

DOI:

10.1038/nrg2245

 MILESTONE 10

Randomness versus order

Whereas randomness is avoided in most experimental techniques, it is fundamental to sequencing approaches. In the race to sequence the human genome, research groups had to choose between the random whole-genome shotgun sequencing approach or the more ordered map-based sequencing approach.

When Frederick Sanger and colleagues sequenced the 48-kb bacteriophage λ genome in 1982, the community was still undecided as to whether directed or random sequencing strategies were better. With directed strategies, DNA sequences were broken down into ordered and overlapping fragments to build a map of the genome, and these fragments were then cloned and sequenced. With the shotgun approach, DNA sequences were broken randomly, cloned, sequenced and then pieced together by analysing the overlap. Sanger *et al.* compared these strategies while sequencing bacteriophage λ and reported that the random approach was faster than any directed method.

One problem with the random approach, however, was that of filling gaps when the sequence was nearly complete (or closure), as randomly selected clones were often redundant. For instance, Sanger *et al.*

used — in their opinion mistakenly — direct sequencing strategies to finish the last 10% of the bacteriophage λ sequence. In 1991, Al Edwards and Thomas Caskey proposed a method to maximize efficiency by minimizing gap formation and redundancy: sequence both ends (but not the middle) of a long clone, rather than the entirety of a short clone.

Although the shotgun approach was now accepted for sequencing short stretches of DNA, map-based techniques were still considered necessary for large genomes. Like the directed strategies, map-based sequencing subdivided the genome into ordered 40-kb fragments, which were then sequenced using the shotgun approach. In 1995, however, Robert Fleischmann and colleagues used a whole-genome shotgun approach to sequence the 1,800-kb genome of *Haemophilus influenzae* — the first complete genome of a free-living organism. The authors had randomly generated large 40-kb fragments and had thereby bypassed the mapping stage. In doing so, they had proved that genome-assembly programmes that matched overlap were reliable and that whole-genome shotgun sequencing worked, in principle.

The *H. influenzae* genome, however, was a mere DNA fragment compared with the 1,500-fold longer ~3 billion base-pair human genome. In 1996, Craig Venter and colleagues proposed that the whole-genome shotgun approach could be used to sequence the human genome owing to two factors: its past successes in assembling genomes and the development of bacterial artificial chromosomes (BAC) libraries, which allowed large fragments of DNA to be cloned.

A showdown ensued, with the biotechnology firm Celera Genomics wielding whole-genome shotgun sequencing and the International Human Genome Sequencing Consortium wielding map-based sequencing. Yet, when the dust settled, it was a draw — both groups published their initial drafts of the human genome concurrently in 2001.

Asher Mullard, Assistant Editor,
Nature Reviews Microbiology and
Nature Reviews Molecular Cell Biology

ORIGINAL RESEARCH PAPERS Sanger, F. *et al.* Nucleotide sequence of bacteriophage λ DNA. *J. Mol. Biol.* **162**, 729–773 (1982) | Edwards, A. & Caskey, C. T. Closure strategies for random DNA sequencing. *Methods* **3**, 41–47 (1991) | Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995) | Venter, J. C. *et al.* A new strategy for genome sequencing. *Nature* **381**, 364–366 (1996)

“In the race to sequence the human genome, research groups had to choose between the random whole-genome shotgun sequencing approach or the more ordered map-based sequencing approach.”