1 Introduction

2 Insertion box

For insertion variants the algorithm takes the last filled column before the insertion's position and speculatively continues the computation of the alignment on the insertion string. For the reference base after the insertion's position, the algorithm decides, whether he uses the insertion string for optimization or just the reference string. The relevant predecessors of each entry are indicated by arrows.

3 Deletion box

Deletion variants can be handled by skipping the affected columns for the computation. For each entry the algorithm decides, whether he uses the previous reference base or the reference base before the deletion variant. For the sake of simplicity some arrows are left out for this variant and only indicated by the large arrow above the matrix.

4 SNP box

The algorithm uses the IUPAC alphabet for the reference string, which allows it to encode arbitrary SNPs within the reference sequence. The semantic of the IUPAC chars is considered when comparing reference bases with read bases.

5 Example

Without variants:

Reference:	A	G	A	A		C	T	A	A	T	G
Read:				A		C	G	C	T	T	

Using variants:

Reference:	A	G	A	A	C	G	C	T	A	A	T	G
Read:				A	C	G	C	T			T	

6 NW-Formel

The NW-matrix M is calculated column by column. Each entry is calculated by the recursion formula

$$M_{i,j} = \min \begin{cases} M_{i-1,j-1} & +\text{dist}(p_i, t_j) \\ M_{i-1,j} & +1 \\ M_{i,j-1} & +1 \end{cases}$$

where p_i is the i-th char of the read und t_j the j-th char of reference section.