## Supporting Methods

### Statistical Analysis

To assess the statistical properties of the observed set of insertion sites, we used a neutral-candidate model in which every gene is assumed nonessential and equally likely to be hit, and a neutral-base pair model, in which every base pair is assumed equally likely to define an insertion site. The neutral-candidate model was rejected because variation in gene size had a large effect on the number of times a gene was hit. In both models, the number of times an ORF is hit follows a multinomial distribution with parameters *n*, $p_1,\ldots, p_k$ where *n* is the number of transposon insertions, $p_j$ is the probability of landing in the *j*th ORF, and *k* is the number of ORFs.

For the neutral-base pair model, the expected number of ORFs hit is

$k-\sum_{j=1}^{k}(1-p_j)^n$, and the variance is

$$k-\sum_{j=1}^{k}(1-p_j)^n-[k-\sum_{j=1}^{k}(1-p_j)^n]^2+\sum_{1\le i\ne j\le k}[1-(1-p_i)^n-(1-p_j)^n+(1-p_i-p_j)^n].$$ The parameters for this model are $n$ = 30,100 and $k$ = 5,571. All of the ORFs are included along with an extra “false ORF” that represents the entire noncoding region. The $p_j$ are determined by dividing the length of the ORF in bp by the total length of the *Pseudomonas* genome, 6,264,403 bp. For this model, the expected number of ORFs missed is 307 with a variance of 235.2 and an upper bound on the *P* value of $1.74 \times 10^{-3}$.

Based on the analytic form of the expected value of the model above, one can fit the function $f_1(n) = b_0 - b_1 \exp(-b_2 n)$ to the cumulative plot of the number of ORFs hit (Fig. 2). The parameters $b_0$, $b_1$, and $b_2$ are chosen to minimize the root-mean-square error between the function and the data. One can interpret the parameter $b_0$ as the number of nonessential genes. The best fitting model has $b_0 = 4887.3$, $b_1 = 4814.0$, and $b_2 = 1.3736 \times 10^{-4}$, and the fit is very good with a residual standard error of 45.33 on 31,086 df.

To compute the bias and variance of the estimated parameters, we assumed that parameters fitted to the observed cumulative plot have the same bias and variance as when they are fitted to the neutral-base pair model. Then the bias and variance can be estimated by a parametric bootstrap. For each simulation, we drew a simple random sample of size 30,100 from the population 1,…, 6,264,403. If $b_0^{(j)}$ is the estimate fitted to the *j*th simulation and $\bar{b}_0 = {}^1\!/_m \sum_{1 \le j \le k} b_0^{(j)}$, then $\text{Bias}\, b_0 \cong \bar{b}_0 - k$, where *k* is the number of ORFs in the genome and $\text{Var}\,(b_0) \cong {}^1\!/_{(m-1)} \sum_{1 \le j \le m} (b_0^{(j)} - \bar{b}_0)^2$. The bias corrected estimate of the number of essential genes is 377 with a standard deviation of 77.3.