A key computational step for resequencing data is to distinguish technical errors and genetic variation from the population. This is accomplished by utilizing manually-curated databases to recalibrate base-level quality scores reported by the machine. Here, we find that this reliance on variant databases introduces an unexpected bias and consequently leads to suboptimal results in variant calling. We present systematic guidelines for constructing a "pseudo-"database as a scalable and portable solution for any species and strain. Applying this approach for the reanalysis of human, rice, sheep, and chickpea data revealed an unforeseen diversity of the non-coding genome.