



**Figure 2.** Properties of binding motifs for bacteria, yeast and multicellular eukaryotes. (a) The bar chart displays the minimum required information content for bacteria, yeast and multicellular eukaryotes (red), and the mean information content of TF-binding motifs (blue) for 98 bacterial [21], 124 yeast [22] and 123 multicellular [20] eukaryotic motifs. The error bars are  $\pm 1$  SD for the information content and for  $I_{min}$  the error bars represent the variability in that quantity because of the range of genome sizes  $N$ . The blue circles on top of the bar for bacteria indicate the average information content from several other TF-binding motif databases (Table S6). Below each series in the bar chart, we display an example of the sequence logo for a binding motif with close to average information content. The chart demonstrates that bacterial TF-binding motifs are informative enough to make spurious hits to the genomic background unlikely, in contrast to yeast and multicellular eukaryotic motifs. (b) The distributions of information content of motifs from the three representative databases cited above. The ranges of required information ( $I_{min}$ ) are marked in red. Most bacterial motifs have  $I > I_{min}$ , whereas almost all eukaryotic motifs do not. (c) The average properties of TF-binding motifs, and the expected number and spacing between the spurious sites per genome in bacteria, yeast and multicellular eukaryotes.

- 20 Vlieghe, D. *et al.* (2006) A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res.* 34, D95–97
- 21 Kazakov, A.E. *et al.* (2007) RegTransBase – a database of regulatory sequences and interactions in a wide range of prokaryotic genomes. *Nucleic Acids Res.* 35, D407–412
- 22 Zhu, C. *et al.* (2009) High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.* 19, 556–566