

HUMAN GENETICS

Individual genomes diversify

Samuel Levy and Robert L. Strausberg

The link between a person's genetic ancestry and the traits — including disease risk — that he or she exhibits remains elusive. Routine sequencing of the genomes of an African and an Asian individual offer a step forward.

The rapid progress in genetic screening assays and DNA sequencing techniques promises to increase our understanding of the complex relationship between the human genetic make-up (the genotype) and its associated traits (the phenotype).

For example, using the composite human genome sequences¹⁻³, genome-wide association studies have identified regions that control specific traits through single nucleotide polymorphisms (SNPs) — the most common form of genetic variation. In this issue, Bentley *et al.*⁴ (page 53) and Wang *et al.*⁵ (page 60) detail the development and application of a high-throughput technology for sequencing DNA to decipher the genomes of two people, one of West African descent and the other of Han Chinese descent. This advance provides a technology that might eventually relate specific sequences and regions of DNA directly to human phenotypes.

Although genome-wide association studies can establish a link between a genetic locus marked by adjacent SNPs and its associated phenotype, they do not automatically identify the implicated nucleotide's position, as they



use only a fraction of human SNPs. Genome-wide association studies were used because of their relatively low cost compared with the technological challenge and high cost of sequencing genomes in large human populations. Sequencing the

genomes of many individuals would overcome the problem of identifying which nucleotide(s) are implicated in a phenotype, as long as the procedure could be performed accurately and completely. From such data sets, DNA variants can be identified, and the frequency with which they occur in humans who carry a particular trait — such as a disease — can then be compared with their frequency in people who lack that trait. Thus, all genetic variants contributing to the trait can be identified, giving a more complete picture of the biology involved.

The genomes of the anonymous African and Asian individuals supplement the existing sequenced genomes of two people of European origin, Craig Venter⁶ and James Watson⁷. Both teams involved in the latest work^{4,5} used the Illumina GA sequencing instrument, in which sequencing is performed by synthesizing fluorescently detectable DNA molecules, using

the DNA from the genome being sequenced as a template. In a single cycle, this platform can produce more than 40 million discrete 'reads' of 35 nucleotides from either end of a 200- or a 2,000-nucleotide DNA fragment. Compared with the instruments used to complete the initial human genome sequence¹⁻³, the Illumina GA generates three to four orders of magnitude more sequence per operation cycle. This instrument therefore joins the 454 Life Sciences sequencer⁷ as yet another 'next generation' technology for sequencing individual human genomes.

How do the two new genome sequences allow a better understanding of human genetics? Both studies^{4,5} confirm that it is possible to routinely sequence the genome of an individual to discover the wide spectrum of DNA variations that it harbours. Of course, this process is greatly facilitated by having a reference human genome against which to compare sequence data from the two individuals. This allows the identification of SNPs, as well as insertion/deletion polymorphisms and structural variations (Fig. 1, overleaf). Extensive validation of the SNPs detected shows that sequencing accuracy is high. A strength of this

OPTICS

Metamaterial Persian carpets

Metamaterials gained renown as a way of creating invisibility cloaks — devices that could make an object 'disappear' before one's eyes. Less well known is that they can also act as detectors for biological compounds. Writing in *Optics Express*, Bingham *et al.* describe two-dimensional metamaterials designed so that, when exposed to electromagnetic radiation, their resonant frequencies coincide with those of vitamin H (C. M. Bingham *et al. Optics Express* **16**, 18565-18575; 2008). The resonant frequencies of vitamin H occur in the terahertz range, and these results thus provide an example of biodetection in that frequency regime.

The properties of metamaterials lie in their structure rather than their chemical composition. One asset of these man-made materials is that they can be engineered to possess a

precise response to electromagnetic radiation. Bingham and colleagues created metamaterials with designs that mimic several types of symmetry observed in nature, using both square and hexagonal tiles. Their tiles, the unit cells of metamaterial structures (shown on left of picture), consist of up to three different subunits. The overall structures (shown on right) look rather like a Persian carpet.

To maximize the electromagnetic response of a metamaterial, the unit cells must be tightly tessellated — that is, the gaps between tiles must be minimized. But why incorporate more than one subunit into a tile? The advantage is that the metamaterial preserves the different electromagnetic properties of each subunit: a material formed with three distinct subunits is resonant at three different frequencies. A triple-resonator metamaterial allows a

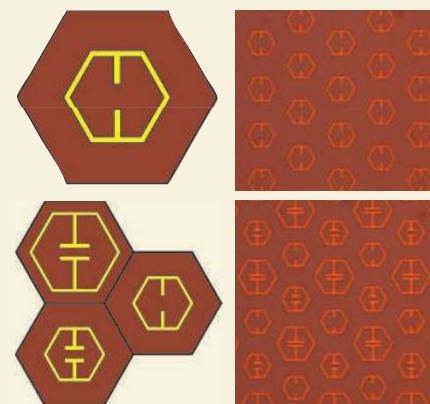
biological compound to be identified more accurately because there are three frequency-match points of comparison.

With this in mind, the authors simulated metamaterial structures computationally to find the best materials for the job. They then made the best designs, shone terahertz radiation on them and recorded the electromagnetic response. As predicted, metamaterials with structures that combined three distinct subunits (such as that pictured on the lower right panel) resonated at three distinct frequencies, the individual frequencies of the different subunits.

As the authors had hoped, the simulated and experimental

resonances of their metamaterials were a good match for those of vitamin H. This match could therefore form the basis of a biodetector. Bingham *et al.* have found that their multi-subunit tiling techniques can create multi-resonator metamaterials that can be used as biodetectors. But that is not all. Their metamaterials could potentially detect hazardous chemicals.

Ana Lopes



05A

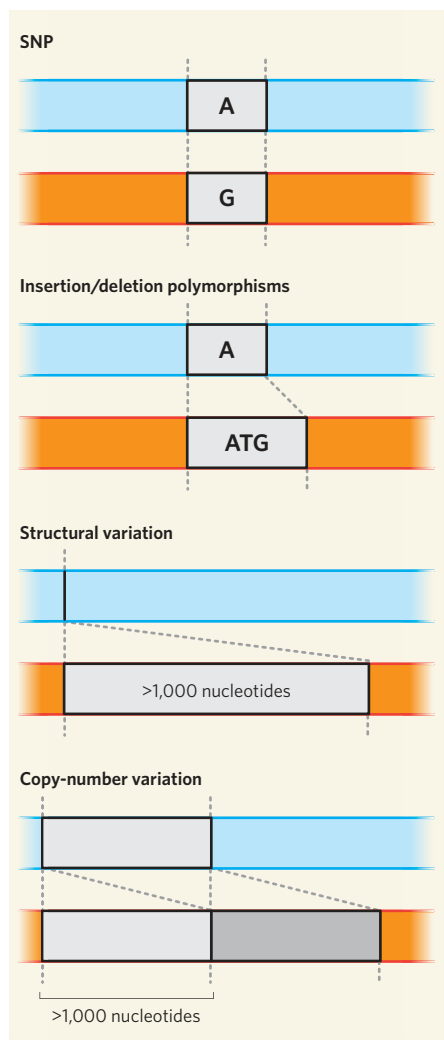


Figure 1 | Genomic variations. The latest whole-genome sequences of two humans confirm^{4,5} that individual genomes vary in several respects. The types of variability in inheritance include: variations in single nucleotides (SNPs); insertion or deletion of several nucleotides; insertion or deletion of thousands of nucleotides (structural variation); and duplication or multiplication of DNA segments more than 1,000 nucleotides long (copy-number variation).

latest approach is the extent of deep sequencing achieved, which aids SNP identification.

The advantages of obtaining these two genomes, such as the identification of DNA variations, indicate that their usefulness will ultimately be much broader than simply demonstrating the technological milestone of relatively low-cost sequencing. But some goals remain. As the genomes were reconstituted on the basis of alignments with existing reference genomes, the set of non-SNP variants that are absent in the reference genome will be incomplete. For example, in these studies, the detection of structural variants — insertions or deletions of thousands of nucleotides at any one position on a chromosome — is preferential for deletions. This is because such insertions come from sequenced reads that will not overlap with the existing reference genome. There are two

possible solutions to this detection bias. One would be to sequence larger DNA fragments whose ends overlap with sequences on the reference genome⁸. Alternatively, all sequenced reads could be assembled independently, before mapping them to a reference human genome⁶.

Another deficiency of the four genomes⁴⁻⁷ is that they do not accurately define copy-number variants at the nucleotide level. These forms of genetic variation arise from the insertion of multiple copies of DNA segments that may include whole genes and that have been increasingly implicated in, among other disease phenotypes, neurological disorders^{9,10}.

Our genomes are not just collections of DNA variation: parental inheritance also dictates specific associations between neighbouring variations. Knowledge of these associations will ultimately help us discover whether and how much of an aberrant protein is produced by each of our cells and how these events contribute to observed phenotypes. The association between neighbouring variations across all 23 pairs of human chromosomes is referred to as haplotype assembly, and has not yet been completely achieved in any of the individual genomes sequenced.

These limitations notwithstanding, the approach of Bentley⁴, Wang⁵ and their colleagues represents a substantial advance in the sequencing of individual human genomes. Together with the other two genomes sequenced^{6,7}, they reinforce the catalogue of variants that exist in human genomes — SNPs in the millions, insertion/deletion polymorphisms in the hundreds of thousands and structural variants in the thousands. The numbers of these variants do not directly tell us how such polymorphisms contribute to the wide spectrum of human traits. But they do provide a necessary step towards accurately defining genomic loci that are likely to be implicated in those traits.

With such rapid advances in next-generation technologies, and with 'third generation' technologies emerging, this is just the beginning of the era of the individual genome. Soon, association studies using complete individual genomes will become the approach of choice for understanding the complexity of human biology and disease. The latest advances have broad implications for expediting that goal. ■ Samuel Levy and Robert L. Strausberg are at the J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850, USA. e-mail: slevy@jcv.org

1. Lander, E. S. *et al.* *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).
3. International Human Genome Sequencing Consortium *Nature* **431**, 931–945 (2004).
4. Bentley, D. R. *et al.* *Nature* **456**, 53–59 (2008).
5. Wang, J. *et al.* *Nature* **456**, 60–65 (2008).
6. Levy, S. *et al.* *PLoS Biol.* **5**, e254 (2007).
7. Wheeler, D. A. *et al.* *Nature* **452**, 872–876 (2008).
8. Kidd, J. M. *et al.* *Nature* **453**, 56–64 (2008).
9. Marshall, C. R. *et al.* *Am. J. Hum. Genet.* **82**, 477–488 (2008).
10. Walsh, T. *et al.* *Science* **320**, 539–543 (2008).

See Editorial, page 1.



50 YEARS AGO

Nobel Prize for Chemistry:

Dr. F. Sanger, F.R.S. — The award has been made for his researches on the structure of the protein hormone insulin ... When he began his investigations on insulin, Dr. Sanger first devised the use of dinitrofluorobenzene for the identification and estimation of the free amino-groups of proteins or peptides, and this method has since been widely adopted ... Dr Sanger's methods and example have stimulated much research in the investigation of protein structure, the limits of which have yet to be visualized, and they make clear the possibility that insulin may be completely synthesized in the laboratory, although this is unlikely to occur for some time to come.

From *Nature* 8 November 1958.

100 YEARS AGO

Windmills and Water-Wheels.

By R. S. Ball — As is natural, the author commences his book with a reference to the, said to be, not distant day when all the coal, and all the oil, in the world will have been used up, and mankind, in order to sustain itself, will have to rely wholly upon the water-wheel and the windmill for that tremendous amount of energy which will be necessary to keep the immense population of the earth in the state of comfort which it has, with the progress of civilization, attained.

ALSO:

A meeting of the Child Study Society was held on October 29, when a paper was read by Miss Alice Ravenhill on the results of an investigation into hours of sleep among elementary-school children ... The evil of insufficient sleep is widespread. Parents must be roused to a sense of the importance of the subject, and the enforcement of the laws on the employment of children should be rendered obligatory upon local authorities.

From *Nature* 5 November 1908.

50 & 100 YEARS AGO