

Research article

Open Access

## A multi-treatment experimental system to examine photosynthetic differentiation in the maize leaf

Ruairidh JH Sawers<sup>1,4</sup>, Peng Liu<sup>2,5</sup>, Katya Anufrikova<sup>1</sup>, JT Gene Hwang<sup>2,3</sup> and Thomas P Brutnell\*<sup>1</sup>

Address: <sup>1</sup>Boyce Thompson Institute, Cornell University, Tower Road, Ithaca, NY 14853, USA, <sup>2</sup>Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA, <sup>3</sup>Department of Mathematics and Department of Statistics, Cornell University, Ithaca, NY 14853, USA, <sup>4</sup>Department of Plant Molecular Biology, Biophore, University of Lausanne, CH-1015 Lausanne, Switzerland and <sup>5</sup>Department of Statistics, Iowa State University, Ames IA 50011, USA

Email: Ruairidh JH Sawers - Ruairidh.Sawers@unil.ch; Peng Liu - pliu@iastate.edu; Katya Anufrikova - fish\_an@mail.ru; JT Gene Hwang - hwang@math.cornell.edu; Thomas P Brutnell\* - tpb8@cornell.edu

\* Corresponding author

Published: 09 January 2007

Received: 14 September 2006

BMC Genomics 2007, 8:12 doi:10.1186/1471-2164-8-12

Accepted: 09 January 2007

This article is available from: <http://www.biomedcentral.com/1471-2164/8/12>

© 2007 Sawers et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The establishment of C<sub>4</sub> photosynthesis in maize is associated with differential accumulation of gene transcripts and proteins between bundle sheath and mesophyll photosynthetic cell types. We have physically separated photosynthetic cell types in the leaf blade to characterize differences in gene expression by microarray analysis. Additional control treatments were used to account for transcriptional changes induced by cell preparation treatments. To analyse these data, we have developed a statistical model to compare gene expression values derived from multiple, partially confounded, treatment groups.

**Results:** Differential gene expression in the leaves of wild-type maize seedlings was characterized using the latest release of a maize long-oligonucleotide microarray produced by the Maize Array Project consortium. The complete data set is available through the project web site. Data is also available at the NCBI GEO website, series record GSE3890. Data was analysed with and without consideration of cell preparation associated stress.

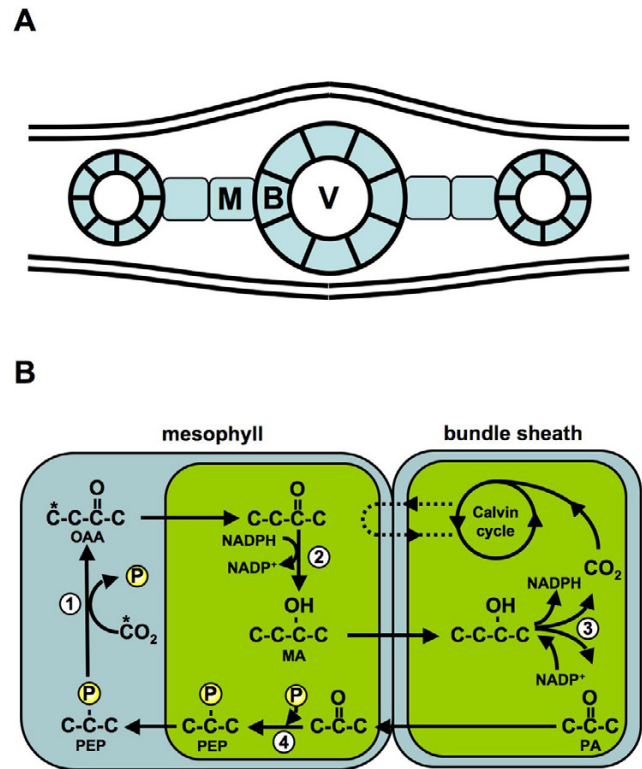
**Conclusion:** Empirical comparison of the two analyses suggested that consideration of stress helped to reduce the false identification of stress responsive transcripts as cell-type enriched. Using our model including a stress term, we identified 8% of features as differentially expressed between bundle sheath and mesophyll cell types under control of false discovery rate of 5%. An estimate of the overall proportion of differentially accumulating transcripts ( $1-\pi_0$ ) suggested that as many as 18% of the genes may be differentially expressed between B and M. The analytical model presented here is generally applicable to gene expression data and demonstrates the use of statistical elimination of confounding effects such as stress in the context of microarray analysis. We discuss the implications of the high degree of differential transcript accumulation observed with regard to both the establishment and engineering of the C<sub>4</sub> syndrome.

## Background

Photosynthesis in the majority of plants occurs in a single photosynthetic cell type ( $C_3$  photosynthesis) [1]. Within the chloroplasts, the enzyme ribulose-1, 5-bisphosphate carboxylase/oxygenase (Rubisco) fixes atmospheric carbon by addition of  $CO_2$  and water to the five-carbon sugar ribulose-1, 5-bisphosphate (RuBP). Rubisco will also catalyze the oxidation of RuBP in a process known as photorespiration that does not fix carbon [2]. The reduction in efficiency associated with photorespiration and the energetic costs of recycling its products has been estimated to limit the performance of  $C_3$  photosynthesis by as much as 30% in hot arid conditions [3]. A number of taxa utilize a two-step carbon fixation process, known as  $C_4$  photosynthesis, to limit the impact of photorespiration upon photosynthetic performance [4]. Plants that utilize  $C_4$  photosynthesis appear to be at a particular fitness advantage under conditions of limited water availability, high temperature and high irradiance light [5]. Interestingly, some of the most promising grasses for biofuel production are  $C_4$  grasses, including *Miscanthus × giganteus* (Giant Miscanthus), *Panicum virgatum* (switchgrass), *Zea mays* (maize), *Sorghum bicolor* (sorghum) and *Saccharum officinarum* (sugarcane).

In  $C_4$  plants, Rubisco accumulation is spatially restricted to  $CO_2$ -rich sites within the leaf so that the carboxylase reaction is favoured over photorespiration. In maize, Rubisco accumulation is restricted to thick-walled bundle sheath (B) cells that surround the leaf veins (Figure 1A). Carbon is initially fixed in adjacent mesophyll (M) cells and subsequently transported, by a multi-enzyme carbon shuttle, into the B, where decarboxylation elevates local  $CO_2$  levels and generates an environment for efficient Rubisco function (Figure 1B).

Cell-type specific differences in morphology and physiology are fundamental to  $C_4$  photosynthesis [1,6]. Detailed analysis of B and M differentiation in maize has shown that Rubisco, enzymes of a  $C_4$  carbon shuttle and components of the light-harvesting machinery accumulate to different levels in B and M cells [7]. B cell chloroplasts are predominately agranal and do not accumulate key components of the water oxidizing complex of photosystem II (PSII) [8,9]. Consequently, a number of processes requiring chemical reduction, including portions of Calvin cycle [10,11], synthesis of antioxidants [12] and nitrogen assimilation [13] are localized to the M cells. Despite detailed understanding of certain metabolic pathways utilized in  $C_4$  photosynthesis, the molecular mechanisms governing cell differentiation and the full extent of metabolic partitioning are still to be fully characterized. Promoter fusion, methylation assays and transient expression studies have identified a number of *cis* acting elements in the promoter sequences of  $C_4$ -related genes [14-17].



**Figure 1**

### $C_4$ photosynthesis in the maize (*Zea mays*) leaf. A.

Schematic of a longitudinal cross section through a maize leaf showing Kranz anatomy. Thick-walled bundle sheath (B) cells surround longitudinal veins (V). Mesophyll cells (M) occupy the leaf space between vascular bundles. **B.** Major reactions of the  $C_4$  carbon shuttle. 1) Carbon is initially added to phosphoenolpyruvate (PEP) to form oxaloacetate (OAA) by the enzyme PEP carboxylase (PEPC). 2) OAA is transported to the M chloroplasts where it is reduced to malate (MA) by NADP-specific malate dehydrogenase (NADP-MDH). 3) MA is transported to the B where it is decarboxylated by NADP-malic enzyme (NADP-ME) to yield pyruvate (PA) and release  $CO_2$  to the Calvin cycle. 4) PA is returned to the M where it is phosphorylated by phosphoenolpyruvate dikinase (PPdK) to regenerate PEP and continue the cycle.

Much less is known about *trans* acting factors that may drive the  $C_4$  differentiation process [18,19]. Genetic approaches have resulted in the isolation of maize mutants characterized by B cell-specific defects, but these mutants have not directly identified regulators of cell-specific development [20-22].

Many biochemical and molecular studies of  $C_4$  photosynthetic cell types have made use of techniques for isolation of separated cells. Typically, B cells have been isolated as vascular strands by mechanical disruption and M cells isolated as protoplasts by enzymatic digestion [23,24].

Therefore, different isolation protocols complicate the identification of differences between the two cell types. This is especially true when comparing the accumulation of RNA transcripts because changes can occur rapidly in response to the stresses of protoplast preparation [25]. When small numbers of genes have been analyzed, additional treatments have been used to control for the effect of these stresses (e.g. [22]). In contrast, previous multi-gene profiling studies of  $C_4$  cell types have not accounted for such effects in the initial analysis [26-28]. Following preliminary two-sample microarray experiments, we were especially concerned with the problem of mistakenly identifying stress-induced transcripts as M enriched (Sawers *et al.*, unpublished observations).

In order to control for a stress effect associated with M cell isolation, total leaf and stressed total leaf samples were included in analysis. In general, microarray experiments are based on paired comparisons [29]. Multiple paired comparisons may be linked, as in a time course, or may be cross-referenced, as in a clustering analysis [29]. In the case of the isolated  $C_4$  cell types, the situation is somewhat different in that cell type and effects of the separation protocol are partially confounded between treatments. To formally describe this relationship, we have developed an analytical model to describe  $C_4$  gene expression and to allow elimination of the stress effect resulting from protoplast isolation. Using this approach, we have identified 1,280 features in the Maize Array Project oligonucleotide array that are predicted to be B or M enriched. We also present an analysis of the same data without consideration of the stress effect. Comparison of these two analyses demonstrates the importance of considering the stress effect and the application of a statistical modelling approach to control confounding factors in microarray experiments.

## Results and discussion

### Experimental design and data collection

B strands ( $T_B$ ) and M protoplasts ( $T_M$ ) were isolated from 10 day-old W22 inbred maize seedlings by mechanical disruption and enzymatic digestion, as previously described [30] (Figure 2A). Isolation of M protoplasts required a 3 h enzymatic incubation that was not performed during isolation of B strands. To control for transcriptional differences arising from these different treatments, additional total leaf ( $T_T$ ) and total leaf stress ( $T_S$ ) samples were isolated. Both  $T_T$  and  $T_S$  samples contain a combination of B and M cells. Leaves for the  $T_T$  sample were harvested and cut into thin strips as for the  $T_B$  sample and then frozen directly in liquid nitrogen rather than subjected to mechanical disruption. Leaves for the  $T_S$  sample were harvested and cut as for the  $T_M$  sample, subjected to a 3 h mock enzymatic digestion and then frozen. An interwoven loop design [31] was used to directly com-

pare all combinations of the four treatment groups ( $T_B$ ,  $T_M$ ,  $T_T$ ,  $T_S$ ) on two-label microarrays (Figure 2B). To maximize biological replication, a unique set of seedlings was used for each  $T_B$ ,  $T_M$ ,  $T_T$  and  $T_S$  replicate. The resulting design used 24 independent RNA samples analysed over 12 array sets.

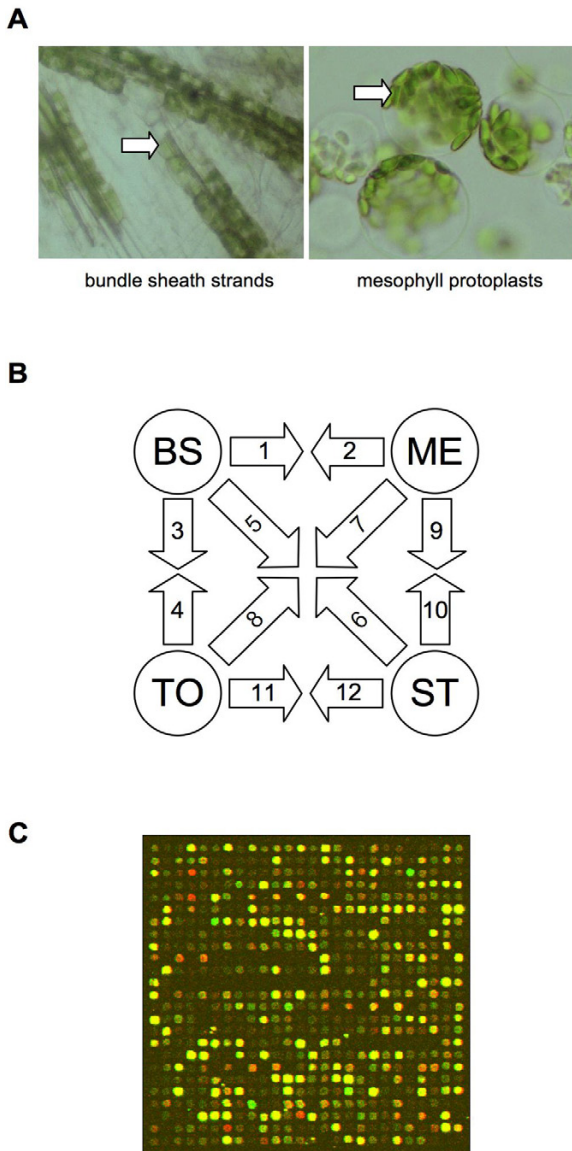
Samples were analysed by hybridisation to a long-oligonucleotide microarray produced by the Maize Array Project consortium [32]. Maize sequences for design of this chip were predominantly selected from the The Institute for Genomic Research (TIGR) maize gene index [33]. To manufacture this microarray, approximately 50,000 sequences were positively orientated and identified as suitable for oligonucleotide design. This initial sequence set was supplemented with further EST sequences, organellar sequences, repeat sequences and community requests to provide a final design data set of 57,441 features. Single 70-mer oligonucleotides were designed for each sequence in the design set and printed over two glass slides.

Array detection (Figure 2C) and image analysis were performed as described in Methods. Images and intensity data may be accessed in a MIAME compliant form at the Maize Oligonucleotide Array Project website [34]. Data are also available in the NCBI GEO database, series record GSE3890. The data are presented in full on both the Maize Array Project and NCBI websites.

Feature intensity values were log-transformed and corrected for local background signal and a LOWESS procedure [35] was used to normalize between channels for each slide (see Methods). On the basis of the  $T_T$  treatment (which contains both B and M), features with either low or saturating signal intensity were discarded (see Methods). A stringent filtering of low expression values was used to reduce the dimensionality of the data set in light of the complexity of the experimental design. High expression filtering was less stringent to avoid elimination of previously characterized, high abundance,  $C_4$  'marker' transcripts. Following filtering, 15,988 unique features were considered for subsequent analysis (Additional file 1). We considered this reduced data set appropriate for model development while maintaining sufficient and meaningful biological information to allow general conclusions on the extent of differentiation between B and M cell types.

### Construction of a model to include a stress term in the analysis of the B and M data set

The experimental design required the analysis of four, partially confounded treatment groups; *i.e.* the  $T_T$  treatment contains both B and M cell types found in  $T_B$  and  $T_M$  treatments, while the  $T_M$  treatment combines the effects of cell-



**Figure 2**  
**Experimental design.** **A.** B stands and M protoplasts were isolated by mechanical disruption and enzymatic digestion, respectively. Representative chloroplasts are marked with an arrow. **B.** Interwoven loop design of microarray labelling. Four treatments (B, M, S, T) were compared using two-dye microarrays. Each arrow represents hybridization. RNA isolated from the treatment at the tail of the arrow was used to synthesize Cy3-labelled cDNA, and RNA isolated from the treatment at the head of the arrow was used to synthesize Cy5-labelled cDNA. Independent biological replicates were used for each labelling, giving a total of six per treatment. Numbers on arrows refer to I. **C.** Pseudo-colour overlay of Cy3 and Cy5 images from a representative hybridization. A single 26 × 26 features sub-grid is shown. The complete University of Arizona oligonucleotide array consists of 96 such grids arrayed over two slides.

specificity and the stress effect that is also seen in  $T_S$ . In the following discussion, the term 'stress' is used to refer specifically to the effect of protoplast isolation on transcript levels. Importantly, although the aim of the experiment is to compare expression levels between B and M cell types, the level of gene expression within the intact M could never be directly measured because of the stress effect of M cell isolation. A model was constructed in order to formally describe this situation and to allow statistical elimination of the stress effect.

Let  $V_{1f}$  and  $V_{2f}$  represent the  $\log_2$  transformed expression level of a given feature ( $f$ ) in B and M cell types, respectively. The aim is to estimate  $(V_{1f} - V_{2f})$  on a feature-by-feature basis. We use  $c$  to label the leaf sample, whether B cell prep ( $c = 0$ ), M cell prep ( $c = 1$ ) or total leaf sample ( $c = 2$ ). The parameter  $j$  indicates the presence ( $j = 1$ ) or absence ( $j = 0$ ) of stress. For any given feature  $f$ , we fit the normalized signal  $Y_{fcjr}$  with the model,

$$Y_{fcjr} = (1 - a_c)V_{1f} + a_cV_{2f} + jS_f + \epsilon_{fcjr} \quad (1)$$

where  $a_c$  is the proportion of M cells in the sample,  $S_f$  represents the effect of stress on gene expression,  $r$  represents the replicate number from 1 through 6 and  $f$  represents the feature number from 1 to  $F$ , the number of features in the array. In this model, the effect of preparative stress is assumed to act additively and uniformly in both cell types. Below, the performance of the model will be assessed with respect to analysis in which the stress effect is not considered.

If we assume there is no contamination of the other cell type for the single cell treatments ( $T_B$  or  $T_M$ ), the values of  $a_0$  and  $a_1$  (the proportion of M in the sample) are set to be 0 and 1, respectively. In practice, two types of cellular contamination might be recognized. First, a proportion of contaminating M cells will be present in the B prep and of B cells in the M prep. The level of this cross-contamination was estimated at below 5% as determined by semi-quantitative PCR using known markers for B (*RbcS* and *ChlMe*) and M cell identity (*Pepc* and *Mdh1*) (data not shown). The level of cross-contamination was considered to be sufficiently low as to be ignored, thereby simplifying the model by elimination of  $V_{1f}$  and  $V_{2f}$  terms in expressions describing M and B, respectively. The values of  $a_0$  and  $a_1$  could be adjusted to allow consideration of such contamination if desired. The presence of additional leaf cell types constitutes a second source of cellular contamination, perhaps most notably the inclusion of epidermal and vascular cells. For simplicity and economy, we do not consider this second source of contamination in our model. We estimated the value of  $a_2$  (the proportion of M cells in  $T_T$  and  $T_S$  preparations), by examination of leaf sections and by marker gene expression, at between 0.7 and

0.5 (data not shown). For the analysis presented here, the value of  $a_2$  was set at 0.5. The model (1) can thus be simplified as,

$$\mu_{TBf} = V_{1f}$$

$$\mu_{TMf} = V_{2f} + S_f$$

$$\mu_{Tf} = (V_{1f} + V_{2f})/2$$

$$\mu_{TSf} = (V_{1f} + V_{2f})/2 + S_f$$

where  $\mu_{TBf}$ ,  $\mu_{TMf}$ ,  $\mu_{Tf}$  and  $\mu_{TSf}$  represent expected expression levels indexed as appropriate to the four treatments  $T_B$ ,  $T_M$ ,  $T_T$  and  $T_S$  for feature  $f$ . Since not all treatments are observed within the same array, the spot-specific array effect is corrected as a random effect to the model (1) and estimated by REML [36]. The normalized values (see Methods) for each feature, corresponding to the 12 hybridizations, are listed in Additional file 1.

Two versions of the analysis were performed. First, an analysis as described above was used to generate estimates of  $(V_{1f} - V_{2f})$  and  $S_f$  for each feature (referred to as the stress model). Second, we repeated the analysis with the stress effect ignored (referred to as the simple model). A second set of estimates of  $(V_{1f} - V_{2f})$  were calculated and compared with those obtained from the stress model. For both models, the deviation of each  $(V_{1f} - V_{2f})$  estimate from 0 was investigated using a t-statistic. The q-value procedure of Storey et al. [37] was used to control the false discovery rate (FDR). The estimates of  $(V_{1f} - V_{2f})$  and  $S_f$  for the stress model are listed in Additional file 2 together with their test statistics. The estimates of  $(V_{1f} - V_{2f})$  obtained from the simple model are listed in Additional file 3.

#### Comparison of analyses using the stress and simple models

Lists of differentially accumulating features identified by the stress and simple models were compared. Under FDR control at the 5% level with q-values  $\leq 0.05$ , the stress model identified 1,280 differentially accumulating unique features. At the same level of control, the simple model identified 4,384 unique features. 1,043 features were common to the gene lists obtained from the two models. Therefore, the simple model identified the majority of features identified by the stress model (approximately 80%). This is shown graphically in Figure 3A. For the features identified by the stress model, the q value obtained from the stress model ( $q_{\text{stress}}$ ) is plotted against the q value obtained from the simple model ( $q_{\text{simple}}$ ). Features are coloured red and blue for predicted M and B enrichment, respectively. The threshold values of  $q = 0.05$  are shown by dashed lines. Among the features identified by the stress model, but not the simple model, were three annotated as components of PSII (MZ00023434,

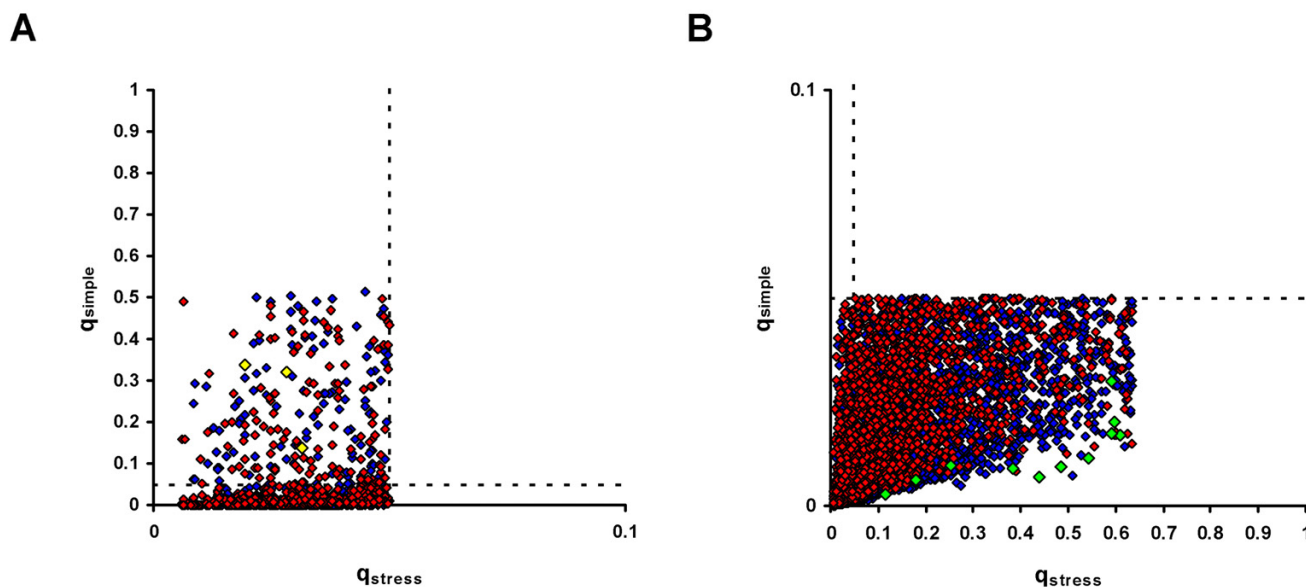
MZ00044083, MZ00040590) (shown in yellow in Figure 3A). Previous analyses have shown that PSII components predominantly accumulate in the M cells of the maize leaf [8,9]. The failure of the simple model to identify these PSII transcripts is consistent with reduction in RNA levels associated with M cell preparation. Under the stress model, accumulation of all three of these RNAs was predicted to be reduced by stress.

Although the simple model identified the majority of features identified by the stress model, it also identified many more features in total than the stress model (Figure 3B). Preliminary microarray analysis of B and M cell types using a simple paired comparison had suggested that one confounding effect of preparation stress might be the mistaken identification of stress induced transcripts as M enriched (Sawers *et al.*, unpublished observations). Among the features predicted to be strongly M enriched only under the simple model were a number annotated as chaperones or heat shock proteins (HSPs) (MZ00000354, MZ00034301, MZ00035916, MZ00035984, MZ00036574, MZ00038036, MZ00039146, MZ00040123, MZ00040558, MZ00040980 and MZ00042904), a class of transcripts known to be induced by stresses [38,39]. These HSPs are shown in green in Figure 3B. Additional stress-related annotations included a glutathione peroxidase (MZ00041338, MZ00041463) and a wounding-associated chymotrypsin inhibitor (MZ00037253, MZ00041005). These HSPs and other stress-related features are likely to have been mis-identified as M enriched on the basis of stress-related increases in accumulation. Consistent with this interpretation, the stress model predicted that accumulation of all of these transcripts increased following stress.

Empirical comparison of simple and stress models demonstrated the requirement for the control of the stress effect and the applicability of our stress model to this problem and is, therefore, the model we have chosen in the analysis of differential gene expression between B and M cells. Below, we describe the use of the stress model to identify potentially differentially accumulating transcripts and briefly discuss the resulting gene list.

#### Statistical identification of B and M enriched transcripts using the stress model

The estimate  $(V_{1f} - V_{2f})$  was calculated for all 15,988 features passing data filtering. The mean estimate of  $(V_{1f} - V_{2f})$  was 0.02 suggesting that the estimates were distributed around a value close to 0, i.e. equivalent expression in B and M cell types. Analysis of the distribution of resulting p-values allowed estimation of the proportion of non-differentially expressed genes ( $\pi_0$ ). Using the method of Storey et al., [37],  $\pi_0$  was calculated to be 0.823, corresponding to 2,830 differentially expressed features from



**Figure 3**

**Comparison of analysis with and without consideration of the stress effect.** For a given feature the q-value for testing differential expression between B and M cells under the stress model ( $q_{\text{stress}}$ ) was plotted against the q-value obtained under the simple model ( $q_{\text{simple}}$ ). **A** 1,280 unique features selected as  $q_{\text{stress}} < 0.05$ . Dashed lines cut axes at  $q_{\text{stress}} = 0.05$  and  $q_{\text{simple}} = 0.05$ . **B** 4,384 unique features selected as  $q_{\text{simple}} < 0.05$ . Features predicted to be B enriched by stress or simple models, in **A**. or **B**. respectively, shown in red. Features predicted to be M enriched by stress or simple models, in **A**. or **B**. respectively, shown in blue. Four features annotated as PSII components (MZ00023434, MZ00044083, MZ00040590, MZ00041794), not identified as M enriched by the simple model, are shown in yellow in **A**. Eleven features annotated as heat shock proteins (MZ00000354, MZ00034301, MZ00035916, MZ00035984, MZ00036574, MZ00038036, MZ00039146, MZ00040123, MZ00040558, MZ00040980 and MZ00042904), predicted to be strongly M enriched by the simple model, are shown as green in **B**.

the 15,988 in the analysis. Under FDR control at the 5% level, the model identified 1,280 candidate features.

#### **Annotation of features identified as differentially expressed under FDR control**

The majority of features present in the Maize Array Project are designed to EST contigs present in the TIGR maize sequence database [40]. An annotation of the features, based on annotation of the TIGR sequences at the time of design, is available at the Maize Array Project website [41]. It is important to note that a complete maize genome sequence is not yet available and that current gene models may well change as additional data become available. Annotation of oligonucleotide probes is further complicated by difficulty in predicting cross-hybridization between related genes [42,43]. This is especially relevant to a highly polymorphic species such as maize that possesses a highly duplicated genome [44]. Given these caveats, approximately 50% of the identified features were fully or partially annotated in the Maize Array Project database. Of the remainder, approximately 15% were annotated as encoding genes related to hypothetical or

unknown *Arabidopsis* or rice proteins, 10% were annotated by similarity to *Arabidopsis* or rice genomic regions and 25% were not annotated. To further investigate the number of unique genes represented by the 1,280 features, a BLAST search was used to identify maize sequences in the NCBI non-redundant database homologous to the oligonucleotides. For each feature, the best-matched sequence was recorded. Under the search criteria used 1,173 matches were recovered. Of these, 899 unique sequences were represented. In addition, TIGR rice gene model matches were obtained for 792 of the features from the Maize Array Project database. These 792 rice matches represent 730 independent gene models. This estimate is in line with the original design criteria of the microarray used [33]. Annotations are included in Additional file 2.

#### **Validation of candidate gene list using prior knowledge**

Differential gene expression in B and M cells has been the subject of extensive prior investigation [18]. The wealth of previous studies provides a collection of well-characterized marker transcripts that can be used in the preliminary validation of the candidate gene list. In total, we identified

approximately 80 features as markers considered to provide biological validation of the data set and analysis (Additional file 2). Data corresponding to a number of these are shown graphically in Figure 4. Markers were selected on the basis of previous studies and the presence of multiple features representing a gene. For each feature, the ratios of signal intensity ( $M$  values) obtained from the 12 hybridizations are shown. While variation is evident both between and among features, the overall consistency of behaviour is evident. In all examples shown, strong B or M cell enrichment is predicted in accordance with previous observations.

The enzymes of the  $C_4$  carbon shuttle are abundant and cell-type specific (Figure 1B). Previous studies have characterised accumulation of their transcripts by RNA gel-blot analysis [45], *in situ* hybridization [46], real-time PCR [27] and differential screening [26]. Features corresponding to genes encoding the carbon shuttle enzymes, phosphoenolpyruvate carboxylase (PEPC), malate dehydrogenase (MDH), NADP-dependent malic enzyme (NADP-ME) and pyruvate orthophosphate dikinase (PPDK) activities are shown in Figure 4. Although considered markers of  $C_4$  cell identity, the carbon shuttle enzymes are encoded by members of small gene families that contain both  $C_3$  and  $C_4$  isoforms. For example,  $C_4$ -specific malic enzyme is hypothesized to have arisen following the acquisition of a plastid transit peptide sequence by a gene encoding a cytosolic isoform, followed by duplication and divergence of  $C_3$  and  $C_4$  forms [47]. Consequently, maize contains at least three NADP-ME loci [47]. The gene *ZmChlMe1* encodes a leaf-specific, plastid-targeted isoform required for decarboxylation of malate in B cells [47,48] (Figure 1B, reaction 3) while two, nearly identical, *Me2* genes (*ZmChlMe2a* and *ZmChlMe2b*) have been identified that encode cytosolic isoforms [47]. At the nucleotide level, *ZmChlMe2a* and *ZmChlMe2b* are 99% identical to each other and 87% identical to *ZmChlMe1* [47]. A further NADP-ME activity has been characterized from roots and found to be 99% and 98% identical to *ZmChlMe2a* and *ZmChlMe2b*, respectively [49]. Gene duplications of this type could pose a serious difficulty in oligonucleotide analysis if features do not discriminate between paralogous gene copies. In the case of the carbon shuttle enzymes, the  $C_4$  isoforms are typically more abundant than  $C_3$  isoforms [50] and the accumulation patterns we observed suggest that the signal obtained in our experiment corresponds to  $C_4$  cell-specific transcripts.

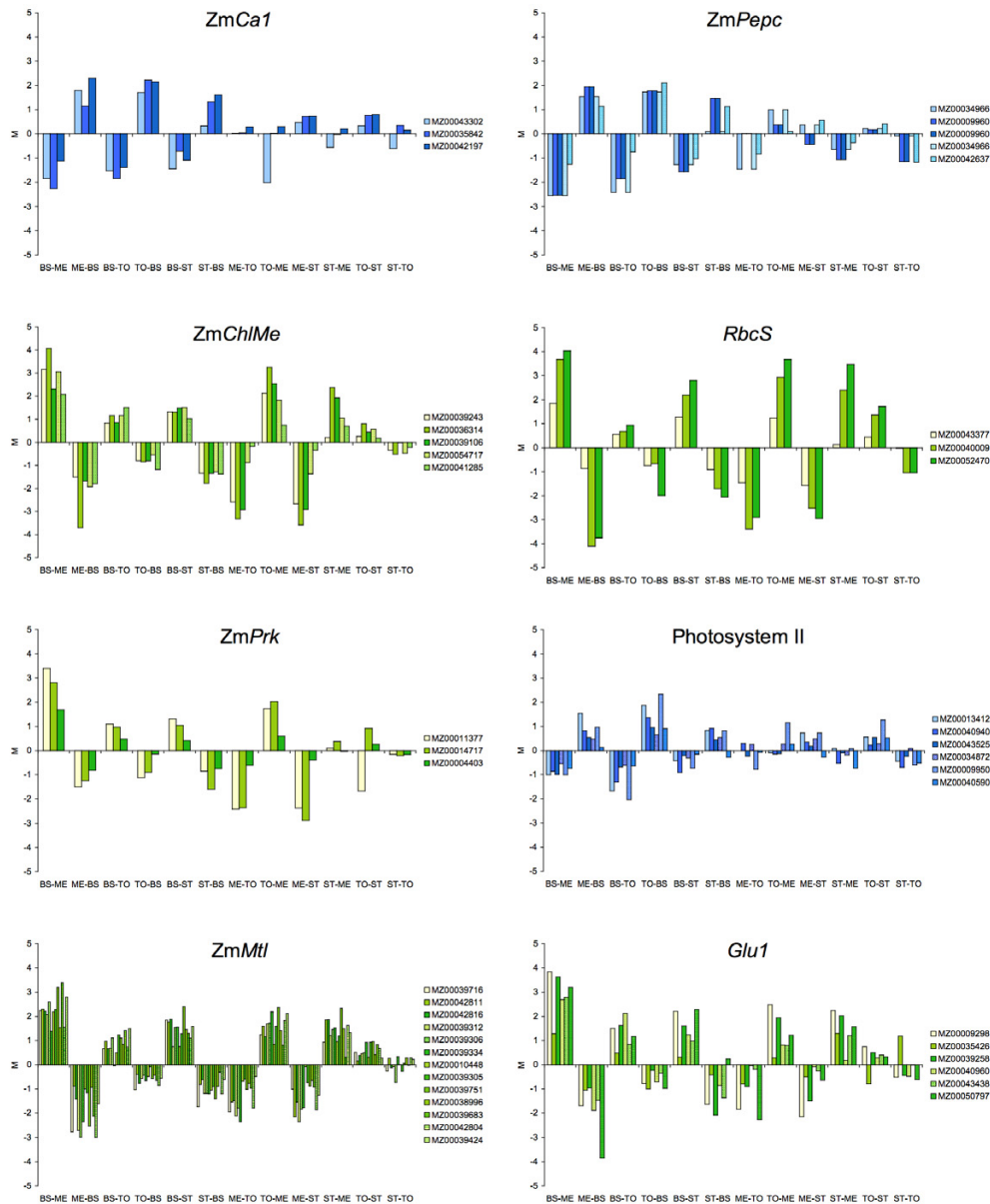
A somewhat different situation is illustrated by the multi-subunit Rubisco holoenzyme. Here, an enzyme that is abundant in all photosynthetic cell types in the  $C_3$  plant is restricted to the B in maize, although the biochemical role of the protein remains unchanged. Rubisco consists of a

number of large (LSU) and small (SSU) subunits. LSU is encoded by the chloroplast gene *rbcL* [51], while SSU is encoded by a family of nuclear *RbcS* genes [52-55]. In maize, both LSU and SSU are restricted to the B cells of mature leaf tissue by regulation of transcript accumulation [54-57]. In this experiment, three features were identified corresponding to *RbcS* and each showed the expected B enrichment. Two additional Calvin cycle enzymes, carbonic anhydrase (CA) and phosphoribulokinase (PRK) were also represented by multiple features on the array and displayed the predicted B-enriched patterns of expression (Figure 4).

Previous gene profiling of leaf cell-types in maize [26,58] and sorghum [28] have identified a number of metallothionein (MT) genes that are expressed preferentially in B cells. MTs are a family of small, metal-ion binding proteins that are found in many taxa and are hypothesized to play important roles in metal tolerance and homeostasis [59]. Consistent with these studies, we identified a number of B-enriched features corresponding to MT-like proteins (Figure 3). There are three annotated MT genes in maize, designated *ZmMtl1* [60], *ZmMtl2* [61] and *ZmMtl3* [62]. Although the features we identified showed homology to these genes, the majority were most similar to a non-characterized EST sequence (Gen Bank: [CF023010](#)) that we tentatively annotate as *ZmMtl4*. The role of MT proteins in the B is not immediately apparent. However, the analysis of Nakazono and colleagues [58] suggests an involvement in the functioning of the vasculature.

Proteomic analysis of maize B and M chloroplasts has identified  $\beta$ -glucosidase as among the most strongly B-enriched proteins [63]. Immunocytochemical studies have also demonstrated B enrichment of the major plastidic isoform of  $\beta$ -glucosidase in maize leaves [64,65]. We found six B-enriched features corresponding to  $\beta$ -glucosidase in our candidate gene list.  $\beta$ -glucosidase is proposed to function in plant defence by conversion of hydroxamic acid glucosides to toxic benzoxazolinones [66]. In addition,  $\beta$ -glucosidase activity has been implicated in cytokinin signalling [67]. Although providing a further marker for B identity, the significance of the localization of the enzyme in maize leaves remains unresolved.

In summary, approximately 8% (1,280 of 15,988) of the features analysed were identified as accumulating differentially between B and M when FDR is controlled at 5%. Approximately 50% of these features were fully or partially annotated by the maize array project database. Searching of maize sequences databases suggested that that these features represent at least 899 unique genes. An estimate of the overall predicted proportion of differentially accumulating features ( $1-\pi_0$ ) suggests that as many



**Figure 4**

**Accumulation of transcripts encoded by representative C<sub>4</sub> marker genes.** The ratios of signal intensity (*M* values) for each feature across the twelve hybridizations. Hybridizations are listed as Cy5 sample – Cy3 sample. Positive *M* values correspond to higher signal in the Cy5 channel. Features in green are predicted to be B enriched whereas features in blue are predicted to be M enriched. Features are listed according to MZ identifiers [87]. *ZmCa1* (carbonic anhydrase), *ZmPepc* (phosphoenolpyruvate carboxylase), *ZmChlMe* (malic enzyme), *RbcS* (Rubisco SSU), *ZmPrk* (phosphoribulokinase), *ZmMtl* (metallothionein), *Glu1* ( $\beta$ -glucosidase). The chart 'Photosystem II' shows features corresponding to a number of transcripts: MZ00013412, *psbH*; MZ00040940, *psbH*; MZ00043525, *psbS*; MZ00034872, *psbT*; MZ00009950, *psbW*; MZ00040590, *oee3*.



as 18% of features may be differentially expressed. Approximately 80 features identified in our gene list correspond to previously characterized marker genes and provide convincing evidence of the validity of our data set and analysis. Below we consider the significance of differential expression in the establishment and potential engineering of the C<sub>4</sub> syndrome.

#### **General comments on the differentiation of B and M cell types**

It is estimated that the C<sub>4</sub> syndrome has been derived at least 45 times in 19 families of angiosperms [5] and that a small number of regulatory changes are sufficient to establish a functional C<sub>4</sub> type [68]. By contrast, our observations suggest that differentiation of B and M cell-types in maize involves the cell-specific regulation of many thousands of genes. A number of these differences likely pre-date the acquisition C<sub>4</sub> photosynthesis. Ancestrally, the M would have been the major site of photosynthesis while the B would have had, and presumably retains, functions associated with proximity to the vascular tissue [69]. Indeed, C<sub>3</sub> isoforms of certain C<sub>4</sub>-related enzymes show specialised patterns of accumulation in the leaves of C<sub>3</sub> plants [70]. Additionally, a number of C<sub>4</sub> cell-specific promoters have been shown to be functional when introduced into C<sub>3</sub> species [71-76]. These observations suggest that spatial information distinguishing B and M also predates the shift to C<sub>4</sub>.

Although it is likely there were ancestral differences between B and M, our data and previous studies suggest that the establishment of the C<sub>4</sub> syndrome resulted in many additional changes to the accumulation of transcripts within these cell types. We distinguish two modes of regulation that might establish these differences (shown graphically in Figure 5). First, modification of *cis*-acting elements and the recruitment of transcription factors may directly change patterns of gene expression (genes A and B in Figure 5). Studies of C<sub>4</sub> gene regulation in maize have successfully identified such elements associated with a number of genes [16,17,19,77,47]. Although it has been assumed that regulatory changes of this type drive the establishment of the C<sub>4</sub> state, it appears unlikely that novel regulatory elements could be recruited on the scale required to explain the number of differentially expressed genes we have observed. As a second mechanism, we suggest that pre-existing regulatory mechanisms established in the C<sub>3</sub> state respond in a cell-specific manner to the creation of novel environments in the C<sub>4</sub> leaf (genes C and D in Figure 5). The B and M cells of a C<sub>4</sub> leaf differ in their complement of protein complexes, concentration of sugars, the redox poise of the photosynthetic electron transport chain and the availability of reducing equivalents [78]. More broadly, we suggest that a small number of changes in gene expression, when superim-

posed on ancestral cellular differences, would further induce secondary changes in transcript accumulation and thereby generate the complex pattern we have observed. It should be noted that such a model does not suggest the presence of regulatory 'master-switches' but rather the rebalancing of a complex system in response to key alterations. We have recently initiated transcriptional profiling of maize mutants with defects in phytochrome signalling [79], Calvin cycle function [22] and tetrapyrrole biosynthesis [80,81] to assess the effect of disrupting cellular conditions on cell-specific patterns of gene expression.

#### **Engineering C<sub>4</sub> photosynthesis**

The ability to express C<sub>4</sub>-associated transcripts in C<sub>3</sub> plants has stimulated interest in the molecular engineering of a C<sub>4</sub> state in C<sub>3</sub> plants to achieve high photosynthetic performance and water and nitrogen use efficiencies [82,83]. Unfortunately, the physiological effects of over-expressing individual carbon shuttle proteins in C<sub>3</sub> plants have been limited and difficult to interpret [82]. The next steps in C<sub>4</sub> engineering will require that multiple enzymes be expressed together. To achieve this, knowledge of the mechanisms that coordinate the distribution of activities in the C<sub>4</sub> leaf will be essential. In addition, an appreciation of the extent of *de novo* regulation required for the establishment of the C<sub>4</sub> state will aid the selection of targets for manipulation. Many of the most promising attempts to transfer individual C<sub>4</sub> activities to a C<sub>3</sub> plant have been achieved by transfer of maize genes to rice [71-74]. Similarly, the applicability of information regarding genome wide regulatory events will likely be greatest between closely related species. In this regard, comparative genomic and proteomic studies of maize and rice offer perhaps the greatest potential for understanding the establishment of the C<sub>4</sub> syndrome.

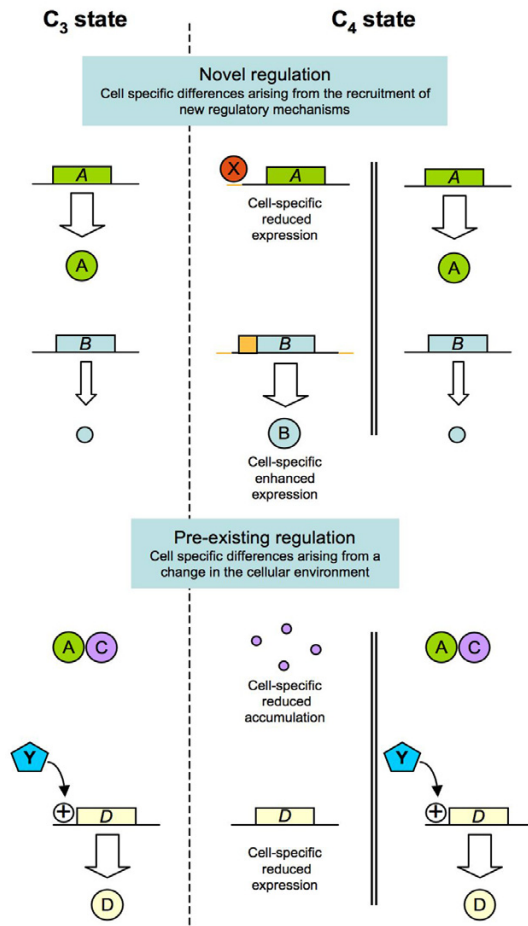
#### **Conclusion**

Differential gene expression was examined in separated B and M cell types from maize leaf blade tissue. To control for stress effects generated during the isolation process we developed a model that includes a stress term and compare the results of the analysis with a model lacking the stress term. These results suggest that gene expression changes are induced during the M cell isolation process and that this confounding effect can be reduced using the stress model. Our analysis indicates that 8% of features detected on the maize long-oligonucleotide microarray produced by the Maize Array Project consortium and up to 18% of genes expressed in the leaf transcriptome are differentially expressed between B and M cell types.

#### **Methods**

##### **Plant material and growth conditions**

Wild-type W22 inbred maize seedlings were grown in a growth chamber at 28 °C, 500 μmol m<sup>-2</sup> s<sup>-1</sup> light, 16 h



**Figure 5**  
**Mechanisms of differential regulation.** The C<sub>4</sub>-specific regulation of four hypothetical genes and their protein products is illustrated. During the establishment of the C<sub>4</sub> syndrome, novel regulatory mechanisms that were absent in the C<sub>3</sub> state control the accumulation of proteins A and B. In the case of A, a protein that accumulates in all photosynthetic cell types in the C<sub>3</sub> state is restricted to a specific cell type, by a *trans* acting factor X, without any change to biochemical function. Gene B, expressed at low levels, accumulates in a specific cell type. Proteins C and D also accumulate differentially in the C<sub>4</sub> state, but, in these instances, differential accumulation is the result of a novel cellular environment. In the case of C, protein accumulation is regulated post-translationally at the level of assembly or stability through interaction with A. In the C<sub>4</sub> state and the absence of A, product C fails to accumulate. Such a scenario does not require differential accumulation of C transcripts, but could affect transcription of other genes. Transcription of gene D is linked, either directly or indirectly, to the presence of a metabolic signal Y. The accumulation of Y is governed by the activities of proteins A and B. In the C<sub>4</sub> state, changes in the accumulation of A and B affect the levels of Y and subsequently alter the accumulation of D. Boxes represent genes and spheres represent gene products.

days, 8 h nights. Light was provided by a combination of 400-W metal halide and 100-W halogen lamps. Ten-day old seedlings were harvested 2 h after the start of the light period for bundle sheath and mesophyll preparation.

**Preparation of bundle sheath strands and mesophyll protoplasts**

Bundle sheath strands and mesophyll protoplasts were prepared as previously described [22,84]. Approximately 5 g of tissue were harvested from the second and third leaves of 10-day-old maize seedlings for each mesophyll preparation. Approximately 4 g of tissue were harvested from the second and third leaves for each bundle sheath preparation.

**Preparation of RNA**

RNA was prepared as previously described [85]. DNase treatment was performed using amplification grade DNase I (Invitrogen, Carlsbad, CA) in the presence of RNase OUT RNase Inhibitor (Invitrogen). Following treatment, RNA was extracted first with phenol:chloroform:IAA (24:1:1) and then with chloroform:IAA (24:1). Following extraction, RNA was ethanol precipitated, washed twice in 70% ethanol and re-suspended in DEPC-treated dH<sub>2</sub>O.

**Microarray detection**

Microarray detection was performed using the Genisphere 3DNA 900 MPX two-stage labelling kit (Genisphere, Hatfield, PA). 8 µg of DNase-treated total RNA were used per labelling reaction and resulting cDNA products split between a two-slide set covering a total footprint area of ~3000 mm<sup>2</sup>. cDNA synthesis was primed using oligo dT and random primers according to the manufacturer's protocol. Maize oligonucleotide microarrays were obtained from the Maize Array Project (University of Arizona) as described [32]. Arrays were imaged using the Scan Array 5000 system (Perkin Elmer, Wellesley, MA). Intermediate laser gain (60–70%) was used to detect a majority of features while minimizing the problem of signal saturation from highly expressed genes.

**Preliminary data processing and background correction**

Preliminary segmentation and data extraction were performed using Imogene software (Biodiscovery, El Segundo, CA). For each feature, median signal (SMD) and background (BMD) values were extracted. To correct for background noise, the difference between the log values of SMD and BMD was calculated (Corrected intensity = log<sub>2</sub>(SMD) - log<sub>2</sub>(BMD)) [86]. Corrected data sets were examined graphically by plotting the difference between Cy5 and Cy3 (M = corrected intensity Cy5 - corrected intensity Cy3) against the average intensity of Cy5 and Cy3 signals [A = (corrected intensity Cy5 + corrected intensity Cy3)/2] for each slide.

### LOWESS normalization and data filtering

A LOWESS procedure [35] was used to normalize signal intensity between channels for every slide. The difference ( $M$ ) and average intensity ( $A$ ) values were calculated as described above. A LOWESS regression was applied to  $M$  against  $A$  and the resulting trend-line was used to centralise  $M$  values around 0 and to correct for any dependence of  $M$  on  $A$ . Unreliable or uninformative data were discarded to reduce the dimensionality of the analysis. Data were discarded if intensity measurements were considered either too low (i.e. detection failed) or too high (i.e. signal saturation). Filtering criteria were applied to data obtained from  $T_1$  hybridizations to retain features showing extreme expression profiles only under certain conditions. The background corrected intensities were averaged across the six  $T_1$  hybridization data sets. A feature was discarded if this average was less than 1 (i.e. geometric average  $SMD < 2 \times BMD$ ). Additionally, a feature was discarded as saturating if 3 or more out of the 6  $T_1$  signal intensities read at the maximum. Following data filtering, 47,591 features were discarded from an original 64,896 because of low expression. Of the remainder, a further 178 features were removed because of saturation. Excluding the control spots, we have a total of 15,988 unique features (25% of the original set) for subsequent analysis. The normalized  $M$  values for these 15,988 features are provided in Additional file 1.

### Authors' contributions

RS designed the study and analytical models, performed experimental procedures and data extraction and drafted the manuscript. PL and GH performed data analysis, including development and implementation of analytical models. KA assisted in the preparation of isolated cell-types and the extraction of RNA. TB contributed to the experimental design and drafting of the manuscript. All authors have read and approved the final manuscript.

### Additional material

#### Additional file 1

*Normalised M values for features passing data-filtering. Feature list spreadsheet. M values are provided for each of 12 slides. See Methods for details.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-12-S1.xls>]

#### Additional file 2

*Estimates of (v1-v2) under stress model and annotation for features passing data-filtering. Feature list spreadsheet. (v1-v2) estimate of the log normalised difference between B (v1) and M (v2) expression derived from the stress model described in the text. FOLD is the degree of enrichment as calculated by back-transformation of the (v1-v2) estimate.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-12-S2.xls>]

#### Additional file 3

*Estimates of (v1-v2) under simple model and annotation for features passing data-filtering. Feature list spreadsheet. (v1-v2) estimate of the log normalised difference between B (v1) and M (v2) expression derived from the simple model described in the text without stress term. FOLD is the degree of enrichment as calculated by back-transformation of the (v1-v2) estimate under simple model or stress model as labeled.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-12-S3.xls>]

### Acknowledgements

We would like to acknowledge Paul Debbie (CGEP, Boyce Thompson Institute) for technical assistance with microarray analysis and Drs. Rob Alba, Wojciech Majeran, Klaas van Wijk and Qi Sun for helpful discussions. We want to thank Dr. Chong Wang for discussions about this data set and help with the REML calculation. We would also like to thank Sarah Covshoff for helpful comments on the manuscript, Nicole Markelz for contributing the image of maize B strands and M cell protoplasts and the Maize Array Project for providing the oligonucleotide arrays. This work was supported by a National Science Foundation grant to T.P.B. (DBI- 0211935).

### References

1. Sage RF, Monson RK: **C4 plant biology**. San Diego, Academic press; 1999.
2. Bowes G, Ogren WL, Hageman RH: **Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase**. *Biochem Biophys Res Commun* 1971, **45(3)**:716-722.
3. Ogren WL: **Photorespiration: pathways, regulation and modification**. *Annu Rev Plant Physiol* 1984, **35**:415-442.
4. Moore P: **Evolution of photosynthetic pathways in flowering plants**. *Nature* 1982, **295**:647-648.
5. Sage RF: **The evolution of C4 photosynthesis**. *New Phytol* 2004, **161**:341-370.
6. Brown NJ, Parsley K, Hibberd JM: **The future of C4 research--maize, Flaveria or Cleome?** *Trends Plant Sci* 2005, **10(5)**:215-221.
7. Meierhoff K, Westhoff P: **Differential biogenesis of photosystem II in mesophyll and bundle sheath cells of monocotyledonous NADP-malic enzyme type C4 plants: The non-stoichiometric abundance of the subunits of photosystem II in the bundle sheath chloroplasts and the translational activity of the plastome encoded genes**. *Planta* 1993, **191**:23-33.
8. Schuster G, Ohad I, Martineau B, Taylor WC: **Differentiation and development of bundle sheath and mesophyll thylakoids in maize. Thylakoid polypeptide composition, phosphorylation, and organization of photosystem II**. *J Biol Chem* 1985, **260(21)**:11866-11873.
9. Sheen J, Bogorad L: **Differential expression in bundle sheath and mesophyll cells of maize of genes for photosystem II components encoded by the plastid genome**. *Plant Physiol* 1988, **86**:1020-1026.
10. Hatch MD: **C4 photosynthesis: an unlikely process full of surprises**. *Plant Cell Physiol* 1992, **33**:333-342.
11. Leegood RC: **The regulation of C4 photosynthesis**. *Adv Bot Res* 1997, **26**:251-316.
12. Doullis AG, Debian N, Kingston-Smith AH, Foyer CH: **Differential localization of antioxidants in maize leaves**. *Plant Physiol* 1997, **114**:1031-1037.
13. Rathnam CK, Edwards GE: **Distribution of nitrate-assimilating enzymes between mesophyll protoplasts and bundle sheath cells in leaves of three groups of C(4) plants**. *Plant Physiol* 1976, **57(6)**:881-885.
14. Langdale JA, Taylor WC, Nelson T: **Cell-specific accumulation of maize phosphoenolpyruvate carboxylase is correlated with demethylation at a specific site greater than 3 kb upstream of the gene**. *Mol Gen Genet* 1991, **225(1)**:49-55.

15. Schaffner AR, Sheen J: **Maize rbcS promoter activity depends on sequence elements not found in dicot rbcS promoters.** *Plant Cell* 1991, **3(9)**:997-1012.
16. Sheen J: **Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes.** *Plant Cell* 1991, **3(3)**:225-245.
17. Schaffner AR, Sheen J: **Maize C4 photosynthesis involves differential regulation of phosphoenolpyruvate carboxylase genes.** *Plant J* 1992, **2(2)**:221-232.
18. Sheen J: **C4 Gene Expression.** *Annu Rev Plant Physiol Plant Mol Biol* 1999, **50**:187-217.
19. Xu T, Purcell M, Zucchi P, Helentjaris T, Bogorad L: **TRM1, a YYI-like suppressor of rbcS-m3 expression in maize mesophyll cells.** *Proc Natl Acad Sci U S A* 2001, **98(5)**:2295-2300.
20. Hall LN, Rossini L, Cribb L, Langdale JA: **GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf.** *Plant Cell* 1998, **10(6)**:925-936.
21. Hall LN, Roth R, Brutnell TP, Langdale JA: **Cellular differentiation in the maize leaf is disrupted by bundle sheath defective mutations.** *Symp Soc Exp Biol* 1998, **51**:27-31.
22. Brutnell TP, Sawers RJ, Mant A, Langdale JA: **BUNDLE SHEATH DEFECTIVE2, a novel protein required for post-translational regulation of the rbcL gene of maize.** *Plant Cell* 1999, **11(5)**:849-864.
23. Sheen J: **Methods for mesophyll and bundle sheath cell separation.** *Methods Cell Biol* 1995, **49**:305-314.
24. Nelson T: **Preparation of DNA and RNA from leaves: expanded blades and separated bundle sheath and mesophyll cells.** In *The maize handbook* Edited by: Freeling M, Walbot V. New York, Springer-Verlag; 1994:541-544.
25. Ishii S: **Factors influencing protoplast viability of suspension-cultured rice cells during isolation process.** *Plant Physiol* 1988, **88(1)**:26-29.
26. Furumoto T, Hata S, Izui K: **Isolation and characterization of cDNAs for differentially accumulated transcripts between mesophyll cells and bundle sheath strands of maize leaves.** *Plant Cell Physiol* 2000, **41(11)**:1200-1209.
27. Hahnen S, Joeris T, Kreuzaler F, Peterhansel C: **Quantification of photosynthetic gene expression in maize C(3) and C(4) tissues by real-time PCR.** *Photosynth Res* 2003, **75(2)**:183-192.
28. Wyrlich R, Dressen U, Brockmann S, Streubel M, Chang C, Qiang D, Paterson AH, Westhoff P: **The molecular basis of C4 photosynthesis in sorghum: isolation, characterization and RFLP mapping of mesophyll- and bundle-sheath-specific cDNAs obtained by differential screening.** *Plant Mol Biol* 1998, **37(2)**:319-335.
29. Clarke JD, Zhu T: **Microarray analysis of the transcriptome as a stepping stone towards understanding biological systems: practical considerations and perspectives.** *Plant J* 2006, **45(4)**:630-650.
30. Markelz NH, Costich DE, Brutnell TP: **Photomorphogenic responses in maize seedling development.** *Plant Physiol* 2003, **133(4)**:1578-1591.
31. Kerr MK, Churchill GA: **Experimental design for gene expression microarrays.** *Biostatistics* 2001, **2(2)**:183-201.
32. **Maize Oligonucleotide Array Project** [<http://www.maizearray.org/index.shtml>]
33. Gardiner JM, Buell CR, Elumalai R, Galbraith DW, Henderson DA, Iniguez AL, Kaeppler SM, Kim JJ, Liu J, Smith A, Zheng L, Chandler VL: **Design, production, and utilization of long oligonucleotide microarrays for expression analysis in maize.** *Maydica* 2005, **50**:425-435.
34. **Maize Oligonucleotide Array Studies** [[http://www.maizearray.org/tigr-scripts/maizearray/studmaize\\_study\\_hybs.pl?study=4&user=&pass=&sort=id&order=asc](http://www.maizearray.org/tigr-scripts/maizearray/studmaize_study_hybs.pl?study=4&user=&pass=&sort=id&order=asc)]
35. Dudoit S, Yang YH, Callow MJ, Speed TP: **Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments.** *Statistica Sinica* 2002, **12**:111-139.
36. Cui X, Hwang JT, Qiu J, Blades NJ, Churchill GA: **Improved statistical tests for differential gene expression by shrinking variance components estimates.** *Biostatistics* 2005, **6(1)**:59-75.
37. Storey JD, Taylor JE, Siegmund D: **Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach.** *Journal of the Royal Society, Series B* 2004, **66**:187-205.
38. Oono Y, Seki M, Nanjo T, Narusaka M, Fujita M, Satoh R, Satou M, Sakurai T, Ishida J, Akiyama K, Iida K, Maruyama K, Satoh S, Yamaguchi-Shinozaki K, Shinozaki K: **Monitoring expression profiles of Arabidopsis gene expression during rehydration process after dehydration using ca 7000 cDNA microarray.** *Plant J* 2003, **34**:868-887.
39. Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K: **Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray.** *Plant J* 2002, **31**:279-292.
40. **The TIGR Maize Database** [<http://maize.tigr.org/>]
41. **Maize Oligonucleotide Array Search** [[http://www.maizearray.org/maize\\_search\\_basic.shtml](http://www.maizearray.org/maize_search_basic.shtml)]
42. Lee I, Dombkowski AA, Athey BD: **Guidelines for incorporating non-perfectly matched oligonucleotides into target-specific hybridization probes for a DNA microarray.** *Nucleic Acids Res* 2004, **32(2)**:681-690.
43. SantaLucia Jr.: **A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.** *Proc Natl Acad Sci U S A* 1998, **95(4)**:1460-1465.
44. Ma J, Morrow DJ, Fernandes J, Walbot V: **Comparative profiling of the sense and antisense transcriptome of maize lines.** *Genome Biol* 2006, **7(3)**:R22.
45. Sheen JY, Bogorad L: **Differential expression of C4 pathway genes in mesophyll and bundle sheath cells of greening maize leaves.** *J Biol Chem* 1987, **262(24)**:11726-11730.
46. Langdale JA, Rothermel BA, Nelson T: **Cellular pattern of photosynthetic gene expression in developing maize leaves.** *Genes Dev* 1988, **2(1)**:106-115.
47. Tausta SL, Coyle HM, Rothermel B, Stiefel V, Nelson T: **Maize C4 and non-C4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor.** *Plant Mol Biol* 2002, **50**:635-652.
48. Rothermel BA, Nelson T: **Primary structure of the maize NADP-dependent malic enzyme.** *J Biol Chem* 1989, **264(33)**:19587-19592.
49. Saigo M, Bologna FP, Maurino VG, Detarsio E, Andreo CS, Drincovich MF: **Maize recombinant non-C4 NADP-malic enzyme: a novel dimeric malic enzyme with high specific activity.** *Plant Mol Biol* 2004, **55**:97-107.
50. Svensson P, Blasing OE, Westhoff P: **Evolution of the enzymatic characteristics of C4 phosphoenolpyruvate carboxylase--a comparison of the orthologous PPCA phosphoenolpyruvate carboxylases of Flaveria trinervia (C4) and Flaveria pringlei (C3).** *Eur J Biochem* 1997, **246(2)**:452-460.
51. Coen DM, Bedbrook JR, Bogorad L, Rich A: **Maize chloroplast DNA fragment encoding the large subunit of ribulose biphosphate carboxylase.** *PNAS* 1977, **74**:5487-5491.
52. Matsuoaka M, Kano-Murakami Y, Tanaka Y, Ozeki Y, Yamamoto N: **Nucleotide sequence of cDNA encoding the small subunit of ribulose-1,5-bisphosphate carboxylase from maize.** *J Biochem (Tokyo)* 1987, **102(4)**:673-676.
53. Lebrun M, Waksman G, Feyssinet G: **Nucleotide sequence of a gene encoding corn ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (rbcs).** *Nucleic Acids Res* 1987, **26**:4360.
54. Sheen J, Bogorad L: **Expression of the ribulose-1,5-bisphosphate carboxylase large subunit gene and three small subunit genes in two cell types of maize leaves.** *Embo J* 1986, **5**:3417-3422.
55. Ewing RM, Jenkins GI, Langdale JA: **Transcripts of maize RbcS genes accumulate differentially in C3 and C4 tissues.** *Plant Mol Biol* 1998, **36(4)**:593-599.
56. Link G, Coen DM, Bogorad L: **Differential expression of the gene for the large subunit of ribulose biphosphate carboxylase in maize leaf cell types.** *Cell* 1978, **15(3)**:725-731.
57. Huber SC, Hall TN, Edwards GE: **Differential Localization of Fraction I Protein between Chloroplast Types.** *Plant Physiol* 1976, **57**:730-733.
58. Nakazono M, Qiu F, Borsuk LA, Schnable PS: **Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize.** *Plant Cell* 2003, **15(3)**:583-596.

59. Robinson NJ, Tommey AM, Kuske C, Jackson PJ: **Plant metallothioneins.** *Biochem J* 1993, **295 ( Pt 1)**:1-10.
60. de Framond AJ: **A metallothionein-like gene from maize (Zea mays). Cloning and characterization.** *FEBS Lett* 1991, **290**:103-106.
61. White CN, Rivin CJ: **Characterization and expression of a cDNA encoding a seed-specific metallothionein in maize.** *Plant Physiol* 1995, **108(2)**:831-832.
62. Charbonnel-Campaa L, Lauga B, Combes D: **Isolation of a type 2 metallothionein-like gene preferentially expressed in the tapetum in Zea mays.** *Gene* 2000, **254**:199-208.
63. Majeran W, Cai Y, Sun Q, van Wijk KJ: **Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics.** *Plant Cell* 2005, **17(11)**:3111-3140.
64. Nikus J, Jonsson LMV: **Tissue localization of B-glucosidase in rye, maize and wheat seedlings.** *Physiol Plant* 1999, **107**:373-378.
65. Nikus J, Daniel G, Jonsson LMV: **Subcellular localization of B-glucosidase in rye, maize and wheat seedlings.** *Physiol Plant* 2001, **111**:466-472.
66. Niemeier HM: **Hydroxamic acids (4-hydroxy-1,4-benzoxazin-3-ones) defense chemicals in the Gramineae.** *Phytochem* 1988, **27**:3349-3358.
67. Brzobohaty B, Moore I, Kristoffersen P, Bako L, Campos N, Schell J, Palme K: **Release of active cytokinin by a beta-glucosidase localized to the maize root meristem.** *Science* 1993, **262(5136)**:1051-1054.
68. Ku MS, Kano-Murakami Y, Matsuoka M: **Evolution and expression of C4 photosynthesis genes.** *Plant Physiol* 1996, **111(4)**:949-957.
69. Dengler N, Nelson T: **Leaf structure and development in C4 plants.** In *C4 Plant Biology* Edited by: Sage RF, Monson RK. San Diego, Academic Press; 1999:133-172.
70. Hibberd JM, Quick WP: **Characteristics of C4 photosynthesis in stems and petioles of C3 flowering plants.** *Nature* 2002, **415(6870)**:451-454.
71. Matsuoka M, Kyoizuka J, Shimamoto K, Kano-Murakami Y: **The promoters of two carboxylases in a C4 plant (maize) direct cell-specific, light-regulated expression in a C3 plant (rice).** *Plant J* 1994, **6(3)**:311-319.
72. Matsuoka M, Sanada Y: **Expression of photosynthetic genes from the C4 plant, maize, in tobacco.** *Mol Gen Genet* 1991, **225(3)**:411-419.
73. Ku MS, Agarie S, Nomura M, Fukayama H, Tsuchida H, Ono K, Hirose S, Toki S, Miyao M, Matsuoka M: **High-level expression of maize phosphoenolpyruvate carboxylase in transgenic rice plants.** *Nat Biotechnol* 1999, **17(1)**:76-80.
74. Tsuchida H, Tamai T, Fukayama H, Agarie S, Nomura M, Onodera H, Ono K, Nishizawa Y, Lee BH, Hirose S, Toki S, Ku MS, Matsuoka M, Miyao M: **High level expression of C4-specific NADP-malic enzyme in leaves and impairment of photoautotrophic growth in a C3 plant, rice.** *Plant Cell Physiol* 2001, **42(2)**:138-145.
75. Svensson P, Blasing OE, Westhoff P: **Evolution of C4 phosphoenolpyruvate carboxylase.** *Arch Biochem Biophys* 2003, **414(2)**:180-188.
76. Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, Westhoff P: **cis-Regulatory elements for mesophyll-specific gene expression in the C4 plant Flaveria trinervia, the promoter of the C4 phosphoenolpyruvate carboxylase gene.** *Plant Cell* 2004, **16(5)**:1077-1090.
77. Metzler MC, Rothermel BA, Nelson T: **Maize NADP-malate dehydrogenase: cDNA cloning, sequence, and mRNA characterization.** *Plant Mol Biol* 1989, **12**:713-722.
78. Kanai R, Edwards G: **The biochemistry of C4 photosynthesis.** In *C4 Plant Biology* Edited by: Sage RF, Monson RK. San Diego, Academic Press; 1999:49-87.
79. Sheehan MJ, Kennedy LM, Costich DE, Brutnell TP: **Subfunctionalization of PhyB1 and PhyB2 in the control of seedling and mature plant traits in maize.** *Plant J* 2007, **49**:338-53.
80. Sawers RJ, Linley PJ, Gutierrez-Marcos JF, Delli-Bovi T, Farmer PR, Kohchi T, Terry MJ, Brutnell TP: **The Elm1 (ZmHy2) gene of maize encodes a phytochromobilin synthase.** *Plant Physiol* 2004, **136(1)**:2771-2781.
81. Sawers RJ, Viney J, Farmer PR, Bussey RR, Olsefski G, Anufrikova K, Hunter CN, Brutnell TP: **The maize Oil yellow1 (Oy1) gene encodes the I subunit of magnesium chelatase.** *Plant Mol Biol* 2006, **60(1)**:95-106.
82. Matsuoka M, Furbank RT, Fukayama H, Miyao M: **Molecular Engineering of C4 Photosynthesis.** *Annu Rev Plant Physiol Plant Mol Biol* 2001, **52**:297-314.
83. Mitchell PL, Sheehy JE: **Supercharging rice photosynthesis to increase yield.** *New Phytol* 2006, **171**:688-693.
84. Westhoff P, Offermann-Steinhard K, Hofer M, Eskins K, Oswald A, Streubel M: **Differential accumulation of plastidic transcripts encoding photosystem II components in the mesophyll and bundle sheath cells of monocotyledonous NADP-malic enzyme-type C4 plants.** *Planta* 1991, **184**:377-388.
85. Sawers RJ, Linley PJ, Farmer PR, Hanley NP, Costich DE, Terry MJ, Brutnell TP: **Elongated mesocotyl1, a phytochrome-deficient mutant of maize.** *Plant Physiol* 2002, **130(1)**:155-163.
86. Zhang D, Zhang M, Wells MT: **Multiplicative background correction for spotted microarrays to improve reproducibility.** *Genet Res* 2006, **87(3)**:195-206.
87. **Maize Oligonucleotide Array Overview** [[http://www.maizearray.org/maize\\_search\\_overview.shtml](http://www.maizearray.org/maize_search_overview.shtml)]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

