1   Introduction

Consensus sequences are sequences of nucleotides (on DNA and RNA) or amino acids (in protein), which are comprised of the most commonly encountered letters at that postion (Pierce, 2012). These sequences are generally associated with inter- or intramolecular interactions (Liljas, 2001). A prime example is the Shine-Dalgarno (SD) sequence in prokaryotes, which is involved in the binding of the ribossome to the mRNA.

Even though these sequences are highly conserved, they do present some variations in between organisms and thus, mearly writing out the consensus sequence leaves out information about the frequence and variation of each of the nucleotides or aminoacids in the sequence, which can be very important in Molecular Biology or Bioinformatics analysis.

Let's take the example of the aforementioned consensus sequence. In *E. coli* the SD sequence is **5'-AGGAGG-3'**. This sequence, however, has been shown to have slight variations in different bacteria (Ma, Campbell, & Karlin, 2002) and thus a bioinformatician, when programming a gene prediction tool, may be misled to believe that that's the full extent of the SD sequence, when it's not. Moreover, one may argue that studying the variety of consensus sequences is important in undertanding intermollecular interactions.

We then arrive at the logical conclusion that the visualisation of consensus sequences plays a very important role in various Molecular Biology and Bioinformatics studies and is therefore relevant to study better ways to display the available data in a way that's informative, visually appealing and easy to understand.

2   Sequence logos

Faced with the challanges of creating a visual representation of consensus sequences that provides more information than just the sequence of letters, Schnider and Stephens came forward with a proposal, for which they called Sequence Logos
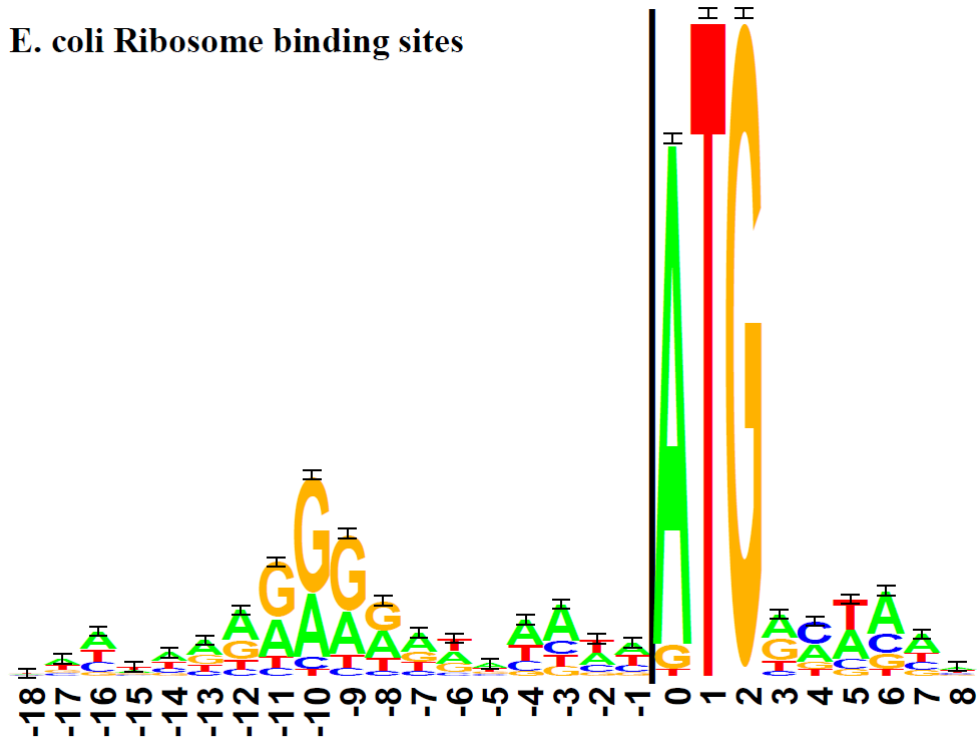
E. coli Ribosome binding sites

Figure 1: The Sequence Logo representation of the *E. coli* binding site (Schneider & Stephens, 1990, p. 2).

(Fig. 1) (Schneider & Stephens, 1990).

References

Liljas, A. (2001). Consensus Sequence. In *Encyclopedia of Genetics* (pp. 457–458). Elsevier. doi: 10.1006/rwgn.2001.0270

Ma, J., Campbell, A., & Karlin, S. (2002, October). Correlations between Shine-Dalgarno Sequences and Gene Features Such as Predicted Expression Levels and Operon Structures. *Journal of Bacteriology*, *184*(20), 5733–5745. doi: 10.1128/JB.184.20.5733-5745.2002

Pierce, B. A. (2012). *Genetics: A conceptual approach* (4th ed ed.). New York: W.H. Freeman.

Schneider, T. D., & Stephens, R. (1990). Sequence logos: A new way to display consensus sequences. *Nucleic Acids Research*, *18*(20), 6097–6100. doi: 10.1093/nar/18.20.6097