DISEASE SUSCEPTIBILITY

That damned elusive polygene

Common diseases such as diabetes, heart disease and asthma are polygenic — disease susceptibility is influenced by several genes (referred to as polygenes), each thought to have a small effect. The identification of polygenes holds great promise for our understanding of disease aetiology and for the development of new therapies. However, efforts to apply positional cloning to the identification of polygenes have met with little success — until now. Graeme Bell and colleagues have used this approach to identify a gene that is strongly implicated in susceptibility to type 2 diabetes, the most common form of diabetes.

Bell's group had previously mapped a susceptibility locus for type 2 diabetes in Mexican Americans to a 12-cM region on chromosome 2. The latest work began by refining the locus to a 7-cM interval, which was then cloned into a contig and shown to encompass 1.7 Mb. This relatively high ratio of recombination rate to physical distance was one piece of good fortune and is thought to be caused by the region's proximity to the telomere. The region contained seven genes and 15 expressed sequence tags (ESTs).

The next phase of the work was to characterize new polymorphisms in the 1.7-Mb interval and to carry out association studies, with the aim of refining the location of the genetic determinant (or determinants) that is associated with disease susceptibility. A huge amount of work culminated in the identification of a single, intronic polymorphism (UCSNP-43) that was significantly associated with disease susceptibility. UCSNP-43 is in a gene (*CAPN10*) that encodes a calpain-like protease. But does this polymorphism itself cause increased disease susceptibility or is it merely associated with a nearby causative variant?

To address this question, Horikawa *et al*. resequenced the 66-kb region that encompasses *CAPN10* in 10 Mexican Americans and uncovered 179 polymorphisms, 63 of which were typed in 100 diabetic cases and 100 controls. Haplotype analysis led to the conclusion that significant elevated risk for diabetes is conferred by heterozygosity for two haplotypes — homozygosity for either haplotype had no association with altered disease susceptibility. This twist to the story suggests that there is a complex interplay of genetic variants in the vicinity of *CAPN10* that leads to altered diabetes susceptibility in Mexican Americans. However, the identity of the causative variants is still not known, and might even lie outside *CAPN10*.

Like Baroness Orczy's Scarlet Pimpernel, the polygene is proving to be an elusive and tantalizing quarry. The study by Horikawa *et al*. is the most successful report to date, and it provides a novel line of inquiry that could be pursued by diabetologists — how might a calpain-like protease be involved in diabetes? But the work also demonstrates just how difficult it is to make the jump from evidence for a disease-susceptibility locus to the gene, and ultimately to the molecular defect that underlies that susceptibility.

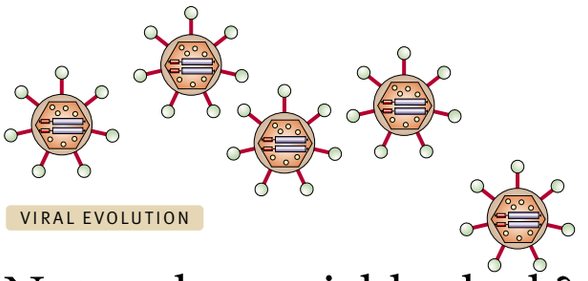
Mark Patterson

References and links

ORIGINAL RESEARCH PAPER Horikawa, Y. *et al*. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nature Genet.* **26**, 163–175 (2000)

FURTHER READING Altshuler, D., Daly, M. & Kruglyak, L. Guilt by association. *Nature Genet.* **26**, 135–137 (2000) | Weiss, K. M. & Terwilliger, J. D. How many diseases does it take to map a gene with SNPs. *Nature Genet.* **26**, 151–157 (2000)

WEB SITE Graeme Bell



VIRAL EVOLUTION

Not such a variable clock?

Viruses are always changing — forced to do so by the constant pressure to dodge the host immune system. RNA-based viruses, such as HIV and influenza, are infamous for evolving very rapidly, and so understanding the forces that shape viral evolution could have significant benefits for public health.

One important question to consider is whether the rate of viral evolution is constant, and valuable insight into this question has recently been gained by two Danish researchers.

Dating where branches split on an evolutionary tree of viral sequences (or any sequence, for that matter) relies upon a constant rate of molecular evolution, known as the 'molecular clock'. The intrinsic rate of the molecular clock, and the degree to which two sequences differ, allows one to infer the time at which the sequences diverged in the past. For viruses, this can help to pinpoint the origins of a particular epidemic.

However, statistical analyses of virus sequence evolution are inconsistent with the molecular clock model, and this has often been cited as evidence that the viral mutation rate changes in response to selection pressures. An alternative explanation — one that is tested by Schierup and Hein — is that the apparent variation in mutation rate is just an artefact caused by ignoring viral recombination.

The authors simulated, *in silico*, the evolution of 1,000-base-pair viral sequences. In each data run, the progenitor viral sequence was allowed to evolve up to an average divergence of 20%, but at a different recombination rate. The family of virus sequences that evolved in each run was used to build two phylogenetic trees — one with, the other without, assuming a molecular clock. In each experiment, the tree that is based on a molecular-clock model of evolution is taken as the null hypothesis. In this way, the two trees can be compared (using a likelihood ratio test) to measure the lowest amount of recombination at which the molecular clock is rejected (that is, the recombination rate above which the two trees are significantly different).

The results of the test are striking: as soon as the total number of recombinations in the history of ten sequences exceeds six (much fewer than actually occur), the molecular clock is rejected in more than 50% of runs. So viral sequences may well be evolving at a constant rate, but this isn't detectable because it is masked by recombination.

Can we still safely say, then, that the origin of the HIV1 pandemic can be dated to 1959, on the basis of a phylogenetic analysis? The pressure is on to invent ways of analysing viral sequence evolution in which the recombination rate is included as part of the method.

Tanita Casci

References and links

ORIGINAL RESEARCH PAPER Schierup, M. H. & Hein, J. Recombination and the molecular clock. *Mol. Biol. Evol.* **17**, 1578–1579 (2000)

FURTHER READING Huelsenbeck, J. P. & Rannala, B. Phylogenetic methods come of age: Testing hypotheses in a phylogenetic context. *Science* **276**, 227–232 (1997) | Schierup, M. H. & Hein, J. Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**, 879–891 (2000)

WEB SITES Mikkel Schierup's lab | Statistical analysis software