GENOME EVOLUTION

Adaptive human regulatory variation

> *Arbiza et al.* have found evidence for a pervasive role of binding site mutation in human evolution
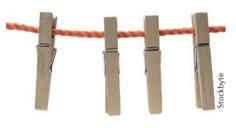
The contribution of mutations in transcription factor binding sites (TFBSs) to mammalian evolutionary adaptation has been much debated. Now, Arbiza *et al.* have found evidence for a pervasive role of binding site mutation in human evolution since the split from chimpanzees.

Previous studies of the role of *cis*-regulatory mutations in human evolutionary adaptation have been hampered by a lack of experimentally defined TFBSs and information about polymorphisms in these regions. The authors used data from the ENCODE project to annotate binding sites for 78 transcription factors, which are estimated to constitute ~5% of transcription factors in the human genome. They combined these with genome sequence data from 54 individuals from the Complete Genomics project and with synteny-based alignments from the chimpanzee, orangutan and rhesus macaque genomes to identify the patterns of polymorphism and divergence.

Arbiza *et al.* then applied a novel probabilistic method termed INSIGHT ('inference of natural selection from interspersed genomically coherent elements') to these data. The method compares TFBSs with flanking neutrally evolving sites to detect signatures of selection in binding sites. It is an improvement on previous related methods as it is able to identify signatures of weak negative, strong negative and positive selection, to integrate population data with data from phylogenetic analyses and to carry out these analyses on short elements.

The authors found strong signatures of selection in the TFBSs. For most sites, this was negative selection, but some signatures of positive selection were seen. Furthermore, they were able to correlate the extent of selection with binding affinity. The fraction of sites under selection was also correlated with the level of expression of the gene and with the number of cell types in which the transcription factor was bound, giving hints as to the source of the selective pressure.

As regards the relative contributions from TFBSs and coding regions to adaptive substitutions, they are on the same order, and the authors estimate that the contributions to deleterious polymorphisms are substantially greater from TFBSs. The higher frequencies of deleterious mutations in TFBSs suggest that these mutations are less deleterious, on average, which may have implications for disease causing mutations.

The findings here pave the way for consideration of the importance of non-coding variation in adaptive evolution. Although this study focused on single-nucleotide polymorphisms (SNPs), it will be interesting to analyse contributions from other types of variants as well.

Hannah Stower

ORIGINAL RESEARCH PAPER Arbiza, L. *et al.* Genome-wide inference of natural selection on human transcription factor binding sites *Nature Genet.* http://dx.doi.org/10.1038/ng.2658 (2013)

Stockbyte