Multiple 3D RNA Structure Superposition Using Neighbor Joining

Alexander Kaiser

CSCE-221-501

September 27, 2016

As opposed to double stranded DNA, RNA is a single stranded form of matter that is free to form with other RNA strings within the same vacinity. RNA are compared to one another by their stems, loops and necks. The RNA can be aligned using an algorithmic function called SETTER, which uses generalized secondary structure units in order to match these three properties in order to find two non-overlapping 3D RNA strands that are most alike. The SETTER program is a series of else if statements that starts with comparing the two RNA strands' position of nucleotides, it's neighboring nucleotide type and the strands' raw length. The algorithm returns a value called an *S-score,* which is a measurement of the superposition of two RNA structures. The lower the *S-score* the higher the superposition of two RNA structures. Recently, a program by the name of MultiSETTER has become a more effective alternative to the original SETTER program. The MultiSETTER program uses an algorithm orignially used for the multialignment of protiens called ClustalW. This algorithm creates a dendrogram, or guide tree, for the neighboring RNA molecules to align with one another until the most identical structures are matched, very much like an else if statement used in programming. This new advancement in RNA matching has increased the capability of matching long structures significantly as well as decreased runtime of SETTER by around seventy-five percent.

This advancement in genetic matching impacts society and the economy in a plethora of ways. As for both categories, the decrease in runtime that comes from the multi-core processing power of MultiSETTER eliminates downtime that can be used for other research purposes. The improved matching ability MultiSETTER brings to the field of genetics will also allow for many new discoveries that would not be possible if it not were for the precision that the new program displays. Examples of these discoveries include knowledge of the human species as a whole, increased research in genomic medicines and prevention and research of genetic disorders. Finally, since MultiSET-

TER is a compilation of C++ source code it can be seen that the functionality as well as the uses for C++ are rapidly increasing day after day. The ever increasing list of uses for C++ expands with every new idea that is brought to the table.

References

[1] Hoksza, David, and Daniel Svozil. "Multiple 3D RNA Structure Superposition Using Neighbor Joining." IEEE/ACM Trans. Comput. Biol. and Bioinf. IEEE/ACM Transactions on Computational Biology and Bioinformatics 12.3 (2015): 520-30. Web.