Nature Reviews Genetics | AOP, published online 30 April 2013; doi:10.1038/nrg3499

POPULATION GENOMICS

Noah's Ark arrives

ImageSource

> "[this] study shows how transcriptomics can be used to extend population genomics research across a far greater range of organisms"

Population genomics — in which genetic variation across the genome is studied in many individuals — has so far been limited to species for which reference genomes are available. A recent study shows how transcriptomics can be used to extend population genomics research across a far greater range of organisms. This advance promises important new insights into the factors that influence genetic diversity and evolutionary change.

To bypass the need for a reference genome, Gayral, Melo-Ferreira and colleagues developed a pipeline in which transcriptomic data from next-generation sequencing is used to characterize genetic variation. In this process, predicted cDNAs are assembled, reads are mapped, single-nucleotide polymorphisms (SNPs) are identified, and population genomic statistics are inferred.

Many challenges had to be addressed in assaying genetic variation in this way, such as differences in expression levels between genes and the need to differentiate sequencing errors from genetic variation. The authors note that a particular challenge arises from cases in which reads from distinct loci map to the same predicted cDNA. For example, in the case of paralogous genes, it is hard to distinguish differences between paralogues from genetic variation. A new method was introduced to deal with such situations.

Using this pipeline, the authors generated population genomic data for five species. Two of these were vertebrates (a turtle and a hare), and three were invertebrates (an oyster, a tunicate and a termite). For two of these species — the tunicate and the hare — a reference genome is available. This allowed a comparison of population genomic data generated both with and without use of a genome sequence. Encouragingly, they found only small differences between the population genomic statistics that were inferred using the two approaches.

The authors examined some key questions in population genomics using their newly generated data. Among animal species, population genomics had previously been limited to fruitflies and humans. These studies suggested a sharp division between vertebrates and invertebrates in terms of levels of genetic variation and rates of adaptive evolution, with both being higher in invertebrates. However, the newly studied species showed levels of genomic diversity that are intermediate between those for fruitflies and humans, suggesting a less marked division. Importantly, in terms of the ratio of nonsynonymous to synonymous polymorphisms — an indicator of inability to purge deleterious mutations — no significant difference was seen between vertebrates and invertebrates. This calls into question the accepted hypothesis that effective population size (which is generally higher in invertebrates) is the main determinant of the ability to remove harmful genetic variation.

The road ahead for population genomics promises to be a fruitful one as the door is now open for a much greater range of species to be explored in this way. As indicated by this initial study, some key concepts may need to be rethought as a result.

Louisa Flintoft

ORIGINAL RESEARCH PAPER Gayral, P. *et al.* Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. *PLoS Genet.* **9**, e1003457 (2013)