Bioinformatic Investigation of the Peculiarities of Long Intron Splicing in *Hominidae*

Saniya J. Gaitonde | W. Tresper Clarke High School | 2020

Computational Biology / Bioinformatics

Research Plan

a. RATIONALE:

Recursive splicing is an RNA maturation phenomenon wherein transcripts are matured during multiple splicing reactions, first discovered in long *Drosophila melanogaster* introns of genes involved in morphogenesis and development. Many basic biological, physiological, and neurological properties are conserved among *D. melanogaster* and mammals, and most human disease-causing genes have a functional homolog in the fly. Human chromosome 2 was formed during an end-to-end ancestral fusion event of two smaller chromosomes present in bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*). Researching the peculiarities of long intron splicing in *Hominidae*, with a focus on recursive splicing, has the potential to provide novel insight into the basis of evolutionary developmental variation among the Great Apes and disorders resulting from implicated RNA splicing.

b. RESEARCH QUESTIONS:

1. What do the recursive sequences of each species look like?
2. How many recursive sites are present in each species?
3. Is the location of these sites conserved across species?

HYPOTHESIS:

Motifs for recursive sites, which mediate multi-step RNA splicing, are enriched and conserved in the introns of genes with key developmental functions, consistent with known evolutionary relationships.

ENGINEERING GOAL:

The overarching goal of this project is to offer novel computational data on the largely unspecified topic of long intron splicing in mammals from the evolutionary perspective, which can eventually be applied to gene disorder research.

EXPECTED OUTCOMES:

It is expected that recursive splicing will be less enriched in the three *Hominidae* species than in *Drosophila melanogaster*, but that the sequences and location recursive sites will

be consistent with known evolutionary relationships, providing additional evidence of the ancestral fusion event which formed human chromosome 2.

c. PROCEDURES:

Recursive splicing can be detected by computationally parsing genomic DNA for recursive sequences. A python program written to scan and score genomic sequences for matches to recursive sites will be used. The script will accept a 26 base pair position-specific scoring matrix (PSSM) file which represents a recursive motif numerically, in addition to genomic samples in FASTA file format from human chromosome 2, chimpanzee chromosomes 2a and 2b, and bonobo chromosomes 2a and 2b, individually. All genome samples were obtained from the publically available UCSC database, affiliated with the Human Genome Project. For each sample, the script will then output a BED file delineating 26 base pair segments of the FASTA within which there is potential for the PSSM's recursive motif to fall, along with the DNA strand from which the gene was transcribed and the score of the region's match to the recursive PSSM. *D. melanogaster* data obtained in a previous investigation will be used as control.

RISK & SAFETY:

There are no risks and no safety precautions necessary for an investigation of this nature.

DATA ANALYSIS:

Using the BED file outputs, the average match score will be calculated for each sample. The BED files will then each be inputted into WebLogo, a software which will materialize the appearances of the recursive sequences in each species to determine whether they look similar. Finally, the BED files will be uploaded to the online LiftOver tool which will convert the genomic coordinates of each species' sequences to a common reference. The converted BED files will then be executed in the pre-programmed python LiftOver script which will return the total number of motif matches in each species and how many overlapped in both directions for any combination of two species, revealing the extent to which the location of recursive sites has been conserved.

d. BIBLIOGRAPHY:

Gallego, A., Melé, M., Balcells, I., García-Ramallo, E., Torruella-Loran, I., Fernández-Bellon, H., … Espinosa-Parrilla, Y. (2016, April 22). Functional Implications of Human-Specific Changes in Great Ape microRNAs. Retrieved January 18, 2020, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0154194

Mooney, C. (2017, June 24). You share 98.7 percent of your DNA with this sex-obsessed ape. Retrieved from

https://www.motherjones.com/politics/2014/02/evolution-creationism-bonobos-neanderthals-denisovans-chromosome-two/

Pai, A. A., Paggi, J. M., Yan, P., Adelman, K., & Burge, C. B. (2018). Numerous recursive sites contribute to accuracy of splicing in long introns in flies. *PLOS Genetics*, *14*(8). doi: 10.1371/journal.pgen.1007588

Pandey, U. B., & Nichols, C. D. (2011). Human Disease Models in Drosophila melanogaster and the Role of the Fly in Therapeutic Drug Discovery. *Pharmacological Reviews*, *63*(2), 411–436. doi: 10.1124/pr.110.003293

Shepard S, McCreary M, Fedorov A (2009) The Peculiarities of Large Intron Splicing in Animals. PLoS ONE 4(11): e7853. https://doi.org/10.1371/journal.pone.0007853

No changes were made to this Research Plan.