CURRENT STUDIES IN COMPARATIVE EDUCATION, SCIENCE AND TECHNOLOGY, VOLUME 4, NUMBER 1, 2017, PP. 221-238

Sequence Alignment As A Method of Bacterial Identification

DISEGHA G.C. & JEAPUDOARI T. F
Rivers State University, Port Harcourt, Nigeria

Abstract

This work is aimed at giving a simple literature of nucleic acid sequence alignment (NASA). Sequence alignment (SA) is a method used to analyze strands of DNA, RNA or proteins to determine the similarities between the strands implying significance of function, structure or phylogeny. Methods of SA include Global, Local, and semi-global sequence alignment. In similarity search between sequences, the Basic Local Alignment Search tool (BLAST) can be used for identification of structure and phylogeny between two sequences by comparing proteins or nucleotide sequences retrieved from databases, and then calculating the statistical significance of the matches that occur. Processes carried out prior to proper SA include isolation of pure culture from the organism, DNA extraction, Primer design and synthesis, PCR amplification and purification and DNA sequencing. In the past, some methods of identifying bacteria have relied on phenotypic methods like colony morphology, Gram reaction, and biochemical tests. These methods have their limitations due to unculturable bacteria and strains which display distinctive biochemical characteristics, making it difficult to characterize them into taxonomic groups. Molecular techniques employed over the years have become useful in overcoming some of these limitations in bacterial identification, thus advocated for more advanced biomedical research and development. Therefore, improved and more efficient algorithms for sequence alignment are required for fast, easy to use, and precise target bacterial identification in this rapidly growing field of Biotechnology. It is hoped that cost of equipment and processes leading to sequence analysis proper will decline as time goes on.

Keywords: Bioinformatics, Sequence, alignment, identification, phylogeny.

Introduction

Bacteria are ubiquitous and constitute a major component of the microbial world. Identification and characterization of bacteria has been a challenging task in research and diagnostic laboratories (Fournier *et al*., 2013). Until recently this task was undertaken by traditional or phenotypic-based methods. These methods depend on cultural characteristics, colony morphology, Gram staining reactions, and biochemical tests. These processes are both cumbersome, and most times, unreliable and less

accurate due to the associated conditions and practices. One major limitation of phenotype-based method is that it can only be used for *culturable* bacteria. In addition, some biochemical patterns observed may not fit into existing patterns already used for identification of known bacteria (Premier Biosoft, 2015). There are however, more modern techniques which have proved more rapid and accurate than the historical methods (Tshikhudo *et al.,* 2013). In genotypic-based identification, one can easily detect bacterial strains directly from clinical samples or from small amounts of cultured bacterial cells, thus improving the sensitivity and decreasing the time required for diagnoses and identification of bacteria. PCR has been useful in this respect as it relies on primer sequences designed to facilitate bacterial identification at any level of specificity (Premier Biosoft, 2015). Genotypic-based methods of bacterial identification are more reliable, specific and sensitive. These methods, however, are more expensive due to the type of equipment required to carry out the tests (Tang *et al*., 1998; Adzitey *et al*., 2013).

Sequence alignment is a method whereby DNA, RNA, or proteins are arranged or ordered in sequences to identify regions of similarity. Sequence alignment methods are used to find the best matching sequences (Mount, 2004). Bacteria identification can be applied in a wide variety of studies including Environmental studies, Microbial Forensics, Analytical studies even in Criminal investigations (Ortet, 2010).

In the past, some methods of identifying bacteria have relied on using phenotypic schemes such as colony morphology, Gram staining reactions, and biochemical tests. These methods, however, have their limitations in regard to existence of artificially unculturable bacteria. Some strains display distinctive biochemical characteristics, which make it difficult to characterize them into any known taxonomic groups and classes (Blazewicz *et al*., 2009)

Molecular techniques developed over the years have become useful in overcoming some of these limitations in bacterial identification. This is because many non-culture based methods, employing the tools of genomics and bioinformatics, have been developed over the years. Sequence alignment one of such key aspects in handling bacterial identification (Kim and Lee, 2008).

In Bioinformatics, sequence alignment is a method used generally to analyze strands of DNA, RNA or proteins in order to determine the similarities between the strands which may depict a significance of functional, structural or evolutionary relationships between the sequences (Mount, 2004).

Though there are other methods of bacterial identification, including conventional methods, the present method under discussion lays focus on:

a) How sequence alignment method is used for bacterial identification.
b) Tools and software used in sequence alignment,
c) The advantages and disadvantages associated with sequence alignment technique of identifying bacteria,
d) Its general application in various areas or studies of life.
e) And that the bio-scientist must be computer literate, and have some basic skills in information communication technology (ICT).

Types of Sequence Alignment

There are varying kinds of sequence alignment but the most common types are:

i. Pairwise alignment
ii. Multiple sequence alignment
iii. Structural alignment

Pairwise Alignment

Two sequences are used when carrying pairwise alignment. In this type of alignment, one of the sequences is inscribed over the other such that the overlapping element chains are then observed and noted. Since the chains are not all of the same length, gaps and insertions are usually distinguished. Recurring chains of elements in the sequence usually occur because they were a favorite evolutionary adaptation (Agraval *et al*., 2008). Some software used for pairwise alignment are BLAST and Mega, YASS, Base by Base, etc.

Multiple Sequence Alignment

Multiple sequence alignment is related to the pairwise alignment in mechanism. However, the multiple sequence alignment uses more than two sequences, that is, it uses three or more sequences. It can sometimes be presented in a tree-shaped form (Elias and Isaac, 2006). Some software used in multiple sequence alignment include MUSCLE, T-coffee, ClustalW, etc

Structural Alignment

Structural alignment compares shapes of two or more sequences thereby establishing homology between them. Structural alignment takes a look at the entire sequence as an individual unit, of which, pairwise and multiple alignment deals only with the elements within the arrangements. Structural alignment usually makes comparison in a three dimensional format (Bourne and Shindyalov, 2003). Some software used for structural alignment include Vector Alignment Search Tool (VAST), Local-Global alignment (LGA), FSA, Expresso, DIALIGN, MAFFT, BALIi-phy, POSA, , FATCAT, *etc*. (Omic Tools, 2015).

Methods of Sequence Alignment

There are two major methods of carrying out sequence alignment and they are:
i. Global alignment
ii. Local alignment
iii. Semiglobal alignment

Global Alignment

Global alignment is a sequencing method that tries to align all the residues from the beginning to the end of the sequence to be able to find out the best possible alignment (Brudno *et al.,* 2003). The Global alignment method is very much appropriate for closely related sequences that are of the same length (Brudno *et al*., 2003). This technique is carried out using the Needleman-winsch algorithm, i.e. based simply on dynamic programming (Mount, 2004)

Local Alignment

Local alignment is method of sequence alignment used for comparing sequences suspected to have a similarity or even dissimilarity by finding local regions on a sequence having high level of similarity. The local alignment method used is usually the Smith-Waterman algorithm, which just likes the global alignment method, is also based on dynamic programming (Polyanovsky *et al.*, 2011).

Semiglobal Alignment

This is a Combined or hybrid method of sequence alignment developed from the combination of the global and local methods, which leads to the expression "semi-global" (Brudno, *et al.*, 2003). This method is used in finding the best alignment that consists of the start and end of one sequence or the other. This method is best used when the downstream part of one of the sequences intersects with the upstream part of the other (Brudno, *et al*., 2003).

Global and Local alignments are clearly characterized by their different algorithms (that is, a dynamic programming approach), which aligns two different series of sequence by using scoring matrices (Polyanovsky *et al.,* 2011). The table below gives major differences between global and local sequence alignment (Major Differences, 2015).

Table 1: Differences between Global and Local Sequence alignment

Global Sequence Alignment	Local Sequence Alignment
• Attempt is made to align the entire sequence from beginning to end.	Finds local regions with highest level of similarity between the two sequences
• Contains all the letters from both query and target sequences.	Aligns a substring of the query sequence to a substring of the target sequence.
• Two sequences of similar length are fitted for global alignment.	Any two sequences can be locally aligned as local alignment finds stretches of sequences with high level of matches disregarding the alignment of the rest of the sequence regions.
• Suitable for aligning two closely related sequences	Suitable for aligning more divergent sequences or distantly related sequences.
• Usually carried out for comparing homologous genes like comparing two genes with same function (in bacteria versus mushroom) or comparing two proteins with similar function.	Used for finding out conserved patters in DNA sequences or conserved domains or motifs in two proteins.
• A general global alignment technique is the Needleman-Wunsch algorithm	A general local alignment method is Smith -Waterman algorithm
• One example of Global Alignment Tool: EMBOSS Needle	One example of Local Alignment Tool: BLAST.

Source: Major Differences (2016)

Pre Requirements For Sequence Alignment

Before sequence alignment is carried out on a sample, some processes have to be carried out. These processes include:

a) Isolation of pure culture from the organism
b) DNA extraction
c) Primer design and synthesis
d) Polymerase Chain Reaction (PCR) amplification and purification
e) DNA sequencing

(a) Isolation of Pure Culture From The Organism

In this process, a pure culture of the specific bacteria is isolated on a culture medium using the traditional culture method. DNA extraction from the cellular components of the organism is a method of purification of DNA from samples by means of combining the physical and chemical methods (Dahm, 2008).

(b) DNA Extraction

The cells for DNA extracted are first collected, then the cell membranes are treated with appropriates detergents and/or surfactants to cause cell lysis, exposing the DNA along with the other cytoplasmic components (Elias and Isaac, 2006).
The solution is then treated again using a salt solution (concentrated) which clumps together broken proteins, RNA and Lipids from the cell membrane.
The treated solution is then subjected to centrifugation which helps to separate the DNA from the other cellular debris. (Elias and Isaac, 2006)
The separated DNA is then purified from the detergents and salts, including reagents used during the cell lysis using such methods as:

i. Ethanol precipitation
ii. Phenol-Chloroform extraction or
iii. Minicolumn purification

An extracted DNA is not completely pure due to the presence of cellular and histone proteins which remain bound to the DNA. Removal of these proteins is done either by adding a protease or by having precipitating the proteins using sodium or ammonium acetate (Claassen *et al*., 2013), DNA purifications kits are available for procurement from different producers such as Zymo, Promega, Sigma-Aldric, Epicentre, etc. DNA sequencing is the last process before alignment of the sequences. DNA sequencing as an identification technique is used to determine the nucleotide sequence of a DNA segment.

(c) Polymerase Chain Reaction (PCR) Amplification and Purification

Various procedures in molecular biology require enough amounts of DNA for identification processes, using a process called Polymerase Chain Reaction (PCR). The Polymerase Chain Reaction is relatively simple and it is an inexpensive tool used to focus in on a segment of DNA and copies it billions of times over (Valones *et al*., 2009).

Steps in PCR

The steps of PCR are vividly described by Valones *et al.,* (2009) as follows:

"Three key steps in PCR operation are denaturation, annealing, and extension (Valones *et al*., 2009). To amplify a segment of DNA using PCR the sample is first heated (from 90 to 97^{o}C) so the DNA denatures, or separates into two pieces of single-stranded DNA. Next, an enzyme called "Taq polymerase" synthesizes - builds - two new strands of DNA, using the original strands as templates. This process results in the duplication of the original DNA, with each of the new molecules containing one old and one new strand of DNA. Then each of these strands can be used to create two new copies, and so on, and so on. The annealing phase happens at a lower temperature, 50-60°C. This allows the primers to hybridize to their respective com-

plementary template strands, a very useful tool to forensic chemistry. The newly-formed DNA strand of primer attached to template is then used to create identical copies off the original template strands desired. Taq polymerase adds available nucleotides to the end of the annealed primers. The extension of the primers by Taq polymerase occurs at approx 72°C for 2-5 minutes. DNA polymerase I cannot be used to elongate the primers as one would expect because it is not stable at the high temperatures required for PCR" (Valones *et al*., 2009; Joshi1 and Deshpande, 2011).

The National Centre for Biotechnology Information (NCBI)

NCBI, also known as the National Centre for Biotechnology Information was founded in 1988 and located in Bethesda, Maryland via the sponsored legislation by Senator Claude Pepper. It is a part of the United States National Library of Medicine (NML). The NCBI is a collection of numerous databases that have a relevance to biotechnology and even biomedicine. These databases also serve as an important source for bioinformatics tools and services. Some of the major databases includes, GenBank for DNA sequences and PubMed which is a database for biomedical literature. Some others include, the NCBI Epigenomics database, which are all available online via the Entrez search engine (Ostell, 2002). David Lipman, one of the original authors of the BLAST sequence alignment program, and a well-respected person in bioinformatics is the director of the NCBI.

NCBI has had responsibility for making available the GenBank DNA sequence database since 1992. (NCBI Handbook). The GenBank coordinates with individual laboratories and other sequence databases such as those of the European Molecular Biology Laboratory (EMBL) and the DNA Date Bank of Japan (DDBJ). (Maglott and Pruitt, 2005). Since 1992, the National Centre for Biotechnology Information has provided some other data bases adding to the GenBank. The NCBI provides Gene, online Mendelian Inheritance in man, Molecular modeling database, database of single-nucleotide polymorphisms (dbSNP), Reference Sequence collection, the human genome map, including a browser for taxonomy and coordinates with the National Cancer Institute to provide the Cancer genome anatomy project. NCBI has also assigned an identifier which is unique to each species of organism using a taxonomy ID number. (NCBI Taxonomy) (Altschul *et al*., 2009).

The NCBI also uses software tools available online, via browsing to help researchers collate data and information that will be useful for further research. An example of such tool is the Basic Local Alignment Tool developed by David Lipman and his co-researchers. The BLAST tool is a program that searches for similarities that exists between sequences. BLAST is able to carry out sequence comparisons between sequences against the GenBank DNA database as quick as 15seconds. (Maglott and Pruitt, 2005).

The NCBI contains other databases such as:

a) The NCBI Bookshelf
b) Entrez
c) Gene
d) Protein database for protein re-source
e) PubchemBioAssay database etc

Basic Local Alignment Search Tool (BLAST)

The Basic Local Alignment Search Tool can be used in sequence alignment in finding similarity regions between sequences by comparing proteins or nucleotide sequences using databases and then calculates the statistical significance of the matches that occur. All these are carried out using the BLAST program (Ochmen and Baxter, 2013)

Relationship exist between sequences, therefore BLAST becomes an appropriate tool for inference of these functional and evolutionary relationships, as well as aid in the identification of the members of gene families. BLAST is one of the most widely used bioinformatics programs for sequence searching The BLAST algorithm and the computer program that implements it were developed by Stephen Attschul, Warren Gish and David Lipman at the U.S. National Center for Biotechnology Information (NCBI), Web Miller at the Pennyslyvania State University and Gene Myers at the University of Arizona and it is available on the web on the NCBI website (Ochmen and Baxter, 2013).

There are different types of BLAST programs as shown in table 2.1;

Table 2.1. Different BLAST programs and Algorithms

Type of blast	Program	Algorithm
• Nucleotide blast Nucleotide Query	Searches nucleotide database using a discontinuous MegaBLAST	BLASTn, MegaBLAST
• Protein Blast, Protein Query	Search protein database using a, delta-BLAST	BLASTp, psi-BLAST, phi- BLAST
• BLASTx	Search protein database using a translated nucleotide query.	BLASTx
• TBLASTn	Search translated nucleotide data base using a protein query	TBLASTn
• tBLASTx	Search translated nucleotide data base using a translated nucleotide query.	tBLASTx

Source; (NCBI, 2016)

Blast Procedure for Bacterial Sequence Alignment

The procedure for using the BLAST tool in sequence alignment for identifying bacteria is a simple one and it includes five different steps:

1. Select the BLAST program (if you are already on NCBI website.)
The type of BLAST program is usually specified since they are more than one, by selecting from the data base like; BLASTp, BLASTn, BLASTx, tBLASTn, tBLASTx (Li *et al.* 2009).

2. Query entering or file upload
A query sequence is entered and this is done by pasting the sequence in the query box. Or, the file containing the sequence for similarity search is uploaded in a FAS-TA format. (Li *et al.* 2009).

Figure 2.1: A dialog box for entering a query sequence or uploading file containing sequence. (Source; NCBI, 2016)

3. Data base selection to search
Using the query sequence uploaded, the user, searches for similar sequences from databases. But before this search, the data bases available and the type of sequences available in those databases must be known (Madden, 2002).

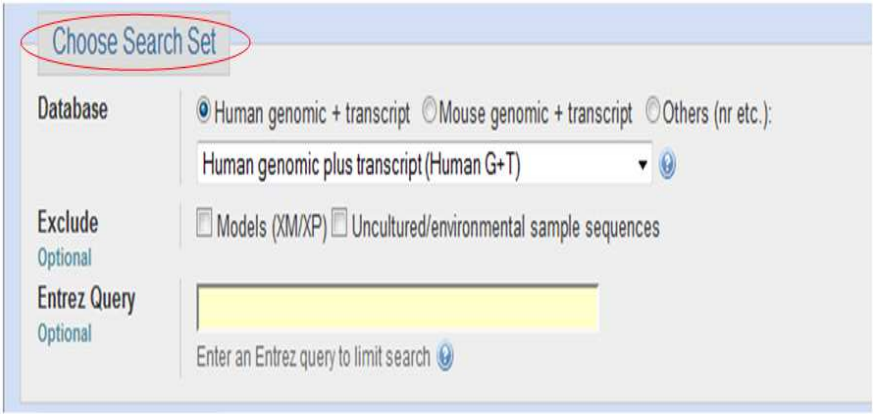


Figure 2.2: A dialog box for selecting database to search
(Source; NCBI, 2016)

Algorithm selection and parameters of the algorithm

A unique algorithm must be specified for the BLAST program to use, since different BLAST programs have different algorithms. Example; the Nucleotide BLAST uses the megaBLAST algorithm. There are different algorithm parameters used, and they are; Target sequences, word size, query range, short queries, E-value, filters (filter and mask) and scoring parameters (match/mismatch scores and gap penalties). All these are required in other to run BLAST programs. Most of the time, default values are provided, the user can make necessary adjustments to suit his desired search criteria (Rastogi *et al*.,2013).

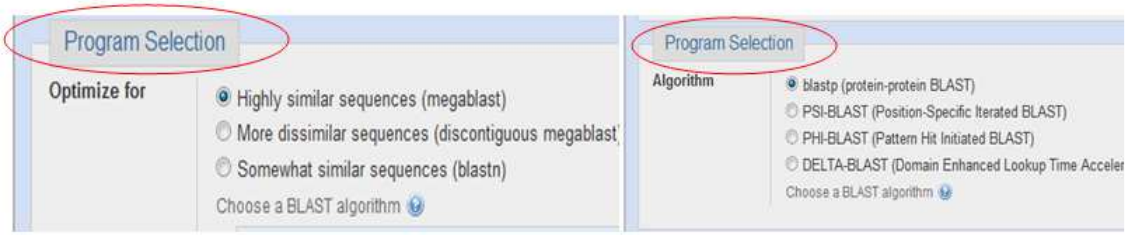


Figure 2.3: A dialog box for Algorithm and the parameters
(Source: NCBI, 2016)

Running the BLAST program

After inputting the required values, the program is submitted or in this case, can be run by clicking the "BLAST" button at the end section of the page. Shortly after submitting the query sequence for a sequence similarity search, a result page will pop-up, displaying the information like; Query id, Description, length of sequence, molecule type, database name and BLAST program, also showing the intended domains that have been detected while it was undergoing sequence similarity search. (Madden, 2002).

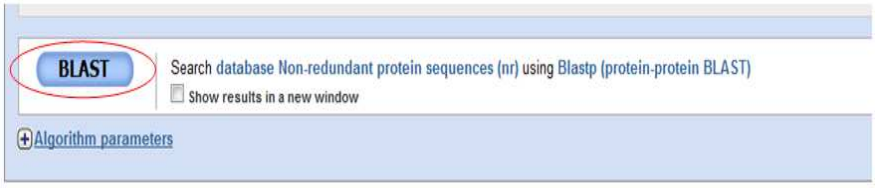


Figure 2.4: A dialog box for running the BLAST program
(Source: NCBI, 2016)

BLAST Search Output

After submitting the query sequence for the sequence similarity search, the result page will appear along with the information like Query id, Description, Molecule

type, Length of sequence, Database name and BLAST program. It shows the predicted conserved domains that have been detected while undergoing sequence similarity search. The query sequence represented as a numbered red bar below the color key. Database hits are shown below the query (red) bar according to the alignment score. Among the aligned sequences, the most related sequences are kept near to the query sequence. User can find more description about these alignments, by dragging the mouse to the each colored bar which is shown in figure 2.5 (Madden, 2002; Wheeler and Bhagwat, 2007).

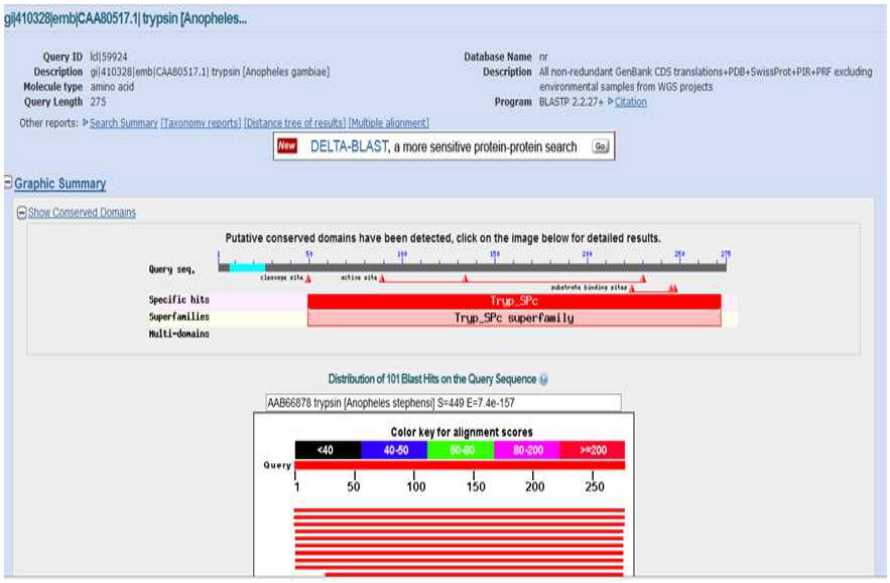


Figure 2.5: A dialog box showing the BLAST result
(Source; NCBI, 2016)

The alignment is preceded by the sequence identities, along with the definition line, length of the matched sequence, followed by the score and E-value. The line also contains the information about the identical residues in alignment (identities), number of positivity's, number of gaps used in the alignment. Finally, it shows the actual alignment, along with the query sequence on the top and database sequence below the query. The number on either sides of the alignment indicates the position of amino acids/nucleotides in sequence (Madden, 2002; Wheeler and Bhagwat, 2007).

CGATTAAAGATAGAAATACACGATGCGAGCAATCAAATTTCATAACATCACCATGAGTTTGATCCAAAGC
ATGAGTGTTTACAATGTTTGAATACCTTATACAGTTCTTATACATACTTTATAAATTATTTCCCAAGCTG
TTTTGATACACACACTAACAGATACTCTATAGAAGGAAAAGTTATCCACTTATGCACACTTATACTTTTT
AGAATTGTGGATAATTAGAAATTACACACAAAGTTATACTATTTTTAGCAACATATTCACAGGTATTTGA
CATATAGAGAACTGAAAAAGTATAATTGTGTGGATAAGTCGTCCAACTCATGATTTTATAAGGATTTATT
TATTGATATTTACATAAAAATACTGTGCATAACTAATAAGCAGGATAAAGTTATCCACCGATTGTTATTA
ACTTGTGGATAATTATTAACATGGTGTGTTTAGAAGTTATCCACGGTTGTTATTTTTGTGTATAACTTAA
AAATTTAAGAAAGATGGAGTAAATTTATGTCGGAAAAAGAAATTTGGGAAAAAGTGCTTGAAATTGCTCA
AGAAAAATTATCAGCTGTAAGTTACTCAACTTTCCTAAAAGATACTGAGCTTTACACGATCAAAGATGGT
GAAGCTATCGTATTATCGAGTATTCCTTTTAATGCAAATTGGTTAAATCAACAATATGCTGAAATTATCC
AAGCAATCTTATTTGATGTTGTAGGCTATGAAGTAAAACCTCACTTTATTACTACTGAAGAATTAGCAAA
TTATAGTAATAATGAAACTGCTACTCCAAAAGAAGCAACAAAACCTTCTACTGAAACAACTGAGGATAAT
CATGTGCTTGGTAGAGAGCAATTCAATGCCCATAACACATTTGACACTTTTGTAATCGGACCTGGTAACC
GCTTCCCACATGCAGCGAGTTTAGCTGTGGCCGAAGCACCAGCCAAAGCGTACAATCCATTATTTATCTA

Figure 2.6: A BLAST output nucleotide Sequence of Staphylococcus aureus. (Source; NCBI, 2016)

Advantages of Sequence Alignment

There are advantages or merits of sequence alignment in identifying bacteria and other organisms. This includes:

1. Sequenced RNA, such as expressed sequence tags and full-length mRNA's, can be aligned to a sequenced genome to find where there are genes and get information about alternative splicing and RNA editing. (Kim and Lee, 2008)
2. Sequence alignment is a part of genome assembly, where which sequences are often aligned to find overlap so that contigs which are long stretches of sequences can be generated. (Blazewicz, *et al*., 2009; Wheeler, D., and Bhagwat, M. (2007).
3. Multiple sequence alignments can be used to create a phylogenetic tree. This is because functional domains known in annotated sequences can be used in carrying out alignment in non-annotated sequences. Also, the regions that are conserved and are known to have functional importance can be found (Budd and Aidan, 2009).
4. Sequence alignment, especially multiple sequence alignment methods can be used to identify sites that are functionally important like; active sites and binding sites by locating or finding conserved domains.
5. Sequence alignment uses similarity in sequences to find common ancestry even to species level. This has played a very vital role.
6. Sequence alignment far outweighs the traditional methods of identification since it is faster, more accurate, precise, and more reliable

Disadvantages of Sequence Alignment

Due to the advantages enumerated above and the high demand of molecular or genomic approaches to bacterial identification, sequence alignment, has over recent times, become in vogue in biosciences**.** However, several current disadvantages poses a limitation on the use of genomics and sequence alignment for bacterial identification (Sentausa and Fournier, 2013). Klenk and Göker (2010) reported that com-

pletely sequenced genomes for many of the major lineages of prokaryotes (bacteria inclusive) are not available (Klenk and Göker, 2010). Because of the vast number of bacteria yet to be discovered open domains of databases cannot possibly have storage of nucleic acid sequences of such bacteria. According to Sentausa and Fournier (2013), "the currently available genome sequences have been obtained mostly from three phyla (Proteobacteria, Firmicutes, and Actinobacteria). Thus, many phyla are poorly represented in genomics (http://www.genomesonline.org/cgi-bin/GOLD/index.cgi)". Furthermore, Klenk and Göker (2010) noted that, "even if the genome sequences of the species of interest are available, in many cases they are not type strains, and, therefore must be used with caution, as prokaryote taxonomy is based on type strains only" (Tindall *et al.,* 2010).

Another disadvantage is that existing genomic sequences vary greatly in their finished quality, often being available only as unfinished draft assemblies that, according to Ricker *et al.* (2012) and Klassen *et al*. (2012), may be less informative than finished whole genome sequences. These inadequacies have lead to giving conditions for minimal sequencing quality which are required for genomes to be incorporated in taxonomic analyses. For example, the guidelines developed by the Next-generation Sequencing: "Standardization of Clinical Testing work group" intended for use in open domains such as genbank (Gargis *et al.,* (2012).

Another drawback in sequence based identification is that results obtained through sequence analysis often do not correspond with existing taxonomic categories and related levels. This is based on the fact that prokaryotes do not have one universal method classification for prokaryotes (AI-Ozen and Vesth, 2012).

Sequence alignment generally does not yield single-specific results, rather cumbersome alignment scores are given for the analyst to choose what is best fitted with respect to maximum scores, total scores, query content, E-values , percentage identify, and the total matches displayed (figure 2.7; Figure 2.8).

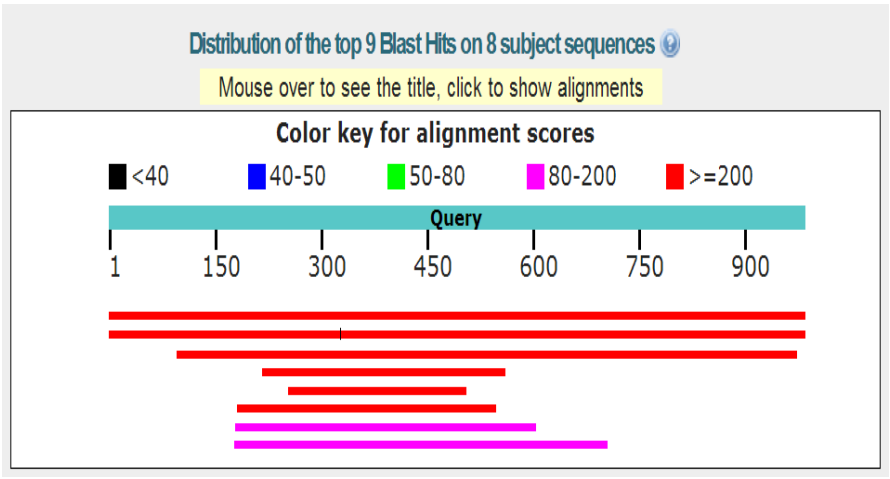
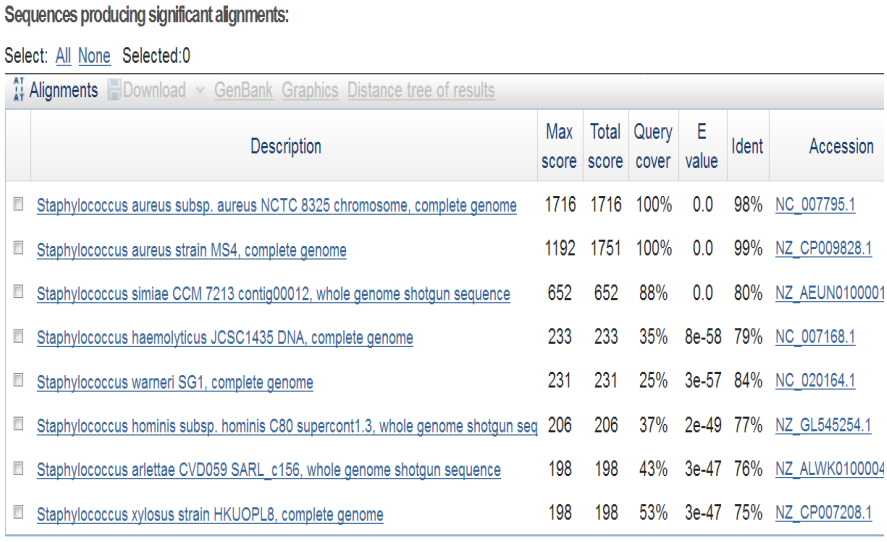


Figure 2.7: Graphic display of the distribution of the top 9 Blast hits on 8 subject sequences



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
	Staphylococcus aureus subsp. aureus NCTC 8325 chromosome, complete genome	1716	1716	100%	0.0	98%	NC_007795.1
	Staphylococcus aureus strain MS4, complete genome	1192	1751	100%	0.0	99%	NZ_CP009828.1
	Staphylococcus simiae CCM 7213 contig00012, whole genome shotgun sequence	652	652	88%	0.0	80%	NZ_AEUN0100001
	Staphylococcus haemolyticus JCSC1435 DNA, complete genome	233	233	35%	8e-58	79%	NC_007168.1
	Staphylococcus warneri SG1, complete genome	231	231	25%	3e-57	84%	NC_020164.1
	Staphylococcus hominis subsp. hominis C80 supercont1.3, whole genome shotgun seq	206	206	37%	2e-49	77%	NZ_GL545254.1
	Staphylococcus arlettae CVD059 SARL_c156, whole genome shotgun sequence	198	198	43%	3e-47	76%	NZ_ALWK0100004
	Staphylococcus xylosus strain HKUOPL8, complete genome	198	198	53%	3e-47	75%	NZ_CP007208.1

Figure 2.8: BLAST display with descriptions of Sequences producing significant alignments

Yet another disadvange of sequence alignment hinges on the fact that, determining phylogeny of an organism based on sequences, poses a difficulty in aligning distantly related sequences using pairwise, alignments without errors creeping in.

Sequence alignment is also a very expensive method of identifying bacteria, This is so because the series of steps, processes, and equipment required prior to sequence alignment proper are high-priced. Equipments such as PCRs or thermocyclers, gene analyzers or gene sequencing machines are sold at very high prices. In addition, there are various tools that are best described as molecular laboratory apparatus that add more to the general cost of molecular identification.

Conclusion

Some methods of identifying bacteria have relied on using phenotypic identifications like Gram staining, bacterial culture and biochemical methods. But these methods however have their limitations as regards to some artificially unculturable bacteria and even some strains which display some distinctive biochemical characteristics making it difficult to characterize them into any known taxonomic groups. Molecular techniques however generated over the years have become useful in overcoming some of these limitations in bacterial identification. This is because many non-culture based methods have been developed over the years. One of these methods is the use of Sequence alignment. This affords the researcher the opportunity of exploring bioinformatics tools as a means of surveying and utilizing newer technologies in teaching, research, and learning, while bearing in mind the cost implications in the use of this specialized field of biotechnology.

Correspondence
Disegha G.C.
Department of Microbiology
Rivers State University, Port Harcourt
Nigeria
Email: gabrieldisegha2@gmail.com

Jeapudoari T. F.
Department of Microbiology
Rivers State University, Port Harcourt
Nigeria
Email: tonyejeaps@gmail.com

References

Adzitey, F., Huda, N., Gulam, R. R., and Ali , G. E. E. (20`3). Molecular techniques for detecting and typing of bacteria, advantages and application to foodborne pathogens isolated from ducks. Journal of Biotechnology. https://www.ncbi.nlm.nih.gov/pmc/articles/

Agrawal, A., Brendel, V.P., and Huang, X. (2008). Comparative analysis of Bioinformatics. *InternationalJournal of Computational Biology and Drug Design*. 1(4):347-367.

Al-Ozen, T. and Vesth, D.W. (2012). Ussery from genome sequence to taxonomy—a skeptic's view. E Rosenberg, E.F. DeLong, E. Stackebrandt, S. Lory, F Thompson (Eds.), The prokaryotes (4th edn.), Springer: Berlin (2012)

Altschul, S.F., Gish, W., Miller, W., and Myers E.W., Lipman, D.J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215:403–410.

Blazewicz, J., Bryja, M., and Figlerowicz, M. (2009). Whole genome assembly from 454 sequencing output via modified DNA graph concept. *Journal of Computational Biology Chemistry,*.33(3)224-30.

Bourne P.E and Shindyalov, I.N. (2003): *Structure Comparison and Alignment*. In: Bourne, P.E., Weissig, H. (Eds): *Structural Bioinformatics*. Hoboken: Wiley-Liss.

Brudno, M., Malde, S., and Poliakov, A. (2003). The Protein Structure. *Journal of Bioinformatics*. 19 (9): 151-62.

Budd, I. and Aidan, D. (2009). Multiple Sequence alignment exercises and demonstrations. *European Molecular Biology Laboratory*. 1(1) 24-30.

Claassen, S, du Toit, E, Kaba, M, Moodley, C, Zar, H, and Nicol M. (2013) A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. J Microbiol Methods, 94:103-10

Dahm, R. (2008).Discovering DNA. *Journal of Nucleic Acid Research*. 122(6): 565-81.

Elias, T., and Isaac, C. (2006).Whole genome assembly. *Journal of Computational Biology*. 1323-1339

Fournier, P.E., Drancourt, M., Colson, P., J Rolain, J.M., La Scola, B., and Raoult, D. (2013) Modern clinical microbiology: new challenges and solutions *Nature Reviews Microbiology,* 11, 574–585

Gargis, A.S., Kalman, L., and Berry, M.W. (2010). Assuring the quality of next-generation sequencing in clinical laboratory practice. Journal of Natural Biotechnology, 30 (12): 1033-1036.

Joshi1, M. and Deshpande, J. D. (2011). Polymerase chain reaction: methods, principles and application. *International Journal of Biomedical Research,* www.ssjournals.com

Kim, N., and Lee, C. (2008). Bioinformatics detection of alternative splicing. *Journal onMethods in Molecular Biology*. 452:179-97.

Klassen, J. and Currie, C. (2012). Gene fragmentation in bacterial draft genomes: extent, consequences and mitigation. Journal of BMC Genomics, 13 (2012):14.

Klenk, H. P. and Göker, M. (2010). En route to a genome-based classification of Archaea and Bacteria?Syst Journal of Applied Microbiology, 33 (10), 175-182.

Li, J.B., Levanon, E.Y., and Yoon, JK.(2009). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* 342 (5932):1210-3.

Madden, T. (2002). *The NCBI Handbook 2nd edition Chapter 16*. The BLAST sequence Analysis Tool. www.ncbi.nlm.nih.org.

Maglott , D.1., Ostel,l J., Pruitt, K.D., and Tatusova, T. (2005). Nucleic Acids Research, 1;33

Major Differences (2016). *Diffeences* between global and local sequence alignment. www.majordifferences.com

Mount, D.M. (2004). Genome Analysis. *Journal ofBioinformatics Sequence.* 2(1)54-60.

NCBI (2016). Different BLAST programs. www.ncbi.nlm.nih.org

Oehmen, C.S., and Baxter, D.J. (2013). Multiple Sequence alignment exercises. *Journal of Bioinformatics*. 29(6):797-798.

Omic Tools (2015). *Multiple structure alignment software tools and Protein data analysis*. https://omictools.com/multiple-protein-structure-alignment-category

Ortet, P., and Bastien, O. (2010). Bioinformatics detection.*Journal of Evolutionary Bioinformatics*. 6:159-187.

Ostell, J. (2002. The Entrez search and retrieval system. IN The NCBI handbook [Internet], National Library of Medicine (US), National Center for Biotechnology Information, Bethesda, MD, Chapter 14. Available from the Entrez Books database

Polyanovsky,V.O., Roytberg, M.A., and Tumanyan, V.G. (2011). Comparative analysis of the quality of a global algorithm and a local algorithm for alignment of two sequences. *Journal of Algorithms for Molecular Biology.*6(1): 25.

Premier Biosoft (2015) *Identifying bacteria.* http://www.premierbiosoft.com/tech

Rastogi, S.C. Mendiratta, N. and Rastogi, P.(2013). Bioinformatics – Methods and applications: Genomics, Proteomics and drug discovery. Delhi: PHI Learning.

Ricker, N, Qian, H., and Fulthorpe, R.R. (2012). The limitations of draft assemblies

for understanding prokaryotic adaptation and evolutionGenomics, 100 (12), 167-175.

Sentausa, E. and Fournier, P.-E. (2013). Advantages and limitations of genomics in prokaryotic taxonomy. *Journal of Clinical Microbiology and Infection*, 19 (9), 790-795.

Tang, Y., Ellis, N.M., Hopkins, M.K., Smith, D.H., Dodge, D.E. and Persing, D.H. (1998). Comparison of Phenotypic and Genotypic Techniques for Identification of Unusual Aerobic Pathogenic Gram-Negative Bacilli.

Tindall, B.J., Rosselló-Móra, R. Busse, H.J. Ludwig, W. and Kämpfer, P.(2010). Notes on the characterization of prokaryote strains for taxonomic purposes International *Journal of Systematic Evolution in Microbiology*, 60 (10), 249-266.

Valones, M,A.A., Guimarães, R.L., Brandão, L.A. C., de Souza, P. R. E., Carvalho, A.A. T., and Crovel, S. (2009) Principles and applications of polymerase chain reaction in medical diagnostic fields: a review. *Brazilian Journal of Microbiology,* 40:1-11

Wheeler D, and Bhagwat M. (2007). BLAST QuickStart: Example-Driven Web-Based BLAST Tutorial. In: N.H, Bergman (ed.).. *Comparative genomics*: Volumes 1 and 2. Totowa: Humana Press. Available from: https:// www.ncbi.nlm.nih.gov.

Yuan, S.1, Cohen, D.B, Ravel ,J, Abdo, Z, and Forney, L.J. (2012). Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One,*. 10 (3)137

Zhang, Y., and Skolnick, J. (2005). The protein structure prediction problem could be solved using the current PDB library. *Journal of Proceedings National Academy of Science,* USA. 102 (4): 1029-34.