

# Figure S1. Score-based Clustering Strategy

## Children

### 1. a. Backwards:

A shorter k-mer is a **parent** of a longer k-mer if it:  
 (1) is a subsequence and  
 (2) scores higher than the longer k-mer

### b. Forward:

A shorter k-mer is a **child** of a longer k-mer if it:  
 (1) is a subsequence,  
 (2) scores lower, and  
 (3) it is not already a parent from (1a).

## Parents

2) Families are combined if they share more than half of their children

The higher scoring parent is the new parent

## Collapsed Parents

### 3) Families are combined if:

(a) Parent sequences are similar  
 (b) Children of the lower scoring parent match the sequence of the higher scoring parent

## Final Parents: Core Motifs

4) If a child has more than one final parent, it is uniquely associated with the higher scoring parent

### Legend:

52 T T G G T T

TTGGTT: Conserved element  
 52: Over-enrichment chi-square value

