

# Use of Display File Structure and Operation Code for DNA Sequence to Visualise and Presentation for Dot Matrix Graph for DNA Sequence Alignment

Chintankumar Dayalal Gohel\*

Assistant Professor, Department of Computer Science, Gujarat Vidyapith,  
Ashram Road, Ahmedabad, India; chintan.kumar@gmail.com

## Abstract

The management and analysis of bio information using computer application is study that related with the use of computing facility for collection, visualization, analysis and presentation of bio information. The DNA sequence achieving in lab is stored in the computer files like as ordinary text files. This file includes information about the sequences. This sequence alignment is comparison and finding relatedness of two sequences using dot matrix graphical method. Display file structure and operation code assigned to nucleic acid used for generating this graph and this will make computing faster. So goal of this concept is used to made visualisation and presentation faster as compared to traditional method. The similarities in graph of two sequences represent that the two sequences must have derived from a common ancestor and evolutionary origin.

**Keywords:** ASCII, Display File, DNA Sequence, Operands, Operation Code, Sequence Alignment

(Date of Acceptance: 04-10-2015; Plagiarism Check Date: 24-10-2015; Peer Reviewed by Three editors blindly: 26-11-2015; Reviewer's Comment send to author: 20-12-2015; Comment Incorporated and Revert by Author: 25-12-2015; Send for CRC: 30-12-2015)

## 1. Introduction

This Management and analysis of bio information using computer application is study that related with the use of computing facility for collection, visualization, analysis and presentation of bio information. The DNA sequence achieving in lab is stored in the computer files like as ordinary text files. This file includes information about the sequences<sup>1</sup>.

Dot matrix method also known as the dot plot method is a graphical way of comparing two sequences in a two dimensional matrix. In this presented work it is recommended and suggested to use primitive operations, display file structure and generated operation code based graph so it occupy less amount of memory and processor. Speed for processing display file is also high.

## 2. About Sequence Alignment

This Sequence alignment is comparison and finding relatedness of two sequences. It is process to finding similarities between two sequences<sup>2</sup>. If the two sequences have major similarity then we can assume that the two sequences must have derived from a common ancestor and evolutionary origin. When a sequence

alignment is processed and displayed correctly it represents the evolutionary relationship of the two sequences and there is possibility of common ancestor is also very high.

## 3. Method for Sequence Alignment

Dot matrix method is used to compare two sequences where sequence letter are written in the horizontal and vertical axes of the matrix. In this the comparison is performed by matching and scanning every sequence letter of one sequence with sequence letter of another one for similarity.

If a sequence letter match is found a dot is placed within the graph else the matrix row and height positions are left blank. When the two sequences have same regions of similarity many dots connected to form line and it shows the sequence alignment. If there are break in the middle of a diagonal line they indicate insertions or deletions in sequences<sup>3</sup>.

A Sequence alignment method works with comparison of sequences. This method explained as under:

Sequence alignment method: The pairwise sequence alignment is to find the best pairing of two sequences so there is maximum relatedness will be finding. To achieve this one

sequence needs to be shifted relative to the other to find the position where maximum matches are found. There are two different alignment concept used for aligning sequences: (1) Global Alignment (2) Local Alignment. In Global alignment two sequences that we want to align considered as similar over their entire length. In this alignment is performed from beginning to end of both sequences to find the best similarity and alignment towards the entire length between the two sequences. This method is suitable for applying when there are closely related and probably similar sequences of the same length. For sequences of variable lengths this method is not suitable for generating good results because it does not recognize highly similar local regions between the two sequences.

While in local alignment it does not consider that the two sequences have similarity in their entire length. This method only finds local regions with the high level of similarity between the two sequences and aligns these regions without alignment of the other part of the sequences and their regions.

#### 4. Use of Display File and Operation Code for Dot Matrix Method

The computer graphics user will not have to start from scratch in preparing for a project. Instead they have a graphics system available. This system will include special hardware for output and input of pictorial information and software routines for performing the basic graphics operations. The purpose of a graphics system is to make programming easier for the user<sup>4</sup>. For that we use primitive commands in graphics are those that regardless of the differences in display devices offer a set of graphics primitive commands. Suppose the first primitive command we shall consider is that for drawing a line segment so we operation code 1 is assigned to line drawing. Where a line segment may be specified by its two endpoints and can be used for dot matrix method to connect two sequences where they are similar. For movement of pen operation code 2 is assigned. This procedure is fine for producing a connected string of line segments for two different sequences and if these two line segment aligns and match with each other we consider it that are similar sequences and match with each other.

In this process we can construct a line drawing by a series of lines and move commands. For that display file structure will be created and interpreted for matching two sequences<sup>5</sup>. The structure of display file consists of mainly two parts: (1) Operation Code (opcode) and (2) Operands (coordinates of a point (x,y) The method for storing these instructions is to use three separate arrays as under.

Operation code assign as per following manner. This is negative ASCII value of every alphabet in the DNA sequence.

This negative value will be identified operation code for that particular sequence letter and it will be stored in display file. Then operands will be assigning for presenting this particular letter in x axe or y axe. Then any line drawing algorithm used for connecting the dot where the sequence is matched in dot matrix method.

Display file consists of three arrays. First array will use to store operation code for primitive operation like character generation, line drawing and movement of pen position on screen. The another two array store information about operands that is helpful for deciding where particular character will be displayed on the screen or drawing line between two endpoints or moving current pen position on the screen. Here operation code 1 is for movement of current pen position and 2 is for line drawing while for DNA sequence letter negative ASCII value of that particular letter will be used for remembering and processing operation code easily.

Display file will be created for particular sequence as mentioned in Table 2.

In Table 2 sample DNA sequence GCCTGGA will presented in x axe and first DF-OP code e.g. 2 is used to move pen to position 10, 10 which is operands in display file. The letter of DNA sequence will be displayed at every location presented in the operands because operands represents x and y direction of the particular letter.

**Table 1.** Operation code assigned to DNA sequence

Symbol	Meaning	Description	Operation Code
G	G	Guanine	-71
A	A	Adenine	-65
T	T	Thymine	-84
C	C	Cytosine	-67
R	A or G	puRine	-82
Y	C or T	pYrimidine	-89

**Table 2.** Display file structure for DNA sequence

DF-OP	DF-X	DF-Y
2	10	10
-71for G	20	10
-67 for C	30	10
-67 for C	40	10
-84 for T	50	10
-71for G	60	10
-71for G	70	10
-65 for A	80	10

Now take another DNA sequence like TTAAGCC<sup>6</sup> and this sequence represented in y axes as mentioned in Table 3 using display file. In this table first pen position moved to x,y location 10,10 and then next operation code will be displayed at x,y location 10,20. This will be displayed in y axe because y coordinate value in operands is same for entire sequence letter. Now comparison will be made using dot matrix method and if letter in x axe and y axe are similar then that location e.g. operands value will be saved and this operands will pass to line drawing algorithm. This line drawing algorithm generates line among different similar sequence and if the line is straight then it concluded that both sequences are similar and they are originated from common ancestor and if the line is not straight then that sequence letter will not match with other sequence. This analysis will useful for disease prediction, drug discovery and vaccine design.

The display file will be interpreted using display file interpreter<sup>7</sup>. Display file interpreter work as interface between our graphics programme and the display devise. It is software which can be developed for interpreting display file structure<sup>8</sup>. It interprets every entry in display file line by line and calls particular primitive operation command for drawing line or for<sup>9</sup> moving pen on screen or foe displaying particular letter on the screen.

## 5. Suggested Algorithm for Presenting DNA Sequence Letter in DOT Matrix Algorithm

For assigning negative ASCII value to the every letter of DNA sequence following algorithm will be suggested<sup>10</sup>. This algorithm read the sequence and gets the character one by one and assign negative ASCII value to them as operation code<sup>11</sup>.

DNA (SEQUENCE) is algorithm for placing instructions for DNA sequence for printing<sup>12</sup>.

Argument STRING the sequence to be entered<sup>13</sup>  
Global DF-PEN-X, DF-PEN-Y the pen position

**Table 3.** Display file structure for DNA sequence

DF-OP	DF-X	DF-Y
2	10	10
-84 for T	10	20
-84 for T	10	30
-65 for A	10	40
-65 for A	10	50
-71for G	10	60
-67 for C	10	70
-67 for C	10	80

SPACE the sequence spacing  
Local LEN the length of the sequence  
X, Y the character position<sup>14</sup>  
I an index to count off the sequence  
CHR the sequence letter to be stored  
OP the operation-code for sequence letter  
BEGIN  
determine the length of the string  
LEN<-LENGTH(SEQUENCE)  
save the pen position for restoration after string is entered<sup>15</sup>  
X <- DF-PEN-X;  
Y <- DF-PEN-Y;  
enter the string  
FOR I = 1 TO LEN DO  
BEGIN  
consider the ith character of the string  
CHR <- GETCHR(SEQUENCE, I);  
form its character code  
OP <- - DECODE(CHR);  
enter it into the display file  
DISPLAY-FILE-ENTER(OP);  
DF-PEN-X <- DF-PEN-X + SPACE;  
DF-PEN-Y <- DF-PEN-Y + SPACE;  
END;  
END;

For storing DNA sequence into display file following algorithm can be used. In this the negative ASCII value<sup>16</sup> of particular letter will be stored and also at which location that letter will be displayed is also stored.

CREATE\_DF(OP, X, Y) This algorithm will place an instruction into the display file<sup>17</sup>  
Arguments OP, X, Y the instruction to be entered  
Global DF-OP, DF-X, DF-Y the three display-file arrays  
FREE the position of the next free cell  
Constant DFSIZE the length of the display-file arrays  
BEGIN  
IF FREE > DFSIZE THEN RETURN ERROR 'DISPLAY FILE FULL';  
DF-OP[FREE] <- OP;  
DF-X[FREE] <- X;  
DF-Y[FREE] <- Y;  
FREE <- FREE + 1;  
END;

For displaying sequence letters on the screen following algorithm will be suggested. In this algorithm the display file will be interpreted using display file interpreter and then as per operation code which is negative ASCII value of sequence letter is encoded into its actual letter and will be displayed on the

screen as per operands value which is x and y coordinate value. Following algorithm interprets the display file and perform character generation algorithm.

INTERPRET(I,LEN) Scan the display file for performing the instructions

```
Arguments I the starting index of the display-file scan
LEN the number of instructions to be interpreted
Local NTH the display-file index
OP, X, Y the display-file instruction
BEGIN
FOR NTH = I TO LEN
DO
BEGIN
GET-POINT(NTH, OP, X, Y); IF OP < -31 THEN
GENERATECHAR(OP, X, Y)
END;
END;
```

## 6. Conclusion

Display file structure is used to present DNA sequence letter in dot matrix sequence comparison algorithm. The operation code assign to every letter of DNA sequence which is negative ASCII value of particular letter for easily remembering the operation code. The sequence can be presented in x axe and y axe. This will improve the speed of computer graphics processing because every biological information stored in the form of numbers which will easily processed by computer. The output of this two sequence comparison will determine that the DNA sequences are from common ancestor or not. Due to this comparison this graphical presentation of sequence analysis will useful for disease prediction, drug discovery and vaccine design.

## 7. References

1. Attwood TK, Smith DP. Introduction to Bioinformatics. 1st ed. Pearson Education; 2001.
2. Gibas C, Jambeck P. Developing Bioinformatics Computer Skills. 1st ed. O'Reilly and Associates; 2001.
3. Higgins D, Taylor W. Bioinformatics – Sequence, Structure and Databanks. 2nd ed. Oxford University Press; 2003.
4. Hearn D, Baker P. Computer Graphics. 2nd ed. Prentice Hall; 2000.
5. Harrington S. Computer Graphics – A Programming Approach. 2nd ed. McGraw Hill Book Company; 1987.
6. Tisdall J. Beginning Perl for Bioinformatics. 1st ed. O'Reilly and Associates; 2001.
7. Nair AS, Sunitha P, Aswathi BL. Bioinspired Computing– the Evolving Scenario. CSI. 2011; 35(9):6–8.
8. Govindan G, Goli B, WEKA – A powerful free software for implementing Bio – inspired Algorithms. CSI. 2011; 35(9):9–11.
9. Tripathi KK. Bioinformatics: The Foundation of Present and Future Biotechnology. Current Science. 2000; 79(5).
10. Evolutionary Genetic Analysis and Sequence Alignment. Briefings in Bioinformatics. 2004; 5(2):150–63.
11. Mukhopadhyay A, Chattopadhyay A. Computer Graphics and Multimedia. 2nd ed. 2007.
12. Anderson G. Bioinformatics Tutorial rev. 9-2012. Cellular and Molecular Biology. 2012; 1–10.
13. Newman W, Sproull R. Principles of Interactive Computer Graphics. 2nd ed. New Yorl: McGraw-Hill; 1979.
14. Pavlidis T. Algorithms for Graphics and Image Processing. Rockville, Md: Computer Science Press; 1982.
15. Rogers DF, Adams JA. Mathematical Elements for Computer Graphics. New York: McGraw-Hill; 1976.
16. Scott JE. Introduction to Interactive Computer Graphics. New York: Wiley Interscience; 1982.
17. Sherr S. Vide and Digital Electronics Displays: A user's Guide. NJ: Wiley, Somerset; 1982.

### Citation:

Chintankumar Dayalal Gohel  
 “Use of Display File Structure and Operation Code for DNA Sequence to Visualise and Presentation for Dot Matrix Graph for DNA Sequence Alignment”  
 Global Journal of Enterprise Information System. Volume-7, Issue-4, October-December, 2015. (www.gjeis.org)

### Conflict of Interest:

Author of a Paper had no conflict neither financially nor academically.  
 In-fact the Authors are highly indebted and acknowledge to the Hessian LOEWE excellence initiative within CASED, and a Fraunhofer Attract grant.