

Greedy Info

Christian Otto

The greedy approach to motif discovery is very straight forward. We begin by randomly selecting a motif of length k in each of the DNA sequences (in our case the first motif in each sequence) and calculate the score. We then take a k -mer in the first DNA sequence and calculate the profile-matrix of this one motif. We then use this profile-matrix to choose a motif in the second DNA sequence and update the profile-matrix based on these two chosen motifs. We then continue this process for the remaining DNA sequences. After we have a set of motifs, we calculate the score and compare it to the current best collection of motifs. If this collection of motifs has a lower score than the previous collection, we set the lower scored motifs as the best motifs. We then repeat the process starting with a different k -mer in the first DNA sequence. We continue this until we have repeated the process for all k -mers in the first DNA sequence. At the end we keep the lowest scoring collection of motifs as our estimated motifs for the DNA sequences. (Compeau and Pevzner 2018)

References

Compeau, Phillip, and Pavel Pevzner. 2018. *Bioinformatics Algorithms: An Active Learning Approach*. 3rd ed. Active Learning Publishers.