



**Figure 1 | The good side of reactive oxygen species.** Blood-cell development in *Drosophila melanogaster* takes place in the larval lymph glands and is controlled by reactive oxygen species (ROS). Multipotent progenitor cells (pink) in the medullary zone of the primary lobe of the lymph gland differentiate into mature blood cells (orange), which reside in the cortical zone and are subsequently released into the open circulation. Blood-cell differentiation is influenced by the posterior signalling centre, a collection of cells that produce signalling molecules. Owusu-Ansah and Banerjee<sup>1</sup> show that ROS in blood-cell progenitors in the medullary zone are crucial for blood-cell development. Increased amounts of ROS activate the JNK signalling pathway. JNK activation in turn activates the FoxO pathway, leading to differentiation of plasmatocytes and crystal cells. Alternatively, repression of Polycomb proteins (PC) by the JNK pathway causes differentiation of lamellocytes. The secondary and tertiary lobes of the lymph gland are not shown in this schematic.

vertebrates, give rise to the myeloid lineages (red blood cells, monocytes and granulocytes). An obvious question is whether there is any connection between the amounts of ROS and haematopoiesis. When Owusu-Ansah and Banerjee<sup>1</sup> decreased the amounts of ROS by inducing the expression of antioxidant scavenger proteins in the lymph gland, the differentiation of progenitor cells to plasmatocytes was indeed significantly decreased. Conversely, induction of ROS production led to premature differentiation of all types of *Drosophila* blood cell.

As is seen in other systems, such as mammalian cells<sup>5</sup>, increased amounts of ROS activated a signal-transduction pathway that is often induced by stress, involving an enzyme known as Jun N-terminal kinase (JNK). JNK-pathway activity in turn activated the FoxO pathway, which is known to mediate signalling by insulin and other growth factors. The combined activity of the JNK and FoxO pathways induced differentiation of plasmatocytes and crystal cells in the *Drosophila* lymph gland (Fig. 1). Conversely, reduced expression of genes encoding Polycomb proteins downstream of JNK induced the specific production

of lamellocytes (Fig. 1). (Polycomb proteins affect gene expression by modifying the chromatin complex in which DNA is packaged.) Although some of the molecular details of the haematopoietic differentiation pathway in *Drosophila* remain to be elucidated, Owusu-Ansah and Banerjee's findings establish ROS as a major intrinsic effector of blood-cell development.

Evidence that ROS might have a similar role in vertebrates comes from work in adult mice showing that the transcription of genes involved in ROS production is also elevated in myeloid progenitor cells in the bone marrow<sup>6</sup>. This points to an evolutionarily conserved role for ROS during haematopoiesis, although the function of FoxO seems to be less widely conserved<sup>6</sup>, perhaps because the *Drosophila* larval lymph gland represents a developmentally earlier stage than vertebrate bone marrow, corresponding to the fetal liver or other sites of vertebrate embryonic haematopoiesis<sup>7</sup>. The relative contribution of sessile blood cells versus lymph-gland cells to larval or adult immunity in *Drosophila* is also not fully resolved<sup>4</sup>.

So are ROS good or bad? The possibly unsatisfying answer is: it depends. Yes, ROS can do damage if they mix with molecules and cells in the wrong place and at too high a concentration, but they also have many benefits, as highlighted by this study<sup>1</sup>. As is often observed in nature, it seems that we must pay an evolutionary price for recruiting potent effector mechanisms to our immune system. The present study comes at a time when the importance of endogenous antioxidant enzymes and the

long-term benefits of the use of antioxidant supplements are increasingly being questioned<sup>8–10</sup>. In addition, other previously unrecognized functions of ROS, such as a role for hydrogen peroxide in wound healing, have been discovered<sup>11</sup>. Perhaps the usefulness of antioxidants will have to be judged individually for each chemical and each target, and combinations thereof. For this purpose, studies at the level of the organism are indispensable, as they allow us to trace effects in different organs. Therefore, Owusu-Ansah and Banerjee's study<sup>1</sup> in the fly, and studies in other model organisms such as the nematode worm *Caenorhabditis elegans*<sup>8,9</sup> and the zebrafish<sup>11</sup>, should inspire further work on the many roles of ROS and their impact on human health. ■

Ulrich Theopold is in the Department of Molecular Biology and Functional Genomics, Stockholm University, 10691 Stockholm, Sweden. e-mail: uli@molbio.su.se

- Owusu-Ansah, E. & Banerjee, U. *Nature* **461**, 537–541 (2009).
- Crozatiere, M. & Meister, M. *Cell. Microbiol.* **9**, 1117–1126 (2007).
- Martinez-Agosto, J. A., Mikkola, H. K., Hartenstein, V. & Banerjee, U. *Genes Dev.* **21**, 3044–3060 (2007).
- Märkus, R. et al. *Proc. Natl Acad. Sci. USA* **106**, 4805–4809 (2009).
- Essers, M. A. et al. *EMBO J.* **23**, 4802–4812 (2004).
- Tothova, Z. et al. *Cell* **128**, 325–339 (2007).
- Cumano, A. & Godin, I. *Annu. Rev. Immunol.* **25**, 745–785 (2007).
- Doonan, R. et al. *Genes Dev.* **22**, 3236–3241 (2008).
- Van Raamsdonk, J. M. & Hekimi, S. *PLoS Genet.* **5**, e1000361 (2009).
- Bjelakovic, G. et al. *J. Am. Med. Assoc.* **297**, 842–857 (2007).
- Niethammer, P., Grabher, C., Look, A. T. & Mitchison, T. J. *Nature* **459**, 996–999 (2009).

## HUMAN GENETICS

# Tracing India's invisible threads

Aravinda Chakravarti

**One measure of the extraordinary level of human diversity found in India is the use of 15 languages on its banknotes. The genetic underpinnings of that population diversity are yielding to whole-genome analysis.**

The idea and shape of modern India was an invention of its twentieth-century political leaders, who crafted citizenship defined by civic and universalist, rather than ethnic or religious, criteria precisely because that citizenship is so diverse<sup>1</sup>. As Jawaharlal Nehru, the nation's first prime minister, wrote<sup>2</sup>: “[India] is four hundred million separate individual men and women, each differing from the other ... a bundle of contradictions held together by strong but invisible threads.” Who are these diverse peoples separated by caste, customs and language? Where did they come from, and when? What are the “invisible threads”, beyond claims on the state, that bind them? Studies of biological kinship, which search for the stories of ancestry marked indelibly in a

person's genome, help to provide answers to these questions because they illuminate that unwritten past<sup>3</sup>. The latest addition to our attempts to understand India through genes comes from Reich, Singh and their colleagues (page 489 of this issue)<sup>4</sup>, who arrive at some bold conclusions about its past population history from genome-variation studies.

The earliest occupation of the subcontinent was by Austro-Asiatic people about 60,000 years ago. They were dispersed and driven into smaller enclaves with the arrival of the Dravidian speakers around 3000 BCE (Before the Common Era, the Common Era marking the same divide as BC and AD). The latter people were themselves driven south with the arrival of the Indo-European speakers in about 1500 BCE.

These early events shaped the growth of an indigenous civilization, with much later conquests by Persians (543 BCE), Alexander III of Macedon (325 BCE), numerous colonial Europeans starting with the Portuguese (1510 CE), and the Mughals (1526 CE). They all came and they were all absorbed — their cultures and their genes — to create the current stew. Although there has been a preoccupation, by both native and foreign scholars, with understanding caste in India and the genetic differences it engenders, there is great diversity at every level: geography, language, caste and customs.

Studies of human variation in India started with the seminal anthropometric surveys of P. C. Mahalanobis in 1941. Subsequently, numerous investigators used various genetic markers (blood groups, serum proteins, enzymes and, later, DNA) to make sense of the vast diversity within the subcontinent. In the genomic era, the Indian Genome Variation Consortium<sup>5</sup> published a study of 420 single nucleotide polymorphisms (SNPs — base-pair variations in DNA) in 75 genes in 1,871 individuals. The consortium's sample was drawn from 55 groups representing all four language families (Austro-Asiatic, Dravidian, Indo-European, Tibeto-Burman), geography (north, south, east, west), social levels (caste, tribe, religion) and group abundance (small, large), to document the great genomic diversity and the clustering of variation by ethnicity and language (but see ref. 6). The results implied that genetic studies of disease in 'Indians' are hopelessly inadequate unless they account for their specific ancestry. This feature is genetic proof of a population structure first described by the social anthropologist Irawati Karve<sup>7</sup> as a "patchwork quilt where bits of material of the same colour and shape may be used in a pattern, but where each bit may be of an origin different in place and time".

Reich, Singh and colleagues<sup>4</sup> instead examine entire genomes' worth of 560,000 SNPs in 132 individuals from 25 groups representing the breadth of social, language and geographic variation in India (see Table 1 and Fig. 1 of their paper on page 490). They sample, in addition, two small groups (the Onge and Great Andamanese) from the Andaman Islands in the Bay of Bengal.

First, the authors show that Indian populations bear the genetic imprint of European, Asian and even, though rarely, African genomes<sup>5</sup>. Second, they find that diversity within India is three to four times greater than that observed within Europe, from which they conclude that many Indian populations, although currently large, were founded by small numbers of individuals with subsequent limited migration<sup>7</sup>. These founder events are dated by the genomic data to between 750 and 2,500 years ago, and therefore occurred well after the arrival of the putative Indo-European speakers.

Third, and most importantly, the authors clearly demonstrate that most of the Indian populations they sampled are mixtures of two

groups that they term ANI (Ancestral North Indians) and ASI (Ancestral South Indians). The degree of ANI:ASI mixture varies between 39% and 71% across India, and is evident in all caste and even tribal groups, and in both extant Indo-European and Dravidian speakers. However, greater ANI ancestry is significantly associated with Indo-European speakers and with traditionally 'higher' caste membership, even after controlling for language. This provides a model of how diversity within India came about. As such, its details are imperfect and will surely be contested, revised and improved; but its implications are significant.

Genetically, the ANI are closest to current-day Europeans whereas the ASI are closest to the disappearing Onge, but neither of these shared ancestries is recent. Reich, Singh *et al.* speculate that the ancestor to both Europeans and ANI spoke a proto-Indo-European language ancestral to both Sanskrit and European languages; the Onge-ASI ancestry is even more remote, and it is unclear whether the ASI were Dravidian speakers. Thus, Indians seem to have a unique set of ancestries for which each population is the same with respect to common descent from two major peoples, but different by virtue of its ancestry proportions and specific genomic content inherited — much like the many hands that can be dealt from a deck of cards. These interpretations are now possible because the authors<sup>4</sup> have developed new statistical methods to assess specific hypotheses regarding population relationships and ancestry, and also because comparable genomic-variation data on many additional worldwide samples are now available<sup>8–10</sup>.

The suggestion that each Indian population had small numbers of founders implies strong 'random genetic drift', whereby current frequencies of gene variants depart strongly from their ancestral frequencies simply by chance, thereby increasing genome similarities between members of the same group. This drift effect is largely a result of the early demographic history being shaped by limited numbers of founders, and creates an 'inbreeding' effect whereby genetic variation is lost. This aspect is independent of the additional loss of variation from consanguinity that is found in many parts of India. The cumulative effect is that gene variants may have quite distinct frequencies in India compared with that expected in many other 'related' populations, and that Indians bear the imprint of this very recent local shared ancestry.

This drift and differentiation has four implications. First, studies of relatively few individuals from any Indian population can characterize their common genomic variation adequately. Second, one predicts a high burden of genetically recessive disorders in India, many unique to each population, estimated to

be greater because of local shared ancestry than consanguinity. Third, some diseases will have elevated frequencies in many regions of India owing to shared ANI or ASI ancestry<sup>5</sup>. Fourth, without accounting for local ancestry, genetic association studies can suffer from numerous false positives arising from systematic differences in ancestry between cases and controls. Indeed, language and caste membership may not be adequate control factors.

To a cynic, the existence of the ANI or ASI, their unique and remote ancestry within India, or their suggestive identities as Indo-European and Dravidian speakers, are already common knowledge. But the precise definition of their

**"This provides a model of how diversity within India came about. As such, its details are imperfect, but its implications are significant."**

ancestral genomic content, their mixture throughout India and the importance of genetic drift are new and have serious implications for both human biology and medicine — and Indian society as well.

Nevertheless, the current analysis<sup>4</sup> is only a beginning.

The next stage will require samples from a much wider array of populations, including a better sampling of tribal populations and Tibeto-Burman speakers to understand their specific contributions. Indeed, sampling Indians, in the face of their diversity, is a challenge similar to that faced in Africa<sup>10</sup>. There is a strong impression that endogamy, the practice of preferring marriage within a group, in India has maintained genetic diversity. However, for this, endogamy must act locally where diverse populations interact<sup>7</sup>: its role can be assessed only by sampling humans locally, not populations distant from one another.

A more comprehensive analysis will require sampling Indians across a grid, assessing both their cultural and genetic diversities, for a deeper understanding of local population structure and the genetic effects of endogamy. Caste and custom may be strong barriers between groups, perhaps even today. But the common shared ancestry and rampant ANI:ASI mixture may be the strong, invisible thread that binds all Indians. ■

Aravinda Chakravarti is at the Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.  
e-mail: aravinda@jhmi.edu

1. Khilnani, S. *The Idea of India* (Farrar, Straus & Giroux, 1999).
2. Nehru, J. *The Discovery of India* (Oxford Univ. Press, 1946).
3. Chakravarti, A. *Nature* **457**, 380–381 (2009).
4. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. *Nature* **461**, 489–494 (2009).
5. Indian Genome Variation Consortium *J. Genet.* **87**, 3–20 (2008).
6. Rosenberg, N. A. *et al.* *PLoS Genet.* **2**, e215 (2006).
7. Karve, I. *Hindu Society: An Interpretation* 2nd edn (Deshmukh Prakashan, Poona, 1968).
8. The International HapMap Consortium *Nature* **449**, 851–861 (2007).
9. Li, J. Z. *et al.* *Science* **319**, 1100–1104 (2008).
10. Tishkoff, S. A. *et al.* *Science* **324**, 1035–1044 (2009).