We introduce a new model of genetic diversity which summarizes a large input dataset into an epitome, a short sequence or a small set of short sequences of probability distributions capturing many overlapping subsequences from the dataset. The epitome as a representation has already been used in modeling real-valued signals, such as images and audio. The discrete sequence model we introduce in this paper targets applications in genetics, from multiple alignment to recombination and mutation inference. In our experiments, we concentrate on modeling the diversity of HIV where the epitome emerges as a natural model for producing relatively small vaccines covering a large number of immune system targets known as epitopes. Our experiments show that the epitome includes more epitopes than other vaccine designs of similar length, including cocktails of consensus strains, phylogenetic tree centers, and observed strains. We also discuss epitome designs that take into account uncertainty about T-cell cross reactivity and epitope presentation. In our experiments, we find that vaccine optimization is fairly robust to these uncertainties.