Junk DNA and Jumping Genes as Computational Analogs in Nature.

Tariq Khan
Omaha, NE USA

Genetic code elements known as junk DNA (non-coding DNA) and jumping genes (transposable elements) are proposed as, like DNA, following a computational framework. This informal essay proposes computer and data scientists in this context revisit these entities versus modern data compression, storage, registry, blockchain, and encryption software methodologies.

> *"The amount of non-coding DNA varies greatly among species. Often, only a small percentage of the genome is responsible for coding proteins, but an increasing percentage is being shown to have regulatory functions. When there is much non-coding DNA, a large proportion appears to have no biological function, as predicted in the 1960s. Since that time, this non-functional portion has controversially been called "junk DNA"."[1]*
> *-- Science*

> *"The system of nature, of which man is a part, tends to be self-balancing, self-adjusting, self-cleansing."*
> *-- E. F. Schumacher, Small is Beautiful[2]*

> *"...I will be free, and the world will be different, because I have understanding... I shall have understanding of computers. And when I have understanding of computers, I shall be the Supreme Being!"*
> *-- Evil, Time Bandits[3]*

In the realm of genetic inheritance, the concepts of both "jumping genes," or transposable elements and "junk DNA," or non-coding DNA, contain characteristics analogous to computational processing strategies. The human genome and the structure and function of DNA already are large and complex data storage and processing mechanisms to build and maintain an organism and the survival of a gene pool. Information processing, replication, signaling, and fidelity are inherent to the very core of life and the genome. In an attempt to extend this paradigm, this short essay proposes examining relatively recently identified, and still somewhat mysterious, areas of the genome in the same context and to suggest further research into the possibility that "jumping genes" and "junk DNA" are actually, in a similar manner, core information processing functions. As Dr. Atma Ivancevic and team at Adelaide University noted in 2018 research that "[t]hese jumping genes are actually small pieces of DNA that can copy themselves throughout a genome and are known as transposable elements" and that "cross-species transfers, even between plants and animals, have occurred frequently throughout evolution." [4]

If we examine a genome of an individual organism, versus its gene pool or the given population of that gene pool that we might consider as a sort of "distributed intelligence," then the most "valuable commodity" to that gene pool could very well be new genetic information it encounters a la "modules" of DNA that the gene pool could utilize or utilize in the future vis a vis a human's tool shed.

Some circumstantial evidence to support this proposal includes behaviors seen in the human immune system. Modern research has shown that it is via exposure and acclimation to genetic material from a male potential mate that the female system sets the stage for a more viable or successful embryo. This implies a sense of "tracking" of genetic material as well as a "memory" of past encounters and obviously a potential "value" in genetic material whether that be as a threat (e.g., virus), benefit (reproduction), or, per this theory, as a potential "tool" to use in the future by the gene pool for survival (i.e., an adaptation). In essence, could Nature or living organisms be using strategies analogous to data archiving, long-term storage, and compression or encryption beyond even known DNA and RNA information processing? The suggestion in this essay is that computer and data scientists revisit junk DNA genetic code versus modern data compression, storage, registry, blockchain, and encryption software methodologies.

Nature uses two great "strategies" to optimize and survive or prosper. They are diversity and fecundity. But, as seen in warfare, economics, biology, and game theory, there are other strategies that can optimize survival including replication/redundancy, modularization, flexibility, partnership, retention/memory, analysis, scalability, etc. Thus, perhaps (similar to the paper-work hoarder that kept every bit of homework and thus had the necessary information from years prior to answer that single rare question and pass the difficult exam), organisms analyze and store ("remember") "useful" modules or snippets of DNA they encounter for potential future use for itself or, more than likely, for the aggregate gene pool. Perhaps the genome stores DNA sections, or at least "potentially useful" sections, like transactions stored in a blockchain?

Now consider a house cat that eats a plate of tuna that accidently had a piece of human hair (DNA) on it. Obviously neither the cat or its gene pool will develop traits of a human or human intelligence or brain size any time soon. However, just perhaps, sections of that "different" DNA that has been encountered could be stored in the cat's DNA but not evident to researchers if it is compressed or encrypted when stored or "archived." Now DNA molecules are not silicon or glucose molecules, and the body's digestive and immune systems appears to know or recognizes this. In fact, from the discussed "systems perspective," snippets or modules of DNA, in terms of survival, are a literal "gold mine." Those modules or DNA snippets may have no value to this particular organism but a million years from now, after additional fitness-driven evolution, they may very well come in handy for the aggregate cat gene pool. To extend the analogy, when a friend looks at your home PC computer's files, she sees many items including apparently useless archive files and compressed zipped folders and even perhaps items stashed in a "junk" Recycle Bin location.

While the analogies noted are extremely speculative and lack demonstrated evidence, the logic makes sense especially given the likelihood of the genome "wasting spacing" with just random information that would have no future potential value at all. To a given gene pool, the most valuable item in the world are bits of useful hard-learned DNA instructions. Why would a gene pool want to get rid of that! Thus, perhaps we are far from junk indeed and perhaps the computational and data processing analogies already seen in DNA coding can be extended further into aspects and behaviors of the genome.

References

[1] Pennisi, E (September 2012). "Gen omics. ENCODE project writes eulogy for junk DNA". Science. 337 (6099): 1159–1161. https://doi.org/10.1126/science.337.6099.1159

[2] Schumacher, E. F., & McKibben, B. (2014). *Small is beautiful: Economics as if people mattered*. Harper Perennial.

[3] Gilliam, T. (Director). (1981). *Time Bandits* [Film]. Handmade Films & Janus Films.

[4] Ivancevic, A.M., Kortschak, R.D., Bertozzi, T. et al. Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol* **19**, 85 (2018). https://doi.org/10.1186/s13059-018-1456-7