CURRENT SCIENCE

Volume 103 Number 6 25 September 2012

EDITORIAL

DNA: Finding Uses for Junk

'Junk' is a common word. As an inveterate accumulator of unusable, unserviceable and useless material, including books, papers, gadgets and odds and ends, I am usually surrounded at the workplace and at home by large, untidy piles of junk. Throwing away things that cannot be put to good use seems a painful act; sentiment invariably overwhelms common sense. In our laboratories, junk accumulates rather quickly, as computers and equipment are replaced with an alarming regularity. The discarded items stand forlorn and abandoned in dark corners. Junk is distinguishable from garbage. The former accumulates in our homes and offices; the latter is disposed off outside, often resulting in unsightly and smelly piles when clearance mechanisms fail; a not uncommon occurrence in our ever expanding cities. The reasons why junk is not as readily discarded as garbage are not easy to enumerate; most often it is sheer lethargy. The energetic (and financial) costs of clearing junk often overweigh the costs of storage. Junk is a word that caught my attention a few days ago in a news report in *Science*, headlined 'ENCODE Project Writes Eulogy for Junk DNA' (Pennisi, E., *Science*, 2012, **337**, 1160). The provocation for this story, and many others that have appeared in the popular press, is the publication last week of as many as thirty papers announcing the results of one of biology's mega-projects, to establish the function of vast amounts of DNA in the human genome, which at first glance appear to be largely 'non-coding'. The sheer scale of the project, succinctly summarized in the *Science* report, should attract the attention of any casual observer of the scientific scene. A financial outlay of $288 million supported 442 researchers at 32 institutions worldwide, to carry out a genome wide survey of functional attributes of the entire genome length of over 3 billion nucleotides. In common parlance, this amounts to deciphering a biochemical significance for long stretches of genomic DNA, which previously were thought to have little or no specific functional role; barren wastelands of DNA that were termed as 'junk DNA'.

The term 'junk DNA' appears to have first surfaced in two articles written by the geneticist Susumu Ohno in the early 1970s, at a time when molecular biology was a young discipline and DNA sequencing was a technology that had not appeared over the horizon. Curiously, Ohno may have used the term twice in the titles of two conference presentations, which later appeared in printed form (Ohno, S., So much 'junk' DNA in our genome. In *Evolution of Genetic Systems* (ed. Smith, H. H.), Gordon and Breach, New York, 1972, pp. 366–370; Ohno, S., Evolutional reason for having so much junk DNA. In *Modern Aspects of Cytogenetics: Constitutive Heterochromatin in Man* (ed. Pfeiffer, R. A.), F. K. Schattauer Verlag, Stuttgart, Germany, 1973, pp. 169–173). A thoughtful commentary on the internet draws attention to the fact 'that most people who cite these papers have not read them'; often drawing the incorrect conclusion that 'non-coding DNA was totally unimportant' (Gregory, R. T., 2008; http://www.genomicron.evolverzone.com). This web posting reproduces an interesting discussion which followed Ohno's 1973 presentation. One of the participants notes that 'the word "junk" is a powerful word' and appears uncomfortable with a notion that large, conserved segments of DNA have no function. This is emphasized by another participant who adds: '…this DNA must have a functional value since nothing is known so widespread and universal in nature that has proved so useless'. The final interjection by a perceptive participant may strike a chord: 'Well there is an exception to that rule. A lot of us have permanent positions at the university but are considered by others (mainly by students) meaningless and of no utility whatsoever.'

The internet, an almost infinite source of information, hosts Ohno's first paper. Appropriately, the site on which the paper is displayed lists the 'archives of the life and death of junk DNA 1972–2008' (www.junkdna.com). Junk DNA's demise may be traced to the appearance, in 2007, of the preliminary results of the ENCODE project, which examined about 1% of the human genome (The ENCODE Project Consortium, *Nature*, 2007, **447**, 799). The impending disintegration of a 'gene centric' approach to genome analysis was proclaimed, once the results established that a great deal of 'junk DNA' was indeed transcribed into RNA and such sequences were 'evolutionarily conserved across mice and humans' (Pennisi, E., *Science*, 2007, **316**, 1556). Did 'junk DNA' as a non-functional, apparently useless element of genomes become an accepted view immediately after Ohno's presentations of 1972 and 1973? Ohno derived his arguments from the traditional viewpoints of genetics: 'The mammalian genome (haploid chromosome complement) contains roughly 3.0×10^{-9} mg of DNA which represents 3.0×10^{9} base pairs. This is at least 750 times the genome

size of *E. coli*. If we take the simplistic assumption that the number of genes contained is proportional to the genome size, we would have to conclude that 3 million or so genes are contained in our genome. The falseness of such an assumption becomes clear when we realize that the genome of lowly lungfish and salamanders can be 36 times greater than our own' (Ohno, 1972). Using arguments based on the probability of acquiring deleterious mutations in functional genes, Ohno argued that 'there seems to be a strict upper limit for the number of gene loci which we can afford to keep in our genome'. Ohno noted that gene duplication created multiple copies, allowing 'redundant copies to accumulate formerly forbidden mutations and thereby to acquire new functions'. Ohno's concluding paragraph bears reproduction at a time when the ENCODE project has focused attention on what was once thought to be a vast stretch of desert in our genomes. There is eloquence in Ohno's writing: 'The chance of acquiring a new function by unrestricted accumulation of mutations, however, should be as small as that of an isolated population emerging triumphant as a new species. Degeneracy is the more likely fate. The creation of every new gene must have been accompanied by many other redundant copies joining the ranks of silent DNA sequences, and these silent DNA base sequences may now be serving the useful but negative function of spacing those that have succeeded. Triumphs as well as failures of nature's past experiments appear to be contained in our genome' (Ohno, 1972). Ohno's arguments may have been less than persuasive in convincing his audience, that the overwhelming majority of DNA bases in human genomes were contributing to 'junk'. A comment from the audience (Boyer) is illustrative: 'It thus seems to me that the permissible number of structural loci is – as yet – a somewhat suspect way to arrive at figures of 1% structural utility to 99% junk.'

The community of molecular geneticists in the 1970s was clearly grappling with the uncomfortable thought that higher organisms appeared to waste a great deal of effort in maintaining large genomes, even though much of the enormous length of DNA sequence had no ascribable function. Genomes seemed to resemble inefficient public institutions which harbour large numbers of staff with little by way of work. Nature's extravagance has attracted imaginative explanations, most notably Richard Dawkin's highly influential exposition of the concept of 'selfish genes' (Dawkins, R., *The Selfish Gene*, Oxford University Press, 1976). A 1980 review declared selfish DNA as 'the ultimate parasite' (Orgel, L. E. and Crick, F. H. C., *Nature*, 1980, **284**, 604). Selfish DNA, according to Orgel and Crick, is characterized by its ability to spread 'by forming additional copies of itself within the genome' and by making 'no specific contribution to the phenotype'. They go on to conclude that 'the conviction has been growing that much of this extra DNA is "junk", in other words, that it has little specificity and conveys little or no selective advantage to the organism…. There is a large amount of evidence which suggests but does not prove, that much DNA in higher organisms is little better than junk'. Orgel and Crick repeatedly use the word 'junk', undoubtedly influencing much of the later thinking on the human genome. Ohno's presentations of 1972 and 1973 are conspicuous by their absence in the list of cited references, while a 1972 paper which does not explicitly introduce the word 'junk' into the vocabulary of molecular biology is mentioned in passing.

In advancing the concept of selfish DNA and considering its implications for the role of Darwinian natural selection on genomes of higher organisms, Orgel and Crick range far and wide, but reach a cautious conclusion: 'While proper care should be exercised both in labelling as selfish DNA, every piece of DNA whose function is not immediately apparent and in invoking plausible but unproven hypotheses concerning the details of natural selection, the idea seems a useful one to bear in mind when exploring the complexities of the genomes of higher organisms.' Genome sequencing and the ENCODE project provide an unprecedented and bewilderingly large amount of information on DNA sequences and function, holding out the promise of providing illuminating insights on the manner in which evolution has shaped genomes.

The 1980 issue of *Nature* which carried the Orgel–Crick review also contained a perceptive overview of evolutionary selective pressures and their role in shaping genomes. W. F. Doolittle and C. Sapienza argued that 'natural selection operating within genomes will inevitably result in the appearance of DNAs with no phenotypic expression whose only "function" is survival within genomes'. The authors note that 'when all attempts to assign a given sequence or class of DNA functions of immediate phenotypic benefit to the organism fail we resort to evolutionary explanations'. Indeed one hypothesis often advanced is almost impossible to test – sequences of undefined function are 'yet-to-be eliminated by-products of past chromosomal rearrangements of evolutionary significance'. Doolittle and Sapienza note that 'the phenotype paradigm is almost tautological; natural selection operates on DNA through organismal phenotype, so DNA structure must be of immediate or long-term (evolutionary) phenotypic benefit, even when we cannot show how'. Molecular biologists have often turned to evolutionary imperatives as a crutch to hold up hypotheses that are hard to test. This has not always been viewed kindly by evolutionary biologists. Stephen Gould and Richard Lewontin are quoted by Doolittle and Sapienza: 'The rejection of one adaptive story usually leads to its replacement by another, rather than to a suspicion that a different kind of explanation might be required. Since the range of adaptive stories is as wide as our minds are fertile, new stories can always be postulated' (Gould, S. J. and Lewontin, R. C., *Proc. R. Soc. London*, 1979, **205**, 581). Even as the ENCODE results are analysed, new stories about genomes and evolution will surely emerge, even as 'junk DNA' is laid to rest.

P. Balaram