

Part II. Selection at silent sites in introns and exons

Regional variation in rates of evolution (Part I) occurs at the scale of megabases, and so cannot be explained by differences between autosomes. If silent sites evolve neutrally, this means that I effectively show that the mutation rate is heterogeneous across autosomes, and hence between genes. But do silent rates of evolution necessarily reflect the mutation rate? Chapter 3 is an indirect test for neutrality, whereby one compares two groups of silent sites and asks whether they differ.

Previously, several groups have found that pseudogenes evolve faster than synonymous sites in their functional counterparts, which suggests the action of purifying selection at synonymous sites (Miyata & Hayashida 1981; Bustamante, Nielsen & Hartl 2002). Unfortunately, however, this suffers from several confounding factors that render interpretation difficult. First, only transcribed genes will experience biases associated with transcriptional-coupled mutation and repair (Green et al. 2003; Majewski 2003). Second, the two groups of sequence may be present in different isochores. Differences in GC content will, for example, affect the relative abundance of hypermutable CpG dinucleotides. Third, as I described in Part I, evolutionary rates vary across the genome.

Consequently, it is desirable to carry out pairwise tests within the same gene. Iida and Akashi (2000), for example, compared constitutively and alternatively expressed exons. They hypothesised that, because constitutive exons are translated more frequently than alternative exons, a difference in nucleotide content, reflecting the use of optimal codons, would indicate selection. Indeed, they found in mammals that both GC₃ (GC content at the mostly synonymous codon third sites) and the rate of synonymous evolution are higher in exons that are expressed constitutively.

In Chapter 3, I compare patterns of evolution at two classes of silent sites within murid genes, in introns and at four-fold degenerate (synonymous) sites. Before making the comparison, I first ask whether there is putatively neutral sequence that we expect *a priori* to be conserved. Hence the chapter is divided into two parts.

Although there are plentiful descriptions in the molecular biology literature of transcriptional control elements within first introns, only a couple have compared their activity relative to the other introns within a gene (Palmiter et al. 1991; Jonsson et al. 1992). As these anecdotes might reflect a reporting bias, I asked whether it was generally true that first introns contain disproportionately more control elements and evolve more slowly as a result. While I found no difference between the densities of transcription factor binding sites, first introns do possess a higher density of CpG

islands. Although several authors have shown that first introns evolve slowly (Majewski & Ott 2002; Keightley & Gaffney 2003), none have systematically shown that this is because they contain more control elements. I show that substitution rates are lower in first introns and that this is consistent with purifying selection to preserve the activity of control elements. Additionally, I find that substitutions are increasingly less frequent as one approaches the intron-exon junction, which I assume reflects selection for splice site recognition.

In the second part of Chapter 3, after eliminating selectively constrained intronic sites, I compare the rates of substitution between introns and four-fold sites. Consistent with the predictions of the neutral theory, these two classes of putatively neutral sites evolve at equivalent rates. Although at first sight this does not deviate from the null expectation, neutrality also predicts that the two classes of sites not only evolve at the same rate, but also in the same manner. Upon closer inspection, however, this is not observed. By examining the patterns of substitution at the first site within dinucleotides, it can be seen that As and Ts are rarely conserved, but C is particularly stable given its relative abundance. Similarly, while A and T content are both higher in introns than at four-fold sites, G content is the same while C content is higher at four-fold sites. Overall, if one assumes that the majority of the remaining intronic sequence is less likely to be under selective constraint, these results suggests that the C preference at four-fold sites provides evidence for selection at synonymous sites.

References

- Bustamante, C. D., Nielsen, R. & Hartl, D. L. (2002) A maximum likelihood method for analyzing pseudogene evolution: Implications for silent site evolution in humans and rodents. *Mol. Biol. Evol.* **19**: 110-117.
- Green, P., Ewing, B., Miller, W., Thomas, P. J., Nc, N. & Green, E. D. (2003) Transcription-associated mutational asymmetry in mammalian evolution. *Nat. Genet.* **33**: 514-517.
- Iida, K., & Akashi, H. (2000) A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes. *Gene* **261**: 93-105.
- Jonsson, J. J., Foresman, M. D., Wilson, N. & McIvor, R. S. (1992) Intron requirement for expression of the human purine nucleoside phosphorylase gene. *Nucleic Acids Res.* **20**: 3191-3198.
- Keightley, P. D., & Gaffney, D. J. (2003) Functional constraints and frequency of deleterious mutations in noncoding DNA of rodents. *Proc. Natl Acad. Sci. USA* **100**: 13402-13406.

- Majewski, J., & Ott, J. (2002) Distribution and characterization of regulatory elements in the human genome. *Genome Res.* **12**: 1827-1836.
- Majewski, J. (2003) Dependence of mutational asymmetry on gene-expression levels in the human genome. *Am. J. Hum. Genet.* **73**: 688-692.
- Miyata, T., & Hayashida, H. (1981) Extraordinarily high evolutionary rate of pseudogenes: evidence for the presence of selective pressure against changes between synonymous codons. *Proc. Natl Acad. Sci. USA* **78**: 5739-5743.
- Palmiter, R. D., Sandgren, E. P., Avarbock, M. R., Allen, D. D. & Brinster, R. L. (1991) Heterologous Introns Can Enhance Expression of Transgenes in Mice. *Proc. Natl Acad. Sci. USA* **88**: 478-482.