Erin Howell

Computational Biology Final Project: Affine Gap and Alternative Splicing

The data was acquired from UCSC genome browser and stored manually in text files. Each gene was examined by looking at its protein form for four alternative splicing possibilities. The alignments were represented as below, with dashes for gaps.

The Program:

- Affine Gap
  - Made three tables: one for movements down (i.e. gaps in the first string), movements right (gaps in the second string) , and one for diagonal movements (matches and mismatches).
  - The program filled in the tables node by node, also making a backtracking matrix for each movement matrix.
- Computing The BackTrack Path
  - Program looked through the backtracking matrices and recorded a path
- Print Alignment
  - Program printed string1 and string2 based on the path

Table 1 shows some of the data gleaned from ATF3 and PCDH15, which were chosen for analysis base on the amount of alternative splicing possibilities they had. The former was shorter and the latter longer.

Gene	String1 length	String2 length	Percent lack of length	Gaps in 1	Gaps in 2	Total Gaps	Alignment Score
ATF3	181	124	31.49171270	0	1	1	615
	181	106	41.43646408	1	3	4	344
	181	135	25.41436464	1	2	3	531
	124	106	14.51612903	2	3	5	228
	124	135	8.148148148	2	2	4	228
	106	135	21.48148148	1	0	1	517
PCDH15	**1677**	**1681**	**0.237953599**	1	1	**2**	**8629**
	1677	1888	11.17584745	16	1	17	6821
	1677	1915	12.42819843	15	2	17	7028
	1681	1888	10.96398305	15	1	16	6829
	1681	1915	12.21932114	14	3		

Erin Howell

There was a slight tendency for more gaps to indicate lower alignment score, as shown in figures 1 and 2. The tendency was expected since gaps are fairly heavily penalized and more gaps means more chance that the rearrangement would include large chunks of added or missing exons. .

(upper, lower, then middle), cannot be changed because the middle makes use of the upper and lower coordinates to determine a path.

This method is easily generalized to other alternative splice sites or comparisons between DNA and mRNA. One interesting study could be the difference in alternative splicing for homologous genes across species. Another might be a cross-species study of number and type of alternative splicing. This last would be very interesting since splicing is thought to account for the difference in complexity between species (Downes). The method could also be further improved in making judgments as to what sort of rearrangement occurred by comparing DNA and mRNA information to the protein sequence.

Work Cited

Downes, S. (2004). Alternative splicing, the gene concept, and evolution. History & Philosophy of the Life Sciences, 26(1), 91-104. Retrieved December 15, 2015, from http://0-www.jstor.org.catalog.multcolib.org/stable/23333382