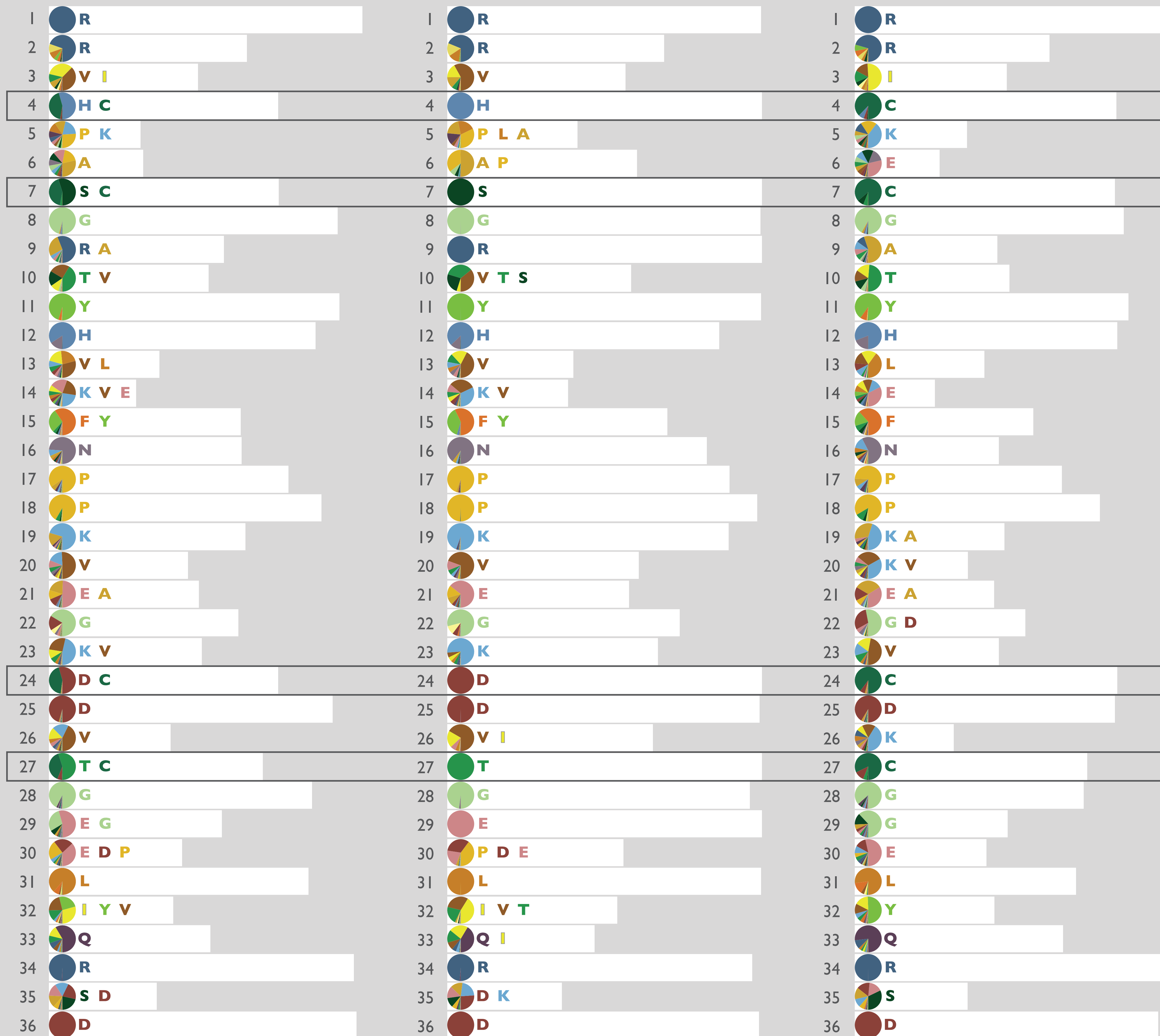


compariSeq: rethinking sequence logos

Sean McKenna
P. Samuel Quinan
Alex Bigelow

overview

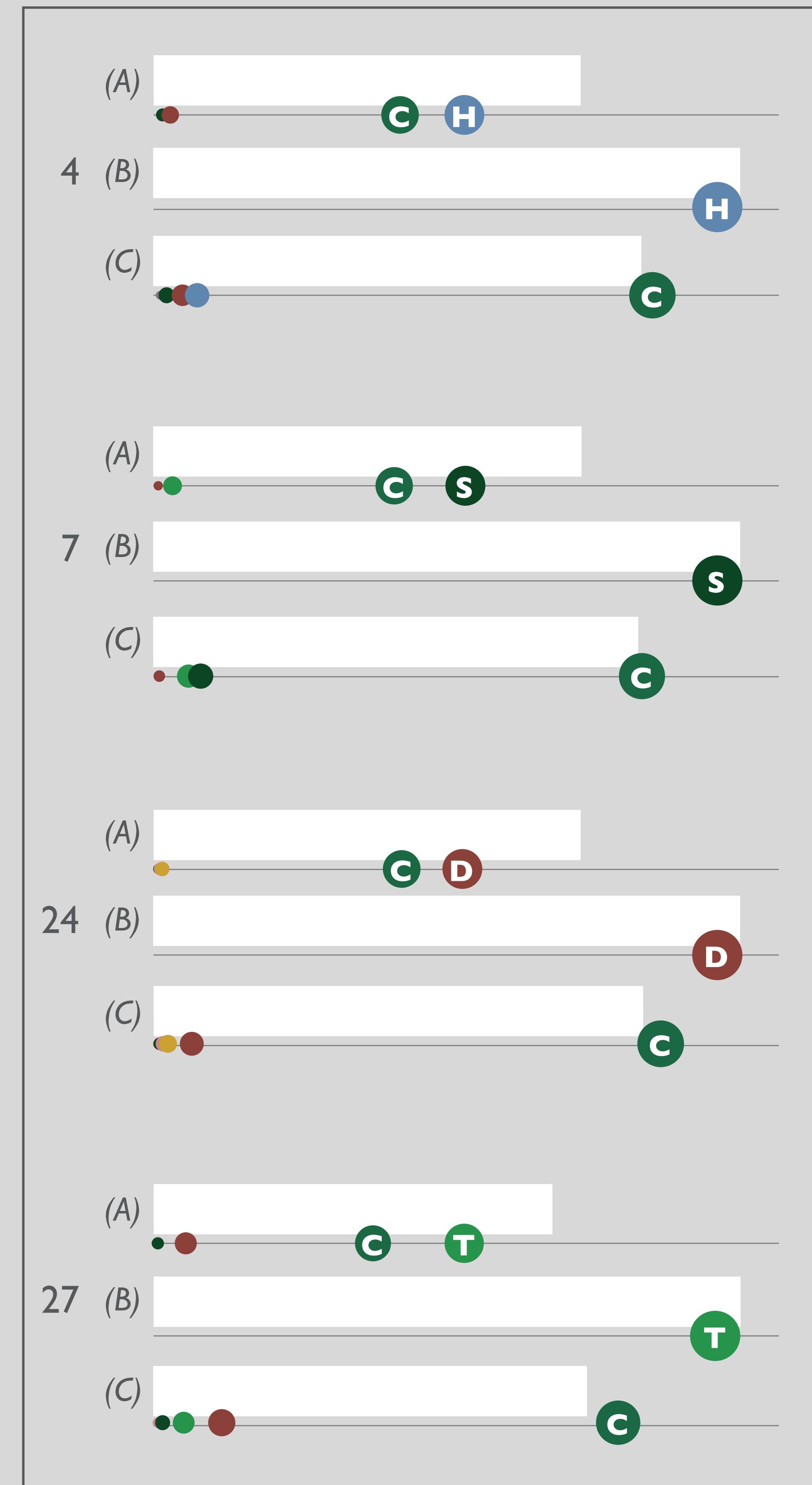


(A) all organisms

(B) gram-negative bacteria

(C) gram-positive bacteria

detail view_{linked}



introduction

Sequence logos are one of the most common methods to display biological sequences. Accurately comparing residue frequencies using sequence logos, however, is extremely difficult due to the combination and vertical stacking and scaling relative to information content. Using observations garnered from informal interviews with biologists of varying expertise, our proposed redesign, compariSeq, was specifically formulated to better support such comparison tasks.

our design

The main view of compariSeq shows the full sequence, allowing users to see general trends and locate interesting features. We vertically orient the sequences, allowing multiple sequences to be compared simultaneously. At each location, the bar length encodes conservation, the pie chart encodes relative frequencies of all amino acids, and the most common amino acids are also listed left to right. This allows *consensus sequences* to easily be constructed through a vertical scan.

We also provide a linked view that enables a direct comparison of frequency information across sequences for selected locations. Relative frequencies are encoded through the position along a common axis, while size simultaneously denotes conservation.

justification

We identified the comparison of specific locations across sequences as a primary task, observing that domain knowledge, coupled with conservation and approximate frequencies, is sufficient to determine which locations are most interesting. As such, we made conservation the most salient feature in our design, encoding it in a bar chart to enable users to quickly cue into locations of interest. Once interesting locations have been identified, users care more about comparison. The pie charts and detail views enable such comparisons while our overview also provides explicit access to the *consensus* through a vertical scan.

acknowledgements

We would especially like to thank Miriah Meyer, Heidi Schubert, Matthew Sdano, and Venkatesh Rajamanikkam for the information and feedback they provided on our design.

Our redesigned color palette associates amino acid groups with convention-informed hues, while encoding each amino acid with a perceptually distinct color

Q N Basic

D E Neutral

R H K Acidic

S C T Y G Polar

V L F A P I W M Hydrophobic