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.
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 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 regulatory elements. A huge part of the eukaryotic genome is repeated sequences, for most of which the function is not clear.
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 conserved the functional region by reducing randomness of variabil-

<table>
<thead>
<tr>
<th>Class of Organism</th>
<th>Variability Creators</th>
<th>Suppressors of Deleterious Random Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viruses</td>
<td>1) Superinfection</td>
<td>1) Mismatch repair system</td>
</tr>
<tr>
<td></td>
<td>2) No proofreading for reverse transcriptase (for retroviruses)</td>
<td>2) DNA polymerase proofreading</td>
</tr>
<tr>
<td></td>
<td>3) Template switching by reverse transcriptase (for retroviruses)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4) Host genome carryover after integration</td>
<td></td>
</tr>
<tr>
<td>Prokaryotes</td>
<td>1) Transduction</td>
<td>1) Mismatch repair system</td>
</tr>
<tr>
<td></td>
<td>2) Transformation</td>
<td>2) DNA polymerase proofreading</td>
</tr>
<tr>
<td></td>
<td>3) Conjugation</td>
<td></td>
</tr>
<tr>
<td>Eukaryotes</td>
<td>1) Chromosome mixing by sex</td>
<td>1) Mismatch repair system</td>
</tr>
<tr>
<td></td>
<td>2) Crossing over</td>
<td>2) DNA polymerase proofreading</td>
</tr>
<tr>
<td></td>
<td>3) Segmental duplication</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4) Gene conversion</td>
<td></td>
</tr>
</tbody>
</table>

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.
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...
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
(5) Zhang L, Lu HH, Chung WY, Yang J, Li WH. Patterns


