NEWS AND COMMENTARY

Genome evolution

Let's stick together

M Nei

Heredity (2003) 90, 411–412. doi:10.1038/sj.hdy.6800287

We know that genes with related functions tend to cluster together in genomes but the cause of this pattern is less clear. Now, Csaba Pál and Laurence Hurst, in a study published in *Nature Genetics*, show that in yeast such genes not only tend to occur together but also tend to be inherited together, indicating that natural selection may favour functionally related genes that stick together.

Functionally interacting genes are often closely located on a chromosome. A good example is the vertebrate major histocompatibility complex (MHC), comprising 20–100 functional genes. These genes are located in the same chromosome region and are tightly linked in most classes of vertebrates (Kulski *et al*, 2002).

The MHC is not an isolated example. Overall, genes with related functions tend to cluster together in baker's yeast (Cohen *et al*, 2000), *Caenorhabditis elegans* (Blumenthal *et al*, 2002) and humans (Lercher *et al*, 2002). There are two possible explanations for these clusters of functionally related genes. First, they merely reflect the origin of functionally related genes via tandem gene duplication. Second, natural selection generates and maintains linked gene blocks.

To distinguish between these possibilities, Pal and Hurst (2003) studied the relationship between clusters of genes essential for viability and the level of recombination within clusters in the genome of baker's yeast. They divided the genome into a large number of contiguous nonoverlapping blocks of 10 genes and examined the relationships between the number of essential genes per block and the recombination value as measured by meiotic double-strand breaks. The results were surprising; whenever the number of essential genes per block was high, the recombination value in the region was low, and this negative correlation was very strong.

How can we explain such a strong negative correlation? The authors argue that if the cluster of essential genes is a consequence of tandem gene duplication there should not be a negative correlation and therefore the most likely

explanation is natural selection. They then refer to my 1967 paper (Nei, 1967) for the explanation of how natural selection could enhance the linkage intensity between functionally interacting genes.

By around 1965, there was a substantial amount of data indicating that the recombination value between different loci is under genetic control, but it was not clear under what circumstances natural selection would increase or decrease the recombination value. Subsequently, my work indicated that in the presence of gene interaction (epistasis), the recombination value between loci should almost always be reduced by natural selection or the linkage of interacting genes is maintained (Figure 1). Essentially, my finding was that if particular alleles at

different loci work well together then selection should favour them staying together.

This finding prompted me to examine whether the recombination frequency has decreased in the course of long-term evolution. This question was based on my conjecture that higher organisms (eg, mammals) should have more interacting genes to form complex characters than lower organisms (eg, bacteria). I studied this problem by computing the recombination value per unit length of DNA for various organisms. Although the results obtained were crude, it was clear that the overall recombination frequency decreased from lower organisms to higher organisms (Nei, 1968).

This finding first appeared to support my mathematical prediction, but later when the genomes of higher organisms were found to contain a large amount of 'junk DNA', I started to wonder whether the apparent decrease in recombination frequency in these organisms could simply reflect the presence of this junk. The fact that the selection intensity for reducing recombination value is generally small, except in the Y chromosome (Nei, 1969), also bothered me. Therefore, when I saw Pal and Hurst's finding, it was a revelation. However, does Pal and Hurst's finding

	Haplotype	General fitness	Case 1	Case 2
			No epistasis	Epistatic selection
Original		W_1	1	1
Recombinant		W_2	0.8	0.8
Recombinant		W_3	0.8	0.8
Original		W_4	0.6	1

Figure 1 Diagram showing the importance of epistatic selection for reducing recombination. Loci A and B are linked and form four different haplotypes with alleles A_1 , A_2 , B_1 and B_2 . Original haplotypes are assumed to have produced recombinant haplotypes. W_1 , W_2 , W_3 , and W_4 are the relative fitnesses of the four haplotypes. Gene interaction or epistatic parameter is defined by $E = W_1 - W_2 - W_3 + W_4$. If $E = 0$, there is no epistasis. In case 1, $E = 0$, so there is no advantage of reducing recombination. In case 2, $E = 0.4$, and recombinant haplotypes are less fit than original haplotypes. Therefore, reduction in recombination is advantageous. In this haploid model, all haplotypes mate at random after selection and maturity, and offspring haplotypes are produced through meiosis and recombination. The recombination frequency is controlled by another recombination-modifying locus, which is polymorphic (Nei, 1967).

really prove that epistatic selection has reduced the recombination values between interacting loci and that some sets of linked genes on a chromosome are organized so as to enhance or maintain gene interaction?

My answer to this question is a conditional 'Yes'. Actually, there is a missing link in Pal and Hurst's argument. As mentioned before, what the authors found is the negative correlation between the proportion of essential genes and the recombination value. Essential genes are certainly expected to be subject to strong natural selection, but this is not sufficient for reducing recombination values. For recombination values to be reduced, there must be a sufficient amount of epistatic selection.

For simplicity, let us consider two linked loci each with two alleles, A_1 and A_2 , and B_1 and B_2 , respectively, in a randomly mating haploid population. We then have four different haplotypes. Let us designate the relative fitnesses of the four haplotypes by W_1 , W_2 , W_3 , and W_4 , as shown in Figure 1. In population genetics, the extent of epistatic selection is measured by the quantity $E = W_1 - W_2 - W_3 + W_4$. The selection intensity for reducing recombination values is a function of the absolute value of E . Therefore, what is important is the magnitude of E rather than the relative fitnesses (Nei, 1967, 1969). Linkage

disequilibrium (LD) is also important, but since LD is generated by epistatic selection, E is the most fundamental quantity. If the essential genes used in Pal and Hurst's study show E values close to 0, the recombination values would not decrease very much. Therefore, to demonstrate epistasis favouring clustering at specific loci, it is necessary to measure not only the fitnesses of mutant haplotypes but also the quantity E .

Nevertheless, one would expect that the absolute value of E is generally greater for essential genes than for weakly selected genes, because the E value is likely to deviate from 0 more often for strongly selected genes than for weakly selected genes. Furthermore, as mentioned earlier, functionally interacting genes tend to be clustered on a chromosome. Therefore, the most parsimonious explanation of the negative correlation observed appears to be epistatic selection.

What are the implications of Pal and Hurst's finding? If their finding generally holds, we would expect that the gene order of functionally related genes is maintained by epistatic selection and that there should be many such gene clusters in the genome. In other words, there is a higher order of gene organization in the genome than a mere collection of different genes or tandemly duplicated genes. In this respect, it is

interesting to note that the gene order of member genes of the Hox gene family controlling the segmentation of animal body is virtually the same for most vertebrate species (Stellwag, 1999): natural selection has apparently maintained this gene order for a very long time. However, the gene orders of some multigene families (eg, MHC and histone gene families) are not necessarily conserved. To prove Pal and Hurst's hypothesis, it would be necessary to examine long-term conservation of various gene clusters in the genome. It is also important to study the relationships between gene clusters and recombination values in many other groups of organisms.

M Nei is at the Department of Biology, 328 Mueller Lab, Pennsylvania State University, University Park, PA, 16802, USA.

e-mail: nxm2@psu.edu

Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J *et al* (2002). *Nature* **417**: 851–854.

Cohen BA, Mitra RD, Hughes JD, Church GM (2000). *Nat Genet* **26**: 183–186.

Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002). *Immunol Rev* **190**: 95–122.

Lercher MJ, Urrutia AO, Hurst LD (2002). *Nat Genet* **31**: 180–183.

Nei M (1967). *Genetics* **57**: 625–641.

Nei M (1968). *Nature* **218**: 1160–1161.

Nei M (1969). *Genetics* **63**: 681–699.

Pal C, Hurst LD

(In press) *Nat Genet* (2003).

Stellwag EJ (1999). *Cell Dev Biol* **10**: 531–540.