Supporting information for Boldrick *et al*. (2002) *Proc. Natl. Acad. Sci. USA* **99** (2), 972–977. (10.1073/pnas.231625398)

Methods

Bacterial Strains and Stimuli. The *Bordetella pertussis* strains, including the virulent wild-type laboratory strain Bp338 (1), the clinical isolate Minnesota 1 (gift from the Minnesota State Department of Health), and isogenic mutant strains (2) derived from Bp338, Bp537 (3), BpTox6 (4), and BpA2-6 (5) were grown at 37°C on Bordet-Gengou agar (Difco), supplemented with sheep blood (13% vol/vol), and then transferred to Stainer-Scholte liquid medium (6), and grown to late logarithmic phase. For experiments with heat-killed bacteria, the liquid culture was heated to 56°C for 30 min. The other bacterial strains, an *Escherichia coli* clinical isolate and two *Staphylococcus aureus* clinical isolates, were isolated from patient blood cultures (Veterans Affairs Palo Alto Health Care System clinical lab), grown in the same medium, and heat killed for parallel experiments in the same manner as the *Bordetella* strains. *B. pertussis* lipopolysaccharide (LPS) (List Biologicals), phorbol 12-myristate 13-acetate (PMA; Sigma), and ionomycin (Sigma) were diluted in fresh Stainer-Scholte media for treatment time courses.

Cells and Media. Human primary peripheral blood mononuclear cells (PBMCs), including all circulating non-red cell and non-granulocyte populations (primarily lymphocytes and monocytes), were acquired by apheresis or from whole blood from healthy donors, and purified by using endotoxin-tested Ficoll-Paque PLUS (Amersham Pharmacia). Before stimulation, cells were cultured overnight at a density of 2×10^7 per ml in RPMI medium 1640 supplemented with 10% endotoxin-free fetal bovine serum (Life Technologies). Cells of the U937 monocyte/macrophage cell line (American Type Culture Collection) were also maintained in RPMI medium 1640 supplemented with 10% fetal bovine serum and were induced to differentiate into macrophage-like cells with 10 ng/ml PMA for 48 hr, washed with phosphate-buffered saline (PBS), and incubated in RPMI medium 1640 without PMA for an additional 24 hr before infection.

Infection Scheme. After heat killing, bacteria were diluted in their original growth medium and added to 1×10^8 PBMCs at a ratio of 4 to 0.002 bacteria per cell, as described in the figure legends. LPS was added to the same number of PBMCs at a concentration ranging from 1 μg/ml to 0.01 μg/ml, and ionomycin was added at 1 μM in combination with PMA at 25 ng/ml. For each of the two large sets of parallel time courses, at least four replicate samples were harvested at the beginning of treatment (0 hr) and subsequent samples were harvested, after incubation at 37°C, at 0.5, 1, 2, 4, 6, 12, and 24 hr. For the comparison of live versus heat-killed *B. pertussis*, the bacterial culture was split, half of the bacteria were heat killed and added to PBMCs, and the other half were added live to PBMCs; samples were harvested at 0, 0.5, 2, 4, 6, and 12 hr. At harvesting, cells were scraped from the flask, diluted in cold PBS, pelleted at $400 \times g$ for 5 min, and snap-frozen in liquid nitrogen. U937 cells were infected at a concentration of 50 live bacteria per cell for all time courses with *Bordetella* mutant and wild-type strains.

Microarray Procedures. Poly(A)$^+$ mRNA was extracted from cell pellets with the FastTrack 2.0 kit (Invitrogen). Poly(A)$^+$ mRNA from samples for the data set comparing the live versus heat-killed bacteria was amplified according to the method of Wang *et al.* (7). The microarrays employed for the bacterial diversity and dose-response data sets comprised 18,432 cDNAs, representing ≈7,619 unique genes (estimated based on NCBI UniGene build no. 136), and were constructed as previously described (8). Fluorescently labeled cDNA probes were prepared by incorporation of Cy3- and Cy5-conjugated dUTP during cDNA synthesis using a standard reference pool of mRNA and the experimental sample of mRNA, respectively. Equal amounts of the two probes were pooled and hybridized to microarrays. The reference pool of mRNA was prepared from several immune cell lines (U937, THP-1, HL-60, VDSO, MonoMac6, and Jurkat) under a variety of stimulatory conditions, including PMA treatment and bacterial infection. Comparison of all experimental samples to the same reference allowed the relative expression level of each gene to be compared across all of the experiments (9). Hybridized arrays were

scanned with a GenePix 4000A microarray scanner (Axon Instruments), and the images were analyzed and tabulated with the GenePix Pro software package.

Data Analysis. Microarray elements (i.e., spots representing unique arrayed cDNA clones) were considered for analysis in two stages. In the first stage, we applied a high-stringency data selection, allowing only elements for which at least half the measurements within a set of experiments had fluorescence intensity in both channels at least 4-fold over background intensity. In the second stage, additional array elements with low signal intensity but high quality were selected, if their expression profile had a correlation of at least 0.7 to any gene selected in the first stage. Only cDNAs that met these criteria at 80% or more of the measured time points across all of the experiments were used in the analysis. Measurements from the multiple samples taken at the $t = 0$, pretreatment time point for each time course hour were averaged, with the exclusion of those measurements that fell beyond 1 standard deviation from the mean. This average $t = 0$ measurement was then subtracted from each subsequent time point measurement to depict the temporal response patterns of expression relative to $t = 0$ as the baseline. We then filtered the data on the basis of the variation of each element from the baseline across all experiments, as described in the figure legends. Hierarchical clusters and self-organizing maps were generated by the Cluster program and analyzed with the TreeView program (http://genome-www.stanford.edu/hostresponse/; ref. 10). For the comparison of live versus killed bacterial stimuli, the Euclidean distance was calculated as $\sqrt[2]{\sum_{i=1}^{5}\left(X_{live,i} - X_{killed,i}\right)^2}$, where $X_{live,i}$ and $X_{killed,i}$ represent $\log_2$ of the ratio of expression level of a given gene at time t_i, in the live and killed time series, respectively (t_i = 0.5, 2, 4, 6, 12 hr).

The time-averaged induction was calculated from the area under the curve (AUC) for expression time courses, by normalizing the AUC over the total time of exposure for each condition (i.e., $\frac{AUC}{\text{total time}}$). The AUC was calculated by using the trapezoid method,

based on the relationship $\sum_{i=1}^{n-1} \frac{1}{2}(X_i + X_{i+1})(t_{i+1} - t_i)$, where t_i represents exposure time, and X_i represents the $\log_2$ of the ratio of expression level at time t_i (t_i = 0, 0.5, 2, 4, 6, 12 hr for PBMC, and t_i = 0, 0.5, 1, 2, 4, 6 hr for U937 experiments).

A Linear Model for Dose Response. We used the following function to model simultaneously the effects of bacterial type, treatment dose, and exposure time on the expression response: the variables *b*, *d*, and *t* denote, respectively, bacterial type (*b* = 0 for *B. pertussis* and *b* = 1 for *S. aureus*), $\log_{10}$ treatment dose (*d* = 0, 1, 2, 3), and exposure time (*t* = 0.5, 2, 4, 6, 12 hr, omitting the 1- and 24-hr time points, which were measured only for the 100× dose). For a particular gene, let

$$X(b,d,t) = \mu + \beta b + \delta d + (\tau_1 t + \tau_2 t^2) + \alpha bd + b(\gamma_1 t + \gamma_2 t^2) + d(\theta_1 t + \theta_2 t^2) \quad \textbf{[1]}$$

represent the expected gene expression response for given levels of the three factors. This model specifies a linear log-dose response and a quadratic time response, and also allows for two-way interactions between each of the factors. Using the model in Eq. **1**, the difference in the expected response for the two types of bacteria is

$$\Delta(d,t) = X(b=1,d,t) - X(b=0,d,t) = \beta + \alpha d + (\gamma_1 t + \gamma_2 t^2) \quad \textbf{[2]}$$

for a given treatment dose *d* and exposure time *t*. We used the method of least squares to fit the function in Eq. **1** to the zero-transformed expression data of a subset of genes with multiple representations on the array. For each gene, this method identified parameter values $\alpha, \beta, \gamma_1, \gamma_2, \delta, \tau_1, \tau_2, \theta_1, \theta_2$, which minimize the sum of squared deviations between the observed responses and the fitted responses.

Enrichment of Gene Ontology (GO) Annotation Terms in the Bacterial Diversity Data Set. From the bacterial diversity data set, 1,208 unique genes were selected on the basis of having a LocusLink assignment (http://www.ncbi.nlm.nih.gov/LocusLink). *p* values for the frequency of 184 GO annotation terms (those that had at least 5 occurrences in the data set) were calculated on the basis of the hypergeometric distribution. Those annotation terms found to have *p* values less than 1% within the

common induction (red) and common repression (green) clusters are displayed with their GO identification numbers and example genes from the 1,208-gene data set.

The GO annotations for genes in the entire bacterial diversity data set were extracted in batch from SOURCE (http://genome-www.stanford.edu/source), using the LocusLink identification number for each gene. The resulting 1,208 unique annotated genes were used for the analysis, among which 184 annotation terms occurred for at least 5 genes. We compared the frequency of these 184 annotations in the data set as a whole to that in the common induction and common repression clusters, using the hypergeometric distribution to calculate *p* values (11) (see Eq. **3** below). Let N = 1,208 denote the total number of genes under consideration and A, the number of these genes with a particular annotation. The chance of observing at least x genes with that annotation in a random subset of n genes is given by

$$p(x;N,A,n)=1-\sum_{i=0}^{x-1}\frac{\binom{A}{i}\binom{N-A}{n-i}}{\binom{N}{n}}, \text{ where } \binom{N}{n}=N!/n!(N-n)!. \qquad \textbf{[3]}$$

1. Stibitz, S., Aaronson, W., Monack, D. & Falkow, S. (1989) *Nature (London)* **338**, 266-269.

2. Boschwitz, J. S., Batanghari, J. W., Kedem, H. & Relman, D. A. (1997) *J. Infect. Dis.* **176**, 678-686.

3. Relman, D. A., Domenighini, M., Tuomanen, E., Rappuoli, R. & Falkow, S. (1989) *Proc. Natl. Acad. Sc.i USA* **86**, 2637-2641.

4. Relman, D., Tuomanen, E., Falkow, S., Golenbock, D. T., Saukkonen, K. & Wright, S. D. (1990) *Cell* **61**, 1375-1382.

5. Gross, M. K., Au, D. C., Smith, A. L. & Storm, D. R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4898-4902.

6. Stainer, D. W. & Scholte, M. J. (1970) *J. Gen. Microbiol.* **63**, 211-220.

7. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T. & Marincola, F. M. (2000) *Nat. Biotechnol.* **18**, 457-459.

8. Alizadeh, A., Eisen, M., Davis, R. E., Ma, C., Sabet, H., Tran, T., Powell, J. I., Yang, L., Marti, G. E., Moore, D. T., *et al*. (1999) *Cold Spring Harbor Symp. Quant Biol.* **64**, 71-78.

9. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., *et al.* (2000) *Nature (London)* **403**, 503-511.

10. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 14863-14868.

11. Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. & Church, G. M. (1999) *Nat. Genet.* **22,** 281-285.