Analysis of whole genome mRNA expression by Affymetrix Droso2.0 chips

RNA pools from 60 guts of 5 days old females were isolated, purified with RNA clean up purification kits (Macherey Nagel), and DNase treated. All RNA quantities were assessed by NanoDrop®ND-1000 spectrophotometer and the quality of RNA was controlled on Agilent 2100 bioanalyzer chips. For each sample, 1 μg of total RNA was amplified and labeled using the GeneChip IVT labelling kit according to the protocol provided by the supplier. Affymetrix *Drosophila* Genome 2.0 arrays were hybridized with 30 μg of labeled cRNA, washed, stained and scanned according to the protocol described in Affymetrix GeneChip® Expression Analysis Manual (Fluidics protocol EukGeWS2v5_450).

Data analysis

Three independent repeats were performed for each time point and the gene expression profiles from challenged flies were normalized to the sucrose fed adults. All statistical analysis were performed using the free high-level interpreted statistical language R (R Core, 2004, http://www.R-project.org) and various Bioconductor packages (http://www.Bioconductor.org). Hybridization quality was assessed using Bioconductor "affy" and "affyPLM" packages (Bolstad et al., 2003; Gautier et al., 2004). Clustering analysis using all 18,952 Probe sets on the array revealed that biological repeats (triplicates) cluster together with the exception of the unchallenged wild-type where samples from the 2 time points cluster together (Figure S5). This shows that in wild-type unchallenged flies, samples are stable over time and that the within group variability is always lower than the between group variability. Normalized expression signals were calculated from Affymetrix CEL files using the RMA normalization method implemented in the "affy" package (Bolstad et al., 2003; Irizarry et al., 2003). Differential hybridized features were identified using Bioconductor package "limma" that implements linear models for microarray data (Smyth, 2004). In "limma", *P* values are obtained from moderated t-statistics or F-statistics using Empirical Bayesian methods. *P* values are then adjusted for multiple

testing with Benjamini and Hochberg's method to control the false discovery rate (FDR) (Benjamini and Hochberg, 1995). Probe sets showing at least 2-fold change and a FDR < 0.01 were considered significant. RT-qPCR of independent gut extracts from control and *Ecc15* orally infected flies confirmed the inducibility of 20 out of 20 tested genes, validating our dataset (indicated on Table 1). This microarray analysis likely missed some genes, as we did not monitor those regulated before 4 hr after ingestion of *Ecc15* (ex. *Puckered* Figure S3). In addition, we cannot exclude that some genes may be modulated by the potential nutritive value of ingested bacteria, rather than by infection. The complete set of gut *Ecc15* regulated genes as well as the complete underlying data are available at http://lemaitrelab.epfl.ch/ (folder "Resources").

References

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B *57*, 289-300.

Bolstad, B., Irizarry , R., Astrand, M., and Speed, T. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.

De Gregorio, E., Spellman, P.T., Rubin, G.M., and Lemaitre, B. (2001). Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. Proc Natl Acad Sci U S A *98*, 12590-12595.

Gautier, L., Cope, L., Bolstad, B.M., and Iriyarry, R.A. (2004). affy—analysis of affymetrix genechip data at the probe level. Bioinformatic *20*, 307–315.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. Biostatistic *4*, 249-264.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol PMID: 16646809