

Figure 1. Mode-1 HOSVD of the 16S rRNA sequence alignment. Organisms, positions and sequence elements, each represent a degree of freedom in the alignment encoded in a cuboid (Equation 1). Mode-1 HOSVD (Equation 2) separates the alignment into combinations of “eigenpositions” and nucleotide-specific segments of “eigenorganisms,” i.e., patterns of nucleotide frequency variation across the organisms and positions, with increase (red), no change (black) and decrease in the nucleotide frequency (green) relative to the average frequency across the organisms and positions. It was shown that SVD provides a framework for modeling DNA microarray data [32]: The mathematical variables, significant patterns uncovered in the data, correlate with activities of cellular elements, such as regulators or transcription factors. The mathematical operations simulate experimental observation of the correlations and possibly causal coordination of these activities. Recent experimental results [16] demonstrate that SVD modeling of DNA microarray data can be used to correctly predict previously unknown cellular mechanisms [15,36]. We now show that mode-1 HOSVD, which is computed by using SVD (Equation 3), provides a framework for modeling rRNA sequence alignments: The mathematical variables, significant patterns of nucleotide frequency variation, represent multiple subgenomic evolutionary relationships of convergence and divergence among the organisms, some known and some previously unknown, and correlations with structural motifs. Our mode-1 HOSVD analyses of 16S and 23 rRNA alignments support the hypothesis that even on the level of a single rRNA molecule, an organism’s evolution is composed of multiple pathways due to concurrent forces that act independently upon different rRNA degrees of freedom. These analyses demonstrate that entire rRNA substructures and unpaired adenosines, i.e., rRNA structural motifs which are involved in rRNA folding and function, are evolutionary degrees of freedom. These analyses also show that mode-1 HOSVD modeling of rRNA alignments might be used to computationally predict evolutionary mechanisms, i.e., evolutionary pathways and the underlying structural changes that these pathways are correlated, possibly even coordinated with.

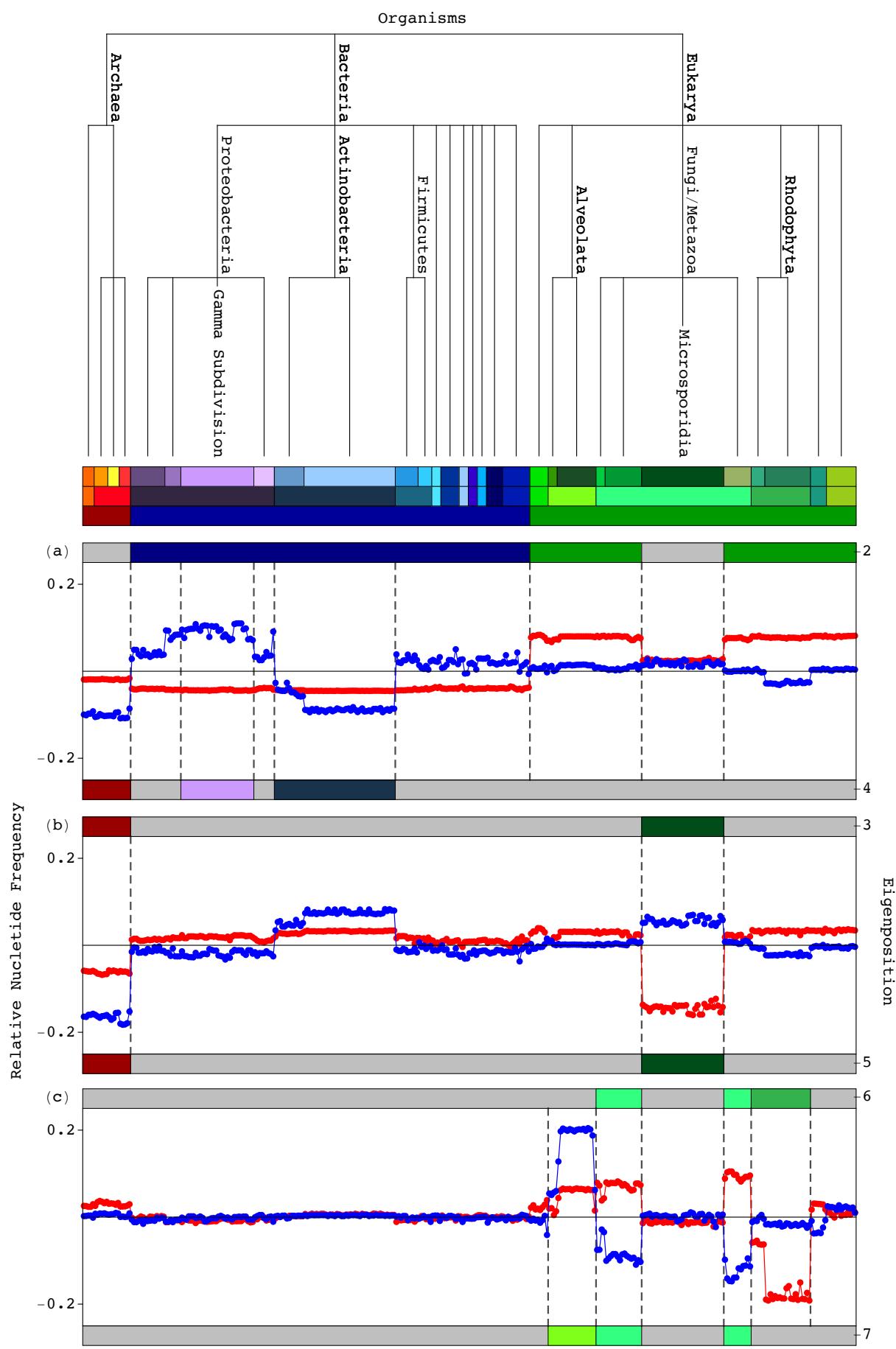


Figure 2 (captions on p. F-3).

Figure 2 (on p. F-2). Significant 16S eigenpositions. Line-joined graphs of the second through seventh 16S eigenpositions, i.e., patterns of nucleotide frequency across the organisms, and their correlation with the taxonomic groups in the 16S alignment, classified according to the top six hierarchical levels of the NCBI Taxonomy Browser [19] (Figure S1 in Supplementary Appendix S1). (a) The second most significant eigenposition (red) differentiates the Eukarya excluding the Microsporidia from the Bacteria, as indicated by the color bar (Table 1). The fourth (blue) distinguishes between the Gamma Proteobacteria and the Actinobacteria and Archaea. (b) The third (red) and fifth (blue) eigenpositions describe similarities and dissimilarities among the Archaea and Microsporidia, respectively. (c) The sixth (red) and seventh (blue) eigenpositions differentiate the Fungi/Metazoa excluding the Microsporidia from the Rhodophyta and the Alveolata, respectively.

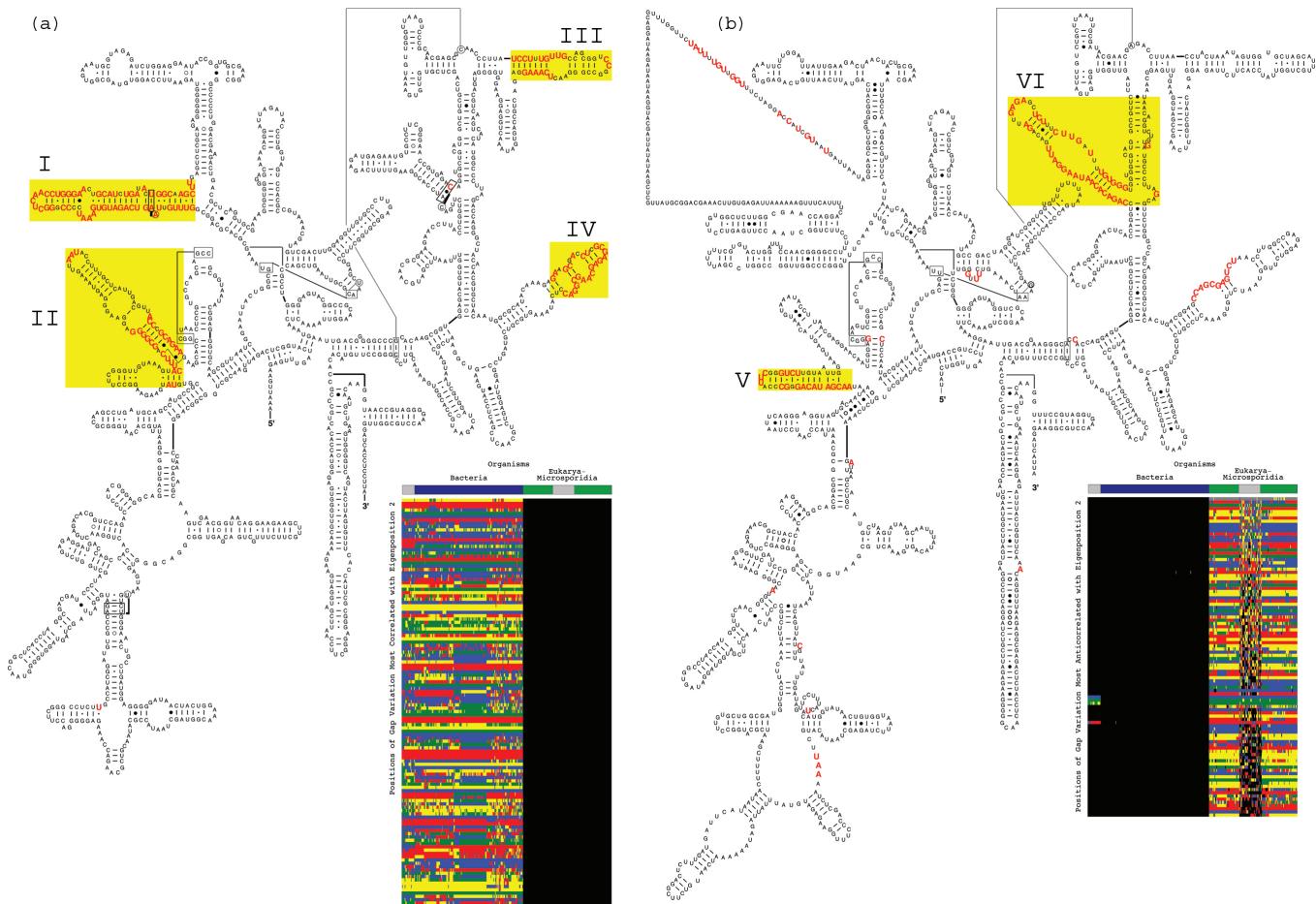


Figure 3. Sequence gaps exclusive to Eukarya or Bacteria 16S rRNAs. The second most significant 16S eigenorganism identifies gaps exclusively conserved in either the Eukarya excluding the Microsporidia or the Bacteria (Table 3) that map out known as well as previously unrecognized, entire substructures deleted or inserted, respectively, in the Eukarya relative to the Bacteria. (a) The 124 positions with largest increase in relative nucleotide frequency in the gap segment of the second eigenorganism, i.e., the 124 positions of gap variation across the organisms most correlated with the second eigenposition, map out the exclusively conserved known substructures [31] I and II and the previously unrecognized substructures III and IV in the secondary structure model of the bacterium *E. coli* [14]. These 124 positions are also displayed in the inset raster, ordered by their significance, with the most significant position at the top. The nucleotides are color-coded A (red), C (green), G (blue), U (yellow), unknown (gray) and gap (black). The color bars highlight the taxonomic groups that are differentiated by the second 16S eigenposition and eigenorganism, i.e., the Eukarya excluding the Microsporidia and the Bacteria. (b) Of the 100 positions of gap variation across the organisms most anticorrelated with the second eigenposition, 99 map out the substructures V and VI in the secondary structure model of the eukaryote *S. cerevisiae*. The 100th position is an unknown nucleotide at the 3'-end of the molecule, which is not displayed. These 100 positions are also displayed in the inset raster.

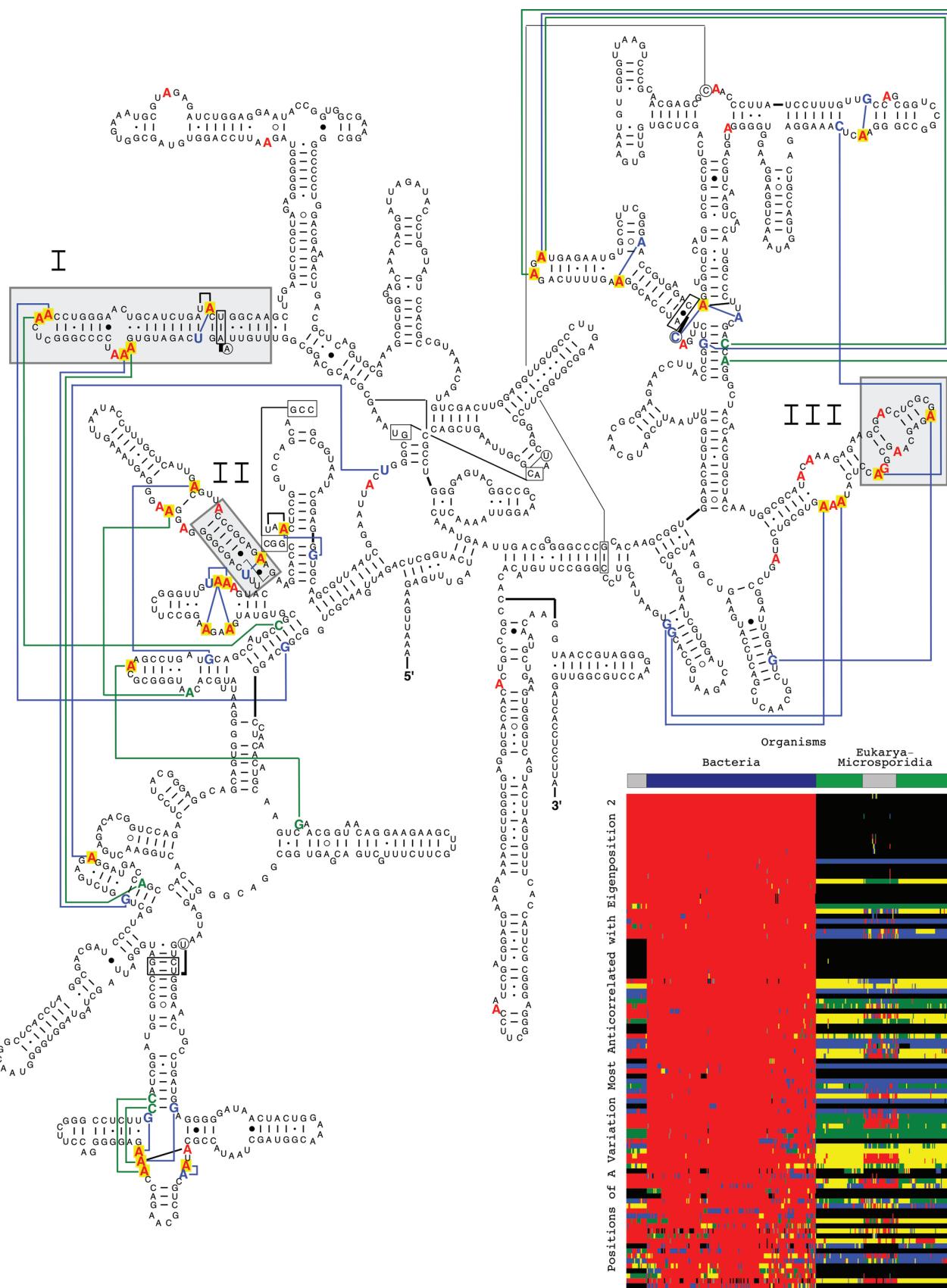


Figure 4 (captions on p. F-5).

Figure 4 (on p. F-4). Unpaired adenosines exclusive to Bacteria 16S rRNAs. The 100 positions identified in the A nucleotide segment of the second 16S eigenorganism with the largest decrease in relative nucleotide frequency include all 50 positions (red) in the alignment with unpaired A nucleotides exclusively conserved in the Bacteria. Of these 50 positions, 28 (yellow) map to known tertiary interactions in the crystal structure of the bacterium *T. thermophilus* [8], plotted on the secondary structure model of the bacterium *E. coli* [14]. These include 22 base-base interactions (blue) and eight base-backbone interactions (green). Of the 50 positions of unpaired A nucleotides exclusively conserved in the Bacteria, 13 correspond to gaps exclusively conserved in the Eukarya excluding the Microsporidia. These 13 positions map to the entire 16S rRNA substructures that are deleted in the Eukarya with respect to the Bacteria (gray), identified by the gap segment of the second eigenorganism (Figure 3). These 100 positions identified in the A nucleotide segment of the second 16S eigenorganism are displayed in the inset raster, ordered by their significance, with the most significant position at the top. The nucleotides are color-coded A (red), C (green), G (blue), U (yellow), unknown (gray) and gap (black). The color bars highlight the taxonomic groups that are differentiated by the second eigenposition and eigenorganism, i.e., the Eukarya excluding the Microsporidia and the Bacteria.

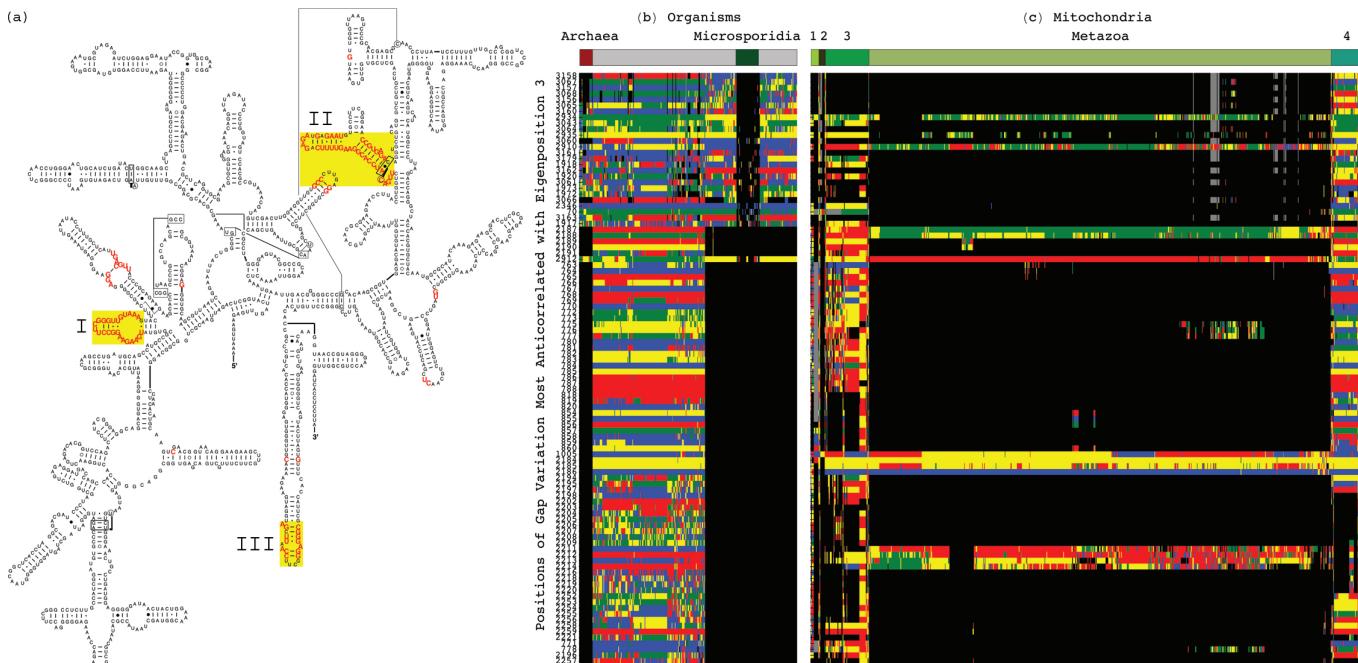


Figure 5. Sequence gaps exclusive to both Archaea and Microsporidia 16S rRNAs. The 100 positions identified in the gap segment of the third 16S eigenorganism with the largest decrease in relative nucleotide frequency map out entire substructures in the Bacteria 16S rRNAs that are convergently lost in the Archaea and the Microsporidia. (a) The 100 gaps conserved in both the Archaea and Microsporidia map to the entire substructures I–III in the secondary structure model of the bacterium *E. coli* [14]. (b) Raster display of the 100 positions of conserved gaps in both the Archaea and Microsporidia across the alignment. (c) Raster display of the same 100 positions across an alignment of 858 mitochondrial 16S rRNA sequences show gaps conserved in most Metazoa. The other groups of Eukarya represented in the mitochondrial alignment are Alveolata (1), Euglenozoa (2), Fungi (3) and Rhodophyta and Viridiplantae (4). The nucleotides are color-coded A (red), C (green), G (blue), U (yellow), unknown (gray) and gap (black). The color bars highlight the taxonomic groups.