

The place and function of non-coding DNA in the evolution of variability

Vishnu Dileep

The function of intergenic and intragenic non-coding DNA, also called 'junk DNA' is a highly debated topic in evolutionary biology. We find an extensive amount of non-coding DNA in eukaryotes. In contrast, the prokaryotic genome has no introns and very few non-coding areas. Researchers have attributed various functions to some parts of the non-coding DNA, but a large part of it has no known function. This hypothesis proposes that non-coding DNA is involved in regulating the amount of random variability in the eukaryotic genome and increases the chance of intact gene transfer during chromosomal crossing over. This article identifies a pattern in the evolution of variability and discusses the hypothesized function of non-coding DNA as a part of this pattern. The known functions of non-coding DNA are also mentioned briefly in this context.

Citation: Dileep V. The place and function of non-coding DNA in the evolution of variability. *Hypothesis* 2009, 7(1): e7.

Introduction

THE JUNK DNA PARADOX is the most puzzling question in evolutionary biology. It was found that 97% of the human genome has no apparent function. Similar observations were made in other eukaryotes. But the prokaryotes have very few non-coding regions, most of which are thought to have regulatory functions (1). In this article I attempt to bring to light a pattern which can be observed in the evolution of variability throughout the history of life, and explain how the presence of non-coding DNA could be a part of this pattern.

Recognizing the pattern

A quick look at the mechanisms that create

variability in life forms exposes a pattern that is being followed from viruses to complex eukaryotes. Viruses have mechanisms that create huge amounts of random variability in them. The high variability observed in RNA viruses is mainly due to superinfections, lack of proofreading by reverse transcriptase, host genome carryover during infection, and reverse transcriptase template switching (2, 3). The DNA viruses also have high variability rates but these do not involve lack of proofreading and reverse transcriptase template switching. However, the very high multiplication rates of viruses offset the deleterious effects of this huge variability. Even if 99% of viruses produced are defective, the successful

Department of Biotechnology, Chemical and Biomedical Engineering, VIT University, Vellore-632 014, Tamil Nadu, India.
Correspondence: vishnudileep2000@gmail.com
Received: 2009/04/23, Accepted: 2009/09/01
Posted online: 2009/09/26

© 2009 Vishnu Dileep. This is an Open Access article distributed by Hypothesis under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1% forms a significant number to colonize its habitat, thus making them extremely efficient.

In the next hierarchy of life forms we see more efficient forms of variability creators. The prokaryotes (bacteria) are more complex and have lower replication rates compared to the viruses. The mechanisms that create variability in prokaryotes are conjugation, transformation and transduction. These

As complexity of organisms increases and replication rates decrease, the randomness and deleteriousness of the variability is reduced.

phenomena involve transfer of larger segments of DNA and thus increase chances of functional gene transfer. Moreover, prokaryotes have additional mechanisms that reduce highly random changes in the genetic material such as errors by DNA polymerase and mutations; these are the DNA mismatch repair system and DNA polymerase proofreading.

But with the evolution of many complex eukaryotes with low replication rates, random variability presents a hazard rather than an advantage. A lethal mutation in any of the hundreds of specialized cells in a eukaryote could severely disable or kill an organism. Thus, eukaryotes need variability creators that have more chances of creating functional gene combinations to effectively overcome selection pressures. Sexual reproduction involves one such very refined form of variability creation. It jumbles the chromosomes, creating millions of different gametes, thus

ensuring variability without disrupting any genes. For instance, humans have 23 chromosomes, so there are mathematically 2^{23} possible combinations of gametes from the father and equal number of possible gametes in the mother. That means a staggering 2^{46} possible combinations of offspring. Eukaryotes also evolved another refined mechanism to create variability, chromosomal crossing over. Apart from these major phenomena that cause variability, eukaryotes also rely on other events like segmental duplication (4-7) and gene conversion to create variability (8).

A distinct pattern can be observed in the mechanisms that create variability (Table 1). As complexity of organisms increases and replication rates decrease, the randomness and deleteriousness of the variability is reduced. We see that the organism achieves this by evolving mechanisms that affect larger segments of DNA and also by mechanisms that suppress high randomness. Non-coding DNA fits into this pattern as a way to increase intact gene transfer during crossing over. This hypothesis is further developed in the following explanation.

Demystifying junk DNA

Some functions have been attributed to non-coding DNA. Most of the non-coding regions flanking a gene have regulatory functions such as regulation of transcription (9-11). Many sequences previously thought to be non-coding regions are now found to code for regula-

A huge part of the eukaryotic genome is repeated sequences, for most of which the function is not clear.

Dileep

tory RNAs such as microRNAs and long intergenic non-coding RNAs (lincRNAs) (12-14). Researchers have found non-translated RNAs that are involved in silencing such as X-chromosome inactivation (15). But a huge part of the eukaryotic genome is repeated sequences, for most of which the function is not clear.

The best way to study the function of an object is to remove it from the system and study the effects. Let's theoretically remove the entire non-coding DNA from human chromosome 1 and speculate what may happen. Imagine chromosome 1 in a germ cell that is about to undergo meiotic division. Now this hypothetical chromosome 1 consists of wall-to-wall coding region; it doesn't have any non-functional DNA. Thus, any crossing over happening in such a chromosome can disrupt a gene, especially if it is an unequal crossover. Such variation involving disruption of a gene has less probability of resulting in a functional gene. But in reality, the intergenic DNA in the chromosome may act as a buffer. Since most of the non-coding region is repetitive, there is a high tendency for unequal crossover to fall in this

region. This keeps the genes intact yet transfers them to the non-sister chromatid, creating new gene arrangements which have more probability of being functional. Human recombination studies have proven that crossing over occurs preferentially outside the genes (16). In accordance, researchers have found that recombination hotspots tend to cluster in the non-coding regions of DNA (17-20).

Now we can analyze the present hypothesis in more detail. We have seen how intergenic DNA can assist shuffling of intact genes. Now we have to analyze the function of intragenic non-coding DNA or introns. The functional units of a gene are its exons that code for the protein. So shuffling the exons between genes without disrupting them can lead to new genes with altered functions (21) (Figure 1). Since the gene is largely composed of non-coding introns, there is a high chance that the crossing over will fall inside the intron. This fits the pattern that we observed, since complex organisms like eukaryotes tend to conserve the functional region by reducing randomness of variability.

Hypothesis Vol. 7, No. 1 | September 2009 | hypothesisjournal.com

Class of Organism	Variability Creators	Suppressors of Deleterious Random Variability
Viruses	1) Superinfection 2) No proofreading for reverse transcriptase (for retroviruses) 3) Template switching by reverse transcriptase (for retroviruses) 4) Host genome carryover after integration	None
Prokaryotes	1) Transduction 2) Transformation 3) Conjugation	1) Mismatch repair system 2) DNA polymerase proofreading
Eukaryotes	1) Chromosome mixing by sex 2) Crossing over 3) Segmental duplication 4) Gene conversion	1) Mismatch repair system 2) DNA polymerase proofreading 3) Non-coding DNA

Table 1 | The evolutionary pattern of variability. The significant variability creators and variability suppressors listed for each class of organisms. Non-coding DNA is indicated in boldface.

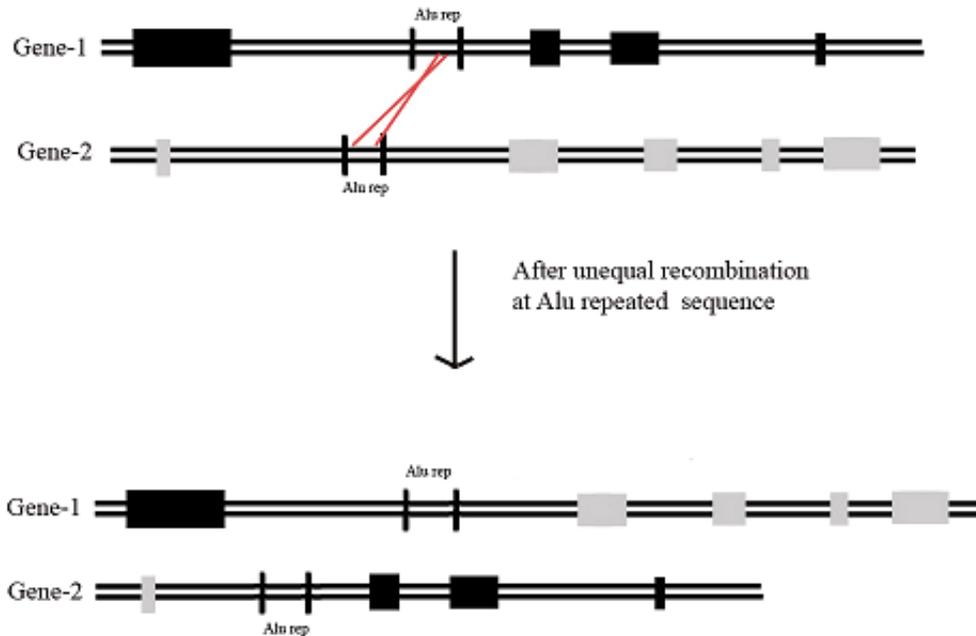


Figure 1 | Exon shuffling by crossing over. Exons can get exchanged between the non-sister chromatids when the chiasmata form inside an intron. Here the region of homology is Alu repeats. This results in the formation of hybrid genes without disrupting any exons. Alu is a class of repetitive DNA which is approximately 300bp long. They are classified as SINES (short interspersed elements) and are found extensively in the eukaryotic genome.

ity. Further, introns contain repeat sequences that can facilitate unequal crossing over.

This raises the question about the mechanism of how such long stretches of non-coding DNA originated in the eukaryotic organism. A large majority of it is attributed to transposable elements; the rest is thought to be sequences left behind following viral infection during the history of the organism's evolution. But quite intriguingly, we don't find such extensive DNA accumulation in prokaryotes, even though transposons and bacteriophage infections are found in prokaryotes. In contrast, it seems that the non-coding sequences are actively being minimized in prokaryotes (1). The hypothesis states that this might be because they don't have crossing over due to their asexual multiplication and thus do not require the 'buffer DNA'. The prokaryotes

only preserve small regions of non-coding intergenic DNA and have no introns. It has been found that these non-coding intergenic DNA sequences have regulatory function.

Therefore, non-coding DNA appears to be a requirement for non-deleterious crossing over. The non-coding DNA might have been introduced into eukaryotes along with the evolution of sexual reproduction. This theory supports the 'introns late' hypothesis, which is one of two schools of thought about evolution of introns. It states that introns were inserted into the eukaryotes in the later part of their evolution (22).

Conclusion

Observation points to a pattern in the evolution of variability in organisms. As complexity of the organism increases, the processes that

Dileep

create the variability tend to conserve the coding regions and focus on creating variability by shuffling the functional regions. In other words, the randomness of variability decreases as complexity of the organism increases (Figure 2).

The non-coding DNA fits into this pattern since it prevents disruption of genes and exons by crossing over and facilitates their shuffling. This means that organisms that use meiotic division to produce gametes require non-coding DNA to offset the deleterious effects of crossing over. The fact that prokaryotes (which reproduce asexually) do not have extensive non-coding DNA supports this hypothesis. Thus, the hypothesis proposes that while non-coding DNA has many known functions, it may also be a necessary requisite for crossing over. Non-coding DNA might have been introduced into the genome when sexual reproduction evolved. This idea therefore supports the 'introns late' hypothesis.

References

- (1) Rogozin IB, Makarova KS, Natale DA, Spiridonov AN, Tatusov RL, Wolf YI, *et al.* Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. *Nucleic Acids Res*, 2002; 30: p. 4264-4271.
- (2) Artzi HB, Shemesh J, Zeelon E, Amit B, Kleiman L, Gorecki M, *et al.* Molecular analysis of the second template switch during reverse transcription of the HIV RNA template. *Biochemistry*, 1996; 35: p. 10549-10557.
- (3) Anderson JA, Teufel RJ, Yin PD, Hu W-S. Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. *J Virol*, 1998; 72: p. 1186-1194.
- (4) Cannon SB, Mitra A, Baumgarten A, Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biology*, 2004; 4: p. 10.
- (5) Zhang L, Lu HH, Chung WY, Yang J, Li WH. Patterns

Hypothesis Vol. 7, No. 1 | September 2009 | hypothesisjournal.com

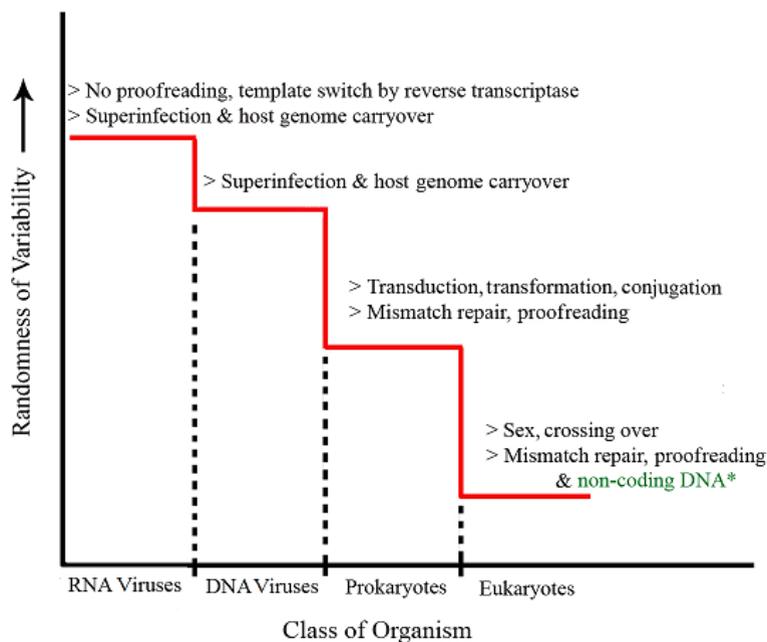


Figure 2 | Evolution of variability and position of non-coding DNA. A graphical representation of the pattern observed in the evolution of variability and non-coding DNA's position in it.

- of segmental duplication in the human genome. *Mol Biol Evol*, 2005; 22: p.135–141.
- (6) Leister D. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. *Trends Genet*, 2004; 20: p.116-122.
- (7) Fiston-Lavier AS, Anxolabehere D, Quesneville H. A model of segmental duplication formation in *Drosophila melanogaster*. *Genome Res*, 2007; 17: p.1458-1470.
- (8) Xu S, Clark T, Zheng H, Vang S, Li R, Wong GK, et al. Gene conversion in the rice genome. *BMC Genomics*, 2008; 9: p. 93.
- (9) Nelson CE, Hersh BM, Carroll SB. The regulatory content of intergenic DNA shapes genome architecture. *Genome Biol*, 2004; 5: R25.
- (10) Sabarinadh C, Subramanian S, Tripathi A, Mishra RK. Extreme conservation of noncoding DNA near HoxD complex of vertebrates. *BMC Genomics*, 2004; 5: p. 75.
- (11) Ludwig MZ. Functional evolution of noncoding DNA. *Curr Opin Genet Dev*, 2002; 12: p. 634-639.
- (12) Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*, 2009; 23: p. 1494-1504.
- (13) Pang KC, Dinger ME, Mercer TR, Malquori L, Grimmond SM, Chen W, et al. Genome-wide identification of long noncoding RNAs in CD8+ T cells. *J Immunol*, 2009; 182: p. 7738-7748.
- (14) Griffiths-Jones S. Annotating noncoding RNA genes. *Annu Rev Genomics Hum Genet*, 2007; 8: p. 279-298.
- (15) Ng K, Pullirsch D, Leeb M, Wutz A. Xist and the order of silencing. *EMBO Rep*, 2007; 8: p. 34-39.
- (16) McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. The fine-scale structure of recombination rate variation in the human genome. *Science*, 2004; 304: p. 581-584.
- (17) Cromie GA, Hyppa RW, Cam HP, Farah JA, Grewal SI, Smith GR. A discrete class of intergenic DNA dictates meiotic DNA break hotspots in fission yeast. *PLoS Genet*, 2007; 3: p. e141.
- (18) Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, 2000; 97: p. 11383-11390.
- (19) Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. *Science*, 2005; 310: p. 321-324.
- (20) Webb AJ, Berg IL, Jeffreys A. Sperm cross-over activity in regions of the human genome showing extreme breakdown of marker association. *Proc Natl Acad Sci U S A*, 2008; 30: p. 10471-10476.
- (21) Patthy L. Genome evolution and the evolution of exon-shuffling- a review. *Gene*, 1999; 238: p. 103-114.
- (22) Koonin EV. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate. *Biol Direct*, 2006; 1: p. 22.