EVOLUTION

A matter of balance

Balancing selection is the retention of advantageous population-wide diversity at genomic loci. The best characterized example in mammals is the polymorphic nature of major histocompatibility complex (MHC) loci, which benefits mammalian hosts in their immune responses to pathogens, but it is unclear how widespread balancing selection is beyond isolated examples. A new study uses a genome-wide cross-species approach to identify and characterize various additional loci that are under balancing selection in humans.

Leffler *et al.* searched for single-nucleotide polymorphism (SNP) alleles that were shared between species in the whole-genome sequences of 59 human and 10 chimpanzee individuals. The resulting set of 33,906 SNPs could contain historic SNPs that have been retained in a polymorphic state by balancing selection since these species split (termed 'identity by descent'). However, confounding the detection of balancing selection, such shared SNPs can also result from independent mutations at the same site in each species subsequent to their split, followed by random genetic drift (termed 'identity by state'). Therefore, the authors applied two alternative

filtering steps to enrich for genuine sites of balancing selection.

First, they analysed shared SNPs in coding regions on the basis of the idea that SNPs in coding regions will largely be non-neutral and thus less subjected to confounding genetic drift. As a validation of the approach, MHC genes showed a clear signature of balancing selection by collectively harbouring 18 shared coding SNPs. Additionally, 335 shared coding SNPs were found across 324 other genes, although most of these SNPs were synonymous (that is, they resulted in no amino acid change).

In the second filtering approach, the authors identified shared haplotype blocks of up to 4 kb that contained at least two shared SNPs; this configuration is unlikely to have randomly arisen through genetic drift, and this filter is expected to eliminate >96% of SNPs that are identical by state. As well as finding expected regions encompassing MHC loci, this approach identified 125 additional candidate regions of balancing selection. Interestingly, these regions were predominantly non-coding: only two of the total SNPs in these regions were in coding DNA, and both of these SNPs were synonymous. Combined with the prevalence of synonymous SNPs from the first approach, this indicates that some balancing selection does not affect protein sequences and thus perhaps functions

instead through gene-regulatory roles. Consistent with this, various candidate regions overlapped putative regulatory elements, such as sites with known chromatin signatures of enhancer activity.

To gain global insight into the types of genes that might be regulated by the identified balancing selection, the authors used gene-set enrichment analysis to find characteristic gene types near to the SNPs identified from both filtering approaches. In both cases, an enrichment was seen for membrane glycoproteins and for enzymes involved in protein glycosylation. Host glycoproteins are known to have key roles in host–pathogen interactions, thus it seems that pathogens have had a more widespread role in shaping human genetic diversity through balancing selection than was previously appreciated.

It will be interesting to dissect further the functional effects of this newly identified balancing selection: in particular, to determine whether the identified SNPs have their proposed regulatory effects on the expression of nearby genes and, if so, whether this can be confirmed to alter aspects of host–pathogen interactions.

Darren J. Burgess

ORIGINAL RESEARCH PAPER Leffler, E. M. *et al.*
Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science* 14 Feb 2013 (doi:10.1126/science.1234070)

“ various candidate regions overlapped putative regulatory elements ”