clauser, J. F. & Shimony, A. *Rep. Prog. Phys.* **41**, 1881 (1978).
5. Kaiser, D. *How the Hippies Saved Physics: Science, Counterculture, and the Quantum Revival* (Norton, 2011).
6. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. *Phys. Rev. Lett.* **23**, 880–884 (1969).
7. Larsson, J. Å. *J. Phys. A* **47**, 424003 (2014).
8. Hensen, B. *et al. Nature* **526**, 682–686 (2015).
9. Giustina, M. *et al. Phys. Rev. Lett.* **115**, 250401 (2015).
10. Shalm, L. K. *et al. Phys. Rev. Lett.* **115**, 250402 (2015).
11. Rosenfeld, W. *et al. Phys. Rev. Lett.* **119**, 010402 (2017).
12. Krinner, S. *et al. Nature* **605**, 669–674 (2022).
13. Google Quantum AI. *Nature* **614**, 676–681 (2023).

The author declares no competing interests.

Forum: Genomics

A collective human reference genome

A pangenome is a collection of DNA sequences that reveals genetic variation between individuals. Four scientists discuss the generation of a human pangenome, and what insights can be gained from it. **See p.312, p.325 & p.335**

Arya Massarat & Melissa Gymrek

Describing genetic diversity with graphs

Reference genomes are crucial coordinate systems for genomic analyses. However, the references that scientists currently work from when studying humans (the draft human genome[1] and its complete, gap-free successor[2], dubbed T2T-CHM13) are both based mostly on single individual genomes. A linear genome sequence of this type cannot adequately represent genetic diversity within our species. Instead, such diversity is more accurately described using a graph-based system of branching and merging paths. On page 312, Liao *et al.*[3] describe the first human reference pangenome — a collection of genome sequences compiled into a single data structure.

The use of human reference genomes from single individuals is problematic because it introduces biases in how sequences from other human genomes are interpreted. For instance, sequences from other genomes are first commonly aligned to the reference (read mapping) and then reduced to a set of differences from that reference (variant calling). Both processes might yield different results if a different person's DNA had been used to generate the original reference. This is particularly true for highly diverse and structurally complex regions of the genome. Furthermore, there are hundreds of megabases of DNA that cannot be captured in a reference based on a single genome, because they exist in only a subset of humans[4,5].

A pangenome representing many genomes from different ancestries could overcome these issues. However, constructing a pangenome is a complex task. Breakthroughs in the past decade in long-read sequencing technology and computational methods have now enabled this vision to be realized.

Liao and colleagues first generated 94 genome assemblies from 47 individuals (one for each of the two sets of chromosomes that each individual carries). The individuals represent diverse ancestries from around the globe. The assembled genomes, which were generated using a combination of long-read and other sequencing technologies, are highly accurate and nearly complete, and include 119 million base pairs of sequence not included in the draft human reference genome.

The authors used three graph-building methods to construct pangenomes from these assemblies. One of these methods aligns all sequences simultaneously; the others use one genome as a reference and align each subsequent sequence iteratively. The result is a set of publicly available pangenome graphs, along with a rich ecosystem of open-source tools and standardized file formats that researchers can use in a similar way to a linear reference genome.

Liao *et al.* demonstrated that using their pangenomes for read mapping and variant calling resulted in 34% fewer errors in calling small variants (those shorter than 50 bases) than did using a linear reference. The difference was particularly pronounced in challenging repetitive DNA regions. Impressively, the pangenomes enabled the authors to identify twice as many large genomic alterations, called structural variants, per person than is possible using a linear reference (Fig. 1).

The human pangenome reference represents a milestone in human genetics. However, challenges remain. Alignment of sequences against highly variable repetitive regions in the pangenome could be improved by more-accurate assemblies or new algorithms. More samples from diverse groups are also needed. Finally, widespread adoption of the pangenome by scientists could take time, because new methods supporting pangenome analysis are continually being developed, and scientists will often require training to use them.

Continued improvements in methods for building and using pangenomes will enable researchers to overcome these challenges. Use

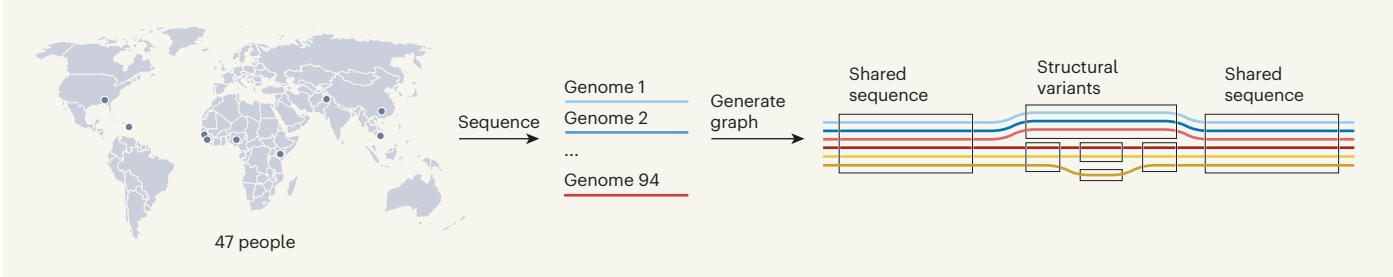

Figure 1 | A human pangenome. The genomes of 47 people of diverse ancestries have been used to generate a draft human pangenome reference[3,6,7]. Two whole-genome sequences were generated from each individual (one for each of their two sets of chromosomes). The 94 sequences were aligned to form a pangenome graph, which is conceptually similar to an underground-railway map. Boxed regions indicate sequences present in one or more genomes that at a given site, with branching paths indicating sequence variation. The graphs reveal large genome alterations called structural variants, and enable easy analysis of how they vary between individuals.

of pangenomes has the potential to transform human genomics. This will ultimately make it easier to discover genetic variants that mediate physical and clinical traits and – it is to be hoped – will eventually lead to better health outcomes for many people.

Arya Massarat is in the Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, California 92093, USA. **Melissa Gymrek** is in the Departments of Medicine, of Computer Science and Engineering and of Biomedical Informatics, University of California San Diego, La Jolla, California 92093, USA.
e-mails: amassara@ucsd.edu; mgymrek@ucsd.edu

Brian McStay & Hákon Jónsson
Untangling repeated sequences

Repetitive DNA regions are challenging to sequence, because it is hard to place them accurately in a genome assembly. These regions include segmental duplications (in which sequences more than one kilobase long are repeated elsewhere in the genome) and the short arms (p-arms) of a subset of chromosomes, dubbed acrocentric chromosomes. Two studies in *Nature* now use Liao and colleagues' pangenomes to systematically explore these regions – Guarracino *et al.*[6] (page 335) to analyse acrocentric p-arms, and Vollger *et al.*[7] (page 325) to investigate segmental duplications. Their work provides a glimpse of the insights that can be gained from a pangenome reference.

The acrocentric chromosomes (chr13, chr14, chr15, chr21 and chr22 in humans) are those in which the p-arm is considerably shorter than the other (q) arm. Acrocentric p-arms are dedicated to one task: forming sites called nucleoli, where the cell's protein-assembling machines are made[8]. P-arms contain nucleolar organizer regions (which encode the RNAs that drive nucleolar formation), highly repetitive DNA and many other shared sequences. This shared, repetitive DNA reflects a phenomenon called heterologous recombination, whereby different acrocentric p-arms pair and cross over to exchange DNA during the cell divisions that generate sperm and eggs. By contrast, in most chromosomes, pairing and crossover are restricted to two copies of the same chromosome (homologous recombination).

In XY sex chromosomes, which also exhibit heterologous recombination, pairing is aided by short regions of homology (near-identical sequences) shared between X and Y. Guarracino *et al.* constructed a variation graph for acrocentric p-arms using Liao and colleagues' sequences, and found that they contain pseudo-homologous regions (PHRs). Each PHR is a patchwork of sequence blocks that – as the authors discovered when they compared their graphs with T2T-CHM13 – often show more similarity to the other four acrocentric p-arms in T2T-CHM13 than to the T2T-CHM13 version of themselves. Presumably, these blocks assist heterologous recombination, ensuring that p-arms evolve in concert to preserve their shared role in nucleolar formation.

Robertsonian translocations (ROBs) are phenomena, usually occurring during egg-cell production, whereby the q-arms of two acrocentric chromosomes fuse and most of the p-arms are lost[8]. Guarracino *et al.* identified sequences in PHRs at which the breaks that lead to ROBs occur – indicating that ROBs are collateral damage arising from heterologous recombination. Given that ROBs occur in one in 800 human births, we surmise that heterologous recombination between acrocentric chromosomes is both ongoing and frequent. We expect that, as more genomes are added to the pangenome reference, it will be possible to quantify the frequency of this recombination.

Vollger *et al.* used the reference to systematically compare variation in segmental duplications with that in non-repetitive parts of the genome (Fig. 1). They found 60% higher sequence diversity in segmental duplications, and showed that these duplications are highly divergent between populations and individuals.

Genes in segmental duplications are susceptible to interlocus gene conversion (IGC) – an exchange of short DNA sequences between non-homologous parts of the duplicated region. Vollger and colleagues identified IGC events by looking for signs of sequence shuffling in the pangenomes, and concluded that these events are probably one of the main reasons that segmental duplications are so diverse. They found that 799 genes had protein-coding regions affected by an IGC.

It is exciting to see accurate characterization of segmental duplications, because duplicated sequences can fuel the evolution of new, specialized roles for a gene. Vollger *et al.* assessed sequence 'constraint' in duplicated genes, with a particular interest in those duplicated during evolution of the human lineage. Constraint is a measure of sequence variability, with less variation indicating that mutations are detrimental to the organism's viability. Thirty-eight genes were constrained, including members of the *NOTCH2* gene family, which has been linked to human-specific changes in brain size during evolution[9]. The repetitive nature of segmental duplications had previously led to difficulties in assessing constraint for at least 40% of the analysed genes. The authors also found that 171 genes were duplicated and relocated intact to new genomic regions, potentially meaning that their regulation would be rewired. In future, the pangenome project should enable researchers to assess constraint in recently duplicated genes in more depth.

Together, these papers provide a taster of how the human pangenome reference can be used. They reveal how sequence exchanges between repetitive regions of our genome contribute to variation in the population and to our evolution. As the scope of the reference expands, we look forward to further insights into these fascinating genomic regions.

Brian McStay is in the Centre for Chromosome Biology, University of Galway, Galway H91 W2TY, Ireland.

Hákon Jónsson is at deCODE genetics, Reykjavik 101, Iceland.
e-mails: brian.mcstay@universityofgalway.ie; hakon.jonsson@decode.is

1. Lander, E. S. *et al. Nature* **409**, 860–921 (2001).
2. Nurk, S. *et al. Science* **376**, 44–53 (2022).
3. Liao, W.-W. *et al. Nature* **617**, 312–324 (2023).
4. Sherman, R. M. *et al. Nature Genet.* **51**, 30–35 (2019).
5. Ebert, P. *et al. Science* **372**, eabf7117 (2021).
6. Guarracino, A. *et al. Nature* **617**, 335–343 (2023).
7. Vollger, M. R. *et al. Nature* **617**, 325–334 (2023).
8. McStay, B. *Annu. Rev. Genom. Hum. Genet.* https://doi.org/10.1146/annurev-genom-101122-081642 (2023).
9. Fiddes, I. T. *et al. Cell* **173**, 1356–1369 (2018).

The authors declare no competing interests.

Palaeoceanography

Salty seas sway global glacial cycles

Kaustubh Thirumalai

Analysis of microfossils in marine sediments spanning the past 1.2 million years suggests that increased intrusion of highly saline waters from the Indian Ocean into the South Atlantic propelled Earth's deglaciations during this period. **See p.306**

Earth's climate fluctuated remarkably over the past one million years, yo-yoing between long-lived ice ages, called glacials, and brief periods of relative warmth known as interglacials. Each maximum in global ice volume was reached several tens of thousands of years after the initial onset of glacial conditions, but transitions into interglacial periods, termed deglaciations, involved rapid global warming and were generally complete within fifteen thousand years or so[1,2]. Delineating the precise order of ocean and atmospheric mechanisms associated with this sawtooth-like pattern of glaciation and deglaciation has remained an outstanding problem in Earth science. On page 306, Nuber *et al.*[3] suggest that intrusions of highly saline waters from the Indian Ocean into the less salty Atlantic Ocean provided the necessary pushes to deglaciate the planet during the past 1.2 million years.

How might the distribution of salt across oceanic basins be related to global climate change? To answer this question, we must first address why certain parts of the ocean are saltier than others. Ocean salinity is defined as the concentration of salt-forming ions dissolved in a given volume of seawater. Although the atmosphere can alter surface-ocean salinity by removing or adding fresh water (through evaporation and precipitation, respectively), salinity can also change as a result of the movement and mixing (advection) of water parcels originating from distinct source regions[4]. For instance, the advection of fresh water derived from riverine run-off or ice melt can cause sharp variations in ocean salinity. And in ocean settings away from the coast, vertical advection along the water column and mixing of waters by oceanic currents can induce changes in the salt content of seawater[4,5].

Surface and deepwater currents in the Atlantic Ocean are pivotal components of the global climate system because they help to regulate hemispheric distributions of oceanic heat and salt (thermohaline fluxes). This system of currents, known as the Atlantic meridional overturning circulation (AMOC), is driven partly by patterns of prevailing winds, and partly by buoyancy forces arising from gradients in seawater density — which, in turn, depend on seawater salinity and temperature[2].

In the modern ocean, the AMOC moves heat and salt from the South Atlantic into the North Atlantic. As portions of the transported water masses journey polewards in the Northern Hemisphere, they become cool and salty enough to reach densities that cause them to sink, forming deep waters[2,6]. These waters then travel southwards at the sea floor and ultimately resurface in upwelling zones, primarily in the Southern Ocean. Owing to its capacity to deliver heat and salt to the high-latitude regions of the North Atlantic, variations in the AMOC are a major factor in mechanisms of glacial-to-interglacial transitions[2,7].

It has been suggested that changes in the thermohaline characteristics of waters flowing into the southern Atlantic can cause swings in AMOC variability[8]. Today, relatively cool and fresh waters move into the South Atlantic gyre (a large, rotating system of wind-driven currents) from the Southern Ocean, whereas warmer and saltier waters from the subtropical

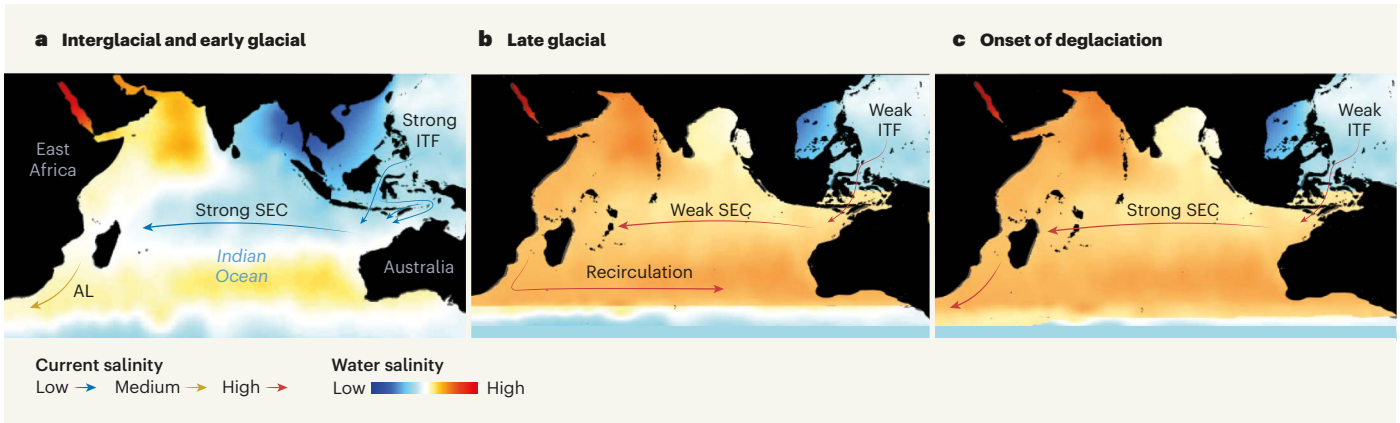

Figure 1 | Salinity in the Indian Ocean affects ocean currents. a, Nuber *et al.*[3] propose that, during warm (interglacial) periods and early in cold (glacial) periods, sea levels are relatively high and the Indonesian archipelago is largely submerged. The Indonesian throughflow (ITF) currents are strong, and carry fresh waters to the Indian Ocean, which contribute to the South Equatorial Current (SEC). A current called the Algulhas Leakage (AL) around the tip of Africa intrudes into the Atlantic Ocean. **b**, In the second half of a glacial period, sea levels are low, exposing more land in the Indonesian archipelago. The ITF is therefore weaker and saltier, lowering the import of fresh water to the Indian Ocean. The SEC is weak and AL reduces, so that saline water recirculates and builds up in the Indian Ocean. **c**, At the onset of deglaciation, sea levels are still low but the AL resumes, releasing highly saline waters to the Atlantic. This would boost the northward flow of warm and salty waters in the Atlantic (not shown), accelerating deglaciation. The coloured shading representing seawater salinity illustrates broad trends, rather than modelled data. (Adapted from Fig. 4 of ref. 3.)