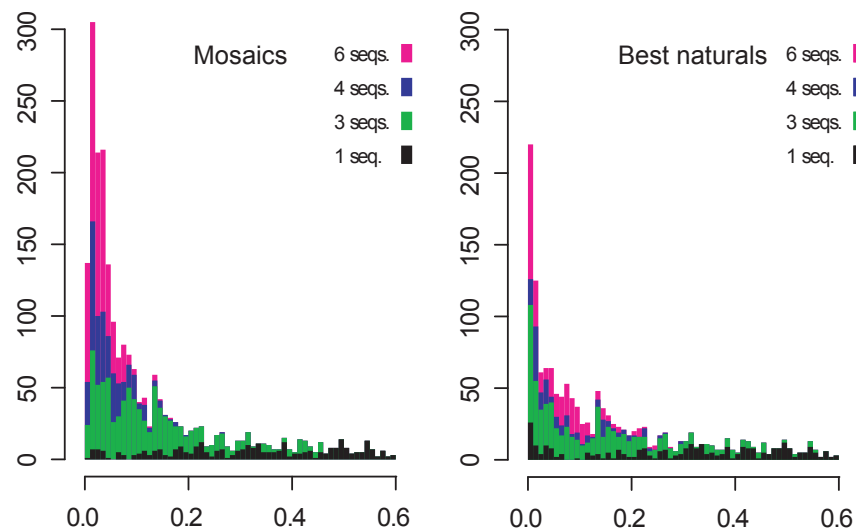
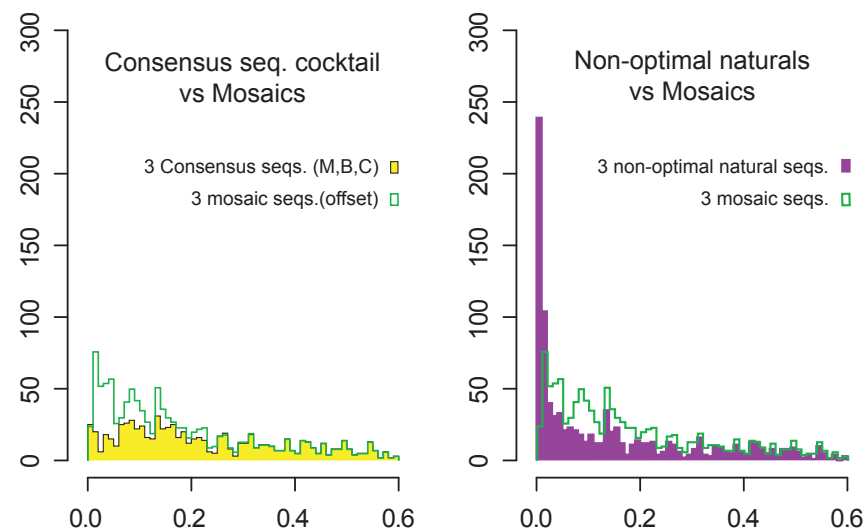


9-mer Frequencies (0–60%)

9-mer counts



9-mer counts



9-mer Frequencies (0–10%)

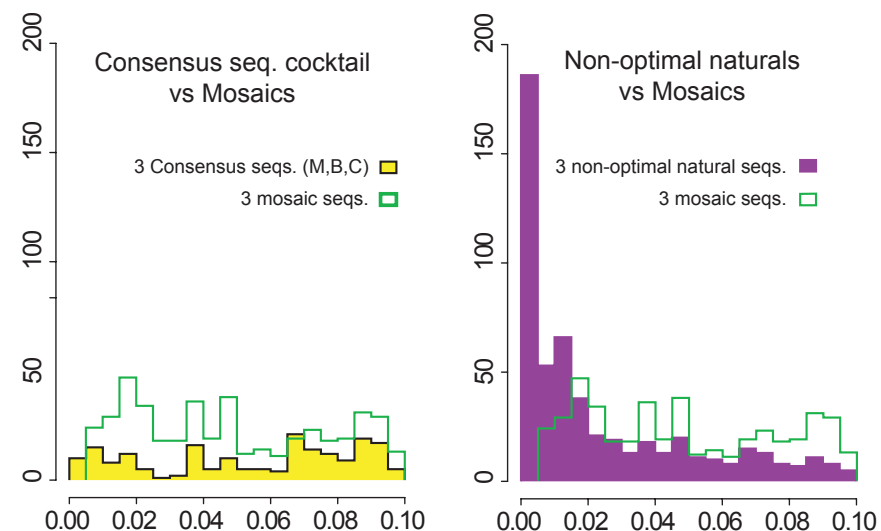
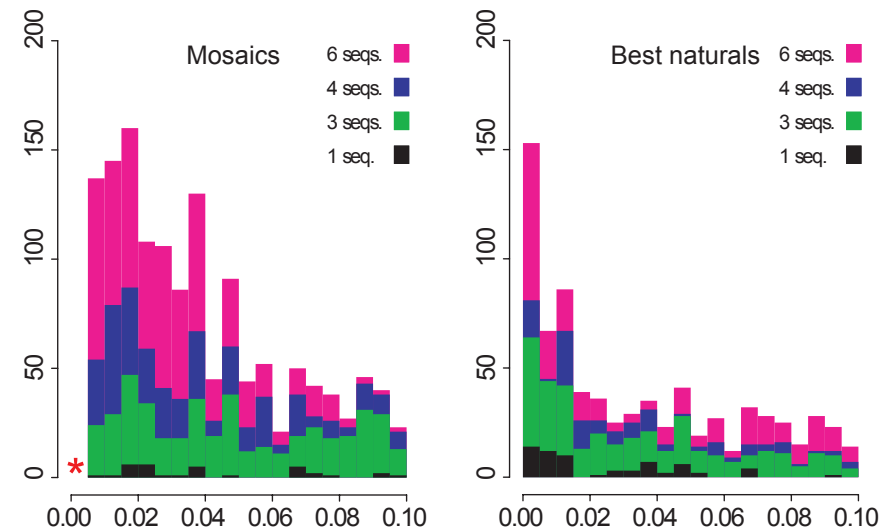


Figure S2. The distribution of nine-mers by frequency of occurrence in natural, consensus, and mosaic sequences. Occurrence counts (y-axis) for different nine-mer frequencies (x-axis) for vaccine cocktails produced by several methods. Left panel: frequencies from 0-60% (for nine-mer frequencies > 60%, the distributions are equivalent for all methods). Right panel: Details of low-frequency nine-mers. Natural sequences have large numbers of rare or unique-to-isolate nine-mers (bottom right, each panel); these are unlikely to induce useful vaccine responses. Selecting optimal natural sequences does select for more common nine-mers, but rare and unique nine-mers are still included (top right, each panel). Consensus cocktails, in contrast, under-represent uncommon nine-mers, especially below 20% frequency (bottom left, each panel). For mosaic sequences, the number of lower-frequency nine-mers monotonically increases with the number of sequences (top left, each panel), but unique-to-isolate nine-mers are completely excluded (top left of right panel: * marks the absence in mosaics of nine-mers with frequencies < 0.005).