## Multiple sequence alignment by consensus

# Michael S.Waterman

Departments of Mathematics and Molecular Biology, University of Southern California, Los Angeles, CA 90089-1113, USA

Received 16 September 1986; Revised and Accepted 20 October 1986

### ABSTRACT

An algorithm for multiple sequence alignment is given that matches words of length and degree of mismatch chosen by the user. The alignment maximizes an alignment scoring function. The method is based on a novel extension of our consensus sequence methods. The algorithm works for both DNA and protein sequences, and from earlier work on consensus sequences, it is possible to estimate statistical significance.

### INTRODUCTION

Sequence alignment is an important problem motivated by molecular biology and many computer algorithms have been devised to accomplish alignments. Most of the results have been for two sequences. The analytical work began with the paper of Needleman and Wunsch(1) who solved the problem of maximum similarity alignment for two sequences. Later Sellers(2) gave a related dynamic programming algorithm for minimum distance alignment of two sequences. These algorithms were extended to cover multiple insertions and deletions by Waterman et al.(3). In Waterman (1984)(4) the subject of sequence comparisons is It is clear that, while the two sequence case has reviewed. been adequately solved for many cases, the situation for alignment of more than two sequences is quite different. Next we review the history of approaches to multiple sequence Then we give a new algorithm for multiple sequence alignment. alignment that is based on a novel extension of our consensus sequence methods.

Sankoff(5) gave the first treatment of which we are aware that considered multiple sequence alignment. His method requires a tree relating the sequences and employs dynamic programming as well as parsimony. For comparing R sequences of length N, the method takes time proportional to  $2^{R}N^{R}$  (O( $2^{R}N^{R}$ )) and storage O( $N^{R}$ ). Only small sequences are practical even when R-3. Sankoff et al.(6) apply the algorithm, modified somewhat, to 5S sequences. In Waterman et al.(3) a similar algorithm is presented which does not assume a tree. The time and space requirements are similar.

More recently, Waterman and Perlwitz(7), apply the geometry of geodesics to sequence alignment. The idea of sequence is extended so that a "nucleotide" is a mixture of A,C,G,T and  $\phi$ (deletions). The basic algorithm is based on dynamic programming, and the sequences are aligned and merged two at a time. When the relationships between the sequences are understood, as when a correct evolutionary tree is available, the method works very well. Still it is a pairwise algorithm and the final alignment is dependent on the order in which the sequences are processed.

Computer science has studied what they call string matching problems, beginning with Wagner and Fischer (they prefer "string" to sequence)(8). In Itoga(1981)(9) the string merging problem is studied, and Hsu and Du(1984)(10) consider the longest common subsequence of a set of strings. None of these authors seem aware of the earlier results described above.

Sobel and Martinez(11) approach sequence matching as a regions problem, where their algorithm is based on locating all exact repeats of patterns which occur in the sequence set. The method of finding repeats takes O(NlogN) operations(12). The best set of regions making up the alignment is found by longest path methods from computer science. This is perhaps the only practical method available for more than three sequences until the present paper.

Some interesting results have been obtained for protein sequences. There it is thought that gaps should receive a constant penalty, regardless of length. Fredman(13) improved the algorithms to  $O(N^3)$  for aligning 3 sequences of length n with the constant gap penalty, whereas the direct extension of Waterman et al. is  $O(N^5)$ . Murata et al.(14) give a similar improvement. Johnson and Doolittle(15) give a method, not guaranteed to be globally optimal, which is based on the progressive evaluation of selected segments from each sequence.

There are two recent methods designed to align multiple sequences that require more discussion. That of Sobel and Martinez(11) is based on aligning segments common to the sequences; it is frequently of interest to align patterns with some degree of mismatch, insertion, and/or deletion. There may be no segment common to all the sequences. The method of Johnson and Doolittle(15) is based on aligning segments found within a window of width W and for R sequences of length N has running time proportional to  $O(R(N-W)W^{R-1})$ . For small R this is a practical method, but is not practical for a larger number of sequences. Next we present our method to overcome some of these difficulties.

# DESCRIPTION OF THE METHOD

We begin with a set of R sequences of length N

These sequences can be taken to be initially aligned on some biologically determined feature. The alignment is, of course, unknown except approximately. It is the purpose of this paper to give an algorithm aligning the sequences by matching or aligning on words of a given size. The usual methods of sequence analysis align on single letters, that is words of length 1.

A concept basic to our algorithm is that of consensus word. The definition has been given in earlier work (Waterman et al. (1984)(16) and Galas et al.(1985)(17)) and will be briefly given here. First, take a fixed word size(length) k and a word w of length k. There are  $4^{k}$  such words in DNA and  $20^{k}$  in proteins. Next, define the window width W. This parameter gives the width of sequence in which a word can be found and thus defines the amount of shifting allowed in matching consensus words. The sequences starting at column j+1 with window width W appear as

<sup>a</sup>1, j+1<sup>a</sup>1, j+2 <sup>...</sup> <sup>a</sup>1, j+W <sup>a</sup>2, j+1<sup>a</sup>2, j+2 <sup>...</sup> <sup>a</sup>2, j+W ... <sup>...</sup> <sup>...</sup>

First, we search the first sequence of the window for matches to our word w. An exact match to w is called a d-0 neighbor while a 1-letter mismatch from w is called a d-1 neighbor, and so on. It is possible to include insertions and deletions in this list of neighbors. We may decide, e.g., to limit the amount of mismatch to d-0,1,2 and not find w in a portion of sequence unless it is within this neighborhood. Let  $q_{w,d}$  equal the number of lines that the best occurrence of w is as a d-th neighbor. Each of these occurrences receives weight  $\lambda_d$ . The score of word w in this window is

```
s(w) = \sum \lambda q.
j+1,j+W ddw,d
```

A best word is word w satisfying

 $s (w^*) = max s (w).$ j+1, j+W w j+1, j+W

The idea of the algorithm is to align on consensus words, attempting to maximize the sum of the scores of the words. Before the practical algorithms are presented, a more general concept of alignment on words is presented.

Now we define a partial order on words. The words  $w_1$  and  $w_2$  satisfy  $w_1 < w_2$  if the occurrences of  $w_1$  in sequence i are to the left of the occurrences of  $w_2$  in sequence i (and do not intersect) for i=1 to R. It is not necessary for  $w_1$  or  $w_2$  to have occurrences in all sequences. Implicit in the definition is a window width W and neighborhood specification. The goal of an optimal alignment is to find words  $w_4$  which satisfy

 $\max\{\sum_{i\geq 1} s(w_i) : w_1 < w_2 < \ldots\}.$ 

(It is frequently desirable to require  $s(w_i) \ge c$  for all i, where c is some cutoff value.) It is not possible to accomplish this

goal in reasonable time, but it is possible to come quite close. We now define two practical algorithms.

Next  $w_1 | w_2$  means that consensus words  $w_1$  and  $w_2$  can be found in non-overlapping windows, each word satisfying as usual the window width and neighborhood constraints. The modified optimization problem is to satisfy

$$\mathbf{T} = \max\{\sum_{i \ge 1} \mathbf{s}(\mathbf{w}_i) : \mathbf{w}_1 | \mathbf{w}_2 | \dots \}.$$

There is a straightforward recursion to find T. Let  $T_i$  be the maximum sum for the sequences from base 1 to base i:

 $a_{1,1} \cdots a_{1,1}$  $a_{2,1} \cdots a_{2,1}$  $\cdots \cdots a_{R,1} \cdots a_{R,1}$ 

Then T, satisfies

$$T_{i} = \max{T_{i} + S_{i+1,i} : i-W+1 \le j \le i-k}$$

and  $T_{-W} = T_{-W+1} = \dots = T_0 = T_1 = \dots = T_{k-1} = 0$ . Also  $s_{x,y} = 0$  if y-x+1 < k. This algorithm runs in time approximately proportional to  $NW^2RB$  where B = neighborhood size. Here the factor of WRB accounts for the consensus word algorithm with a window width W. (This is an overestimate since the actual windows vary from k to W in width.)

If much shifting is necessary to match the sequences, T is an underestimate and misses some of the relevant matching. To overcome this problem, the definition of T is modified to

 $S_{i} = \max\{S_{j} + \hat{S}_{j+1,i} : i-W+1 \le j \le i-k\}.$ 

where  $s_{j+1,i}$  is the largest scoring consensus word in the window from j+1 to i such that all occurrences of the consensus word are to the right of the consensus words for  $S_j$ . This algorithm is not guaranteed to be equal to the global maximum, but it is much more useful than T.

## EXAMPLE

The algorithm is now illustrated with an alignment of 34 5S sequences from <u>E. coli</u> and related organisms that were obtained

: ecumenter agras gogge case accesser rear tree acc eggaagtraagederinkeinkeinkeinkeine tggaikteelekeineekaintee	o GCAAGGG tcatagcaGCAGGG aaac accaGATC ccat tccg aactCGAC ggttaagcCTGCTGCGTATTGCGTTGtactGTATGCCGCGGGGGTAC gggaAGGCGCAATATGCTGTTA	TGCC tggcGGGGGGGGGGgGgGgGGGGGGGGGGC ccat gccg act cagaagtgaaacGCGGTAgcgccgatGG tagtG tgggGTCTCCCCATGC gagaGTagggAACTGCCAGGC	sgtis tyggafrercechtrig gagasfraggaAnreechdochi sgtafretgggafrercechtrig gagasfraggaAnreechdochi sgtafretgggafrercechtsig gagasfraggtAnreechdochi satis tyggafrercechtrig gagasfraggtAnreechdochi satis tyggafrercechtrig gagasfraggtCanreethdocch satis tyggafrercechtrig gagasfraggtCanreethdocch sgtaft tyggafrercechtrig gagasfraggtechtechdoch sgtaft tyggafrercechtos gagasfraggtechtechdoch sgtaft tyggafrercechtos gagasfraggtechtechdoch sgtaft tyggafrercechter gagasfraggtechtechdoch sgtaft tyggafrercechter gagasfraggtechtechdoch sgtaft tyggafrercechter gagasfraggtechtechdoch sgtaft tyggafrercechter gagasfraggtechtechdoch sgtaft ggggAnthAntherchdoch sgtaft ggggAnthechterchdo sgtift ggggAntrercechter gagasfraggtechtechdoch sattreffe ggggafrercechdoch sattreffe ggggafrergeffer sattreffe ggggafrergeffer sattre	<pre>c acctGATC cost group and: usuary systemacoccurrants accactGATTC cost coreg aact cagtagtgaaacGTATGAGGGAggocogtGG ti aactactGATTCCATTCCGAACTCGAACTGAAACGAATTGAACGAATTGAACGAAC</pre>	Incc tggeckecatageg/trappu/ Incr tggeckecatageg/trappu/ Incr tggeckecatageg/trappu/ Incr tgackecatageg/trappu/ Incr tgackecatageg/trappu/ Incr tgackecatageg/trappu/ Incr tgackecatageg/trappu/ Incr tgackecatageg/trappu/ Incr tggtckccatageg/trappu/ Incr tggtckccatageg/tgggg/ Incr tggtckccatageg/tgggggggggggggggggggggggggggggggg	
	or excretectedired agrea great accorded that they and aggrangthangeederrangeederrangededered typaracteredeartece	<pre>r 167: 19906.cerasperformed accellance contrant of aspatuseasers in the second of the second accellance of the second accellance accellanc</pre>				
dop GGCMAGGC tcatagcaGCAGGG asac accaGATC ccat tecg aactCGAC ggttaagcCTGCTGCGTAGCTATCGCTTGtactGTATGCCGCGAGGGTAC gggaAGCGCAATATGCTGTTA		<pre>rior: 1996/SectabageGATEGACCA Section Sequest Sequest Sequest Section States 1996/SectabageGATECCCANST Sequest Section Se</pre>	agtAC tggaGTGTGCGAACCTCTGGGAAAACTGGTTCGCCGCC	C tcccGTAC ccat cccg aacaCGGC agataagcCCGCCAgcgtTCCAGC ga	FTA aggcGGccacagcgGCggggCGi	rrh 1
rrh TTA aggeGGceacagegGGggggGGAC teecGTAC ceat eeeg aaeaGGGC agataageCCGCCAgegtTCCAGC gagtAC tggaCTGTGGCGAACCTCTGGGAAAACTGGTTCGCCGGC do GGCAAAGGG teatageaGCAGGG aaae aceaGATC ceat teeg aaetCGAC ggttaagecTGCTATTGGCTAGTACGACGGGAGGGTAC gggaAACCGCAATATGGTGTA	rrh TTA aggeGeceaeageGeGggggGGAC teeeTAC ceat eeeg aacaGGGC agataageCGGCCAgegtTCCAGC gagtAC tggaGTGTGGGAAAACTGGTTGGGTGGGGCGC	<pre>rrgrr tgredkenapged/httggaccc acctdar ceat geg aat: aggagggaacGMATTARGCAGA tagg taggGrrTCCCCATGT ggagGrTTCCCCATGT ggagGrTTCCCCATGT ggagGrTTCCCCATGT ggagGrTTCCCCATGT ggagGrTTCCCCATGCA arCT tgredkenapgegTtaggACC acctdATT ceat ceg aat: aggagggaacGATGTCGATGTAGGT gggGGTTCCCCATGT gggGTTCCCCATGT gggGTTTCCCCATGT gggGTTTTCCCCATGT gggGTTTTCCCCCATGT gggGTTTTCCCCCATGT gggGTTTTCCCCATGT gggGTTTTCCCCCAGGTTTCCCAGG</pre>	agtAC tggaGTGCGCGAGCCTCTGGGAAATCCGGTTCGCCGCC	C tcccGTAC ccat cccg aaca cggaagataagcCCGCCTgcgtTCCGGT ca	<pre>rTA aggcGGccatagcgGTggggTTi</pre>	lin.
in TTA aggeGGecetagegGTggggTTAC teceGTAC ceat eceg aaca eggaagataageCCGCCTgegtTCCGGT eagtAC tggaGTGCGCGGAGCCTTGGGAAATCCGGTTCGCCGCC crh TTA aggeGGeceaegegGGggggGGAC teceGTAC ceat eceg aaeaCGGC agataageCCGCCAgegtTCCAGC gagtAC tggaGTGTGCGGAACCTTGGGAAAAACTGGTTCGCCGGC do GGCAACGG teatageaGCAGGG aaae aceaGATC ecat teeg aaetCGAC ggttaageCTGCGGGAATTGCSTTGtaetGTATGCCCGGAGGGAGGTAA	.in TTA aggeGGecetagegeTggggTTAC teeeGTAC ceat eeg aaea eggaagataageCCGCCTgegtTCCGGT eagtAC tggaGTGCGCGGGGGGCCTGGGAAATCCGGTTCGCCGCC erh TTA aggeGGeceaegegGGggggGGAC teeeGTAC eeat eeg aaeaCGGG agataageCCGCCAgegtTCCAGC gagtAC tggaGTGTGCGGAACCTCTGGGAAAACTGGTTCGCCGCC	ETCC: typeGeceatageQAAttegrCC actGAT costTyperaeCTTATAGeCGAGE tagte typeGeceatageQAATTCCCAATCA TCC: typeGeceatageQAAttegrCC actGAT costTyperaeCTTATAGAGCGAA TCC: typeGeceatageQAAttegrACC actGAT costTyperaeCAAAtCGAATTAGCGCAGE TCC: typeGeceatageQAAttegrACC actGAT costTyperaeCAAAtCGAAAtCGAATTAGCCAA TCC: typeGeceatageQAAttegrACC actGAT costTyperaeCAAAtCGAAAAtCGAAAtCGAAAtCGAAAtCGAAAtCGAAAAtCGAAAtCGAAAtCGAAAtCGAAAtCGAAAA	TAC tgagGTACGCGAGTCCTCGGGGAAATCATCCTCGCTGCT	AC acccGTTC ccat tccg aaca cggaagttaagaCACCTCACGTggatGACGG1	<pre>ICAA tageGGeceacageaG gtgtGT(</pre>	, but
ng TCAA tageGGecatageaG grgtGTCAC accGTTC cat teeg aaca eggaagttaagacACCTCACGTGggtGACGGGTAC tgagGTACGCGAGTCTTGGGGGAAATCATCCTGCGGTGGT in TTA aggeGGecatageGTGgggTAC teceGTAC cat teeg aaca eggaagtaagecGGCTGggtTCCAG  aggtAC tggaGTGGCGGGGGAGAAATCGTGGGCGGC th TTA aggeGGecatageGGGggGGAC teceGTAC cat teeg aacaGGG agataagecGGCGGGGGGGAGGGGCGGGGGGAAAATGGTGGTGGGCGGC op GGCAACGG teatageaGGGGG aaca cecagaacaGGG ggttaageCTGGGGGGGGGGGGGGGGGGGAGAAATGGTTGGGGAGAAACGGTGGGGGGGG	ng TCAA tageGeceateageag grypOTCA acceNT cat teeg aaca egaagttaageCACCTAcCTgegttAcGGTAC tgagTACGGGATCTCGGGATATATCATCCTGCGTGCT In TTA aggeGeceateageG1gggTAC teecoTAC cat eeg aaca eggaagtaageCCGCTgegtTCCAG aaptAC tgagGTGCGGGACCTCTGGAAAATGATGCTCGCGCCC h TTA aggeGeceateageG1gggGAC teecoTAC cat eeg aacaGGC agataageCCGCCTgegtTCCAG gagTAC tggaGTGCGGACCTTGGAAAATGATGCTCGCGCC	<pre>rr foor typeGaceatageGAFTAFGAFTCCAFTCCAFTAC celt foot act foo</pre>	GTGTACTATGGGTTCCGGTCTAT gggaATTTCATTTAGCTGCGA	ATCACCTGGTCTCGTTTCGATCCCAGAAGTTAAGTCTTT tcgcGTTTTGTTTG	rAGGTTTGGCGGTCATAGCG atggGG	i2 :
12 TAGGTTIGGCGTCATAGCG atgGGGTATCACTGGTTCGTTCGAAGTTAAGTCTT tegeGGTTTGTTGTTGTACTATGGGCTCGGAGTTCGGGTAT gggaATTCATTAGTTAGCTGGGA ng TCAA tagGGGccateggaG gtgGTCGAC acceCTTC cat tecg aaca cggaagttaageCACCTGAGGTGGGTCGGTGGT catAAC tgaGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	12 TAGGTTGGCGCGCATAGCG atgGGGTATCACTCGGTCTGGTTGAAGTTAAGTCTT IN TA tgGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	<pre>rr fictr typeGaceatageGAptropyTecc accGArr care yeeg aaac cagtagtgaacGTAAAAGAAGTAAAAGAAGTAAAAGAAGTAAAAGAAGTAAAAGAAG</pre>	AATTGTTC gggaGACATGGTCAAAGCCCG	cacccGATCCCATTCCGACCTCGATATATGTGGGAATCGTCTTGCGCCATATGTACTGA	AAACCGGGCACTACGGTGAGACGTGAAA	mt ,
me AMACGEGACHAGEGAGENAGAGGAMAAAaacCGATTCGATCGATATATATATGGATGGTCTTGGGGCATATGATATATTGTTC 12 TAGGTTGGGGGGCATAAAGGG atggGGATTGACGTGGATCGATCGATCGATGCGAGGGTTGATGGGGGGTGGGAGTTGTTGGGTTCGGGTGGGA 19 TGAA TagGGGGGGATAGGGGTGGGTTGGGATGCGATGCGAGGGAGGTGGGGGGGG	mt AAACCGGGGGGGGGGGGGGGGGGGGGAAAcaccCGGATCCGATCGGCGCGTAAGTGTGCGCGAATGGTACGGAATGGTCG TAGGTTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	In the transmission of	actGC agggGGGACCCTGT gggaG agtaGGACGCCGCG	aac gcccGGTT acat tccg aacc cggaagctaagcCTTACAgcgccgatGG ta	STTTCGCTGG tcatagcgT gaggG	Ē
<pre>m diffication classeger gase geocoder act terg are gragetargetering group of the sector grage graded graded concorn m Ancornecting concentrocarrecontrocarrecontext and graded grad</pre>	<pre>m drincords textager gage and geocéert varients and text construction provide text of gage apaceAccected m frincords textager gage and geocéert varients and text construction and text and</pre>	<pre>r ICST tggeGecataggeGTTetggACCC acctGAT cargerg aact cagaatgaaacGTMATAgeGecgatG tggtG tgggGTCCCCATGT gaaGTaggAACTGCCAGG r ggeGecataggeGTtetggACC acctGAT cact acgradtgaaacGTAATAGAATTAGGCCAA r TTT tggeGecataggeGTtetggACC acctGATTCCAATTCCAATGATATAGGCCAA r TTT tggeGecataggeGTtetggACC acctGATTCCAATTCCAATGATTAGGCCAA r TTT tggeGecataggeGTtetggACC acctGATTCCAATTCCAATGATGAATGAATTAGGCCAA r TTT tggeGecataggeGTtetggACC acctGATTCCAATTCCAATGATGAATGAATTAGGCCAA r TTT tggeGecataggeGTtetggACC acctGATTCCAATGATTCCAATGATGAAGTCAAGCAATTCGCAATG r TTT tggeGecataggeGTtetggACC acctGATTCCAATGATTCCAATGATGAAGCAATTGGCCAA r TTTT tggeGeCataggeGTTetggACCT r TTC tgacGAtcataggeGTTetggACCTGAT r TTTT tggeGeCataggeGCTGAAA r ATCCCCCCGTGCCAAGC r ggrGGCCataggeGTTetggACCTGCAATGACGAAGTCGAAAGCGCCACGCCAAGGC r ggrGGCCAtaggeGTTetggAGTCCCCAATGATGAATAGCGCCACGCCAAGGC r ggrGGCCataggeGTTetggAGTCCCAATCCCAATCCCAATGAGGCGATGAAAACCCCCCCAGGC r ggrGGCCAtaggeGTTAGG r true tggrGGCCataggeGTTAGGAAGTCGAAACCCCCCCCCCCCAGGCGCGGGGGGGG</pre>		οιά από το	STTA COORGENTATION COORD	ر
c GTTA cggcGGctatagcgT gggg aaac geecGG ccgtATATCGaace cggaagctaagcCCCATAgcgccgatGGTTATCGTAACGGGGGGGTGTT gggad agtaGGTCGCCGCG m GTTTGGGTCG caarageGGTTAGGTTAAAaacgccgcGTTATGTATGAATGAGGGCGGGGGGGGGGGGGGGGGGGG	c GTTA cggcGGCtatagcgT gggg aas gccGG ccgtATATCGaacc cggaagctaagcCCCATAgcgccgatGGTTATCGTAACCGGGAGGTTGT gggaG agtaGGTCGCCGCG m GTTTGGTGC caataggeGT gggG aas gccGGT aast tecg aas c gggaagctaagcCTTAGGAGGAGGTCAATGTCTACTGAGGCGCGCG mt AAACGGGGGAGGGGGAGAGGCGTAAAAAAAAAAAAAAAA	In this the product and the production of the product and the product product and the produ	actTA tgtgAGAAGATAGCAAGCTGCCAGTT	caccTGTTCCCATGCCGAA cacaGAAGTTAAGCTCTATTACGGTGAAAATAT to	TTGG tggtATAGCATAGAGGTC	, Ym
<pre>my TTGG tgitATAGCATAGAGGT acacCTGTTCCCATGCCGGAA cacaGAGTTAAGCTCTATTACGTGAAAATAT tactTA tgitgAGAAGATAGCAGGGTCGCGCGT of TTA oggeGGetatagGT gggd aard gecodd cacaGAGTTAAGGTCATAGGGGGAGCTGT gggdGAGGTCGCGGCGCG m GTTCGGTGG LacatagGT gggd aard gecodd caca teeg aacc oggaagtaagtcCCTAAggGGGGGGTACTGATAGGGGGAGCTGT gggaGGAGCGCGCGCG m AAACGGGGGACTAGGACGTGGGTTCGAGCCGATTATATAGTCTT TA AAACGGGGGACTAGGACGTGGGTTCGAGTCCAATATATAT</pre>	<pre>my TTGG tggtATAGCATAGAGGTC acacGTGTTCCCATGCGGAA cacaGAGTTAAGCTCTATTAGGTGAAATAT tactTA tgtgAGAAGATAGCAAGCTGCGGGTG c GTTA gggGGGtataggGT gggG aaac gccGGG cgtATATGCGCAGGGCGGGGGGGGGGGGGGGGGGGGGGGGGG</pre>	FIGTE tggedkecataggeg/fieldggACCCARGC. FIGTE tggedkecataggeg/fieldggACCCARGC. FIGTE tggedkecataggeg/fieldggACCCARGC. FIGTE tggedkecataggeg/fieldggACC FIGTE tggedkecataggeg/fieldggACC FIGTE tggedkecataggeg/fieldgaCCGARGC. FIGTE tggedkecataggeg/fieldgaCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCCARGC. FIGTE tggedkecataggeg/fieldgaCCCCARGC. FIGTE tggedkecataggaC field gaACCACCCARCCARCCARCCCARGC. FIGTE tggedkecataggaC field gaACCACCCARCCARCCARCCCARCCARCCARGC. FIGTE tggedkecataggaC field gaACCACCCARCCARCCARCCCARCCARCCARCCARCCARC	actGA totoAGAAAATAGCAAGCTGCCAGTT	cacCTGTTCCCATGCCGAA cacaGAAGTTAAGCTCTATTACGGTGA agatAT to	TTGG togtATAGCATAGAGGTC	ca
ca TTGG tggtANGCANGAGGTG acacCTGTTCCCANGCGGAA casaGAGTTAAGCTCTATTACGTGA agatAT tactGA tgtgAGAAANTAGCAAGCTGCCGGT W TTGG tggGATAGGTTG acacCTGTTCCCANGCGGAA casaGAGTTAAGCTCTATTACGGTGAAAANT tactTA tgtgAGAGAAGATGCCAGGTG c TTA ggGGGALAGGTGA acacGGTTG acacGAGTTAAGTCTCATTAGGTGGTGAAAATT tactTA tgtgAGAGAGAGTGCGCGGCG m GTTCGGTGG tcataggT gagG aas gecc6G cegtAAAGGTGAACGTTATTAGGGGGAGGTTGT AcgGAGGGCGCGGGG m CATTCGGTGG tcataggT gagG aas gecc6G acat tcg aacc eggaagttaagcCTTACAGGGGAAGTGGT tactGC aggGGGGACCCTGT gggaG agtaGGTGCGCGGG m AACCGGGGCAAAGGGGGAGAGTAAGAGTCAAAGTAGAATCTTACAGGGGAAAGTTATTG AAGTTGGGGGGACAAGGGGAGGTAAGAGTAAGATCTTT tagGGGGGCAAAGGGGAGAGGTAAGAGTAAGATCTTT tagGGGGGCAAAGGGGGAGGGGGGGGGGGGGGGGGGGGGGG	a TIGG tggtATAGGATAGGAGGT acacGTGTTCCATGCGGAA cacaGAGGTAAGCTCCTATTAGGGTGA agatAT tactGA tgrgAGAAGATGCGGCGGCGGTGAATAT W TGG tggtATAGGATGGGGG cacaGAGGTTAGGCGAA cacaGAGGTTAGGCGGAAAATT tactGA tgrgAGAAGATGCGGCGGGTG c TTA gggGGGALagggT gagg aac gcccGG ccgtAATGGACC cggaggtcaagGCGTAAGGGGGAGGTTGT m fTTCGGGGGGGG tatagggT gagg aac gcccGAT cacaGAGGTTAGGCGCATAGGGGGGGGGGGGGGGGGGGGGGGG	In this transmission of the second second and the second secon	agcCGCAAGGTGAAAATAGGGCAGTGCC aggt	cac acccGTTC ccat cccg aaca cggaagttaagcACTATTacgccgacGA t	IC tggtGCTTATGGCATAgtgg	a I
a TC tggtcCTMAGCANAGTGG acc accedt acced acced acced acced accedtaced tagtcCCAAACTAGCCAAATAGCCAAACTGCC agt art ggtcTMAGCANAGAGGT acceCTGTTCCCATGCCGAA cacaGAGTTAAACTCTTATAGGCGAA agtAGAAATAGGCCAAGGTGCCAGTT my TrGG tggtAMGCANAGAGGT acceCTGTTCCATGCGGAA cacaGAGTTAAATAGGCGAAGATT tactGA tgtgAGAAATAGGCAAGGTGCCAGTT art gggGGGctataggT gggG aac gcccGG ccqTAATGGGTAA gggCGAGGTCGTGTAGTGTAACTGTAACTGCCAAGAGTGCCAGTT a fTACGGTG tacataggT gggG aac gcccGG ccgTAATGGTAAGATTACGTGAAAATAGGCGAAGACTGCTAGT a fTACGGTG catataggT gggG aac gcccGG ccgTAATGGTAAGATTACGTGAAAATAGGCGAAGACTGTG agggGGGGGAAGAGCGGCGAGTGCGATAATATAGTCTTTAGGTGAAGGGGGGGG	a TC tggtcGTMAGGCANAgrgg ace accedTC cat cccg ace cggaagttagGACTATTacgccGatA tagcGCCAAGCTGAAATAGGCGAGTGCC aggt at tecta tggtAAGCANAGGCT acacGTTCCATGCGCAA cacaGAGTTAAGCTCATTAGGGTGAAATAT tacTA tgg4GAAATAGCAAGTGCCAGTT my TGG tggtAAGGATAGAGGT acaCTGTTCCATGCGGAA cacaGAGTTAAGCTCATTAGGGTGAATAT tacTA tgg4GAAATAGCAAGTGCCAGTT c GTTA cggcGGctatagegT gggg ace gccGG ccgtAATCGaacc cggaagctaagcCCCAAAggscgatGGTTATCGGGAGGTTGT ggg4GAAATAGCAGGTGCCGCG mt AAACGGGTGAGAGGGGT ace tccg ace cggaagctaagcCCCAAAggscgatGGTTACTCAACGGGGAGGTTGT gggaG agtaGGTCGCGCGG mt AAACGGGGAGAGAGGGGTAGAA tccg ace cggaagctaagcCTTAGGGGGAATGGTCAACGGGGGGGGGGCGCGG mt AAACGGGGGAAGAGGGTGAAATCCAATGCAAATATAAGGTGAAATGGGGAATGGTCAAGGGGGGGG	In the projection and any second section of a section	.actGC agggGAAGCCCTGTGG aagaGTaggtCGACGCTGGGT	aac actcCTTC ccat tccg aacaGGcaggTT aagcTCTAATgtgctgatGG ta	rccAGTGT ctatGACTTAgagg t	eu '
<pre>but CGARCT tetteGANGTAgagg taag stercg asesGGeagTT asgCTANTTagregatGG tactCG aggGAAACCCTGTGG aggGAAACCTGCGG aggGAAACCTGCGG aggGAAACCTGCGG aggGAAACCTGCAAGGTGAAATAGCAACTGCAAGGTGAAATAGCAACTGCAAGTGAAATAGCAACTAAATAGCAAACTGCAAGTGAAATAGCAACTGCAAGTGAAATAGCAACTAAATAGCAAACTAACCAAGTAAATAGCAACTAACT</pre>	<pre>eurrcMarr tetakorThMagg taa artecrTr cart teeg aaagdeaggttaageArThTracgeegaeds taates aggeAbAcCrTorGo aggeAbArThGGCATHGGG aggeAbArThGGCATHGGG aggeAbArThGGCATHGGGAAGTGAAATTAGGAAGTGAAATTAGGAAGTTAAAATTGAAGTGAA</pre>	In this regedecatageochytigtics acclust cast group aact magaatering acceleration types (aggering aggering ag	actTGCTCCGCAGGGAGCCGG gagaGTaggaCGTCGCCAGGC	cac acccGTTC ccat accg aacaCGGC cgttaagcCCTCCAgcgccaatGG t	<pre>PC tggtGATGATggcgGAgggg </pre>	is,
is TC stycenthydgogdaggg ace accorr cast aceg acadGGC optiageCTCTACAqcgecastGG tactTGCTCCGGGGGGCGCGG gagaCTaggaCTGCGCCGGGG urcCGGTG cataGATAggg ace accCTTC cast aceg acaGGGcagTT aagcTCTAATGGGG tactCG aggGGAGGCCGTGGG aggGGAGGCCTGGG ar TC gggGCTTAGGGT aggGGAGGCCTGGAC acaGGGGAGTT aagcTCTAATAGGGGAGGTGAAAATAGGGCC aggGGAGGCCGTGGG ar TGG tggtGGTAAGGGT acaCTGTTCCCATGCCAA cacaGAGTTAAGCTCTATTACGGTGAAATT tactTA tggtGGAAGCCGGGGGGGGG ar TGG tggtAAGCAAAGGGT acaCTGTTCCCATGCCAA cacaGAGTTAAGCTCTATTACGGTGAAATT tactTA tggtGAAAATAGGGCGAGGTGGGT ar TGG tggtAAGCAAAGGGT acaCTGTTCCATGCCAA cacaGAGTTAAGCTCTATTACGGTGAAATT tactTA tggtGAAAATAGGGCGAGCTGCGGGG acaCTGTTCCATGCGCGA cacaGAGTTAAGCTCTATTACGGTGAAATT tactTA tggtGAAAATAGGCGAGGTGCGGGGG an TTGG tggtGATAGCAAGGGT acat tccg acc cggaagtaagcCTTAAGGGTGAGGTTACTGAAGGTGCGAGGCTGCGGG acaCTGTTCCATGGCGG ccgtAAGGTAAGGTCCAAGGGGAGATAAGCTCTATTAGGGTGGCGGGGGGGG	is TC typeGATAGGATAGGATAGGATAGATAGATAGATAGATAGATA	In the transmission of the second second second second second to the second	agtGA ggggGTTGCCCCTTGT gagaGTaggaCGTCGCCACGC	tac acctGTAA ccat gccg aaca cagaagttaagcTTCTTAgcgccgatTG to	<pre>IG tggtGGcgatagcgAGAAGG</pre>	4
If the trepresent second are accelling cast accords capacitageTICTMAGGGCGGTG aggGTTGCCCCCTCGT gagaGTaggGCGCGCACACC trepresent second second trepresent accords accords accords accords aggaCTTACCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	If the typeGeopageAgeAgeAgeAgeAge are accepted agaagtLaageTTCTAAgeGeocatTG tateA gogGTTGCCCCTTG gagaGTgggaGCTGCCCACGC to typeGeopageAgeAgeAgeAgeAgeAgeAgeAgeAgeAgeAgeAgeAg	In this regeodecatageophytogeocce activation and regentance cast group and control regeoference and regent regeoferences and regentances an	agtTG ggggATCGCCCCCGC gaggATaggaCGTTGCCACGC	tac acctGTTC ccat gccg aaca cagaagttaagcTTCAGCacgccgatAG to	<pre>IG tggtGGcgatagccTGAAGG </pre>	ev.
<pre>ev rG tggtGGegatagecrEMAGG atex acctGTM ceat geeg ace cagagttageTTCMGGGegecgatGG tagtGG gggATGCCCCTCGC ggggATGGGCGAGCG f rG tggtGGGgatagecGAGGGG atex acctGTM ceat geeg ace cagagttageTTCMGGGGegecatGG tagtGGCGAGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</pre>	<pre>evror typtGGGgatageoCFGMAGG atex acctGFM ccat gccg axea cagagitaageTTCAAGCaGCGGCG ggggAFTGCCCCTTGT ggagAFTgGGCTGCCACGG frg typtGGGgatageoGAGAGG atex acctGFM ccat gccg axea cagagitaageTTCAAGGCGTGA ggggGTTGCCCCTTGT ggagGTggGCCCACGG is registCarATTAggg acac acceCTTC ccat accg axeaGGC ggttaageTTCATAFTGGGGT tagtGA ggggGTTGCCCCTTGT ggagGTggGCCACGGG arrc typtGCTTAAggg acac acceCTTC ccat accg axeaGGC ggttaageTCTCAAGGGG tagtGA gggGTTGCCCCTTGT ggagGTggGCCACGGG arrc typtGCTAATAGGGT axea accGTTC ccat accg axeaGGC ggttaageTCTAATAFTGGGTGAGGCGG gggGAGCCCCGGG gggGGCGAGCCGGG arrc typtGCTAATAGGGT acacCTGTTC ccat accg axeaGGC ggttaageCTTAATAGGTGA tgtGAGAGCCGG gggGGCGAGCCGGG arrc gggtCATAGGGTC acacGGTTCCAATCGCAAA cacaGAGTTAAGGCTAATAGGGCGAGGGGGAGCCGG gggGGCGAGCCGGG arrc gggtCATAGGGTC acacGGTTCCCAATCGCGAA cacaGAGTTAAGCCTAATTAAGGTGA tgtGAGAGCCGG gggGGGGCCGGGGGGGGGGGGGGGGGGGGG</pre>	In this regeodecataged hydroc accelerate can easaged acceleration and types (aggerrececkner) aggerrececkance accelerate can easaged acceleration and the aggerrececkance accelerate accelerateraterate acceleraterateraterateraterateraterateraterat	agtTGGAGGATCTTCCTGC gaggATaggaCGTCGCAATGC	cacCTGTTC ccat accg aaca cagaagttaagcTCAATAgcgccgaaAG ta	rGTTGTGATGATGGCATT gaggTC	t
<pre>Instituted and a sequence of a construction cast constructin cast construction cast construction</pre>	<pre>in TritricTantaTracture geg acc cgaagttaageTCAMTagegecgaals tagtTGSAGATTCTECTEC gagATageGGGGGGAGGGGCGCAGGG is TritricTantaTractore geg acc cgaagttaageTCAATAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</pre>	In the transpondent and the second and the second second and the second second topic to	actGGG agggCAGCCTCCTGG gaaaGTaggtCGTTGCCAGGC	aac acccGTAC ccat cccg aaca cggaCGTGaagaCCTCCAgcgccgagAA ta	<pre>IC tggtGActatagcgGAgggg 4</pre>	ö
<pre>oc TC type:GATTAGGATTAGGATTAGGATAGGATAGGATAGGAGGATAGGGAGGA</pre>	<pre>crit repredentagerophage case concrite cast aceg acea cgaagettaagerCriteAproproceapality tactore aggArreseConfronter repredentagerophage case concriter cast aceg acea cgaagettaagerTreAhproproceapality tagtics aggArreseConfronter repredentagerophage area carcefire cast aceg acea cgaagettaagerTreAhproproceating tagtic aggArreseConfronter repredentagerophage area carcefire cast teop acea cgaagettaagerTreAhpropropropredent tagtic aggArreseConfronter repredentagerophaged area carcefire cast teop acea cgaagettaagerTreAhpropropropropredent tagtic aggAAAAATACAAAATACAACAAATAAAAATACAAAATAACAAAATAAAAATACAAAATAAAAAA</pre>	In this regodeneragegrinteggades accelerate cast group aaac magaagement group ggggderescander (aggderescanderescander) a fost gggderescanderescander (aggderescanderescanderescanderescanderes) a fost gggderescanderescanderes (aggderescanderescanderes) a fost gggderescanderescanderes (aggderescanderescanderes) a fost gggderescanderescanderes (aggderescanderes) (aggderesc	agtTGGGGCCAGCGCCCCTGC aagaGTaggtCGTTGCTAGGC	aac acccGTTC ccat cccg aaca cggaagttaagcTCTCCAgcgccgatGG t	CCTAGTGA caatagcgGAgagg	ro Lo
<pre>c crthorth contageogdaggg aaa cocorrect cara corg aaca orgaacttaagerfrechAgorgoegatG tagtrGGGGGCAGCCCCTGC aggaragerGTTGGCAGCC aggaragerGTTGGAGGA caracter corg aaca orgaactGGGGAGGAGGAGGATCTTTGCTGC gaggAraggaCGTTGCCAGGC arger caracter goorg aaca orgaacttaagerTTAAAGOGGGAGAGGA tagtrGGGAGGATCGCTTGCG gaggAraggaCGTTGCCAGCCACGC tggtGGGGgAtaggaCGGTGCATGC gaggAraggaCGTTGCCAGCCACGC arger tagtGGGGAGGATCGATTGCTGC gaggAraggaCGTTGCCACGCC gaggAraggaCGTTGCCACGC gaggAraggaCGTTGCCACGC arger tagtGGGGAtaggaCGTGCCACGC gaggAraggaCGTTGCCACGCCACGC gaggAraggaCGTTGCCACGC gaggAraggaCGTCCCACGC gaggAraCGCCACGC gaggAraggaCGTCCCACGC gaggAraggaCGTCCCACGC gaggAraggaCGTCCCACGC gaggAraCGCCCCCCCCCAGC gaggAraCGCCCCCCCCCCCCCCAGC gaggAraCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC</pre>	<pre>ccrbcits catagoordhagg ase according cost acces area cypaactrangerCircCApproceedatics tagtroGAGGCCACCCCCC asgadraggCGTGTGCTAAGCC cr tigtrGATAGOORT gaggreactoring cat according cost acces area cypaactrangerCircTAATAGOOGCAAGCCCACCCC gagaGTGGGGCAAGCCAAGC cr tigtrGATAGOORT gaggreactoring cat according cat acces area cypaactrangerCircTAATAGOOGCAAGCCGCACCC gagaGTGGGGCAAGCCAAGC cr tigtrGATAGOORT gaggreactoring cat according cat acces area cypaactrangerCircTAATAGOOGCAAGCCGCACCC gagaGTGGGGCAAGCCAAGC cr tigtrGATAGOORT gaggreactoring cat acces area cypaactraagCTTCAATAGOOGCGAAGCGAAGCCACCCC gagGTGGGGGAAGCCAAGCC is tr tigtrGATAGOORT gaggreactoring cat acces area cypaactraagCTTCAAGCGGGGGGGGGGGGGGGGGGGGGGGGGCCCCCCCC</pre>	In tote regeodecategeodatignees accession castagerpaaaconticphergeogatic types (aggeonescategeodatic) regeodecategeodatic) regeodecategeodatic accession control regeodecategeodatic regeodecategeodatic) regeodecategeodecategeodecategeodecategeodecategeodecategeodatic) regeodecategeo	agtTG ggggCTTCCCCCTGT gagaGTaggaCGCCGCCAAGC	cacCCGTTC ccat gccg aaca cggaagttaagcTCTTCAgcgccgatGG t	<pre>IT tggtGGcgatagcgAAgaggTC</pre>	en
In the process process of the pro	en IT tgetGeopatagoGMaggTcacacCCTTC ceat geog area orgaagttaagcTCTCAgogocgatG tagTG gggGCCCCCCCTG gagGTaggCCCCCCCAACC CCTAGTG caatagoGMaggTcacacCTTTC ceat ecor area orgaagttaagcTCCAgogocgagAA tactGGG aggGCGCCCCCCTGG gaagGTaggCCGTCGCAACC CCTGTGTATAATGCCATT gggTCacacCTTTC ceat ecor area orgaagttaagcTCAAgogocgagAA tactGGG aggGCAGCCTCCTGG gagGTaggCCGTCGCAACC tf GggGCGGAtggCGAGTAG area caccGTTC ceat ecor area orgaagttaagcTCAAGGGCGGCACCCCCTGG gagGTGGGCCCCCCCG gagGTGGGCCAACCCCCCGC gagGTaggCCGTGGCAACC tf GggGCGGATAggGGCAATGC ccat gccg area cacgaaca cagaagttaagcTCAAGGGCGAGCCCCCCGC gagGTGGGGCCACCCCCCG gagGTGGGGCCACCCCCCG gagGTGGGGCCCCCCCCG gagGTGGGGCCCCCCCC tf GggGCGGATAggGG area caccGTTC ceat area area cagagttaagcTCAAGGGCGAGCCCCCCCCGC gagGTGGGGCCCCCCCC tf GggCGCGATAggGG area caccGTTC ceat area area cagagttaagcTCAAGGCGCACCCCCCGC gagGTGGGGCCCCCCCCG gagGTGGGGCCCCCCCCC tf GggCGCGATAggGG area caccGTTC ceat area area cagagttaagcTCAAGGCGCACCCCCCCCG gagGTGGGGCCCCCCCCC aggCGCCTTAGGG area cacCGTTC ceat area area cagagttaagcTCTAATTAGCGCGAGCGCACCCCCCCCG gagGTGGGGCCCCCCCCCC aggCGCCTTAGGG area cacCGTTC ceat area area cacgaacac cggaagttaagcTCTAATTAGCGGGAGCTCCCCCCCGC gagGTGGGGCCCCCCCCC aggCGCCTTAGGG area cacCGTTC ceat area area cacgaacaccCCCCCCAAGGCGCAGCCCCCCCCCG gagGTGGGGCCCCCCCCCCCG gagGTGGGGCCCCCCCCCCCCCG gggGTTCCCCCCCCCCCC	In the regedecetageophytophytom accelerate cast asystytaaacenthytopeostated type typeffectochord gagentrecochord gagentreco	agtCG ggggTTTCCCCCTGT gagaGTaggaCGCCGCCAAGC	cacCCGTTC ccat accg aaca cggaagttaagcTCTTCAgcgccgatGG t	<pre>TT tggtGGcgatagcgAAgaggTC</pre>	ц.
11 IT typtGGGgatagoGMAggGTGacaCCGTTC ceat acc9 acea cypagitageTCTCAAGGGG tagTG gggGTTCCCCGT gggaGTaggGCGCCAAGG CCTTATA gggGGGgatagoGMAggGTCacaCCGTTC ceat acc9 acea cypagitageTCTCAAGGGG tagTGGGGGCAAGGGCCCAAGG CCTTATATATATATATATATATATATATATATATATATA	II III typtGGGGAtagGGMAggGTGLacaCCGTTC ceal accg acta cggaagttaagCTCTGAgGGCGAtaGGC tagtCG gggGTTCCCCCTG gagGGTGgGGCCAAGC III III tygtGGGGAtagGGMAggGTCacaCCGTTC ceal accg acta cggaagttaagCTCTGAgGGCGAAGCGCAAGC CCTGTAATCAATCGATT gggGGAtagGTGTGTCCCTG gagGGTGgGGCCAAGCGCAAGC CCTGTGATCATCGATT gggGGGAtagGCTGTTCCCCTG gagGGTGgGGCCAAGCGCAAGC CCTGTGTATCATCGATT gggGGAtagGTGTGTTCAGGGGCAAGCGCCCAGC CCTGTGTATCATCGATT gggGGAtagGTGTGTCCCCCTG gagGGTGgGGCCAAGCGCAAGC CCTGTGTATCATCGATT gggGGAtagGTGTGTGCG CCTGTGTGTATCATCGATT gggGGGGAtagGCTGTGTGGGGCAAGCGCCAAGC CCTGTGTGTATCATCGATT gggGGGGGAtgGCGTGTGCG CGTGTGTGTATCATCGATT gggGGGGGATGGCGTGTG CGTGTGTGTATCATCGATT CGTGTGTGTGTGTGGGGGGGGGG	In thick regeodecatageophytopic accleant cert group aact magaagement aget type toperforce. And the space age age advectance of the space age age. The space age age advectance of the space age age advectance of the space age age. The space age age advectance of the space age age. The space age age advectance of the space age age. The space age age advectance of the space age age. The space age age advectance of the space age age. The space age age age age age advectance of the space age age. The space age age age advectance of the space age age age age advectance of the space age age age advectance of the space age age age age advectance of the space age age age advectance of the space age age age age advectance of the space age age advectance of the space of the space advectance of th	CCTGGGTTGCTGGCT gggaAAATAGCTCGATGCCAGC	accACTTCGATTCCA tcccGAACTCGATTGTGAAACTTTGCTGCGGCTAAGATACTT	<pre>rrcc tggtGTCTCTagcgCTttgg </pre>	ы
IT TTC tggtGGGTAGGGTAGGGTTCGTTCGTTCGTTCGATTCGTTGAGGTCGTCGGCTGGGTGGG	IT TTC tggtGGGTATGGTTCGATTCGATTGATTGATTGCATGGTGGGGTAGATGATTGCTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGG	In TCT typeGecentageg/Apt/gpTCC acctGAT cart grey aact ragaatgaaacGTTGTAgegecgatch typtG typeGecentageg/AcTCCCCAFCT gagadTaggAACTCCCAFCCAFCT are cart grey and the second state and second state are cart grey and the second state transformation of the second state typeGecentage/AcTCCCAFCT gagadTaggAACTCCCAFCT gagadTaggAACTCCCAFCC are cart grey gaaactagagaGATCCCAFCT gagadTaggAACTCCCAFCT gagaTAggAACTCCCAFCT gagadTaggAACTCCCAFCT gagaTAggAACTCCCAFCT gagaTAggAACTCCCAFCT gagaTAggAACTCCCAFCT gagaTAggAACTCCCAFCT gagaTAggAACTCCCAFCT gagaTAggAACTCCCAFCT gagaTAggAACTCCCAFCT gagaTAggAACTCCCAFCT gagaTAggAACTCCAACTCCAFCT gagaTAggAACTCCAACTCCAFCT gagaTAggAACTCCCAFCT gagaTAggAACTCCAACTCCAACTCCAACTCCAAACTCCAAACTCCAAACTCCAAACTCCAAACTCCAAACTCCAAACTCCAAACTCCAAACTCCAAACTCCAAACTCCAAACTCCAAACTCCAAAACCCAAAACCCCAAAACCCCACAAACTCCAAAAACTCCAAAAAA	ageTCCCGGGTAGCCGGTCGCTAAAATAGCTCGACGCCAGGTC	acc actcTGACCccat cccg aact cagtTGTGaaacATACCTgcggCAACGA to	<pre>rcc tggtGTctatggcgGTatgg ;</pre>	
In tric trigitarian production according according control	In tric tggtterictingeograms according cost act of triptaatachikerikarikarikarikarikarikarikarikarikarika	In TCPT tggedecetagegraftggarcc acctGAT cert gega act cagagtgaasCTTGTAggecegatGA tggtG tgggGTCTCCCATCT gagaGTaggAACTGCCAGGC In TCPT tggedecetagegraftggACC acctGAT certTggaracTTGTAACGATTAggecgatGG tagtG tgggGTCTCCCATCT gagaGTaggACTCCCAGGC In tggedecetaggGATtggACC acctGATTCCATTCCAATCGAAGTGAAGCAATTAGGCGGA Inter tggedecetaggGGTtgggACC acctGATTCCAATCGAAGTGAAGCAATTAGGCGGA Inter tggedecetaggGGTtgggACC acctGATTCCAATCGAAGTGAAGCAATTAGGCGCA Inter tggedecetaggGGTtgggarcCAATCGAATCGAAGTGAAGCAATTGGCAGG Inter tggedecetagggGGTGG garc acctGATTCCAATCGAATCGAAGGCAATGGCCAA Inter tgacGAtcataggGGTtgg acccactGAT cere act aggagtgaaacGAATGGCAA cgcgatGG tagtG tgggGTTTCCCAATCT gagaGTaggtCATCGTCAAGA Inter tgacGAtcataggGGGTG garcGacctGAT cere act aggagtgaaacGAAGCAAtgGGCGAAtgGGGGAAGGGCAAGGGGCAAGGCGAAGGCAATGGCAA Inter tgacGAtcataggGGGTG garcGacctGAT cere act aggagtgaaacGAAGCCAAtgGGCGAAtgGGGGAAGGGCAAGGGG in GCC tggtGGcaaaggGGGTG garcGacctGAT cereAAAGCGCCGGGGCAAtgGGCGAAGGGGTGAGGG in tggtGGCGGGGGGGGGGGGGGAAA accCGGATC cereAAAGAGGGCAAGGCCGAGGGGGGAGGG tagtG tgggGTTTCCCCAGTG gagGTGAGCCGGGCCAAGC in tggtGGCCGGGGCGGGGGGGGGGAAAAGAGCGAAGGGCCGGGGGG	ggtAC tgggCGGGCGACCGCCTGGGAG agtaGGTCGGTGCGG	aaccACCCGTTCCCATTCCGAACACGGGAAGTGAAACGCGCCAGCGCCGA to	<b>AATCCCCCGTGCCCATAGCGGCGTGG</b>	E
In the concentrange of the concentration of the con	In tree regreened according according according and the approximation of approximation and according accor	In TGPC regeodecatageogrhytigatics acctedate cart georg aact cargaageacTTGThyrgocegatich tygts tyggefTCCCCATGT gageTTaggAACTGCCAAGG I GCT tyggedAccatageogTTatggACC acctedAT coctTgocg aact cargatageaacGTAAThyrgocegatich tygts tyggeGTCTCCCATGT gageTTaggACTCCCAGGG I ggeGAccatageogTTatggACC acctedATTCCATTCCATTCCAATCAATAAGAATTGCACGA tggts tyggeGTCTCCCATGT gageGTAggACTCCCCAGGG I ggeGAccataggagAttgg acc acctedAT cart cargatagtgaaacGTAATAAGAATTGCCAGG tagt tyggGATCTCCCATT gageGTAggACTCCCAATA TTCT tygeGAccatagagGAttgg acc acctedATC cart corg aact cargatagtgaaacGATGGCAAtegregatics tagt tyggGATTCCCAATGT gageGTAggACTCCCAATG TTCT tygeGAtcatagagGAttgg acc acctedATC cart corg aact cargatagtaacGATCGACGAtegregatics tagt tyggGATTCCCAATGT gageGTAggACTCCCAATGT in TCCT tygeGAtcatagagG tttg gaacCacctGAT ccat corg aact cargatagtaacGACGCAtegregatics tagt tyggGATTCCCAATGT gageGTAggATCCCCAATGT gageGTAggATGTCCCAATGT gageGTAggACTCCCCAATGT gageGTAggATGTCCCAATGT gageGTAggATGTCCCCAATGT gageGTAggATGTCCCCAATGT gageGTAggATGTCCCCAATGT gageGTAggATGTCCCCAATGT gageGTAggATGTCCCCAATGT gageGTAggATGTCCCCAATGT gageGTAggATGTCCCCAATGT gageGTAGGATGAGGTAGGAGGTAGAAGAGTAGGTAGGAGTGAGGTAGGAGG	ggtAC tgggACCGCAGGGTCCTGGGAG agtaGGTCGGTGCGG	a a c c A C C C G T T C C G A A C G C G A A C G C C C C C C C	<b>AATCCCCCGTGCCCTTAGCGGCGTGG</b>	ati
<pre>m Anrecoconfecentinoconseries area construction to provide a production of a product of a p</pre>	<ul> <li>INTECECCENTECCENTECCENTECCENTECCENTECGENERGENERGENERGENERGENERGENERGENERGEN</li></ul>	In TGTC tggedGccatagegGdytggTCC acctGATC cat geog aact cagaagtgaaacGTTGTAAgegecgatGA tggtG tgggGTCTCCCCATGT gagaGTagggAATG6CCA6GCI In TGTC tggedGccatagegTTatggACC acctGAT cectTgceg aact cagagtgaaacGTAATAGegecgatGG tagfG tgggGTCTCCCCATGT gagaGTaggAATG6CCA6GCI In TGT tggeGAccatagagGTttggACC acctGATTCCATTCCAATGCAAATGAGAGTGAAAGCGAATAGGGCGAGGG TGGT tggeGAccatagagGCTttg gaaccactGATC cat eceg aact cagtagtgaaacGATGGCAtcgccgatGG tagfG tgggGTTTCCCCATGT gagaGTaggAATG6CCA6GCi In TGT tgacGAccatagagGCTttg gaaccactGATC ccat eceg aact cagtagtgaaacGATGGCAtcgccgatGG tagfG tgggGTTTCCCCATGT gagaGTaggtCATGGAGGC In TGT tgacGAccatagagG gttg gaaccactGATC ccat eceg aact cagtagtgaaacGATGGCAtcgccgatGG tagfG tgggGTTTCCCCATGT aagaGTaggtCATGGAGGC In TGT tgacGAtcatagagG gttg gaaccacctGAT ccat eceg aact cagtagtgaaacGACGCAtcgccgatGG tagfG tgggGTTTCCCCATGT gagGTaggtCATGGAGCCAAGC CT tgacGAtcatagagG gttg gaaccacctGAT ccat eceg aact cagtagtgaaacGACGCAtcgccgatGG tagtG tgggGTTTCCCCATGT gagGTaggtCATGCAAGC CT tgacGAtcatagaGC gttg gaaccacctGAT ccct eceGaact cagaagtgaaacGACGCAtcgccgatGG tagtG tgggGTTTCCCCATGT gagGTaggtCATGCAAGCCAAGCCAAGCCAAGCGCAAGGCCCCAAGC CT tggcGAtcatagaGC gttg gaaccacctGAT ece accg aact cagaagtgaaacGACGCAAtcgccgatGG tagtG tgggGTTTCCCCATGT gagGTaggtCATGCAAGCCAAGCAAGCACAAGCACAAGCAAAGACTAACAAGACTAACAAGAACTAACAAGACTAACAAGAACTAACAACAAAGACCAAGAGCAAGAACAAATGAAGAACTAACAACAAATGAAGAACTAACAAAGAACTAACAAAAGAACTAACAAAGAACTAACAACAAAAGAACTAACAACAAACA	ggtAC tgcgTCTTAAGGCGTGGgagaGTaggtCGCCGCCAGGC	acacCCGATCCCA teccGAACTCGGCCGTGAAAGAGGCCCTGCGCCAA to	rescreetestcatteceectcaa	Ę
<pre>um reconstructures accontrocant reconstruction accontendence control to get the registre transaction age accontendence acco</pre>	<pre>un rescrictorent mecoancon accedance accedance accedance approximation to accedence accedance accedan</pre>	In TGFC tggeGGceatageGGAbtgGTCC acc16ATC cat gecg aact cagaatgaaacGTTGThAgegecgatGA tggtG tgggGTTCCCCATCT gagaGTaggAATGCCAAGGC, In TGTC tggeGAcceatageGATTCCCAATCCATTCCAATCAAGTGAAGAATTAAGGEGAAAAGACAATTAAGGECGAGGC, TGTC tggeGAcceatageGATttggACCC acc16ATTCCAATTCCAATCAAAGTGAAAGCAATTAAGGECGAGGC, TTCT tgacGAcceatageGATttggACCC acc16ATTCCAATTCCAATTCCAAAGTGAAGCAATTAGGEGCGAGGC, TTCT tgacGAcceatageGCATtg aacc acc16ATC ccat cccg aact cagtagtgaaacGATGGCAAtcgecgatGG tagtG tgggGTTTCCCCATGT gagaGTaggAATTGCCAAGC, TTCT tgacGAcceatageGCAttg aacc acc16ATC ccat cccg aact cagtagtgaaacGATGGAAtcgecgatGG tagtG tgggGTTTCCCCATGT gagaGTaggtCATCGTCAAGA, TTCT tgacGAcceatagegCAttg aacc acc16ATC ccat cccg aact cagtagtgaaacGATGGAAtcgecgatGG tagtG tgggGTTTCCCCATGT gagaGTaggtCATCGTCAAGA, TTCT tgacGAcceatagegC ttg gaacCacceATC cccat acc7 aact cagaattgaaacGACGCAAtcgecgatGG tagtG tgggGTTTCCCCATGT gagaGTaggtCATCGTCAAGA, TTCT tgacGAcceatagegC ttg gaacCacceATC cccat cccg aact cagaattgaaacGACGCAAtcgecgatGG tagtG tgggGTTTCCCCATGT gagaGTaggtCATCGCTCAAGA.	act GCGTCAAAAGACGTGG gagaGTaggt CACCGCCAGAC	aac acccGATC ccat cccg aactCGGC cgttaagtGCCGTAgcgccaatGG t	STC tggtGGccaaagcaCGAGCA	
<ul> <li>cric tigridecensagescinder and according cost act for transport in the second of the second control of the second cost of the second</li></ul>	<ul> <li>GTC tygtGGCCAAAGCGGTG aar secoGATC cat cag aactGGC gttaagtGCCCTAggGGCCAAGCGTGG gaggTGGGCCCAGGC</li> <li>MTCCCCCTGGCCTGGAA</li> <li>MATCCCCCTGGCGTGGA</li> <li>MATCCCCCTGGGTGGA</li> <li>MATCCCCCTGGTGGGA</li> <li>MATCCCCCTGGGTGG</li> <li>MATCCCCCTGGGTGG</li> <li>MATCCCCCTGGTGGGG</li> <li>MATCCCCCTGGGTGG</li> <li>MATCCCCCTGGGTGG</li> <li>MATCCCCCTGGTGGGG</li> <li>MATCCCCCTGGTGGGG</li> <li>MATCCCCCTGGTGGGG</li> <li>MATCCCCCTGGTGGGGG</li> <li>MATCCCCCTGGTGGGGG</li> <li>MATCCCCCTGGTGGGGG</li> <li>MATCCCCCTGGTGGGGG</li> <li>MATCCCCCTGGTGGGGG</li> <li>MATCCCCCTGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGG</li></ul>	In TGFC tggeGGecatageGAtgGTCC acctEATC cat casaagtgaaacGTTGTAggeGeGatGA tggtG tggGTCTCCCATGT gaaGTaggAACTGCCAGGC I TGT tggeGAccatageGTTatggACC acctEAT ccat castagtgaaacGTAATAggeGGGGAGG tagtG tggGTCTCCCATGT gaaGTaggACTCGCCAGGG I TGT tggeGAccatageGTTatggACC acctEATC cat castagtggaaacGTAATAGACTAATAGGCGA tggGTCTCCCATT ggGGCGGGGGTG I TGT tggeGAccatageGATtggACC acctEATCCATTCCAATCAATAGGATGAATGGACGA tggG tggGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	cast 6 togoGTCTCCCCATGT gagaGTaggtCATCGTCAAGC	aacCacctGAT ccct tcccGaact cagaagtgaaacGACGCAtcgccgatGG t	TGCT tracGAtcatagage otto	1
<pre>intert typedetentaged drig gaactaactelin continedaatt gaagatgaastackedintercented tagte typedintercentraceast accounter agaatgaaf agagtentercentraceast accountered tagtent upgtentercentraceast agaatgagtentercentraceast accountered tagtentercentraceast agaatgagtentercentraceast accountered tagtentercentraceast agaatgagtentercentraceast agaatgaggtentercentraceast agaatgaggetentercentraceast agaatgaggetentercenteraceast agaatgaggetentercentercentercenterc</pre>	g) fict tystekteatagic grtf gaaccacted cases of astrono extraordisecting tagic triggered trig	ar TGTC tggcGGccatagcGCAgtggTCCC acctGATC ccat gccg aact cagaagtgaaacGTTGTAgcgccgatGA tggtG tgggGTCTCCCCATGT gagaGTagggAACTGCCAGGC. c TGCT tggcGAccatagcgTttggAACCC acctGAT ccctTgccg aact cagaagtgaaacGTAATAgcgccgatGG tagtG tgggGTCTCCCCATGT gagaGTaggAAATGCCAGGC. c TGCT tggcGAccatagcgTttggAACCC acctGAT ccctTgCGAAGGCAAGAGCAATTAGcGccGAAGG tagtG tgggGCTTCCCCATGT gagaGTaggAATGCCAGGC. a TGCT tggcGAccatagcgTttgGACCC acctGATC ccatTgccg aact cagtagtgaaaCTAATAGcGccGAGG tggtGCTtgGGGTGTGAGAG. a TGCT tggcGAccatagcgTttgGACCC acctGATC ccat cccg aact cagtagtgaaaCGTATAGCGCCGA tggtGGTtgggGCTTCCCCATG gagaGTagtCATGGCAAGC. TGCT tggcGAccatagcgTttgGAT ccctGATC ccat cccg aact cagtagtgaacGTGCGCAGG tagtG tggtGTTCCCCATGT gagaGTagtCATGGCAAGC.	adtG tqqqGTTTCCCCATGT aaqaGTaqqtCATCGTCAGGC	aacCacctGATC ccat cccg aact cagcagtgaaacGACGCAtcgccgatGG t	TGCT tqacGAtcatagagC qttq	e
<pre>m root tacktalagag tit gaacaccith cart corg aatt agaargaacAxochtorcottor tagt taggrinteconners agaargage Antonic taggrinteconners agaargage according taget taggrinte acgording taget taggrinteconners agaargage according taget taggrinteconners agaargage according taget taggrinteconners agaargage according taget taggrinte acgording taget taggrinteconners agaargage according taget taggrinte acgording taget taggrinteconners agaargaged taget taggrinte acgording taget taggrinteconners agaargage according taget taggrinte acgording taget taggrinteconners agaargaged taget taggrintege agairagged taget taggrintege according taget taggrintege actord agarding taget taggrintege according taget taggrintege according taget taggrintege according taget taggrintege according taggrintege according taggrintege according taget taggrintege according taggrintege ac</pre>	<pre>me FCC tgackteatageg gtg gaaccaccidar ceat corg aact capacyaaacGACCAtegocgatGs tagts tgggGTTCCCATTG gaagTagtcATCCCATCA ff fCC tgackteatageg gtg gaaccaccidar ceat corg aact capacyaacGACCAtegocgatGs tagts tgggGTTCCCATTG gaagTagtcATCCCATCA marcCCCGTCCCATCCATCCCATCCCAACCGCGATCACCACCACCACCACCACACA marcCCCGTCCATCCATCCCAATCCCAATCCCAATCCCCACCCCCACACA marcCCCGTCCCATCCAATCCCAATCCCAATCCCAATCCCCACCCCCACACA marcCCCGTCCCATCCAATCCCAATCCCAATCCCAATCCCCACCCCCACACA marcCCCGTCCCATCCCAATCCCAATCCCAATCCCCACCCCCCACACACACACCAC</pre>	ir TGTC tggedecetagegAptggTCCC acctGATC cet gocg aect cagaagtgaaacGTTGTAgegecgaicA tggtG tgggGTCTCCCCATGT gageTaggAACTGCCAGGC. 5 TGCT tggedecetagegTTatggACCC acctGAT ccctTgccg aact cagtggtgaaacGTAATgegecgatG tagtG tgggGTCTCCCATGT gageTaggACATCGCAGGC. 3 TGCT tggedecetagegATttggACCC acctGACTTCCAATTCGAAACTAATGAGGATGAAATGAGCAATAGCGCAGC tggtAGTG5GGAGCCATG gggGCAAATG	<pre>:agtG tgggGTTTCCCCATGT gagaGTaggtCATCGTCAAGA</pre>	acc acctGATC ccat cccg aact cagtagtgaaacGATGCAtcgccgatGG to	<pre>TTCT tgacGAccatagagCAttgg </pre>	
<pre>TTCT tgacdkreatagagC fttg gaaccactGAT ceat ceap aact acquistgaaacGATCATCATCATGAT GagTTTTCCCATFT gaaCTagTCATCATCATCA FTCT tgacdkreatagagC fttg gaaccactGAT ceat ceap aact acquistgaaacGATCATCATCATCATCATTTATTCATTCATTCATCATTTATTCATCA</pre>	<pre>TTCT tpacketaragag(ttg gaaccactGAT ccat cceg aat aggaaggaacGACCAtegecge1G tag1G tgggTTTCCCAATG gagGTTGCCAAGG TGG tgaGGAteatagag(gtg gaaccactGAT ccat cceg aat aggaagGAGCCAAtegecg1G tag1G tgggTTTCCCAATG gagGTaggtCATCGCAAAGC TGC tgaGGAteatagag(gtg gaaccactGAT ccat cceg aat aggaagGAGCCCAAtegecg1GG tag1G tgggTTTCCCAATG gagGTaggtCATCGCAAAGC TGC tgaGGAteatagag(gtg gaaccactGAT ccat cceg aat cagaaggaagGAGCCCAAtegecGAGG tag1G tgggTTTCCCAATG TATCCCGGTGCAGACG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGGTGCGAGCG TATCCCGTGCGAGCG TATCCCGTGCGAGCG TATCCGGTGCGAGCG TATCCCGTGTGCGAGCG TATCCCGTGCGAGCG TATCCGTGTGCGAGCG TATCCGTGTGCGAGCG TATCCGTGTGCGAGCG TATCCGTGTGCGAGCG TATCCGTGTGCGAGCG TATCCGTGTGCGAGCG TATCCGTGTGCGAGCG TATCCGTGTGCGAGCG TATCCGTGTGCGAGCGGTGGG TATCCGTGTGCGAGCG TATCCGTGTGCGAGCGGTGGG TATCCGTGTGCGGGGGGGGG TATCCGTGTGCGAGCGGGGGGGGGG</pre>	ar TGTC tggcGGccatagcGQtgggTCCC acctGATC ccat gccg aact cagaagtgaaacGTTGTAgcgccgatGA tggtG tgggGTCTCCCCATGT gagaGTagggAACTGGCAGGC : TGCT tggcGAccatagcgTTatggACCC acctGAT ccctTgccg aact cagtagtgaaacGTAATAgcgccgatGG tagtG tgggGTCTCCCCATGT gagaGTaggaCATGGCCAGGC	<pre>:ggtAGTGtgggGCTTCCCCAT gtgaG agtaGGACATCGCCA</pre>	C acctGACTTCCGAACTCAGAAGTGAAACGAATTAGCGCCGA to	TGCT tggcGAccatagcgATttggAC	-
a for legechecategentry parce accidant controchartnaccords to geget are degradent geget are control of agar agar agar adartactor as a control of agar agar agar and control of agar agar agar agar agar agar agar aga	a ricci tyacakacatagad quiq gaaccacciant cast cast astransforcakactakan hytococcata tyo hytorecon tyo as a partaget Arroccan to a partage	ar TGTC tggcGGccatagcgCAgtggTCCC acctGATC ccat gccg aact cagaagtgaaacGTTGTAgcgccgatGA tggtG tgggGTCTCCCCATGT gagaGTagggAACTGCCAGGC	<pre>:agtG tgggGTCTCCCCATGT gagaGTaggaCATCGCCAGGC</pre>	C acctGAT ccctTgccg aact cagtagtgaaacGTAATAgcgccgatGG to	TGCT tggcGAccatagcgTTatggAC	0
<ul> <li>Tick: 1999chceatagoghttggacc</li> <li>Tick: 1995chceatagoghttggacc</li> <li>Tick: 1995chceatagoghttggager</li> <li>Tick: 1995chceatagoghtggager</li> <li></li></ul>	<pre>c TGCT tgredkerategageTracgageTracgageTracgageTracceAnTrackA</pre>		:ggtG tgggGTCTCCCCATGT gagaGTagggAACTGCCAGGC	C acctGATC ccat gccg aact cagaagtgaaacGTTGTAgcgccgatGA to	TGTC tggcGGccatagcgCAgtggTC	ar

# **Nucleic Acids Research**

from the collection of Olsen and Pace. They can be found in GenBank or the review paper of Erdmann et al.(18).

An alignment of these sequences by algorithm S with k-4 and W-6, allowing up to 1 mismatch is given in Fig. 1. The parameter  $\lambda_d = (k-d)/k$ , where d equals the number of mismatches. Algorithm T which does not allow overlapping windows has score 304.75, while algorithm S has score 364.00. Some scores for various choices of the parameters are given next:

W = window width	<u>k = word size</u>	<pre># mismatches</pre>	<u> </u>
6	4	0	117.00
6	4	1	364.00
8	4	0	196.00
8	4	1	487.25
8	6	0	23.00
8	6	1	124.00
8	6	2	269.17
10	6	0	82.00
10	6	1	213.67
10	6	2	329.00

# DISCUSSION

The existing multiple sequence alignment programs based on our algorithm are written for DNA sequences, but they can easily be extended to proteins. (We plan to do this.) The largest protein word possible is k-3, since  $20^3$ -8,000, while words with k-8 or 9 is possible with DNA. Whenever the amino acid alphabet is decreased from the usual 20-letter alphabet, k can be increased.

Another feature we have not included in our program is a cost for unmatched letters (deletions/insertions). This could easily be done as Martinez does, for example, but we are not yet persuaded of the necessity here.

Statistical significance is always an issue in sequence alignment. Frequently the sequences are randomly permuted, and the optimal alignment scores obtained from the random sequences. Statistical significance is estimated from these scores. Here we have a more direct approach. It is possible to calculate the probability that a word occurs in the specified neighborhood and in the window of one sequence. Then the binomial distribution or the theory of large deviations (for large R)(16) can be used to calculate the probability of ever seeing a matched word in L out of the R sequences. By requiring this probability to be small (e.g., .01), we can be assured that every matched word in an alignment is significant at that level of significance.

The implementation for DNA sequences is written in the C language and is available from the author. Please inquire for details.

## ACKNOWLEDGMENTS

The 5S sequences were received from Gary Olsen and Norm Pace. Assistance from Mark Eggert and Felicitas Smith is greatly appreciated.

This work was supported by grants from the System Development Foundation and the National Instutes of Health(GM 36230).

### REFERENCES

- 1. Needleman, S.B. and Wunsch, C.D. (1970). J. Mol. Biol. 48, 444.
- 2. Sellers, P. (1974). SIAM J. Appl. Math. 26, 787.
- 3. Waterman, M.S., Smith, T.F., and Beyer, W.A. (1976). Adv. Math. 20, 367.
- 4. Waterman, M.S. (1984). Bull. Math. Biol. 46, 4, 473-500.
- 5. Sankoff, D. (1975). <u>SIAM J. Appl. Math.</u> 78, 35-42. 6. Sankoff, D., Cedergren, R.J., and Lapalme, G. (1976). <u>J.</u> <u>Mol. Evol.</u> 7, 133-149.
- 7. Waterman, M.S. and Perlwitz, M. (1984). Bull. Math. Biol. 46, 4, 567-577.
- (1976). J. of the Assoc. for Computing 8. Wagner and Fischer. Machinery 23, 50-57.
- 9. Itoga, S.Y. (1981). BIT 21, 20-30.
- 10.Hsu, W.J. and Du, M.W. (1984). <u>BIT</u> 24, 45-59. 11.Sobel, E. and Martinez, H. (1986). <u>Nucl. Acids Res.</u> 14, 1, 363-374.
- 12.Martinez, H. (1983). <u>Nucl. Acids Res.</u> 11, 4629-4634. 13.Fredman, M.L. (1984). <u>Bull. Math. Biol.</u> 46, 4, 553-566. 14.Murata, M., Richardson, J., and Sussman, L. (1985). <u>Proc.</u>
- Natl. Acad. Sci. USA 82, 3073-3077.
- 15. Johnson, M.S. and Doolittle, R. (in press). <u>J. Mol. Evol.</u> 16. Waterman, M.S., Arratia, R., and Galas, D. (1984). <u>Bull.</u> Math. Biol. 46, 515-527.
- 17.Galas, D.J., Eggert, M., and Waterman, M.S. (1985). J. Mol. Biol. 186, 117-128.
- 18. Erdmann, V.A., Wolters, J., Huysmans, E., and R. de Wachter. (1985). <u>Nucl. Acids Res.</u>, 13, r105.