

Methods S1. Additional experimental procedures

Maize field trials

The CML444 and SC Malawi RIL population (145 RIL, F7:S6) (Messmer *et al.*, 2009) was planted in December 2008 at Baynesfield Estate in KwaZulu-Natal Province, South Africa. The plot design was a randomized block with three replicates. Each replicate of a RIL was a row of 10 plants. RILs were inoculated at the six leaf stage by placing into the whorls 50g per plant of dry powdered maize leaf material collected from GLS infected maize in the previous season. RILs were also exposed to natural inoculum from the environment since Baynesfield is a hotspot for the disease. GLS disease severity was scored on a per row basis using a 1-9 scale, where 1 and 9 represent no GLS disease and full GLS susceptibility, respectively (Berger *et al.*, 2014). The maturity of the maize RILs was between Reproductive 1 and Reproductive 4 over this period. The PROC GLM procedure of SAS 9.3 (SAS Institute, Cary, NC) was used to calculate least square means of the GLS disease scores for each of the four ratings (92, 99, 109 and 116 days after planting (DAP)) using a mixed model approach considering replication as random effects and genotypes as fixed effects (Berger *et al.*, 2014). These least square mean values were used to perform QTL mapping of the GLS disease phenotype (see later).

For the validation experiment, maize inbred line B73 was planted in December 2012 in three replicate rows of 10 plants each at Hildesheim Research Station, PANNAR SEED Pty Ltd, Greytown, KwaZulu-Natal, South Africa. Hildesheim is situated 96 km from Baynesfield Estate, and is a site of regular GLS disease screening due to favourable conditions for GLS disease, and therefore plants were subject to natural inoculum of *C.zeina*.

RNA extraction and microarray analysis

RNA was extracted from three biological replicates of 100 selected maize RILs at 103 DAP, sampled at Baynesfield Estate, KwaZulu-Natal in March 2009. The RNA from the three biological replicates of each of the 100 RILs was pooled to create 100 samples. Each biological replicate consisted of 10 cm X 10 cm leaf pieces from two different plants in a RIL row. Total RNA was extracted using Qiazol, treated with DNase, and purified with RNeasy Plant Mini kit according to manufacturer's instructions (Qiagen,

Hilden, Germany). Total RNA was shown to be of good quality using an Agilent Bioanalyser (average RIN number = 7.7).

The 100 RILs were selected out of the 145 that were planted using genotype data used for construction of the genetic map Q-Map 2.0 (reported in Berger *et al.*, 2014), such that pairs of RILs with dissimilar genomes could be processed on fifty microarray slides, according to the “distant pair” experimental design for two-colour dye microarrays described by Fu and Jansen (Fu and Jansen, 2006). RNA was amplified from each of the 100 RIL samples using the Amino Allyl MessageAmpII aRNA Amplification kit (Ambion, Dallas, USA), labelled with either Cy3 or Cy5 (100 pmol each), and hybridized to Agilent-016047 maize 4×44 K microarrays. This microarray had previously been annotated based on the maize inbred B73 genome sequence annotations (Coetzer *et al.*, 2011).

The microarrays were scanned using a Tecan LS Re-loaded scanner (Tecan, Mannedorf, Switzerland). Image acquisition and analysis was carried out using the GenePix Pro 6.1 software (Molecular Devices, Sunnyvale USA). The Agilent RNA Spike-in kit (Agilent Technologies, Santa Clara, USA) was used. The quality of each array was determined by evaluating probe spot, background spot and control spot intensity (Methods S1 Table 1). There was high correlation between the expected and observed spike-in controls (0.992, Methods S1 Table 1). Median background variance was low (less than 16%) and foreground spot median values were similar for both channels (2253 and 2456 for Cy5 and Cy3 respectively, excluding flagged spots, Methods S1 Table 1). Normalization of the expression data was performed in the R-based software package *limma* (Smyth, 2004), with a weighting of zero for flagged spots. Background correction was performed using the *normexp* method (offset =50) (Ritchie *et al.*, 2007). The *loess* method was used for normalization within arrays and *Aquantile* for normalization between arrays (Yang and Thome, 2003). After normalization, 50 datasets of M and A values represented the expression data from 100 RILs. Back-conversion of the normalized data was required to obtain separate expression values per reporter for each of the 100 RILs, by solving simultaneously for R (Red intensity; Cy5 channel) and G (Green intensity; Cy3 channel) from the formulas $A = \frac{1}{2} \log_2(RG)$ and $M = \log_2\left(\frac{R}{G}\right)$, yielded $G = \sqrt{\frac{2^{2A}}{2^M}}$ and $R = \sqrt{2^{2A}2^M}$. Out of the 42,034 maize gene reporters on the Agilent arrays, after removal of flagged reporters, back-converted

intensity expression profiles for 30,280 reporters across the 100 RILs were obtained. Microarray data have been deposited in the NCBI Gene Expression Omnibus (Accession # GSE76242).

Methods S1 Table 1. Quality control summary of microarray data from 100 maize RILs

	Min.	1st Quartile	Median	3rd Quartile	Max.
% flagged spots	43	51	54	58	69
Variance in Cy5 background	6.5	9.9	12.6	32.2	175.1
Variance in Cy3 background	3.1	6.3	15.6	20.9	70.4
Mean Cy5 intensity (including flagged spots)	209	897	1115	1477	3808
Mean Cy5 intensity (excluding flagged spots)	512	1959	2253	2969	6808
Mean Cy3 intensity (including flagged spots)	323	823	1192	1498	2653
Mean Cy3 intensity (excluding flagged spots)	795	1821	2456	3094	5018
Cy5 negative controls that are not flagged (%)	0	0	0	0.7	2.6
Cy3 negative controls that are not flagged (%)	0	0	0	0.7	2.6
Correlation coefficient between expected and mean observed Agilent spike-in controls:	0.871	0.988	0.99	1	1

Gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) was performed as described previously (Langfelder and Horvath, 2008) with an R-script modified for this study (Methods S2). The input data matrix consisted of expression values of 30,280 microarray reporters for each of the 100 RILs. Prior to the analysis, the expression values for each reporter were normalized relative to the mean value for that reporter across the 100 RILs, and then \log_{10} -transformed. A filtering step removed reporters with zero variance as well as those with more than 50% missing entries. The best estimate soft-thresholding power (β) for this dataset was 12, since this was the lowest power for which the scale-free topology fit index was above 0.8 (Zhang and Horvath, 2005). The minimum module size was set to 30. The output of WGCNA was the identification of co-expression modules and calculation of an “eigengene” for each module (a summary expression profile representing the module) (Langfelder and Horvath, 2008). In addition, module membership (MM) scores (scale of 0-1) are calculated for each gene, which represent the correlation of each gene with the eigengene for each module. Candidate module hub genes were identified, by ranking the reporters within each co-expression module by their

MM scores. A hub gene is expected to have a similar gene expression profile to the module eigengene, since it is highly co-expressed with many of the genes in a module.

To determine whether module eigengenes were significantly correlated with GLS disease scores, a permutation approach was followed. For each of the 42 modules, resampling (10,000 iterations) without replacement, of the GLS scores across the 100 RILs was performed in R. For each permutation the correlation coefficient between the permuted GLS scores and the module eigengene expression values was calculated. The resulting collection of correlation coefficients (for a specific module) represented the null distribution of the test statistic for that module, and the p-value was calculated as the proportion of these correlation coefficients that was greater than or equal to the absolute value of the observed correlation coefficient (using the actual data). Calculated p-values were adjusted for multiple testing by controlling the false discovery rate (Benjamini and Hochberg, 1995).

Fisher's exact tests (Upton, 1992) were used to determine whether each co-expression module with a significant correlation to GLS disease (Table 1) was enriched for genes that were either significantly induced (FDR<0.05) or significantly repressed (FDR<0.05) in the susceptible interaction RNAseq experiment. Lists of maize gene IDs were extracted per co-expression module, based on the reporter - gene model annotation from the Maize microarray annotation database (Coetzer *et al.*, 2011). Two 2x2 contingency tables were created per co-expression module, with the number of successes defined as the number of genes that were significantly induced or repressed, respectively, in the RNAseq experiment. Fisher's exact tests were performed with the *fisher.test()* function in R (R Core Team, 2014) and the resulting p-values were adjusted for multiple testing by controlling the false discovery rate (Benjamini and Hochberg, 1995).

eQTL data analysis pipeline developed in Galaxy.

An eQTL data analysis pipeline (Figure S5) was developed as part of this study. It was implemented in Python and R, and developed as three consecutive workflows in the online data analysis platform Galaxy (<http://galaxyproject.org>). The eQTL data analysis pipeline is available through Galaxy (Giardine *et al.*, 2005) via the Toolshed (<http://toolshed.g2.bx.psu.edu>).

The first workflow, called the eQTL permutation workflow (Figure S5A) is designed to determine the likelihood ratio (LR) threshold that corrects for genome-wide markers and multiple e-traits. It allows the user to specify the *number of traits to permute* and the *number of permutations per trait*. After randomly selecting the specified number of traits from the e-traits input file, these traits are equally divided into 40 sub-files for independent parallel computing. Zmapqtl (QTL Cartographer's composite interval mapping module) permutes the trait values and the genotypes per selected e-trait, and calculates the global maximum likelihood ratio (LR) per permutation. The pipeline then integrates the global maximum LR results by calculating the 95th percentile LR per e-trait (Table S16); and calculates an estimated experiment-wise LOD threshold (derived from average of all the calculated maximum LR thresholds) (Table S16). To convert LR to LOD values, $LOD=0.217 \times LR$ is applied.

The second workflow, called the eQTL mapping workflow (Figure S5B), splits the entire expression traits (e-traits) file into 40 sub-files; identifies eQTLs by running QTL Cartographer as 40 independent parallel tasks using different nodes on a compute cluster; and concatenates the 40 eQTL result files after the parallel runs. The workflow permits the setting of several parameters, including the model for stepwise regression, the model for interval mapping, the walk speed and the likelihood ratio (LR) threshold (as determined by the permutation workflow; Figure S5A). It also allows the inclusion of "other traits", so that factors that are potential confounders can be eliminated in the regression analysis.

The third workflow, called eQTL backend workflow (Figure S5C), aims to classify eQTLs as *cis* or *trans* and to identify eQTL hotspots. As eQTLs have centimorgan (cM) positions and genes (or microarray reporters) have base pair (bp) positions, firstly, the genetic and physical maps are linked. The resulting *lookup table* is used to proportionally estimate a cM-based position for each gene. Secondly, *cis/trans*-eQTL classification is carried out. An eQTL that is located within a distance of half the average size of an eQTL from the location of its linked gene is classified as *cis*, and an eQTL that is located further than half the average size of an eQTL from the location of its linked gene, often on a different chromosome, as *trans* (see below for discussion of *cis/trans* eQTL classification). The average size of an eQTL was 12.5 cM for the current study so a distance of 6.25 cM from a reporter was used for *cis*-eQTL classification.

Thirdly, to identify *trans*-eQTL hotspots, the frequency of *trans*-eQTLs and genes throughout the genome are calculated per 2 cM bin. eQTL peaks and gene models are counted per sliding window (consisting of two or three 2 cM bins). Fourthly, with this information available, two tests are conducted: (i) whether the *trans*-eQTL frequency is higher than that expected by chance, and (ii) whether gene density is an explanatory factor for *trans*-eQTL hotspots. In (i) mentioned above, the expected maximum number of *trans*-eQTL peaks per cM is calculated with a permutation approach. Each of the identified eQTLs are randomly assigned to 1 cM of the total number of cM on the map, and the resulting maximal number of eQTLs per bin is stored. The procedure is repeated 1000 times and the threshold corresponding to 95% of the obtained distribution established (Potokina *et al.*, 2008). This serves as a threshold to test if the *trans*-eQTL frequency per sliding window is significantly high. Sliding windows for which the number of *trans*-eQTL peaks per cM is above the permutation threshold are marked as potential *trans*-eQTL hotspots. In (ii) mentioned above, the proportion of genes to *trans*-eQTL peaks per cM is calculated for each sliding window. Here the null hypothesis, which is tested for each sliding window, is that the proportion of genes to *trans*-eQTL peaks in a specific sliding window is the same as the proportion of genes to *trans*-eQTL peaks across the whole genome, i.e. that the number of *trans*-eQTLs can be explained simply by local gene density. Sliding windows for which the null hypothesis is rejected, thus with a significant *trans*-eQTL excess or deficiency compared to gene number, are identified (chi-squared test p-value <0.0001). Finally, sliding windows with (i) the number of *trans*-eQTL peaks per cM that are above the permutation threshold and (ii) significant *trans*-eQTL excess compared to gene number, are called “unbiased” *trans*-eQTL hotspots. Adjacent sliding windows that meet the specified hotspot criteria are merged to form larger hotspot regions.

Fifthly, a GO over-representation analysis is conducted on each identified hotspot using the TopGO R package (Alexa and Rahnenfuhrer, 2010). A Fisher’s exact test is applied and significant GO-terms (<http://www.geneontology.org>) per *trans*-eQTL hotspot are listed in an output file. In an additional optional step, each hotspot is split according to the parental allele associated with higher expression and further GO over-representation analyses are performed on the resulting subsets.

eQTL data analysis pipeline input files

The e-traits file consisted of microarray-based gene expression profiles for 19,281 reporters in maize leaf samples across 100 RILs. Functions from Windows QTL Cartographer were used to convert the linkage map and cross information into a *map* file and a *cross* file, respectively, suitable for the QTL Cartographer modules *Rcross*, *SRmapqtl* and *Zmapqtl*. eQTL mapping was performed using the parameters mentioned above for QTL mapping. Sequences of all 10 chromosomes (in FASTA format) of the maize B73 reference genome (RefGen) v2 (Schnable *et al.*, 2009) were downloaded from the maizesequence.org FTP site (<http://ftp.maizesequence.org/current/>), in order to construct the chromosome length file. To generate the markers file, the MaizeGDB locus lookup tool (Andorf *et al.*, 2010) was used to extract the physical positions of most markers. In cases where the physical coordinates on the B73 genome sequence were not available, primer or marker sequences were downloaded from MaizeGDB (<http://www.maizegdb.org>) and located on the maize B73 reference genome v2.0 using the basic local alignment search tool (BLAST) (Altschul *et al.*, 1990). The gene positions file consisted of the reporter start and end bp positions, which was extracted from the Maize Microarray Annotation Database (Coetzer *et al.*, 2011). In order to generate a suitable *gene2GO* mapping file for GO enrichment analysis, the *Zea mays V5a* GO annotation file was downloaded from the AgriGO website (<http://bioinfo.cau.edu.cn/agriGO/>).

***Cis/trans*-eQTL classification implemented in eQTL data analysis pipeline**

Previous global eQTL mapping studies in plants implemented 3.5 – 10 cM windows around gene loci to define *cis*-eQTL (Holloway *et al.*, 2011; Potokina *et al.*, 2008; Swanson-Wagner *et al.*, 2009; West *et al.*, 2007), whereas others called *cis*-eQTL if the gene was within the eQTL support interval or the genetic map bin containing the eQTL peak (Drost *et al.*, 2010; Keurentjes *et al.*, 2007; Shi *et al.*, 2007; Wang *et al.*, 2010). In our study we chose to apply a rule for defining *cis*-eQTL by calculating the average genetic interval size of eQTLs across the study, which was 12.5 cM. We then scored *cis*-eQTL as those that were within half the interval size of the gene exhibiting the *cis*-eQTL (6.25 cM). This approach is therefore dependent upon the resolution of the study at hand and the key parameters, namely population size and statistical test applied to detect eQTL. The principle can therefore be

applied consistently across eQTL studies. We defined *trans*-eQTL the same way as most authors, namely those eQTL on the same or different linkage group that did not overlap with a gene's *cis*-eQTL interval (Holloway *et al.*, 2011; Keurentjes *et al.*, 2007; Li *et al.*, 2013). Other authors chose a stringent definition of *trans*-eQTL as only those on different linkage groups (Kloosterman *et al.*, 2012; Swanson-Wagner *et al.*, 2009), however this creates an additional category of eQTL and does not account for *trans*-regulated genes on the same linkage group.

Meta analysis to integrate co-expression, eQTL and phenotypic QTL data

The centimorgan (cM) positions of the phenotypic GLS QTL and *trans*-eQTL hotspots on Q-Map 2.0 for the CML444 X SC Malawi RIL population (Berger *et al.*, 2014) were used to determine if there was overlap between them. In addition a permutation test was conducted to establish how many overlaps would be significant. The calculation took into account that the average size of a *trans*-eQTL hotspot was 7 cM (which divided the genome in 266 bins), and the average size of a QTL was 14 cM. The calculation was conducted as follows: (i) Each of the 21 hotspots was randomly assigned to one of 266 bins on the genome (without replacement). (ii) The start bin of each of the 8 QTLs was also randomly assigned to one of the 266 bins on the genome and the adjacent bin to the right was also selected as a QTL bin. (iii) The number of bins selected to be both a "hotspot bin" and a "QTL bin" was stored (the number of overlaps). (iv) This process (i)-(iii) was repeated 10,000 times and finally the 95th percentile was calculated (the expected number of overlaps).

Fisher's Exact tests were applied to search for over-representation of genes in each co-expression module with genes in each *trans*-eQTL hotspot. The gene lists for each *trans*-eQTL hotspot were split into two sub-lists based on which allele was associated with higher expression, since we had observed parental allele directional bias (Table S6). Analysis of gene sets based on directional expression is an approach highlighted by Våremo *et al.* (Våremo *et al.*, 2013). Fisher's exact tests were used to determine whether each co-expression module was enriched for genes with eQTLs in a common *trans*-eQTL hotspot using a customized script in R (Methods S3). Thus for a specific co-expression module, the proportion of reporters with eQTLs in a given *trans*-eQTL hotspot was compared to the proportion of reporters in the genome (all reporters included in the analysis, i.e. 19281) with eQTLs in that *trans*-

eQTL hotspot. Therefore, for each of the 8 co-expression modules significantly correlating with GLS severity, 42 tests were performed (21 hotspots were split according to the parental allele associated with higher expression). The p-values were adjusted for multiple testing by controlling the false discovery rate (Benjamini and Hochberg, 1995) (Table S9).

“*A posteriori* network eQTL” analysis (as named by Hansen *et al.*, 2008) was carried out as follows: the module eigengene expression profiles of the eight gene co-expression modules correlating with GLS severity (Table 1) were extracted, using functions from the WGCNA package in R. Module eigengene eQTLs were mapped using QTL Cartographer with parameters described above for the global eQTL analysis. The estimated experiment-wise LOD threshold (LOD=2.8) for eQTL mapping was applied.

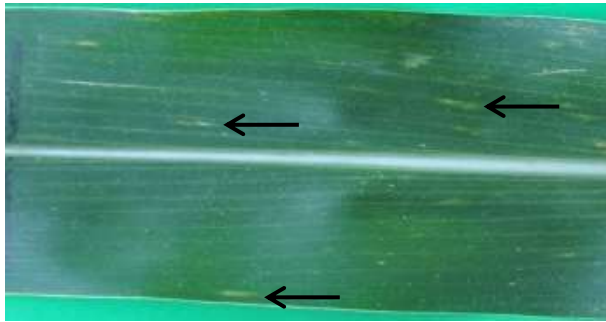
RNAseq analysis of maize inbred line B73

Maize inbred B73 plants were planted in three replicate rows of 10 plants each at Hildesheim Research Station, PANNAR SEED Pty Ltd, Greytown, KwaZulu-Natal in December 2012. The plants developed typical GLS disease symptoms initiating at the lower leaves and progressing to upper leaves through the season. GLS disease was scored for each replicate row of B73 plants at five intervals from 66-103 DAP using the 1-9 scale described above. At 77 DAP, with the plants at VT stage of development, the average GLS scores were between 4 to 5 with well-developed lesions on lower leaves (average of 8% lesion surface area determined by digital image analysis as described in Korsman *et al.*, 2012) but very few lesions and only at chlorotic stage on leaves above the ear (average of 0.2% lesion surface area) (Methods S1 Figure 1a). Therefore, to collect material that reflected a difference between *C. zeina* infected B73 leaves and control B73 leaf material, samples were collected from two lower leaves (second and third leaf internode below ear), and two upper leaves (second and third internode above ear), respectively, from three biological replicate plants. Methods S1 Figure 1a illustrates the difference in symptoms between upper and lower leaf samples. Leaf samples were collected and total RNA was isolated as described above for the maize RILs. However, prior to RNA extraction the leaf powder of each sample which had been crushed in a frozen mortar and pestle in liquid nitrogen, was mixed thoroughly and a sub-sample of 5g was removed for fungal gDNA

quantification using a *C. zeina* species specific qPCR assay previously developed in the group (Korsman *et al.*, 2012). Methods S1 Figure 1b shows that there is a significantly greater amount of *C. zeina* gDNA (as a proxy for fungal biomass) in the lower leaf samples compared to the upper leaf samples, which had very low but detectable amounts of *C. zeina* gDNA. The upper leaves had a few chlorotic spots indicative of early stages of fungal invasion (indicated by arrows in Methods S1 Figure 1ai), however were sufficiently devoid of GLS symptoms to serve as a control for comparison to lower leaves with severe GLS lesions (Methods S1 Figure 1aii).

RNA integrity of the six maize samples was confirmed using an Experion™ Automated Electrophoresis System (Bio-Rad Laboratories Ltd., Hercules, USA) using Experion™ RNA StdSens Analysis Kit (Bio-Rad Laboratories Ltd., Hercules, USA) before sending to BGI Tech Solutions Co., Ltd. (Beijing Genome Institute; Hong Kong) where 200 bp short-insert libraries were prepared. Strand specific RNA-sequencing of libraries was performed on Illumina HiSeq™ 2000 with a 100 bp paired-end module (Illumina Inc., San Diego, USA).

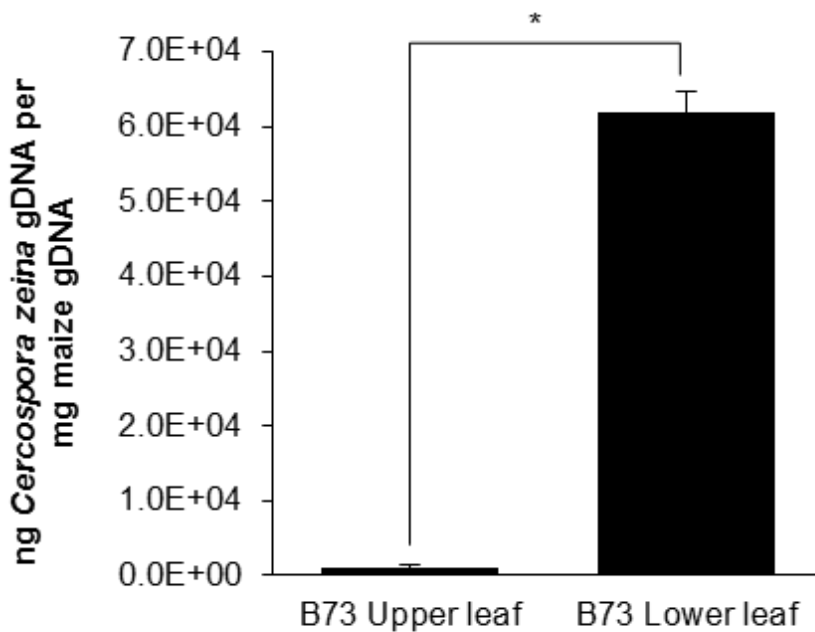
a) i. Upper leaf: early *C. zeina* infection



ii. Lower leaf: severe *C. zeina* infection



b)



Methods S1 Figure 1. Gray leaf spot (GLS) disease symptoms (a) and quantification of *Cercospora zeina* infection in field grown B73 maize plants (b).

(a) Representative upper and lower leaves of field grown B73 maize plants infected with *C. zeina*. Samples were photographed at 77 days after planting. (i) The upper leaf displays small, immature lesions (arrows) characteristic of early GLS disease symptoms. (ii) Rectangular, mature lesions were visible on the lower leaf which indicates a late stage of GLS disease. (b) The qPCR assay was used to quantify fungal gDNA expressed as ng *C. zeina* gDNA per mg maize gDNA (Korsman *et al.*, 2012) of all biological replicates of upper and lower leaves of the same field grown B73 maize plants used for RNA extraction and represented in (a). The asterisk (*) represents a significant difference between samples (T-test, $p < 0.05$). Standard errors are illustrated by error bars.

Differential gene expression analysis of B73 RNAseq data

The quality of the raw sequence data was assessed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Data were then filtered to remove adapters and trimmed for quality using Trimmomatic to remove adapters, to remove the leading low quality 3 bases, to scan the read with a 4-base wide sliding window, cutting when the average quality per base drops below 15, and to drop reads below 36 bases long (Bolger *et al.*, 2014)(v0.30; settings TruSeq3-PE-2.fa:2:30:10 LEADING:3 SLIDINGWINDOW:4:15 MINLEN:36). Methods S1 Table 2 reports the number of filtered RNAseq reads for each sample.

Methods S1 Table 2 RNAseq reads for B73 samples after filtering.

Sample	Insert size (bp)	Read length (bp)	Raw reads	Clean (filtered) reads
B73 (upper leaf)(control) Rep 1	200	90	28,517,956	26,295,478
B73 (upper leaf)(control) Rep 2	200	90	28,516,418	26,408,614
B73 (upper leaf)(control) Rep 3	200	90	28,517,626	26,005,796
B73 (lower leaf) (<i>C. zeina</i> infected) Rep 1	200	90	28,517,456	26,795,094
B73 (lower leaf) (<i>C. zeina</i> infected) Rep 2	200	90	28,516,650	26,502,110
B73 (lower leaf) (<i>C. zeina</i> infected) Rep 3	200	90	25,924,952	24,690,360

Filtered reads were aligned to v5b.60 of the *Zea mays* genome (Andorf *et al.*, 2016) using GSNAP to output in SAM format using 15-kmer length, both genome and the suffix array were loaded into memory (Wu and Nacu, 2010) with the following setting: `-n 3 -A sam --quality-protocol illumina -B 5 -k 15 --maxsearch 50 -t 16 -N 1 --split-output output_file--no-sam-headers --force-xs-dir`. The longest transcript gene structure in the *Z. mays* annotation GFF3 file (Andorf *et al.*, 2016) was used to estimate the transcript abundance in each locus by Htseq (Anders *et al.*, 2015) ([http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html](http://www.huber.embl.de/users/anders/HTSeq/doc/overview.html)).

Read counts were imported into R v3.1.0 (R Core Team, 2014) using the Bioconductor v2.14 (Gentleman *et al.*, 2004) edgeR package v3.6.8 (Robinson *et al.*, 2010) to calculate normalised read counts and perform differential expression analysis. RNAseq samples made up a paired design experiment, with lower leaf tissue (*C. zeina* infected) and matched upper leaf tissue (control) from

three plants. We used the GLM (generalized linear models) function in edgeR to account for any differences between the plants. The R code is available in supplementary file Methods S4.

Glasshouse trial of maize B73 for RT-qPCR and Kauralexin analysis

Maize inbred line B73 plants were grown in pots containing a sand/coir mix (1:1) in a greenhouse set to 21 +/- 5°C and 60% relative humidity. Leaves of plants at the V8-V10 stage were inoculated with *C. zeina* as previously described (Korsman et al., 2012). A 10cm² area surrounding the inoculation point was harvested immediately (to represent 0 dpi) or after 18 days, at which time lesions had developed. Both inoculation treatments were carried out on three plants in order to sample three biological replicates. Leaf samples from each plant were subjected to both RNA extraction (as described above) or kauralexin extraction. Kauralexin extraction, derivitisation and GC-MS analysis was carried according to (Schmelz *et al.*, 2011). The Maxima First Strand cDNA Synthesis Kit for RT-qPCR with dsDNase (Thermo Scientific, Waltham, USA) was used to generate cDNA from one µg of total RNA. Primers targeting *An2* were obtained from (Schmelz *et al.*, 2011). Primers were designed to target AC214360.3_FGP001 (ent-kaurene synthase, *KS*) and GRMZM2G161472 (putative kaurene oxidase, *KO*) as follows: *KS-F*: 5'- ACTCATCTCCGCTCACGAAT-3'; *KS-R*: 5'- ACCGGGAGTTGATCTTCTT-3'; *KO-F*: 5'- GAAGCATCCAGGCAGTGAAC-3'; *KO-R*: 5'- GAGGTACACATGCAACGGGT-3'. RT-qPCR was carried out using KAPA SYBR® FAST qPCR Kit (Kapa Biosystems, Boston, USA) on the Rotor-Gene™ 6000 11 instrument (Corbett Life Science, Sydney, Australia). All samples were analysed by RT-qPCR in triplicate. Relative gene expression was normalized using the reference genes *Lugein* (Manoli *et al.*, 2012) and *RNA Polymerase II* (Ma *et al.*, 2006) using qBase+ (www.biogazelle.org). An unpaired *t*-test was used to determine significantly different expression between the 0dpi and 18dpi samples.

References

- Alexa, A. and Rahnenfuhrer, J.** (2010) topGO: Enrichment analysis for Gene Ontology.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J.** (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–10.
- Anders, S., Pyl, P.T. and Huber, W.** (2015) HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

- Andorf, C.M., Cannon, E.K., Portwood, J.L., et al.** (2016) MaizeGDB update: new tools, data and interface for the maize model organism database. *Nucleic Acids Res.*, **44**, D1195–D1201.
- Andorf, C.M., Lawrence, C.J., Harper, L.C., Schaeffer, M.L., Campbell, D.A. and Sen, T.Z.** (2010) The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps. *Bioinformatics*, **26**, 434–436.
- Benjamini, Y. and Hochberg, Y.** (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.*, **57**, 289–300.
- Berger, D.K., Carstens, M., Korsman, J.N., Middleton, F., Kloppers, F.J., Tongoona, P. and Myburg, A.A.** (2014) Mapping QTL conferring resistance in maize to gray leaf spot disease caused by *Cercospora zeina*. *BMC Genet.*, **15**, 60.
- Bolger, A.M., Lohse, M. and Usadel, B.** (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
- Coetzer, N., Myburg, A.A. and Berger, D.K.** (2011) Maize microarray annotation database. *Plant Methods*, **7**, 31.
- Drost, D.R., Benedict, C.I., Berg, A., et al.** (2010) Diversification in the genetic architecture of gene expression and transcriptional networks in organ differentiation of *Populus*. *Proc. Natl. Acad. Sci.*, **107**, 8492–8497.
- Fu, J. and Jansen, R.C.** (2006) Optimal design and analysis of genetic studies on gene expression. *Genetics*, **172**, 1993–1999.
- Gentleman, R.C., Carey, V.J., Bates, D.M., et al.** (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
- Giardine, B., Riemer, C., Hardison, R.C., et al.** (2005) Galaxy: A platform for interactive large-scale genome analysis. *Genome Res.*, **15**, 1451–1455.
- Hansen, B.G., Halkier, B.A. and Kliebenstein, D.J.** (2008) Identifying the molecular basis of QTLs: eQTLs add a new dimension. *Trends Plant Sci.*, **13**, 72–77.
- Holloway, B., Luck, S., Beatty, M., Rafalski, J.-A. and Li, B.** (2011) Genome-wide expression quantitative trait loci (eQTL) analysis in maize. *BMC Genomics*, **12**, 336.
- Keurentjes, J.J.B., Fu, J., Terpstra, I.R., et al.** (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc. Natl. Acad. Sci.*, **104**, 1708–1713.
- Kloosterman, B., Kumari, A.M., Chibon, P.-Y., Oortwijn, M., Linden, G.C. van der, Visser, R.G.F. and Bachem, C.W.B.** (2012) Organ specificity and transcriptional control of metabolic routes revealed by expression QTL profiling of source–sink tissues in a segregating potato population. *BMC Plant Biol.*, **12**, 17.
- Korsman, J., Meisel, B., Kloppers, F.J., Crampton, B.G. and Berger, D.K.** (2012) Quantitative phenotyping of grey leaf spot disease in maize using real-time PCR. *Eur. J. Plant Pathol.*, **133**, 461–471.
- Langfelder, P. and Horvath, S.** (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
- Li, L., Petsch, K., Shimizu, R., et al.** (2013) Mendelian and non-Mendelian regulation of gene expression in maize. *PLoS Genet.*, **9**, e1003202.
- Ma, J., Morrow, D.J., Fernandes, J. and Walbot, V.** (2006) Comparative profiling of the sense and antisense transcriptome of maize lines. *Genome Biol.*, **7**, R22.
- Manoli, A., Sturaro, A., Trevisan, S., Quaggiotti, S. and Nonis, A.** (2012) Evaluation of candidate reference genes for qPCR in maize. *J. Plant Physiol.*, **169**, 807–815.
- Messmer, R., Fracheboud, Y., Bänziger, M., Vargas, M., Stamp, P. and Ribaut, J.-M.** (2009) Drought stress and

tropical maize: QTL-by-environment interactions and stability of QTLs across environments for yield components and secondary traits. *Theor. Appl. Genet.*, **119**, 913–930.

Potokina, E., Druka, A., Luo, Z., Wise, R., Waugh, R. and Kearsley, M. (2008) Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. *Plant J.*, **53**, 90–101.

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A. and Smyth, G.K. (2007) A comparison of background correction methods for two-colour microarrays. *Bioinformatics*, **23**, 2700–2707.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Schmelz, E.A., Kaplan, F., Huffaker, A., Dafoe, N.J., Vaughan, M.M., Ni, X., Rocca, J.R., Alborn, H.T. and Teal, P.E. (2011) Identity, regulation, and activity of inducible diterpenoid phytoalexins in maize. *Proc. Natl. Acad. Sci.*, **108**, 5455–5460.

Schnable, P.S., Ware, D., Fulton, R.S., et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.

Shi, C., Uzarowska, A., Ouzunova, M., Landbeck, M., Wenzel, G. and Lübberstedt, T. (2007) Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint x Flint maize recombinant inbred line population. *BMC Genomics*, **8**, 22.

Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, 1–26.

Swanson-Wagner, R.A., DeCook, R., Jia, Y., Bancroft, T., Ji, T., Zhao, X., Nettleton, D. and Schnable, P.S. (2009) Paternal dominance of trans-eQTL influences gene expression patterns in maize hybrids. *Science*, **326**, 1118–1120.

Upton, G. (1992) Fisher's Exact Test. *J. R. Stat. Soc.*, **155**, 395–402.

Varemo, L., Nielsen, J. and Nookaew, I. (2013) Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res.*, **41**, 4378–4391.

Wang, J., Yu, H., Xie, W., Xing, Y., Yu, S., Xu, C., Li, X., Xiao, J. and Zhang, Q. (2010) A global analysis of QTLs for expression variations in rice shoots at the early seedling stage. *Plant J.*, **63**, 1063–1074.

West, M.A.L., Kim, K., Kliebenstein, D.J., Leeuwen, H. van, Michelmore, R.W., Doerge, R.W. and St Clair, D.A. (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. *Genetics*, **175**, 1441–1450.

Wu, T.D. and Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.

Yang, Y.H. and Thome, N.P. (2003) Normalization for two-color cDNA microarray data. In D. R. Goldstein, ed. *Statistics and science: a Festschrift for Terry Speed*. pp. 403–418.

Zhang, B. and Horvath, S. (2005) A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.*, **4**, Article17.