Selfish genes, the phenotype paradigm and genome evolution

W. Ford Doolittle & Carmen Sapienza

Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7

Natural selection operating within genomes will inevitably result in the appearance of DNAs with no phenotypic expression whose only 'function' is survival within genomes. Prokaryotic transposable elements and eukaryotic middle-repetitive sequences can be seen as such DNAs, and thus no phenotypic or evolutionary function need be assigned to them.

THE assertion that organisms are simply DNA's way of producing more DNA has been made so often that it is hard to remember who made it first. Certainly, Dawkins has provided the most forceful and uncompromising recent statement of this position, as well as of the position that it is the gene, and not the individual or the population, upon which natural selection acts[1]. Although we may thus view genes and DNA as essentially 'selfish', most of us are, nevertheless, wedded to what we will call here the 'phenotype paradigm'—the notion that the major and perhaps only way in which a gene can ensure its own perpetuation is by ensuring the perpetuation of the organism it inhabits. Even genes such as the segregation-distorter locus of *Drosophila*[2], 'hitch-hiking' mutator genes in *Escherichia coli*[3,4] and genes for parthenogenetic reproduction in many species[4]—which are so 'selfish' as to promote their own spread through a population at the ultimate expense of the evolutionary fitness of that population—are seen to operate through phenotype.

The phenotype paradigm underlies attempts to explain genome structure. There is a hierarchy of types of explanations we use in efforts to rationalize, in neo-darwinian terms, DNA sequences which do not code for protein. Untranslated messenger RNA sequences which precede, follow or interrupt protein-coding sequences are often assigned a phenotypic role in regulating messenger RNA maturation, transport or translation[5-7]. Portions of transcripts discarded in processing are considered to be required for processing[8]. Non-transcribed DNA, and in particular repetitive sequences, are thought of as regulatory or somehow essential to chromosome structure or pairing[9-11]. When all attempts to assign to a given sequence or class of DNA functions of immediate phenotypic benefit to the organism fail, we resort to evolutionary explanations. The DNA is there because it facilitates genetic rearrangements which increase evolutionary versatility (and hence long-term phenotypic benefit)[12-17], or because it is a repository from which new functional sequences can be recruited[18,19] or, at worst, because it is the yet-to-be eliminated by-product of past chromosomal rearrangements of evolutionary significance[9,19].

Such interpretations of DNA structure are very often demonstrably correct; molecular biology would not otherwise be so fruitful. However, 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. As Gould and Lewontin note, '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' (ref. 20).

Non-phenotypic selection

What we propose here is that there are classes of DNA for which a 'different kind of explanation' may well be required. Natural selection does not operate on DNA only through organismal phenotype. Cells themselves are environments in which DNA sequences can replicate, mutate and so evolve[21]. Although DNA sequences which contribute to organismal phenotypic fitness or evolutionary adaptability indirectly increase their own chances of preservation, and may be maintained by classical phenotypic selection, the only selection pressure which DNAs experience directly is the pressure to survive within cells. If there are ways in which mutation can increase the probability of survival within cells without effect on organismal phenotype, then sequences whose only 'function' is self-preservation will inevitably arise and be maintained by what we call 'non-phenotypic selection'. Furthermore, if it can be shown that a given gene (region of DNA) or class of genes (regions) has evolved a strategy which increases its probability of survival within cells, then no additional (phenotypic) explanation for its origin or continued existence is required.

This proposal is not altogether new; Dawkins[1], Crick[6] and Bodmer[22] have briefly alluded to it.

However, there has been no systematic attempt to describe elements of prokaryotic and eukaryotic genomes as products of non-phenotypic selection whose primary and often only function is self-preservation.

Transposable elements in prokaryotes as selfish DNA

Insertion sequences and transposons can in general be inserted into a large number of chromosomal (or plasmid) sites, can be excised precisely or imprecisely and can engender deletions or inversions in neighbouring chromosomal (or plasmid) DNAs[12–16]. These behaviours and, at least in some cases, the genetic information for the enzymatic machinery involved, must be inherent in the primary sequences of the transposable elements themselves, which are usually tightly conserved[12–16,23]. Most speculations on the function of transposable elements concentrate on the role these may have, through chromosomal rearrangements and the modular assembly of different functional units, in promoting the evolution of plasmid and bacterial chromosomes, and thus in promoting long-term phenotypic fitness[12–16]. Most assume that it is for just such functions that natural selection has fashioned these unusual nucleic acid sequences.

Although transposable elements may well be beneficially involved in prokaryotic evolution, there are two reasons to doubt that they arose or are maintained by selection pressures for such evolutionary functions.

First, DNAs without immediate phenotypic benefit are of no immediate selective advantage to their possessor. Excess DNA should represent an energetic burden[24,25], and some of the activities of transposable elements are frankly destructive[12–16]. Evolution is not anticipatory; structures do not evolve because they might later prove useful. The selective advantage represented by evolutionary adaptability seems far too remote to ensure the maintenance, let alone to direct the formation, of the DNA sequences and/or enzymatic machinery involved. A formally identical theoretical difficulty plagues our understanding of the origin of sexual reproduction, even though this process may now clearly be evolutionarily advantageous[1,4].

Second, transposability itself ensures the survival of the transposed element, regardless of effects on organismal phenotype or evolutionary adaptability (unless these are sufficiently negative). Thus, no other explanation for the origin and maintenance of transposable elements is necessary. A single copy of a DNA sequence of no phenotypic benefit to the host risks deletion, but a sequence which spawns copies of itself elsewhere in the genome can only be eradicated by simultaneous multiple deletions. Simple translocation (removal from one site and insertion into another) does not provide such insurance against deletion. It is significant that recent models for transposition require retention of the parental sequence copy[26,27], and that bacterial insertion sequences are characteristically present in several copies per genome[16]. The assumption that transposable elements are maintained by selection acting on the cell does not require that they show these characteristics. The evolutionary behaviour of individual copies of transposable elements within the environment represented by a bacterial genome and its descendants can be understood in the same terms as organismal evolution. Replicate copies of a given element may diverge in sequence, but at least those features of sequence required for transposition will be maintained by (non-phenotypic) selection; copies which can no longer be translocated will eventually suffer elimination. Some divergent copies may be more readily transposed; these will increase in frequency at the expense of others. Transposable elements which depend on host functions run the risk that host mutants will no longer transpose them; it is significant that at least some transposition-specific functions are known to be coded for by the transposable elements themselves[26–29]. It is not to the advantage of a transposable element coding for such functions to promote the transposition of unrelated elements; the fact that given transposable elements generate flanking repeats[16,30] of chromosomal DNAs of sizes characteristic to them (that is, 5, 9 or 11–12 base pairs) may indicate such a specificity in transposition mechanism. It is to the advantage of any transposable element to acquire genes which allow independent replication (to become a plasmid), promote host mating (to become a self-transmissable plasmid) or promote non-conjugational transmission (to become a phage like Mu).

It is certainly not novel to suggest that prokaryotic transposable elements behave in these ways, or to suggest that more frankly autonomous entities like phages have arisen from them[12–16,31]. However, we think it has not been sufficiently emphasized that non-phenotypic selection may inevitably give rise to transposable elements and that no phenotypic rationale for their origin and continued existence is thus required.

Transposable elements in eukaryotes

There has long been genetic evidence for the existence in eukaryotic genomes of transposable elements affecting phenotype[32]. These have been assigned roles in the regulation of eukaryotic gene expression and in evolution, but would have escaped genetic detection had they not had phenotypic effect. More recent evidence for transposable elements whose effects are not readily identified genetically has come fortuitously from studies of cloned eukaryotic DNAs. For instance, the Ty-1 element of yeast (which has no known phenotypic function) is flanked by direct repeats (like some prokaryotic transposons) and is transposable[33]. It is present in some 35 dispersed copies and comprises some 2% of the yeast genome (like a higher-eukaryotic middle-repetitive DNA). The directly repeated δ-sequence elements flanking it are found at still other sites (just as prokaryotic insertion sequences can be found flanking transposons or independently elsewhere in the genome). Cameron *et al.* suggest that 'Ty-1 may be a nonviral "parasitic" DNA' but then go on to suggest, we think unnecessarily, that transposition 'allows adaptation of a particular cell to a new environment' (ref. 33). The repetitive elements *412*, *copia* and *297* of *Drosophila* are physically similar to Ty-1 (and to bacterial transposable elements) and are transposable[34–37]. Strobel *et al.* suggest 'it is possible that the sole function of these elements is to promote genetic variability, and that their gene products may only be necessary for the maintenance and mobility of the elements themselves, rather than for other cellular processes' (ref. 37). But if maintenance and mobility mechanisms exist, then no cellular function at all need be postulated.

A large fraction of many eukaryotic genomes consists of middle-repetitive DNAs, and the variety and patterns of their interspersion with unique sequence DNA make no particular phylogenetic or phenotypically functional sense. Britten, Davidson and collaborators have elaborated models which ascribe regulatory functions to middle-repetitive DNAs, and evolutionary advantage (in terms of adaptability) to the quantitative and qualitative changes in middle-repetitive DNA content observed even between closely related species[17,38–40]. Middle-repetitive DNAs are more conserved in sequence during evolution than are unique-sequence DNAs not coding for protein, and Klein *et al.* suggest that 'restraint on repetitive sequence divergence, either within the repeat families of a given species, or over evolutionary time spanning the emergence of different species, could be due to [phenotypic] selective pressures which prevent free sequence change in a large fraction of the repeat family members. Or perhaps repetitive sequences diverge as rapidly as do other sequences, but the type sequence of the family is preserved by frequent remultiplication of the "correct" surviving sequences' (ref. 41). The evidence for a phenotypically functional role for middle-repetitive sequences remains dishearteningly weak[40–43], and if the calculations of Kimura[44] and Salser and Isaacson[45] are correct, middle-repetitive DNAs together comprise too large a fraction of most eukaryotic genomes to be kept accurate by darwinian selection operating on organismal phenotype. The most plausible form of "remultiplication of the 'correct' surviving sequences" is transposition. If we assume middle-repetitive DNAs in general to be transposable elements or degenerate (and no longer transposable and ultimately to be eliminated) descendants of such elements, then the observed spectra of sequence divergence within families and changes in middle-repetitive DNA family sequence and abundance can all be explained as the result of non-phenotypic selection within genomes. No cellular function at all is required to explain either the behaviour or the persistence of middle-repetitive sequences as a class.

The rest of the eukaryotic genome

Middle-repetitive DNA can comprise more than 30% of the genome of a eukaryotic cell[46]. Another 1–40% consists of simple reiterated sequences whose functions remain unclear[10], and Smith has argued that 'a pattern of tandem repeats is the natural state of DNA whose sequence is not maintained by selection' (ref. 47). Even unique-sequence eukaryotic DNA consists in large part of elements which do not seem to be constrained by phenotypic selection pressures[45]. Some authors have argued that the intervening sequences which interrupt many eukaryotic structural genes are insertion sequence-like elements[6,48,49]. If they are, they are likely to be the degenerate and no-longer-transposable descendants of transposable sequences whose insertion was rendered non-lethal by pre-existing cellular RNA : RNA splicing mechanisms. Such elements, once inserted, are relatively immune to deletion (since only very precise deletion can be non-lethal), and need retain only those sequence components required for RNA splicing. The rest of the element is free to drift and one expects (and observes) that only the position and number of intervening sequences in a family of homologous genes remain constant during evolution. Although evolutionary and regulatory phenotypic functions have been ascribed to intervening sequences[6,49–51], it is unnecessary to postulate any cellular function at all if these elements are indeed degenerate transposable elements arising initially from non-phenotypic selection. Another explanation for the origin and continued existence of intervening sequences, which also does not require phenotypically or evolutionarily advantageous roles, has been suggesed elsewhere[50,51].

Why do prokaryotes and eukaryotes differ?

It is generally believed that prokaryotic genomes consist almost entirely of unique-sequence DNA maintained by phenotypic selection, whereas the possession of 'excess' unique and repetitive DNA sequences whose presence is at least difficult to rationalize in phenotypic terms is characteristic of eukaryotes. However, it is more accurate to say that there is a continuum of excess DNA contents; at least 1% of the *E. coli* genome can be made up of copies of six identified insertion sequences alone[16]. Yeast, whose genome is no larger than that of some prokaryotes, has few repeated sequences other than those coding for stable RNAs, and *Aspergillus* may have none[52,53]. There is in general (but with many exceptions) a positive correlation between excess DNA content, genome size and what we anthropocentrically perceive as 'evolutionary advancement'. Many interpret this as the cause and/or consequence of the increasing phenotypic complexity which characterizes organismal evolution, and attribute to excess DNA a positive role in the evolutionary process [17–19,40]. The interplay of phenotypic and non-phenotypic forces, and the importance of understanding both in attempts to restore the '*C*-value paradox' are discussed more thoroughly by Orgel and Crick in the following article.[54]

There is another, simpler and perhaps obvious explanation. Non-phenotypic selection produces excess DNA, and excess DNA logically must be an energetic burden; phenotypic selection should favour its elimination[24,25]. The amount of excess (and hence total) DNA in an organism should be loosely determined by the relative intensities of the two opposing sorts of selection. The intensity of non-phenotypic pressure on DNA to survive even without function should be independent of organismal physiology. The intensity of phenotypic selection pressure to eliminate excess DNA is not, this being greatest in organisms for which DNA replication comprises the greatest fraction of total energy expenditure. Prokaryotes in general are smaller and replicate themselves and their DNA more often than eukaryotes (especially complex multicellular eukaryotes). Phenotypic selection pressure for small 'streamlined' prokaryotic genomes with little excess DNA may be very strong.

Necessary and unnecessary explanations

We do not deny that prokaryotic transposable elements or repetitive and unique-sequence DNAs not coding for protein in eukaryotes may have roles of immediate phenotypic benefit to the organism. Nor do we deny roles for these elements in the evolutionary process. We do question the almost automatic invocation of such roles for DNAs whose function is not obvious, when another and perhaps simpler explanation for their origin and maintenance is possible. It is inevitable that natural selection of the special sort we call non-phenotypic will favour the development within genomes of DNAs whose only 'function' is survival within genomes. When a given DNA, or class of DNAs, of unproven phenotypic function can be shown to have evolved a strategy (such as transposition) which ensures its genomic survival, then no other explanation for its existence is necessary. The search for other explanations may prove, if not intellectually sterile, ultimately futile.

We thank L. Bonen, R. M. MacKay and M. Schnare for help in development of the ideas presented here, and C. W. Helleiner, M. W. Gray, C. Stuttard, R. Singer, S. D. Wainwright and E. Butz for critical discussions.

We are especially grateful to C.E. Orgel and F.H.C. Crick for discussing with us the ideas presented in the following article before publication and for encouragement and support.

1. Dawkins, R. *The Selfish Gene* (Oxford University Press, 1976).
2. Crow, J. F. *Scient. Am.* **240,** 134–146 (1979).
3. Cox, E. C. & Gibson, T. C. *Genetics* **77,** 169–184 (1974).
4. Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, 1978).
5. Darnell, J. E. *Prog. Nucleic Acid Res. molec. Biol.* **22,** 327–353 (1978).
6. Crick, F. H. C. *Science* **204,** 264–271 (1979).
7. Murray, V. & Holliday, R. *FEBS Lett.* **106,** 5–7 (1979).
8. Sogin, M. L., Pace, B. & Pace, N. R. *J. biol. Chem.* **252,** 1350–1357 (1977).
9. Fedoroff, N. *Cell* **16,** 697–710 (1979).
10. John, B. & Miklos, G. L. G. *Int. Rev. Cytol.* **58,** 1–114 (1979).
11. Zuckerkandl, E. *J. molec. Evol.* **9,** 73–122 (1976).
12. Cohen, S. N. *Nature* **263,** 731–738 (1976).
13. Starlinger, P. & Saedler, H. *Curr. Topics Microbiol. Immun.* **75,** 111–152 (1976).
14. Nevers, P. & Saedler, H. *Nature* **268,** 109–115 (1977).
15. Kleckner, N. *Cell* **11,** 11–23 (1977).
16. Kopecko, D. in *Plasmids and Transposons: Environmental Effects and Maintenance Mechanisms* (eds Stuttard, C. & Rozee, K. R.) (Academic, New York, in the press).
17. Britten, R. J. & Davidson, E. H. *Q. Rev. Biol.* **46,** 111–138 (1971).
18. Ohno, S. *Evolution by Gene Duplication* (Springer, New York, 1970).
19. Hinegardner, R. in *Molecular Evolution* (ed. Ayala, F. J.) 160–199 (Sinauer, Sunderland, 1976).
20. Gould, S. J. & Lewontin, R. C. *Proc. R. Soc.* **B205,** 581–598 (1979).
21. Orgel, L. *Proc. R. Soc.* **B205,** 435–442 (1979).
22. Bodmer, W. in *Human Genetics: Possibilities and Realities,* 41–42 (Excerpta Medica, Amsterdam, 1979).
23. Johnsrod, L. *Molec. gen. Genet.* **169,** 213–218 (1979).
24. Zamenhof, S. & Eichorn, H. H. *Nature* **216,** 456–458 (1967).
25. Koch, A. L. *Genetics* **72,** 297–316 (1972).
26. Shapiro, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76,** 1933–1937 (1979).
27. Arthur, A. & Sherratt, D. *Molec. gen. Genet.* **175,** 267–274 (1979).
28. Gill, R., Heffron, F., Dougan, G. & Falkow, S. *J. Bact.* **136,** 742–756 (1978).
29. MacHattie, L. A. & Shapiro, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76,** 1490–1494 (1979).
30. Huberman, P., Klaer, R., Kühn, S. & Starlinger, P. *Molec. gen. Genet.* **175,** 369–373 (1979).
31. Campbell, A. in *Biological Regulation and Development* Vol. 1 (ed. Goldberger, R. F.) 19–55 (Plenum, New York, 1979).
32. McClintock, B. *Cold Spring Harb. Symp. quant. Biol.* **16,** 13–47 (1952).
33. Cameron, J. R., Loh, E. Y. & Davis, R. W. *Cell* **16,** 739–751 (1979).
34. Finnegan, D. J., Rubin, G. M., Young, M. W. & Hogness, D. S. *Cold Spring Harb. Symp. quant. Biol.* **42,** 1053–1063 (1977).
35. Carlson, M. & Brutlag, D. *Cell* **15,** 733–742 (1978).
36. Potter, S. S., Borein, W. J. Jr, Dunsmuir, P. & Rubin, G. M. *Cell* **17,** 415–427 (1979).
37. Strobel, E., Dunsmuir, P. & Rubin, G. M. *Cell* **17,** 429–439 (1979).
38. Britten, R. J. & Davidson, E. H. *Science* **165,** 349–357 (1969).
39. Davidson, E. H., Klein, W. H. & Britten, R. J. *Devl Biol.* **55,** 69–84 (1977).
40. Davidson, E. H. & Britten, R. J. *Science* **204,** 1052–1059 (1979).
41. Klein, W. H. *et al. Cell* **14,** 889–900 (1978).
42. Sheller, R. H., Constantini, F. D., Dozlowski, M. R., Britten, R. J. & Davidson, E. H. *Cell* **15,** 189–203 (1978).
43. Kuroiwa, A. & Natori, S. *Nucleic Acids Res.* **7,** 751–754 (1979).
44. Kimura, M. *Nature* **217,** 624–626 (1968).
45. Salser, W. & Isaacson, J. S. *Prog. Nucleic Acids Res. molec. Biol.* **19,** 205–220 (1976).
46. Lewin, B. *Cell* **4,** 77–93 (1975).
47. Smith, G. P. *Science* **191,** 528–534 (1976).
48. Tsujimoto, Y. G. & Suzuki, Y. *Cell* **18,** 591–600 (1979).
49. Gilbert, W. *Nature* **271,** 501 (1978).
50. Doolittle, W. F. *Nature* **272,** 581–582 (1978).
51. Darnell, J. E. *Science* **202,** 1257–1260 (1978).
52. Roberts, T. M., Lauer, G. D. & Klotz, L. *CRC Crit. Rev. Biochem.* **3,** 349–451 (1976).
53. Timberlake, W. F. *Science* **702,** 973–974 (1978).
54. Orgel, G.E. & Crick, F.H.C. *Nature* **284,** 604-607 (1980).