Covenant Journal of Physical and Life Sciences (CJPL) Vol. 1, No. 1 (Maiden Edition), August, 2013





A Two-Phase Dynamic Programming Algorithm Tool for DNA Sequences

Oyelade Jelili & Fatumo Segun

Abstract:

Sequence alignment has to do with the arrangement of DNA, RNA, and protein sequences to identify areas of similarity. Technically, it involves the arrangement of the primary sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Similarity may be a consequence of functional, structural, or evolutionary relationships between the sequences in an alignment share a common ancestor, mismatches can be interpreted as mutations, and gaps as insertions. Such information becomes of great use in vital areas such as the study of diseases, genomics and generally in the biological sciences. Thus, sequence alignment presents not just an exciting field of study, but a field of great importance to mankind. In this light, we extensively studied about seventy (70) existing sequence alignment tools available to us. Most of these tools are not user friendly and cannot be used by biologists. The few tools that attempted both Local and Global algorithms are not ready available freely. We therefore implemented a sequence alignment tool (CU-Aligner) in an understandable, user-friendly and portable way, with click-of-a-button simplicity. This is done utilizing the Needleman-Wunsh and Smith-Waterman algorithms for global and local alignments, respectively which focuses primarily on DNA sequences. Our aligner is implemented in the Java language in both application and applet mode and has been efficient on all windows operating systems.

Keywords— sequence, global alignment, local alignment, dynamic programming and java applet.

1. INTRODUCTION

DNA sequencing is a process of determining the exact order of over 3 billion chemical building blocks i.e. the A, C, T, and G bases that make up the DNA. This process was an important scientific discovery, as before sequencing was actually done, it was assumed that DNA consisted hundreds of thousands of genes. Upon performing sequencing, though, it was realized that there are approximately 20000-25000 genes in the human DNA. It is with the foundation of sequencing that sequence alignment has relevance.

According to Batzoglou [1], sequence alignment is the analysis of the sequence of the elements of genes or proteins. It involves the arrangement of DNA, RNA, or protein sequences to identify regions of similarity between them. Such similarity may be a consequence of functional, structural, or evolutionary relationships between the sequences. If two sequences in an alignment share a common ancestor, mismatches can be interpreted as mutations. Aligned sequences of nucleotide or amino acid residues are represented as rows within a matrix. Gaps are introduced between the residues so residues with identical or similar characters are aligned in the successive columns. In situations where two sequences in an alignment share a common ancestor, mismatches can be interpreted as point mutations and gaps as indels (that is, insertion or deletion mutations) introduced in either or both of the mutations in the period within which they diverged from one another. In protein sequence alignment, the degree of similarity between amino acids occupying a particular position in the sequence can be interpreted as a rough measure of how conserved a particular region or sequence motif is among lineages. The absence of substitutions, or the presence of only very conservative substitutions (that is, the substitution of amino acids whose side chains have similar biochemical properties), in a particular region of the sequence, suggests that such a region has structural or functional importance. Though DNA and RNA nucleotide are more similar to each other than to amino acids, the conservation of base pairing still can indicate a similar functional or structural role. It is of course of some importance to note that sequence alignment can be used for non-biological sequences, such as those present in financial data or in natural language [9]. In the light of this, we observe that sequence alignment can be used to identify similarities in a series of letters and words present in human language.

2. PROBLEM STATEMENT

Very short and similar sequences can adequately be aligned by hand, but most interesting problems require lengthy, highly variable or extremely numerous sequences which cannot be aligned solely by human effort. We can see, for example that two 12 residue sequences, would require considering approximately one million alignments, and two 150 residue sequences would require considering approximately 10⁸⁸ alignments (interesting considering that 10⁷⁸ is the estimated number of atoms in the universe). Hence, human effort is thus applied in formulating algorithms to simplify and reduce such a lengthy process, and also occasionally in adjusting results to reflect patterns difficult to express algorithmically. Computational approaches to sequence alignment generally fall into two categories, these being Global and Local Sequence Alignment [2].

2.1 Pair wise Alignment

Pair wise sequence alignment methods, are used to find the best matching local or global alignments of two query sequences. It can only be utilized when considering two sequences at a time, and are often used for methods that do not require extreme precision. The three primary techniques for pair wise alignment are Dot-Matrix methods, Dynamic Programming, and Word Methods, though multiple sequence alignment techniques can be used to solve pair wise alignment cases. All three methods have their strong and weak points, but all have difficulty with highly repetitive sequences of low information content, especially where the number of repetitions differs in the two sequences to be aligned.

2.2 Smith-Waterman algorithm

The Smith-Waterman algorithm was first proposed by Temple Smith and Michael Waterman in 1981. It is a wellknown algorithm for performing local sequence alignment [7]; that is, for determining similar regions between two nucleotide or protein sequences. Instead of looking at the total sequence, the Smith-Waterman algorithm compares segments of all possible lengths and optimizes the similarity measure

2.3 Needleman–Wunsch algorithm

The Needleman–Wunsch algorithm performs a global alignment on two sequences. It is commonly used in bioinformatics to align protein or nucleotide sequences [8]. The Needleman–Wunsch algorithm is an example of dynamic programming, and was the first application of dynamic programming to biological sequence comparison.

2.4 Dynamic Programming

Global alignments, which attempt to align every residue in every sequence, are most useful when the sequences in the query set are similar and of roughly equal size. Local alignments are more useful for dissimilar sequences that are suspected to contain regions of similarity or similar sequence motifs within their larger sequence context. Very common and useful alignment techniques for the implementation of global and local sequence alignment are Needleman Wunsch and Smith Waterman algorithms [8]. These techniques are based on the dynamic programming principle.

Dynamic Programming Principle: The technique of dynamic programming is theoretically applicable to any number of sequences; however, because it is computationally expensive in both time and memory, it is rarely used for more than three or four sequences in its most basic form. This method requires constructing the ndimensional equivalent of the sequence matrix formed from two sequences, where *n* is the number of sequences in the query. Standard dynamic programming is first used on all pairs of query sequences and then the "alignment space" is filled in by considering possible matches or gaps at intermediate positions, eventually constructing an alignment essentially between each two-sequence alignment. Although this technique is computationally expensive, its guarantee of a global optimum solution is useful in cases where only a few sequences need to be aligned accurately.

The dynamic programming principle involves three (3) major steps:

- Initialization: This is the first step in dynamic programming. It involves the creation of a matrix with M+ 1 column and N+ 1 row, where M and N are the sizes of the sequences to be aligned.
- **2.** Matrix fill (scoring): This is where each cell in the created matrix is filled with calculated values.
- **3.** Trace back (alignment): The trace back step simply determines the actual alignment that results in a maximum score.

This work, utilizes the Needleman-Wunsch and Smith-Waterman algorithms, which are examples of dynamic programming, to implement the pairwise sequence alignment of DNA sequences. CU-Aligner is a standalone application but can also be deployed over the internet, thus being available to a much wider audience.

In Table 1, we show list of existing sequence alignment tools available with brief description of the feature they offer.

3. BASIC FEATURES OF CU-ALIGNER

Our tool presents the following features together. It makes the work of a biologist easier.

User-Friendly Nature- CU-Aligner presents what could otherwise be a cumbersome task by hand, in a plain and understandable format with click-of-a-button simplicity in carrying out alignments. Also its specialization in focusing on DNA sequences aids its simplicity. This is a solution to situations where multiple functionality of aligners, adversely affect understandability of such programs.

Portability- As this Aligner has been implemented in the Java language it presents the advantage of being able to be

Table 1: List of different sequence alignment tools available and their features

S/N	N	Destation	Sequence	AF	1	¥
1	Name	Description	Type	Augument Type	Autor	1 ear
	BioPerl dpAlign	programming	Both	Both + Ends-free	Y. M. Chan	2003
2	BLASTZ	Seeded pattern- matching	Nucleotide	Local	Schwartz et al.	2003
3	DNA Baser	Multi alignment	Both	Local/Global+ Post	M. Gabriel	2005
4	DIVA Dasci	Web based dat plat	Dom	processing	NL Gabilei	2005
	DNADot	tool	Nucleotide	Global	R. Bowen	1998
2	DOTLET	Java-based dot-plot tool	Both	Global	M. Pagni and T. Junier	1998
6	Ed'Nimbus	Seeded filtration	Nucleotides	Local	P. Peterlongo et al.	2006
8		Progressive/Iterative				
	Geneious	alignment; ClustalW plugin	Both	Local/ Global	A.J. Drummond et al.	2006
9		Global:Global(GG),				
	COSTADOU	Global:Local(GL)				
	GLSEARCH	statistics	Protein	Global in query	W. Pearson	2007
10		Open source Java				
	LAligner	implementation of	D 4			2005
11	JAngliei	Smith-Waterman	Both	Local	A. Moustafa	2005
	Kalign	Progressive alignment	Both	Global	T. Lassmann	2005
12		Multiple, non-				
		overlapping, local				1001
	LALIGN	algorithm as SIM)	Both	Local non-overlapping	W. Pearson	(algorithm)
13		Progressive/iterative				
	MAFFT	alignment	Both	Local/ Global	K. Katoh et al.	2005
14		Memory-optimized				
		needleman but slow				
		programming (based			I. Longden (modified	
	Matcher	on LALIGN)	Both	Local	from W. Pearson)	1999
15	MAVID	Progressive	Both	Global	N Bray and L. Pachter	2004
16		amlicit models of	bou	010011	1. Ditty and 2. Facility	2001
	MCALIGN2	indel evolution	DNA	Global	J. Wang et al.	2006
17		Seeded pattern-				
	MCALIGN2	matching	Nucleotide	Local	Schwartz et al.	2003
18		Seeded pattern-				
	MCALIGN2	matching	Nucleotide	Local	B. Ma et al.	2004
19		Denomia				1989
	MSA	programming	Both	Local/ Global	D.J. Lipman et al.	(moomed 1995)
20		Dynamic				
		programming/cluster				
21	MULTALIN	Ing Despressive deserve	Both	Local/ Global	F. Corpet	1988
-1		programming				
	Multi-LAGAN	alignment	Both	Global	M. Brudno et al.	2003
22						
	MUMmer	Suffix-Tree based	Nucleotide	Global	S. Kurtz et al.	2004
23		Web-based dot-plot				
-	MUMmer	tool	Nucleotide	Global	R. Bowen	1998
24		Stochastic partition				
		via dynamic				
	MUMmer	programming	Both	Global	U. Mückstein	2002

run on all platforms, this being possible as a result of the portable nature of the Java Language.

Understandable Output- CU-Aligner presents the result of an alignment in a manner not requiring special knowledge to understand. It also presents results as soon as they are computed, as opposed to a situation where output is sent to the user at a later time.

Open Source: CU-Aligner is offered free of charge upon request.

Flexibility- CU-Aligner presents a user with the choice of parameters to use in performing an alignment, these being the gap penalty, the match and mismatch scores.

1.04		1		1	1	
25	MUSCLE	Progressive/iterative alignment	Both	Local/ Global	R. Edgar	2004
26		Needleman-Wunsch dynamic				
27	Needle	programming	Both	Global	A. Bleasby	1999
20	Needle	tool	Both	Global	M. Pagni and T. Junier	1998
20	No. II.	aligns sequence &	Destrict	(label (hundration)	WI DI	2007
29	Needle	logarithmic and	Protein	Global (by selection)	W. L. DeLano	2007
		affine gap costs and explicit models of				
30	Ngila	Global:Global (GG),	Both	Global	R. Cartwright	2007
		Global:Local (GL) alignment with				
31	Ngila	statistics	Protein	Global in query	W. Pearson	2007
32	Ngila	Suffix-Tree based	Nucleotide	Local	S. Kurtz et al.	2001
	PatternHunter	Seeded pattern- matching	Nucleotide	Local	B. Ma et al.	2002-2004
33		Open source Java implementation of				
34	PatternHunter	Smith-Waterman Various dynamic	Both	Local	A. Moustafa M.S. Waterman and P	2005
3.5	PatternHunter	programming	Both	Local or Global	Hardy	1996
35	POA	Partial order/hidden Markov model	Protein	Local/ Global	C. Lee	2002
36		Stochastic partition function sampling				
	ProbA (also propA)	via dynamic programming	Both	Global	U. Mückstein	2002
37		Multiple, non- overlapping, local				
	ProbA (also propA)	similarity (same algorithm as SIM)	Both	Local non-overlapping	W. Pearson	1991 (algorithm)
38		Local similarity with varving gap				
30	ProbA (also propA)	treatments	Both	Local or global	X. Huang and W. Miller	1990-6
	ProbCons	ency	Protein	Local/ Global	C. Do et al.	2005
40		Iterative alignment (especially		Tool State	Y. Totoki (based on O.	1991 and
41	PRRN/PRRP	refinement) Alignment	Protein	Local/ Global	Gotoh)	later
	PSAlign	preserving non- heuristic	Both	Local/ Global	S.H. Sze, Y. Lu, Q. Yang.	2006
42		"align" command aligns sequence &				
43	PyMOL	applies it to structure Memory-optimized	Protein	Global (by selection)	W. L. DeLano	2007
		needleman but slow dynamic				
	PvMOL	programming (based on LALIGN)	Both	Local	I. Longden (modified from W. Pearson)	1999
44	B 1/01					1001
45	PyMOL	Local similarity	Both	Local	X. Huang and W. Miller	1991
46	REPuter	Suffix-Tree based	Nucleotide	Local	S. Kurtz et al.	2001
47	REPuter	indel evolution	DNA	Global	J. Wang et al.	2006
			1			
	REPuter	Seeded pattern- matching	Nucleotide	Local	Schwartz et al.	2003
48	REPuter	Seeded pattern- matching Combines DNA and	Nucleotide	Local	Schwartz et al.	2003
48	REPuter	Seeded pattern- matching Combines DNA and Protein alignment, by back translating	Nucleotide	Local	Schwartz et al.	2003 2003 (newest
48	REPuter RevTrans	Seeded pattern- matching Combines DNA and Protein alignment, by back translating the protein alignment to DNA.	Nucleotide DNA/Protein (special)	Local Local/ Global	Schwartz et al. Wernersson and Pedersen	2003 2003 (newest version 2005)
48	REPuter RevTrans	Seeded pattern- matching ONA and Protein alignment, by back translating the protein alignment to DNA. Sequence alignment	Nucleotide DNA/Protein (special)	Local Local/ Global	Schwartz <i>et al.</i> Wernersson and Pedersen	2003 (newest version 2005) 1996 (new version
48 49 50	REPuter RevTrans SAGA	Seeded pattern- matching Combines DNA and Protein alignment, by back translating the protein alignment to DNA. Sequence alignment by genetic algorithm	Nucleotide DNA/Protein (special) Protein	Local Local/ Global Local/ Global	Schwartz et al. Wemersson and Pedersen C. Notredame et al.	2003 (newest version 2005) 1996 (new version 1998) 1994 (most
48 49 50	REPuter RevTrans SAGA	Seeded pattern- matching Combines DNA and Protein alignment, by back translating the protein alignment to DNA. Sequence alignment by genetic algorithm Hidden Markov	Nucleotide DNA/Protein (special) Protein	Local Local/Global Local/Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al.	2003 2003 (newest version 2005) 1996 (new version 1998) 1994 (most recent version
48 49 50 51	REPuter RevTrans SAGA SAM	Seeded pattern- matching Combines DNA and Protein alignment, by back translating the protein alignment to DNA. Sequence alignment by genetic algorithm Hidden Markov model	Nucleotide DNA/Protein (special) Protein Protein	Local Local/ Global Local/ Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al.	2003 2003 (newest version 2005) 1996 (new version 1998) 1994 (most recent version 2002)
48 49 50 51	REPuter RevTrans SAGA SAM SEQALN	Seeded pattern- matching DNA and Protein alignment, by back translating the protein alignment to DNA. Sequence alignment by genetic algorithm Hidden Markov model Various dynamic Programming	Nucleotide DNA/Protein (special) Protein Both	Local Global Local Global Local Global Local Global Local Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Watrman and P. Hardy	2003 (newest version 2005) 1996 (new version 1998) 1994 (most recent version 2002) 1996
48 49 50 51 52	REPuter RevTrans SAGA SAM SEQALN SEQALN	Seeded partern- matching Combines DNA and Protein alignment by back translating the protein alignment to DNA. Alignment to DNA. Sequence alignment by genetic algorithm Hidden Markov model Markov Various dynamic programming	Nucleotide DNA/Protein (special) Protein Protein Both Nucleotide	Local Global Local Global Local Global Local Global Local Global Local Global Celabal	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Waterman and P. Hardy	2003 (newest version 2005) 1996 (new version 1998) 1994 (most recent version 2002) 1996
48 49 50 51 52 53	REPuter RevTrans SAGA SAM SEQALN SEQALN	Seeded pattern- matching Combines DNA and Porton Alignment to DNA Sequence alignment to DNA Sequence alignment by genetic algorithm Hidden Markov model Various dynamic proprumning Suffix-Tree based	Nucleotide DNA/Protein (special) Protein Protein Both Nucleotide	Local Global Local Global Local Global Local Global Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Wateman and P. Hardy S. Kartz et al.	2003 (newest version 2005) 1996 (new version 1998) 1994 (most recent version 2002) 1996 2004
48 49 50 51 52 53 54	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN	Seeded pattern- matching Combines DNA and Poten alignment to DNA by back transporter alignment to DNA Sequence alignment by genetic algorithm Hidden Markov model Various dynamic proprimming Suffix-Tiree based Web-based dor-plot tool	Nucleotide DNA/Protein (special) Protein Protein Both Nucleotide Nucleotide	Lecal Global Lecal Global Lecal Global Lecal Global Global Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Wateman and P. Hardy S. Kurtz et al. R. Bowen	2003 2003 (networkt version 2005) 1996 (netw version 1998) 1994 (most recent 2002) 1996 2004 2004
48 49 50 51 52 53 54 54	REPuter RevTrans SAGA SAGA SEQALN SEQALN SEQALN SEQALN SEM	Seeded pattern- matching Combines DNA and Protein alignment to DNA by back translating the second second second by genetic algorithm by genetic algorithm Hidden Markov model Various dynamic programming Suffix-Tree based Web-based dor-plot tool Local seminarity	Nucleonde DNA/Protein (special) Protein Protein Both Nucleonde Both Both	Lecal Global Lecal Global Lecal Global Local Global Global Global Lecal	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller	2003 2003 (newest version 2005) 1996 (new version 1998) 2002) 2004 1996 2004 1998
48 49 50 51 52 53 54 55	REPuter RevTraus SAGA SAGA SEQALN SEQALN SEQALN SEM	Seeded pattern- matching Combines DNA and Protein alignment of by back translating the alignment of DNA sequence alignment of DNA sequence alignment by genetic algorithm Hidden Markov model Various dynamic programming Suffix-Tree based Web-based doeplot tool Local similarity costs and affine gap costs and	Nucleonde DNA/Protein (special) Protein Protein Both Nucleonde Both	Local Global Local Global Local Global Local of Global Global Global Local	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller	2003 2003 (newest version 2005) 1996 (new version 1998) 2004 1996 2004 1998 1991
48 49 50 51 52 53 54 55	REPuter RevTrans SAGA SAM SEQALN SEQALN SDM SDM SDM	Seeded partern- matching Combines DNA and Portein alignment, by back translating the protein alignment to DNA. Sequence alignment by generic algorithm Hidden Markov model Markov Markow dynamic programming Suffix-Tree based Web-based dor-plot tool Local similarity logarithmic and explicit models of explosition	Nucleonde DNA-Protein (special) Protein Protein Both Nucleonde Both Both Both	Local Global Local Global Local Global Local or Global Global Local Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright	2003 2003 (netwest version 1995 1996 2004 1998 1991 2007
48 49 50 51 52 53 54 55 56	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN SIM SIM	Seeded pattern- matching Combines DNA and Portein alignment, by back translating the protein alignment to DNA. Sequence alignment by genetic algorithm model Various dynamic programming Suffix-Tree based Web-based dor-plot tool. Local similarity logarithmic and explicit models of midel evolution Global Local (GL)	Nucleonde DNA-Protein (special) Protein Both Nucleonde Both Both Both	Local Global Local Global Local Global Local of Global Global Global Global Global Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright	2003 2003 (netwest version 1995 1996 2004 1998 1991 2007
48 49 50 51 52 53 54 55 56	REPuter RevTrans SAGA SAGA SAGA SEQALN SEQALN SEM SIM SIM	Seeded pattern- matching Combines DNA and Portein alignment by back translating the protein alignment to DNA. Sequence alignment by generic algorithm model Various dynamic programming Suffix-Tree based Web-based doeplot neel Local similarity local similarity clobal Local (GL) alignment with statistics	Nucleonde DNA-Protein (special) Protein Protein Both Nucleonde Both Both Protein Both Protein Both	Local Global Local Global Local Global Local of Global Global Global Global Global Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M. S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson	2003 2003 (newest version 2003) 1996 (new version 1998) 1994 (noss recent version 2002) 1996 2004 1998 1999 2004 2004 2004 2004 2004 2004 2004 2004 2005 2004 2005 2007 2007
48 49 50 51 52 53 54 55 56 57	REPuter RevTrans SAGA SAGA SAM SEQALN SEQALN SEQALN SIM SIM SIM SIM SIM SIM SIM	Seeded pattern- matching Combines DNA and Potrim alignment by back translating the protein alignment to DNA. Sequence alignment by genetic algorithm Hidden Markov model Various dynamic programming Methods and the second web-based dot-plot tool Local similarity Local similarity Local similarity Colour of the second affine app costs and affine a	Nucleonde DNA Protein (special) Protein Both Nucleonde Both Both Both Protein	Local Global Local Global Local Global Local Global Global Global Global Global Global Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson	2003 2003 (nerver) version 2005) 1996 (nerv version 1998) 1998 1998 1998 2002 2004 1998 1998 1999 2004 2004 2004 2007 2007
48 49 50 51 52 53 54 55 56 57 57 58	REPuter RevTrans SAGA SAGA SEQALN SEQALN SEQALN SIM SIM SIM SIM SIM SIM SIM SI	Seeded pattern- matching Combines DNA and Potein alignment by back translating the protein alignment to DNA. Sequence alignment by genetic algorithm Hidden Markov model Various dynamic programming Suffix-Tree based Web-based dorplot tool Local similarity Ingentomic and affine app costs and affine app c	Nucleonde DNA-Protein (upecial) Protein Both Nucleotide Both Both Both Protein Both Both Both	Local Global Local Global Local Global Local Global Local of Global Glob	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Kroph et al. M.S. Waterman and P. Hardy S. Kartz et al. R. Bowen X. Huang and W. Miller W. Pearson X. Huang and W. Miller	2003 2003 (nerver) version 2005) 1996 (new version 1998) 1994 (most recent version 2002) 1996 2004 1998 1991 2007 2007 2007 1996
48 49 50 51 52 53 54 55 56 57 58	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN SIM	Seeded pattern- matching Combines DNA and Porton alignment. Doron alignment to DNA Sequence alignment to DNA Sequence alignment by genetic algorithm model Various dynamic programming Suffix-Tree based Web-based dorplot teol Local similarity Ioparthmic and affine pap costs and affine to the second colour of the second model evolution Global Global (GO). Global Local (GD). Local similarity with varying gap treatments Needlemaw Funcch dynamic	Nucleonde DNA Protein (special) Protein Both Nucleonde Both Both Protein Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local Global Local or Global Global Global Global Global Global Global Global Global Clocal or global Clocal Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Waternan and P. Hardy S. Kutz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A. Blauchen	2003 2003 (nervent ervice) 2005) (1996 (nev version 1996 1998 1994 (nov recent version 2002) 1996 1998 1991 2007 2007 1996 1996 1996 1996 1996 1996 1996 199
48 49 50 51 52 53 54 55 56 57 58 59	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN SIM	Seeded pattern- matching Combines DNA and Poorn a high-most. by back translation and protein alignment to DNA Sequence alignment by genetic algorithm Hidden Markov model Various dynamic programming Suffix-Tree based Web-based dorplot tool Local similarity Digarithmic and affine gap costs and affine fage costs and affine gap costs and affine ga	Nucleonde DNA Protein (special) Protein Both Nucleonde Both Both Both Both Both	Local Global Local Global Local Global Local Global Global Global Global Global Global Global Global Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Wateman and P. Hardy S. Kartz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A. Bleasby	2003 2003 (network retricen 2005) 1996 (new 1998) 1994 (most recent version 2002) 1996 2004 1998 1991 2007 2007 2007 1996 1999
48 49 50 51 52 53 54 55 56 57 58 59 60	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN SIM	Seeded pattern- matching Tratching Combines DNA and Portein alignment, by back translating the protein alignment b DNA. Sequence alignment by genetic algorithm Midden Markov Markow dynamic programming Local similarity logarithmic and explicit models of midel evolution. Clobal Local (GL) alignment with autistics Local similarity pro- gramming Java-based doeplor tool.	Nucleonde DNA-Protein (special) Protein Both Nucleonde Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local of Global	Schwarz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M. S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A. Bleasby M. Pagni and T. Junier	2003 2003 (network version 2005) 1996 (new version 1996 1998 1994 (most recent version 2002) 1996 2004 1998 1991 2007 2007 1996 1999 1999 1999 1999 1999 1999 199
48 49 50 51 52 53 54 55 56 57 58 59 60	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN SIM SIM SIM SIM SIM SIM SIM SIM SI	Seeded pattern- matching Twatching Portein alignment, by back translating the protein alignment bDNA. Sequence alignment by genetic algorithm model Various dynamic programming Suffix-Tree based Web-based dor-plot tool Local similarity logarithmic and explicit models of midel evolution Global Local (GL) alignment with drusting gap treatments. Local similarity with Avaring gap treatments.	Nucleonde DNA-Protein (special) Protein Both Nucleonde Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local of Global Global Global Global Global Global Global Global Both	Schwarz et al. Wernersson and Pedersen C. Notredame et al. A Keogh et al. A Koogh et al. M.S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A Bleasby M. Pagni and T. Junier L. Bloksberg	2003 2003 (newest version 2005) 1996 (new version 1996 1998 1994 2002 1996 2004 1998 2007 2007 1996 1999 1999 2007
48 49 50 51 52 53 54 55 56 57 58 59 60 61	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN SIM SIM SIM SIM SIM SIM SIM SI	Seeded pattern- matching Twatching Portein alignment by back translating the protein alignment to DNA. Sequence alignment by genetic algorithm model Various dynamic programming Suffix-Tree based Web-based doeplot tool Local similarity logarithmic and explicit model evolution Global Cocal (GL) alignment with attitutics Local similarity grap tratimets Needlemax-Wunch dynamic programming Java-based doeplot tool Ultra-fast blocked alignment (Statistics)	Nucleonde DNA-Protein (special) Protein Both Nucleonde Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local of Global	Schwarz et al. Wernersson and Pedersen C. Notredame et al. A. Koogh et al. A. Koogh et al. M. S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A. Bleasby M. Pagni and T. Junier L. Bloksberg	2003 2003 (newest version 2005) 1996 (new version 1996 1998 1994 2002 1996 2004 1999 2007 2007 1996 1999 1998 2004
48 49 50 51 52 53 54 55 55 56 57 58 59 60 61	REPuter RevTrans SAGA SAGA SAAM SEQALN SEQALN SEQALN SIM	Seeded pattern- matching Combines DNA and Potein alignment by back translation by genetic algorithm of the protein alignment to DNA. Sequence alignment by genetic algorithm of the second second second second second second second second second second second Web-based dot-plot teol Local similarity local similarity local similarity clobal Local (GD) alignment with satistics	Nucleonde DNA Protein (percial) Protein Both Nucleonde Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local Global Local Global Local	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Koogh et al. M. S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A. Bleasby M. Pagni and T. Junier I. Bloksberg W. Pearson	2003 2003 (nervest version 2005) 1996 (nerv version 1998) 1994 (noost recent version 2002) 1996 2004 1998 1991 2007 2007 2007 1996 1999 1999 1999 1999 1998 2004 1981
48 49 50 51 52 53 54 55 55 55 55 55 58 59 60 61 62	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN SIM	Seeded pattern- matching Combines DNA and Potrim alignment to DNA. Potrim alignment to DNA. Sequence alignment by genetic algorithm model Various dynamic programming Seguence alignment with the second second programming Local similarity loganthmic and affine app costs and affine ap	Nucleonde DNA Protein (special) Protein Both Nucleonde Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local of Global Glo	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M. S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A. Bleasby M. Pagni and T. Junier L. Bloksberg W. Pearson I. Lengden (modhfied from G. Myres and W.	2003 2003 (nerver) version 2005) 1996 (nev version 1998) 1994 (most recent version 2002) 1996 2004 1998 1991 2007 2007 2007 1996 1999 1998 1999 1998 2004 1981
48 49 50 51 52 53 53 54 55 55 56 57 58 59 60 61 62 62 62	REPuter RevTrans SAGA SAGA SAGA SEQALN SEQALN SEQALN SEQALN SIM SIM SIM SIM SIM SIM SIM SI	Seeded partern- matching Combines DNA and Potein alignment by bock translation to protein alignment to DNA. Sequence alignment by generic algorithm model Various dynamic programming Suffix-Tree based Suffix-Tree based Suffix-Tre	Nucleonde DNA-Protein (upecial) Protein Both Nucleotide Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Waternan and P. Hardy S. Kartz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A. Bleasby M. Pagni and T. Junier L. Bloksberg W. Pearson I. Longden (modified from G. Myers and W. Miller)	2003 2003 (nerver) version 2005) 1996 (nev version 1998) 1994 (most recent version 2002) 1996 2004 1998 1991 2007 2007 2007 1996 1999 1999 2004 1981 1999
48 49 50 51 52 53 55 55 55 55 55 55 57 58 59 60 61 62 63	REPUTER SAGA SAGA SEQALN SEQALN SEQALN SEM SIM SIM SIM SIM SIM SIM SIM SIM SIM SI	Seeded pattern- matching Tratefaing Protein alignment, by back translating the protein alignment to DNA. Sequence alignment by genetic algorithm by genetic algorithm. Hidden Markov model Markov model Markov model Markov Markov dynamic programming Java-based dorplot tool Local similarity alignment with alignment with ali	Nucleonde DNA Protein (upecial) Protein Both Nucleonde Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local of Global Local Global Local Loca Locad Local Local Local Loca Locad Local Locad Lo	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A. Krogh et al. M.S. Waternan and P. Hardy S. Kattz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A. Bleasby M. Pagni and T. Junier L. Bloksberg W. Pearson I. Leagden (modified from G. Myers and W. Miller) C. Numedami in if	2003 2003 (nervent (nervent 2005) 1996 (nev version 1998) 1994 (nov recent version 2002) 2004 1998 1991 2007 2007 1999 1999 1999 2004 1981 1999 2004
48 49 50 51 52 53 54 55 57 58 59 60 61 62 63 64	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN SEM	Seeded partern- matching Transching Protein alignment by back translating the protein alignment b DNA. Sequence alignment by genetic algorithm Hidden Markov model Markov Markov dynamic programming Local similarity logarithmic and explicit models of mide evolution Global Global (GGL) alignment with autistics Local similarity vith dynamic programming Java-based doeplot tool Units fast blocked alignment with adigment adigment with adigment adigment with adigment adigment adigment with adigment adigment adigment adigment adigment adigment adigment adigment adigment adigment adigment adigment adigment adigment adigment adigment adi	Nucleonde DNA-Protein (special) Protein Both Nucleonde Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local Global Local Global Local Global Local Global	Schwarz et al. Wernersson and Pedersen C. Notredame et al. A Keogh et al. M.S. Waterman and P. Hardy S. Kutz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A. Bleasby M. Pagni and T. Junier L. Bloksberg W. Pearson I. Longde (modified from G. Myers and W. Miller)	2003 2003 (newest version 2005) 1996 (new version 1996 1998 1994 2002 1996 2004 1998 1991 2007 2007 1996 2007 1999 1999 1999 1999 2004 1981 1981 1999 2000
48 49 50 51 52 53 54 55 57 58 59 60 61 62 63 64	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN SEM SIM SIM SIM SIM SIM SIM SIM SIM SIM SI	Seeded partern- matching matching Protein alignment. by back translating the protein alignment to DNA. Sequence alignment by genetic algorithm model Various dynamic programming Suffix-Tree based Web-based dot-plot tool Local similarity logarithmic and explicit models of mide evolution Global-Local (GL) alignment with attaines: with Maternav Winch dispared dot-plot tool Ultra-fast blocked of clobal-Local (GL) alignment with attaines: with attaines: with alignment Local similarity with Astronomes and the second programming programming programming More sensitive alignment Aligns molece acids alignment Aligns molece acids alignment Aligns molece acids sequence agina and alignment Aligns molece acids sequence agina acids alignment Aligns molece acids sequence agina acids alignment	Nucleotide DNA/Protein (special) Protein Both Nucleotide Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local of Global Local Global Local Global Local Global Both Local Global Local Global Local Global NA	Schwarz et al. Wernersson and Pedersen C. Notredame et al. A Keogh et al. A Keogh et al. M.S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A Bleasby M. Pagni and T. Junier L. Bloksberg W. Pearson L. Bloksberg W. Pearson L. Bloksberg C. Notredame et al. G. Wijfiams (modified from G. Myers and W. Miller)	2003 2003 (newest version 2005) 1996 (new version 1996 1998 1994 2002 1996 2004 1998 2007 1996 2007 1999 1999 1999 1999 2007 1999 2004 1999 2004 2004 2004 2004 2004 2004 2004 2
48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65	REPuter RevTrans SAGA SAM SEQALN SEQALN SEQALN SIM	Seeded pattern- matching matching Protein alignment by back translation by back translation by genetic algorithm by genetic algorithm by genetic algorithm model Various dynamic programming Suffix-Tree based Web-based dor-plot tool Local similarity logarithmic and explicit models of mide evolution Global Global (GGI) dipant with statistics Local similarity with uranistic gramment Meedemas Wunch dynamic programment Alignment Meedemas with statistics Local (Smith- Waterman alignment Meedemas with statistics Local (Smith- Waterman) alignment Mean company Meedemas with statistics Local (Smith- Waterman) alignment Mean company Meedemas with statistics Local (Smith- Waterman) alignment Mean company Meedemas with statistics Local (Smith- Waterman) alignment Mean company Meedemas with statistics Local (Smith- Waterman) Align model and alignment Mean company Meedemas with statistics Align model and alignment Mean company Meedemas with statistics Align meeter and company Meedemas with statistics alignment Mean company Meedemas with statistics alignment Mean company Meedemas with statistics Align meeter and company Meedemas with statistics Align meeter and company Meedemas with statistics alignment Mean company Meedemas with statistics alignment	Nucleonde DNA-Protein (special) Protein Both Nucleonde Both Both Both Both Both Both Both Both	Local Global Local Global Local Global Local of Global Global Global Global Global Global Global Global Local Global Global Local Global Local Global Local Global	Schwartz et al. Wernersson and Pedersen C. Notredame et al. A Koogh et al. A Koogh et al. M.S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A Bleasby M. Pagni and T. Junier L. Bloksberg W. Pearson I. Longden (modified frig G. Myres and W. Miller) C. Notredame et al. G. Williams (modified	2003 2003 (newest version 2005) 1996 (new version 1996 1998 1994 2002 1996 2004 1991 2007 2007 1996 1999 1998 2004 1981 1999 2004 2004 2004 2004 2004 2004 2004 2
48 49 50 51 52 53 54 55 55 56 57 58 59 60 61 62 63 64 65 66	REPuter RevTrans SAGA SAAM SEQALN SEQALN SEQALN SEQALN SIM	Seeded partern- matching matching Protein alignment by back translation by back translation by genetic algorithm by genetic algorithm collection collection by algorithm collection	Nucleonde DNA Protein Protein Protein Both Nucleonde Both Both Both	Local Global Local Global Local Global Local of Global Global Global Global Global Global Global Global Global Local Global Local Global Local Global Local Global Local Global Local Clobal Clocal Clocal Clobal Clobal Clocal Clobal Clocal Clobal Clocal Clobal Clobal Clocal Clobal Clocal Clobal Clocal Clobal Clocal Clobal Clobal Clocal Clobal Clocal Clobal Clocal Clobal Clocal Clobal Clobal Clocal Clobal Clocal Clobal Clocal Clobal Clobal Clocal Clobal Clocal Clobal Clocal Clobal Clocal Clobal Clobal Clocal Clobal Clocal Clobal	Schwarz et al. Wernersson and Pedersen C. Notredame et al. A. Koogh et al. M. S. Waterman and P. Hardy S. Kurtz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller A. Bleasby M. Pagni and T. Junier L. Bloksberg W. Pearson I. Lengden (modified from G. Myers and W. Miller) C. Notredame et al. G. Williams (modified from B. Pearson) A. Bleasby	2003 2003 (newest yersion 2003) 1996 (new yersion 2002) 1996 2004 1998 2004 1998 2007 2007 2007 2007 1996 2007 1996 2007 1998 2004 1998 2004 1998 2004 1999 2004 2002 2002 2002 2002 2002 2002 2
48 49 50 51 52 53 54 55 55 55 55 55 55 57 58 59 60 61 62 63 64 64 65 66 66	REPuter RevTrans SAGA SAGA SAGA SEQALN SEQALN SEQALN SEQALN SIM SIM SIM SIM SIM SIM SIM SI	Seeded pattern- matching matching Portein alignment by back translation by back translation by genetic algorithm of the protein alignment to DNA. Sequence alignment by genetic algorithm of the programming Suffix-Tree based Web-based dot-plot teol Local similarity Local similarity Local similarity Local similarity Celobal Local (Gb) alignment with attistics Celobal Local (Gb) alignment with statistics Uthra-based dot-plot teol Uthra-based dot-plot teol Local (Smith- Waterman) alignment with statistics	Nucleonde DNA Protein Protein Protein Both Nucleotide Both Both Both	Local Global Local Global Local Global Local Global Cocal Global Local Global Roth Local Local Global Roth Local Global Roth Local Roth Roth Roth Roth Roth Roth Roth Roth	Schwarz et al. Wernersson and Pedersen C. Notredame et al. A. Koogh et al. M. S. Waternan and P. Hardy S. Kurz et al. R. Bowen X. Huang and W. Miller R. Cartwright W. Pearson X. Huang and W. Miller R. Cartwright W. Pearson I. Engles (modified from G. Myrs and W. Miller) C. Notredame et al. G. Williams (modified from B. Pearson) A. Bleasby I. Longden	2003 2003 (nervest version 2005) (nervest 2005) 1996 (nerv version 1998) 1994 (nost recent version 2002) 1996 2004 1998 2007 2007 2007 2007 1996 1999 2007 1996 1999 2004 1981 1999 2004 1981 1999 2000 2002 2002 2002 2002 2002 200

4. MATERIALS AND METHODS

CU-Aligner was implemented in the Java language in both application and applet mode. Deployment on to a web page can be done with simple coding statements once, the web page is built using Macromedia Dreamweaver.

5. DESIGN OF CU-ALIGNER

CU-Aligner is designed based upon the Needleman-Wunsch and Smith-Waterman algorithms (Wunsch and Waterman 1970), for Global and Local alignments, respectively. The aligner program implements each stage of each algorithms implementation: These being, the Initialization, Matrix-Fill and Trace back Stages as depicted in the figure 1, 2 and 3. The flow of the CU-Aligner is given is figure 4.

		G	Α	Α	Т	Т	С	А	G	\mathbf{T}	Т	Α
	0	0	0	0	0	0	0	0	0	0	0	0
G	0											
G	0											
А	0											
т	0											
С	0											
G	0											
А	0											

Figure 1: Initialization

The first step in the global alignment dynamic programming approach is to create a matrix with M + 1 columns and N + 1 rows where M and N correspond to the size of the sequences to be aligned.

		G	Α	Α	Т	Т	С	A	G	Т	Т	A
	0	0	0	0	0	0	0	0	0	0	0	0
G	0	1	1	1	1	1	1	1	1	1	1	1
G	0	1	1	1	1	1	1	1	2	2	2	2
А	0	1	2	2	2	2	2	2	2	2	2	3
т	0	1	2	2	3	3	3	3	3	3	3	3
С	0	1	2	2	3	3	3	4	4	4	4	4
G	0	1	2	2	3	3	3	4	4	5	5	5
А	0	1	2	3	3	3	3	4	5	5	5	6

Figure 2: Matrix-Fill

This is where each cell in the created matrix is filled with calculated values.

In order to find $M_{i,j}$ for any i,j it is necessary to know the score for the matrix positions to the left, above and diagonal to i, j. In terms of matrix positions, it is necessary to know $M_{i-1,j}$, $M_{i,j-1}$ and $M_{i-1,j-1}$. i.e.

$$G[i, j] = \max \begin{cases} G[i - 1, j - 1] + s(x_i, y_j) \\ G[i - 1, j] + g \\ G[i, j - 1] + g \end{cases}$$
$$G[0, j] = j \cdot g, \quad j \in [0, \dots, n]$$
$$G[i, 0] = i \cdot g, \quad i \in [0, \dots, m]$$

		G	А	А	т	т	С	А	G	т	Т	А
	0	0	0	0	0	0	0	0	0	0	0	0
G	0	1	1	1	1	1	1	1	1	1	1	1
G	0	1	1	1	1	1	1	1	2	2	2	2
А	0	1	1	2	2	2	2	2	2	2	2	3
т	0	1	2	2	3	3	3	3	3	3	3	3
С	0	1	2	2	3	3	4	4	4	4	4	4
G	0	1	2	2	3	3	4	4	5	5	5	5
A	0	1	2	3	3	3	4	5	5	5	5	₽ 6

Figure 3: Traceback

The traceback step determines the actual alignment(s) that result in the maximum score. Note that with a simple scoring algorithm such as one that is used here, there are likely to be multiple maximal alignments. The traceback step begins in the M,J position in the matrix, i.e. the position that leads to the maximal score. In this case, there is a 6 in that location.

Traceback takes the current cell and looks to the neighbor cells that could be direct predecessors. This means it looks to the neighbor to the left (gap in sequence #2), the diagonal neighbor (match/mismatch), and the neighbor above it (gap in sequence #1). The algorithm for traceback chooses as the next cell in the sequence one of the possible predecessors. They are all also equal to 5.

Since the current cell has a value of 6 and the scores are 1 for a match and 0 for anything else, the only possible predecessor is the diagonal match/mismatch neighbor.

If more than one possible predecessor exists, any can be chosen. This gives us a current alignment of :

(Seq #2) A

Continuing on with the traceback step, we eventually get to a position in column 0 row 0 which tells us that traceback is completed. One possible maximum alignment is :

And an alternate derived from the same procedure:

G _ A A T T C A G T T A | | | | | | |

G G _ A _ T C _ G _ _ A

There are more alternative solutions each resulting in a maximal global alignment score of 6. Since this is an

exponential problem, most dynamic programming algorithms will only print out a single solution

The same principle is followed in the implementation of the Smith-Waterman algorithm, except that the matrix is filled on the following basis:

$$L[i,j] = \max \begin{cases} L[i-1,j-1] + s(x_i, y_j) \\ L[i-1,j] + g \\ L[i,j-1] + g \\ 0 \\ L[0,j] = 0, \quad j \in [0, \dots, n] \\ L[i,0] = 0, \quad i \in [0, \dots, m] \end{cases}$$



Figure 4: Flowchart of Pairwise Aligner

6. RESULTS

The pairwise aligner accepts two DNA sequences (see figure 5) and produces their global and local alignments, as well as their scoring matrix (see figure 6 and 7 respectively). A deployable version if CU-Aligner on the internet is shown in figure 8.

Applet Viewer: sealapp2	
Appiet	
Secturence Alignment	
control and a second	
Sequence Algument	Output
Input Sequence One in FASTA format:	
incrediptopanes	
Input Sequence Two in FASTA format: reagingate	
Sauce Director	Qutans
	Clear Output View in New Window Print
Score: 2 v Gap Penalty 2 v Mismatch: 1 v	
Align Galoid Align Local 4	
5 👻	(c) Designed By Dureke Paul and Fadahunsi Damilola
Applet stated.	

Figure 5: Pairwise Aligner Interface

Sequence Alignment			
Sequence Alignment	Output		
Input Sequence One in FASTA format:	GLOBAL ALIGNMENT		
Instructions	2 -2 -2 -2 2 -3 -6 -8 2 -3 -7 -2 2 -3 -4 -2 2 -4 -2 2 -4 -2 2 -4 -2 2 -4 -4 2 -2 -4 2 -4 -4 2 -2 -4 2 -4 -4 2 -4 -4 2 -2 -4 2 -4 -4 2 -2 -4 2 -	1.00 + 0 + 2 + 2 0 0 0	
Sequence Properties:	Options		
Score: 2 v Gap Penalty 2 v Mismatch: 1 v	Clear Output View	n in New Window	Print
Align Gisbal Align Local Preset			

Figure 6: Local Alignment

plet Viewer: sealapp2							00
Sequence Alignment							
Sequence Alignment	Output						
Input Sequence One in FASTA format:	LOCAL ALIONNEY	т					
accatgatgasace	-1 -1 -1 0 -1 0 -1 1 -1 0	-1 1 2 1 0	-1 1 3 2 1	-1 0 2 5 4	-1 D 1 4	-1 0 3 6	0 2 5
kput Sequence Two in FASTA format:	1 0	0	0 0 0	3 2 1 0 2 2	6 5 4 3 2	5 8 7 6 5	577687
	CCA_T0_T_AA CCA_T0_T_AA Local Alignment S	cone: 12					
Sequence Properties:	Options						
	Clear Output		Vice	in New Window			Print
Score 2 Gap Penalty 1 Manualch: 1							
Align Global Align Local Reset							
	(c)	Designed By Dureke F	Paul and Fadahu	nsi Damilola			

Figure 7: Local Alignment



Figure 8: Applet Deployed On Web Page

A user, on accessing the pairwise aligner, enters in both sequences in the appropriate fields. Parameters are then selected so as to carry out the alignment as the user desires in terms of the match score, gap penalty and mismatch score.

REFERENCES

- Flaninick J, Batzoglou S. Using multiple alignments to improve seeded local alignment algorithms. *Nicleic Acids Research*, 33(14), pp. 4563-4577, 2005.
- [2]. K. Somasundaram and S. Radhakrishnan: "Nimble Protein Sequence Alignment in Grid (NPSAG)", Journal

of Computer Science 4 (1), pp. 36-41, 2008.

- [3] Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R.C., and Miller, W., "A Web server for aligning two genomic DNA sequences". *Genome Res.* 10, pp. 577-586, 2000.
- [4] Junier, T. and Bucher, P., "A Java applet for browsing molecular sequence data". *In Silico Biol.* 1, pp. 13-20, 1998.
- [5] T. Junier, *M. Pagni* and P. Bucher, mmsearch, "A motif arrangement language and the Eukaryotic Promoter Database (EPD)". *Nucleic Acids Res*, 26, 353-357, 1998.
- [6] P. Peterlongo, N. Pisanti, F. Boyer, A. P. do Lago, and M.-F. Sagot, "Lossless filter for multiple repetitions with Hamming distance". *Journal of Discrete Algorithms*, 19,
 - pp. 71-87, 2008.
- Smith T.F, Waterman M.S, "Identification of Common Molecular Subsequences". *Journal of Molecular Biology*, 147, pp. 195–197, 1981.
- [8] Needleman SB, Wunsch CD, "A general method applicable to the search for similarities in the amino

On clicking the global alignment button, a check on inputted sequences is made to ensure compliance of the inputted sequences to DNA sequence structure (i.e. consisting alphabets A, C, G, and T). Once this check is cleared the two sequences are globally aligned.

In the case of local alignment, the same process takes place, except that the output is the local alignment of the two sequences.

7. CONCLUSION

Sequence alignment can be a useful technique for studying molecular evolution and analyzing sequence-structure relationships. We present a sequence alignment tool in an understandable, user-friendly and portable way, with clickof-a-button simplicity. This is done utilizing the Needleman-Wunsh and Smith-Waterman algorithms for global and local alignments, respectively which focuses primarily on DNA sequences. Our aligner is implemented in the Java language in both application and applet mode and has been tested working efficiently on all windows operating systems.

acid sequence of two proteins". *Journal of Molecular Biology* 48 (3), 443-53, 1970.

- [9] Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S. "Glocal alignment: finding rearrangements during alignment". *Bioinformatics 19* Suppl 1: i54–6, (2003.
- [10] Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. Human-mouse alignments with BLASTZ. Genome Res. 2003;13:103– 107
- [11] Lassmann T, Sonnhammer EL. Quality assessment of multiple alignment programs. FEBS Lett. 2002;529:126–130.doi:10.1016/S0014-5793(02)03189
- [12] Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–3066
- [13] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet.

2000;16:276-27

- [14] Bray N, Dubchak I, Pachter L. AVID: A Global Alignment Program. Genome Res. 2003;13:97–102
- [15] Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Program NC, Green ED, Sidow A, Batzoglou S. LAGAN and Multi-LAGAN: Efficient Tools for Large-Scale Multiple Alignment of Genomic DNA. Genome Res

About the Authors: Oyelade Jelili and Fatumo Segun: Department of Computer and Information Sciences College of Science and Technology, Covenant University, Ota. Email: Ola.oyelade@covenantuniversity.edu.ng, segunfatumo@yahoo.co.uk