NEVER ENDING HUMAN GENOME PROJECT

Lee et al. (BMC Biology) have used a new software (InserTag) to analyze the sequence data obtained from the 2535 individuals of the 1000 Genome Project in search of missing sequences (>50bp) in the human reference assembly (hg19). They found 1696 non-reference insertion variants, re-classified as (i) retention of ancestral sequences* or (ii) novel sequence insertions, based on the ancestral state. Individuals had, on average, 0.92-Mbp sequences missing from the reference genome. 92% of the variants were common (> 5%), and more than half were major alleles (!).

* "retention of ancestral sequences" means that they are present in non-human hominids (chimpanzee, gorilla, orangutan).