Additional file 1: RNA-seq data analysis

Gene expression in *A. thaliana* hypocotyl or roots structures infected or not by clubroot at 16 or 26 days after inoculation with 3 biological replicates was analyzed using an already published dataset accessible at Array Express with the following accession number: E-MTAB-4176 (Malinowski et al. 2016). The paired-end libraries were mapped to the TAIR10 genome annotation and assembly of *A. thaliana* using STAR (Dobin et al., 2013) with the following parameters: --outFilterMultimapNmax 5 --outFilterMismatchNmax 3 --alignIntronMin 10 --alignIntronMax 50000 --alignMatesGapMax 50000. Fragment counts per genes were estimated by Subread featureCounts (Liao et al., 2014) using default parameters. Using R (R Core Team, 2014) with the limma (Phipson et al., 2016; Ritchie et al., 2015) and edgeR (Robinson et al., 2010) packages based on Law et al., 2016, these raw fragment counts were converted to counts per million (CPM). Expressed genes have been filtered by removing genes with a CPM<1 in at least 6 libraries on the 24 analyzed libraries. CPMs have been normalized by using edgeR TMM method for Normalization Factor calculation. Heteroscedasticity from count data has been removed using voom precision weights. Finally, differentially expressed genes analyses have been performed by fitting linear models to compare Infected Tissue vs Control Tissue at a given DPI (16dpi, 26dpi on hypocotyl and root structures). Graphics supporting the validation of the afore-described pipeline can be found in Additional files 2-4.

References


