

APPENDIX

DNA-QL Operator

We proposed and implement the following (9) operators:

1. Location

This operator is used to find out the location of a given subsequence in DNA structure. It takes DNA strand and Subsequence to be found as input and gives an integer array as output. Algorithm of the operator is given below.

Procedure Location (DNA Strand sequence, subsequence)

1. Get the strand sequence. Extract the barcode sequence
2. Find the subsequence in DNA strand by matching subsequence to be found with DNA strand sequence.
3. The index of starting similar subsequence is stored in integer array
4. For multiple existences get all the indexes of subsequences repeating step 2-3.

Algorithm 1. Algorithm of the operator Location

2. Length Of Strand

This operator counts length of a DNA strand. Since DNA strand length can be infinite. The inputs of the operator are strand name, and give an integer as the length of the strand as output. The algorithm of the operator is given below.

Procedure length Of Strand (DNA Strand sequence)

1. Get the DNA Strand sequence
2. Extract the barcode sequence
3. Count all code pairs of bases present in the DNA strand sequence.

Algorithm 2. Algorithm of the operator lengthOfHelix

3. SubSequence

This operator is used to find out the subsequence at specified position in DNA strand sequence, it takes the start and end position of a strand. Algorithm of the operator is given below.

Procedure SubSequence (DNA strand sequence, int start, int end)

1. Get the DNA Strand Structure. Extract the barcode sequence
2. Get the code pair sequence by starting search from the *start* code till *end*.
3. Return the index where the subsequence code pair resides.

Algorithm 3. Algorithm of procedure SubSequence

4. get DTranscript

This operation will take a base present on DNA anti-sense strand which belongs from the Set b equation 4. Operation will generate complementary base present on DNA anti-sense strand which belong from set b' in section 2.1.1. Algorithm of the operator is given below.

Procedure get DTranscript (B)

1. If (Sense strand Base is "C")
Replace it with "G"
// Complement base of "C" is "G"
2. Else If (Sense strand Base is "G")
Replace it with "C"
// Complement base of "G" is "C"
3. Else If (Sense strand Base is "A")
Replace it with "U"
// Complement base of "A" is "U"
4. If (Sense strand Base is "T")
Replace it with "A"
// Complement base of "T" is "A"

Algorithm 4. Algorithm of procedure Transcription

5. no. Of SubSequence

This operator is used to count the sub sequence in DNA strand sequence. The inputs of the operator are DNA strand sequence and give the total count of occurrences of sub sequence in strand sequence as an output. The algorithm of the operator is given below.

Procedure no Of SubSequence (DNA Strand sequence)

1. Get the DNA Strand Structure. Extract the barcode sequence.
2. Get the code pair sequence by starting search from the *start* code till *end*.
3. The count of sub sequence present in DNA Strand sequence stored in integer array.
4. For multiple existences get all the indexes of sub sequence in DNA strand sequence by repeating step 2-3.

Algorithm 5: Algorithm of procedure noOfSubSequence

6. Similarity Percentage

This operator is used to calculate the Similarity percentage by comparing two DNA strand sequence. The inputs of the operator are two Strand sequences and return the similarity percentage as output. The algorithm of the operator is given below.

Procedure similarity Percentage (DNA Strand sequence A, DNA Strand sequence B)

1. Get the DNA Strand A. Extract the barcode sequence.
2. Get the DNA Strand B. Extract the barcode sequence.
3. Set n to the length of DNA Strand A and m to length of DNA Strand B.
4. Initialize variable intersection to 0.
5. Examine the code pair sequence of DNA Strand A by starting search from the *start* code till *end* (*i* from 1 to *n*).
6. Examine the code pair sequence of DNA Strand B by starting search from the *start* code till *end* (*j* from 1 to *n*).
7. If s[i] equals t[j], do increment as intersection++;
10. Return percentage as (2.0 * intersection)/(A.length + B.length)

Algorithm 6: Algorithm of procedure similarity Percentage

7. **get Base Percentage**

This operator is used to calculate the percentage of each base (A, T, C, G) present in DNA strand sequence. The input of the operator is DNA Strand Sequence and base. It returns the percentage of input base as an output. The algorithm of the operator is given below.

Procedure get Base Percentage (DNA Strand sequence, Base residue)

1. Get the DNA Strand Structure. Extract the barcode sequence.
2. Get the code pair sequence by starting search from the *start* code till *end*.
3. The count of Base code pair present in DNA Strand sequence stored in integer array.
4. For multiple existence get all the indexed of Base code pair by repeating step 2-3.
5. Divide the count of base pair by sequence length.
6. Return the percentage by multiplying the output of step 5 by 100 for base pair.

Algorithm 7: Algorithm of procedure getBasePercentage

8. **get Aligned SubSequence**

This operator is used to find the aligned common sub sequence. The inputs of the operator are two DNA Strand sequences and return the aligned sub sequence as output. The algorithm of the operator is given below.

Procedure get Aligned SubSequence (DNA Strand sequence A, DNA Strand Sequence B)

1. Get the DNA Strand A. Extract the barcode sequence.
 2. Get the DNA Strand B. Extract the barcode sequence.
 3. Examine the code pair sequence of DNA Strand A with code pair sequence of DNA Strand B by starting search from the *start* code till *end* in order.
 4. *The similar code pair sequences of both DNA strand sequence are stored in an array.*
 5. *Return array as aligned sub sequence.*
- Algorithm 8: Algorithm of procedure getAlignedSubSequence

9. **no Of Base Residue**

This operator is used to get the no. of nucleotide bases participating in forming a DNA strand sequence. The input of the operator is DNA strand Sequence and Base residue. It returns the no of (input) Bases as an output. The algorithm of the operator is given below.

Procedure no Of Base Residue (DNA Strand sequence, Base)

1. Get the DNA Strand Structure. Extract the barcode sequence.
 2. Find base code pair in DNA Strand Sequence from start code till end.
 3. The count of base code pair present in DNA strand sequence store in integer array.
 4. For multiple existence get all indexes of base code pair by repeating step 2-3.
- Algorithm 9: Algorithm of procedure noOfBases