Differential expression of starch and sucrose metabolic genes linked to varying biomass yield in *Miscanthus* hybrids

Jose J De Vega[1]*, Ned Peel[1], Sarah J Purdy[2#], Sarah Hawkins[2], Iain Donnison[2], Sarah Dyer[1%], Kerrie Farrar[2]*

[1] Earlham Institute, Norwich, NR4 7UZ, UK.

[2] Institute of Biological, Environmental & Rural Sciences (IBERS), Aberystwyth University, Aberystwyth, SY23 3EE, UK.

[#] Current address: NSW Department of Primary Industries, Chief Scientist's Branch, Locked Bag 21, Orange NSW, 2800, Australia.

[%] Current address: NIAB, Cambridge, CB3 0LE, UK.

* Correspondence to JDV (jose.devega@earlham.ac.uk) and KF (kkf@aber.ac.uk).

ABSTRACT

Miscanthus is a commercial lignocellulosic biomass crop owing to its high biomass productivity and low chemical input requirements. Interspecific *Miscanthus* hybrids with high biomass yield were shown to have low concentrations of starch and sucrose but high concentrations of fructose. We performed a transcriptional RNA-seq analysis between selected *Miscanthus* hybrids with contrasting values for these phenotypes to clarify how these phenotypes are genetically controlled. We observed that genes directly involved in the synthesis and degradation of starch and sucrose were down-regulated in high yielding *Miscanthus* hybrids. At the same time, glycolysis and export of triose phosphates were up-regulated in high yielding *Miscanthus* hybrids. Our results evidence a direct relationship between high expression of essential enzymatic genes in the starch and sucrose pathways, high starch concentrations, and lower biomass production. The strong interconnectivity between genotype, chemotype and agronomic traits opens the door to use the expression of well-characterised genes in the starch and sucrose pathway for the early selection of high biomass yielding genotypes from large *Miscanthus* populations.

INTRODUCTION

Miscanthus is a candidate biofuel crop owing to its high biomass yield and low input requirements [1, 2]. It is also naturally adapted to a wide range of climate zones and land types [3, 4]. Currently, *Miscanthus* is mainly used for combustion, but there is keen interest in its development as a sustainable substrate for bioethanol or biomethane production [5, 6].

Miscanthus is a C4 perennial rhizomatous grass crop closely related to sugarcane (*Saccharum* spp.), sorghum (*S. bicolor*) and maize (*Zea mays*). However, unlike these species, *Miscanthus* is a non-food crop and can be grown on lower agricultural grade or marginal land so as not to compete with food production [7, 8].

Natural interspecific hybridisation events occur between several *Miscanthus* species with overlapping geographic distributions [9]. The main commercial *Miscanthus* genotype to date*, M. x giganteus*, is a sterile triploid wild hybrid resulting from the hybridisation between a diploid *M. sinensis* and a tetraploid *M. sacchariflorus. M. x giganteus* has desirable traits, including high yield and early establishment [10-12]. However, *M. x giganteus* must be clonally propagated, which doubles establishments costs compared to a seed-based option [13]. Therefore, several European breeding programmes are aiming to develop a seed-based crop through recreating the hybridisation event between *M. sinensis* and *M. sacchariflorus* to produce new hybrids that out-perform *M. x giganteus* [13, 14]. The hybrids produced at IBERS (Wales, UK) exhibited strong heterosis for several traits and have been characterised in previous publications [15-18].

A major hindrance to the improvement of perennial energy crops through breeding is the long duration for new crosses to reach the maturity stage when their traits can be assessed [18]. This creates a pressing need for the identification of molecular markers to predict yield before maturity is reached. Genetic markers were identified via association mapping for seven-teen traits in *Miscanthus* [19], and metabolic biomarkers successfully predicted the final yield eight months later [17]. The identification of transcriptional predictors in *Miscanthus* could provide a cost-effective tool to accelerate selection either using expression level as a marker or by identifying new target genes.

Miscanthus is harvested for the structural cell-wall polysaccharides, and as a result, multiple studies have focused on its structural carbohydrates [20, 21]. However, it is the processing and storage of non-structural carbohydrates (NSC), such as sucrose and starch, that underpin biomass traits [17].

We have previously shown that high yielding *Miscanthus* genotypes from an interspecific hybrid mapping family had low starch concentrations in the stem and a low ratio of starch to fructose [17]. These distinctive carbohydrate profiles were consistent across years and growing environments; thus, the phenotype is likely to be genetically controlled [17, 22]. Unlike many C3 temperate grasses, C4 species such as *Miscanthus* or maize do not accumulate fructans but instead accumulate starch as a transient form of storage carbohydrate [23, 24]. The concentration of starch in the mapping family was up to 15 % of the dry weight (DW) on average. However, higher values were observed in the lowest yielding lines, raising the

possibility to bred "starch-cane" *Miscanthus* for liquid biofuel or biogas generation [25]. Identifying differentially expressed genes (DEGs) that relate to the carbohydrate profile could further facilitate breeding for such traits.

In this study, we analysed root, stem and leaf RNA-seq data from the hybrid progeny from a cross of a diploid *M. sacchariflorus* genotype and a diploid *M. sinensis* genotype, which had contrasting carbohydrate profiles and yield measurements. We identified differentially expressed genes associated with the observed metabolic profiles, using the recently completed *M. sinensis* reference genome (*Miscanthus sinensis* v7.1 DOE-JGI). Integrating expression and metabolic data is a logical strategy given the strong interconnectivity between genotype, chemotype and phenotype, and the lower genetic complexity of intermediate phenotypes, such as metabolites and yield subcomponents [26, 27].

METHODS

Mapping population establishment and phenotyping

A total of 102 genotypes from a paired cross between diploid *M. sinensis* genotype "*M. sinen* 102" and a diploid *M. sacchariflorus* genotype "*M. sacch* 297" were sown from seed in trays in a glasshouse in 2009. In 2010, individual plants were split to form three replicates of each genotype and then planted out into the field in a spaced-plant randomised block design comprising three replicate blocks at IBERS, Aberystwyth, UK. Details of the phenotyping were previously described [17]. Briefly, the family was harvested in February 2015 following the 2014 growing season. Biomass was dried to a constant weight, and the average DW weight per plant (kg)

was calculated. Soluble sugars were extracted and quantified enzymatically and photometrically from known standard curves on the same plate, as previously detailed [18]. Starch was extracted using a modified *Megazyme* commercial assay procedure and quantified photometrically from known standard curves on the same plate, as previously described [18]. Four hybrid genotypes were selected based on a low or high number of tillers (transect count of tillers). Correlation between concentrations, plant height and biomass phenotypes for the whole mapping population was previously quantified [17]. Pearson's correlation values between the number of tillers and the other phenotypes were determined for the whole family. Differences between the four selected hybrids for all phenotypes were determined with Student's two-tailed t-tests.

RNA sequencing and pre-processing

RNA was extracted from the four selected hybrids, as well as from the two parents of the family. Extraction was performed using RNeasy Plant Mini kit (Qiagen, CA, USA) according to the manufacturer's instructions. Total RNA samples were sent to the sequencing service at the Earlham Institute (Norwich, UK) where standard Illumina RNA-seq libraries were prepared and sequenced using the HiSeq 2000 platform. The raw reads were filtered with Trim Galore [28] using the default options for paired-end reads to remove Illumina adaptor sequences and reads with quality scores below 20. Cleaned reads were aligned to the *M. sinensis* reference genome (*Miscanthus sinensis* v7.1 DOE-JGI, http://phytozome.jgi.doe.gov) downloaded from Phytozome with STAR using the "2-pass" mode [29] and Kallisto using the "quant" mode with default options [30]. In both cases, the reference was indexed using the *M. sinensis* gene annotation (*Miscanthus sinensis* v7.1 DOE-JGI,

http://phytozome.jgi.doe.gov) downloaded from Phytozome in GFF3 format. This same gene annotation was functionally annotated with GO terms and enzyme codes with the command-line version of Blast2GO [31] using BLASTX with an E-value of 1e-10 and the NCBI non-redundant (nr) and EBI InterPro databases.

Differential expression and enrichment analysis

The differential expression and enrichment analysis are fully available in an R notebook (See Data availability), which is also available via Github. Briefly, Kallisto count files, one from each of the 23 libraries, were imported in R using TXimport [32]. Differential analysis was performed using DESeq2 [33] for each tissue (root, stem, leaf) independently. Raw gene counts were obtained from Kallisto alignments and normalised using DESeq2 for the top 1,000 most variable genes to cluster the samples. Genes with a False Discovery Rate (FDR) < 0.05 were considered differentially expressed (DE). We compared two groups of hybrids; each hybrid group was composed of two genotypes (genotypes 112 and 90 against genotypes 18 and 120). We also compared the hybrids against the *M. sacchariflorus* and *M. sinensis* parent, one at the time. A gene only was considered DE between hybrids and parents when it was DE against both parents. The enrichment analysis was based on an F-fisher test (FDR < 0.05) using the library topGO [34] with the "weight01" algorithm. Using the lists of DE genes and functional annotation as inputs, topGO compared the number of DEGs in each category with the expected number of genes for the whole transcriptome. The "weight01" algorithm resolves the relations between related GO ontology terms at different levels. Enriched categories were plotted using ggplot2 [35]. Genes in enriched GO terms were further analysed

158 using the online Phytomine [36] and Thalemine [37] databases. Genes annotated
159 with enzyme codes were plotted using the online KEGG mapper [38].
160
161 **RESULTS**
162
163 **Contrasting carbohydrate metabolism in sequenced genotypes from a**
164 ***Miscanthus* mapping family**
165 A total of 102 genotypes from a paired cross between diploid *M. sinensis* ("*M. sinen*
166 102") and a diploid *M. sacchariflorus* ("*M. sacch* 297") were established in field
167 conditions and phenotyped. Non-structural carbohydrates were sampled in July
168 2014, during the summer growing season, and annual yield was obtained at harvest
169 after the following winter. The distribution of carbohydrate concentrations and
170 biomass yield for 98 hybrids were previously reported [17]. After including additional
171 information about number of tillers for the population (Fig. 1 and Suppl. Table S1),
172 we observed significant correlations between number of tillers and starch ($r = -0.45$,
173 $p < 0.001$), fructose ($r = 0.31$, $p < 0.005$), and total NSC ($r = -0.40$, $p < 0.0001$) for
174 the whole family (Suppl. Table S1). We also observed significant correlations
175 between number of tillers and the ratio of sucrose/starch ($r = 0.37$, $p < 0.001$),
176 fructose/starch ($r = -0.45$, $p < 0.001$), glucose/starch ($r = -0.38$, $p < 0.001$) and
177 sucrose/fructose ($r = -0.32$, $p < 0.01$). We observed a significant positive correlation
178 between biomass yield and number of tillers ($r = 0.62 \pm 0.03$ for three seasons, $p <$
179 0.001).
180
181 Four *M. sinensis* X *M. sacchariflorus* hybrids from this family (Triangles in Figure 1)
182 were selected for RNA sequencing in their fourth growing season (2013), based on a

higher or lower than the average number of tillers. The two parents of the family were also sequenced (Diamonds in Figure 1). When the four sequenced hybrids were divided into two groups (genotypes 112 and 90 against genotypes 18 and 120), we observed significant differences between these groups in the number of tillers ($p < 0.05$), biomass yield quantified as dry weight per plant ($p < 0.05$), and the final canopy heights ($p < 0.05$). We also observed a significant difference between these two groups in the concentrations of starch ($p < 0.005$) and sucrose ($p < 0.05$), but we did not observe significant differences between groups in the concentrations of fructose or glucose. The most significant difference ($p < 0.001$) was observed in the total concentration of non-structural carbohydrates (NSC), which was calculated as the sum of the glucose, fructose, sucrose and starch concentrations. We observed significant differences also in the fructose/starch ($p < 0.05$) and glucose/fructose ratios ($p < 0.01$). However, any other ratio between concentrations was not significantly different between the groups (Suppl. Table S2).

There was a significant difference between the *M. sacchariflorus* and *M. sinensis* parents in NSC ($p < 0.05$) and sucrose concentrations ($p < 0.01$). However, there was no significant difference between the parents in the starch, fructose or glucose concentrations (Suppl. Table S2). It is likely an example of heterosis (transgressive segregation) that significant differences in starch, fructose or glucose concentrations were observed in the hybrid progeny but not the parents.

Differential expression (DE) analysis between hybrids and species

We performed RNA-seq from the leaf, stem and root tissue samples extracted from four *M. saccharifloru*s X *M. sinensis* interspecific hybrids, and their two parents (Table 1). When the normalised counts obtained from DESeq2 [33] were used to cluster the samples (Figure 2), the samples firstly grouped by tissue (PC1) and secondly by species (PC2). As a result, the downstream analysis was performed for each tissue separately. Stem and root samples clustered together, and the clustering of these separately from the leaf tissue explained 64% of the variation. Species explains 17 % of the variation, with the hybrids falling between the two parents, which were furthest apart from each other.

We obtained 1,386 differentially expressed genes (DEG; Suppl. Table S3) in total between the hybrids identified as “High NSC” and “Low NSC” (Figure 1) at FDR < 0.05 (Figure 3A). There were 892 DEGs in stems (598 up-regulated and 294 down-regulated), 741 DEGs in leaves (410 up-regulated and 331 down-regulated), and only 253 DEGs in roots (116 up-regulated and 137 down-regulated). 64 % of the DEGs in roots were DE in both of the other tissues too, but most DEGs in stem or leaves were exclusively DE in either stem or leaves.

We also compared the expression between the hybrids against each parent and considered a gene as DE if it was DE in both comparisons at FDR < 0.05 (Suppl. Table S4). Under these criteria, there were 2,870 DEGs in roots, 1,464 DEGs in leaves, and 729 DEGs in stems (Figure 3B). Only 64 among these DEG were also DE between “High NSC” and “Low NSC” hybrids. There were 16,311 DEGs between the hybrids and *M. sinensis* alone (Suppl. figure S1), and 15,616 DEGs between the

hybrids and *M. sacchariflorus* alone (Suppl. figure S2), this is over a third of the total transcriptome.

Enriched Gene Ontology (GO) terms in DEGs

Enrichment analysis of GO terms over-represented among DE genes allowed us to identify the biological processes (BP) and molecular functions (MF) that are differentially regulated in each tissue. After annotating the reference transcriptome with the homologous proteins and full set of GO terms and (Suppl. Table S5), we simplified the results to the more general “GO slim” terms.

All the significant enrichment “GO slim” terms among DEGs between the “High NSC” and “Low NSC” hybrids were associated with metabolic processes, with the single exception of “response to stress” in stems (Figure 4; Suppl. Table S6). Among the GO terms in the “biological process” category, the most significantly enriched ones ($p < 0.001$) were “Carbohydrate metabolism” and “Secondary metabolism” in stem and leaves, and “Generation of precursor metabolites and energy” and “response to stress” in stems. Among the “molecular process” category, “hydrolysis on glycosyl bonds” and “redox activities” were the most significantly enriched ($p < 0.0001$) in both stems and leaves (Suppl. Table S6).

Thirty-six enzymatic reactions were annotated among DEG in the stem (Table 2). Only six were down-regulated in “High NSC”; four involved in the generation of precursor metabolites and energy, namely 6-phosphofructokinase (EC 2.7.1.11) and Triose-phosphate isomerase (EC 5.3.1.1) in the glycolysis pathway; Malate dehydrogenase NADP(+) (EC 1.1.1.82) in the pyruvate metabolism; and 2-carboxy-

D-arabinitol-1-phosphatase (EC 3.1.3.63); and one each in the other GO categories, namely Beta-N-acetylhexosaminidase (EC 3.2.1.52) and carboxypeptidase (EC 3.4.16.-).

A similar analysis on the enriched GO slim terms among DEGs between hybrids and parents (Suppl. Figure S3; Suppl. Table S7) revealed that the most significantly enriched GO terms ($p < 0.01$) were in the root and associated with RNA/DNA binding and translation (including ribosome biogenesis and equivalent terms), and several biosynthetic processes. Remarkably, there were no enriched GO terms in the stem between hybrids and parents.

DEG associated with the starch and sucrose metabolism

There were 88 DEGs associated with the enriched “Carbohydrate metabolism” GO term (Suppl. Table S8), specifically 57 DEGs in stems (42 up-regulated and 15 down-regulated) and 44 DEGs in leaves (20 were up-regulated and 24 down-regulated). Thirteen DEGs were common to both tissues and showed close fold-change values in both tissues. All but two of these 88 DEGs could be functionally annotated, 52 and 56 of them had a homologous protein in *A. thaliana* or rice, respectively.

Twenty-nine DEGs were involved in enzymatic reactions that were part of the starch and sucrose metabolic pathways (KEGG pathway ath00500; Suppl. Figure S4). Among these, all 20 DEGs in stems were up-regulated in “High NSC”, but half of the DEGs in leaves (which were beta-glucosidases) were down-regulated in “High NSC”. Enzymatic proteins in the starch degradation pathway were DE in root and leaves

(e.g. AMY3, ISA3, BAM1). At the same time, sucrose metabolism genes in the cytosol were only DE in stems (SUS3, SPS5). Similarly, reactions involving ADP-glucose were only DE in stems (e.g. AGP, SS2, SS3, SBE2).

Twenty-nine genes were annotated as involved in the “generation of precursor metabolites and energy” (Suppl. Table S8), 17 of which could be annotated with an enzymatic code (KEGG pathway ath00010; Suppl. Figure S5). Six genes were involved in starch metabolism (ISA3, DBE1, PFK2, SBE2, PHS2). The phosphofructokinase 2 (PFK2) is the only one clearly down-regulated in “High NSC”. Among the others, a malate synthase (MLS) and an aldehyde dehydrogenase 12A1 involved in siRNAs generation, and an Fts protease (FTSH6) in the chloroplast were all highly up-regulated (FC > 5) in “High NSC”. On the other hand, triosephosphate isomerase (TIM) was down-regulated in “High NSC”.

The relation between 32 DEGs involved in the twelve DE enzymatic reactions in starch and sucrose metabolism, plus three of the glycolysis reactions are summarised in Figure 5 and Table 3.

DEG associated with other enriched GO terms

The 72 genes annotated as "Response to stress" were involved in a broad range of responses (Suppl. Table S10). On the other hand, the most significantly enriched GO terms in the "Molecular functions" category were associated with metabolic-related enzymatic reactions, namely “oxidoreductase activities” and “hydrolase activities”. The former included 38 cytochrome P450 proteins.

306 “Secondary metabolism” was enriched in both stems and leaves. 17 of the 19 DEGs
307 in stems were up-regulated, but half of the DEGs in leaves were down-regulated. 16
308 of the 31 genes involved in the “secondary metabolism” were cytochrome P450
309 proteins (Suppl. Table S11). Six were included in benzoxazinoids biosynthesis,
310 which is associated with defence in grasses. Another six were involved in terpenoids
311 and phenylpropanoid biosynthesis (KEGG ath00900 and ath00940).
312
313 Many of the identified DEG in enriched functions showed no homologies in model
314 organisms and consequently remain uncharacterised. This is the case in 36 DE
315 genes involved in the carbohydrate metabolism (over 88 total), whose function was
316 evidenced by the presence of a protein domain, but with an unclear role. A similar
317 case is noted in two genes involved in the "generation of precursor metabolites",
318 twelve genes involved in the “secondary metabolism”, and 17 genes involved in
319 “response to stress”.
320
321 **DISCUSSION**
322
323 We performed a transcriptional RNA-seq analysis between selected *Miscanthus*
324 hybrids with negative correlations between starch and sucrose concentrations and
325 biomass yield.
326
327 Using a mapping family (n = 102) between a diploid *M. sinensis* and a diploid *M.*
328 *sacchariflorus*, we previously demonstrated that high biomass yielding *Miscanthus*
329 hybrids had low starch and high fructose concentrations in the stem, and a lower
330 ratio of sucrose, glucose and starch to fructose under peak growing conditions [17].

Here, we selected four hybrids from this mapping family based on the number of tillers (transect count), which was shown to be an accurate predictive phenotype for biomass yield [15]. These four hybrids could be divided into two groups (Table 1), which showed significant differences in the concentrations of starch and sucrose, but not of hexose. The most significant differences were observed for total NSC because of the cumulative effect of the differences in starch and sucrose.

Approximately 10 % of the total genes were DE between these two groups of hybrids in stems and leaves, but not in roots. Among these DE genes, there was an enrichment of genes involved in carbohydrate and secondary metabolism in stem and leaves, and in the “generation of precursor metabolites and energy” in stem only. However, these GO terms for biological processes were not regulated similarly in both tissues. While the DEGs in the enriched categories were predominantly up-regulated in stems, they were evenly up-regulated and down-regulated in leaves. The DE of carbohydrate metabolising genes between the leaf, stem and root is to be expected as it has been previously reported that carbohydrates are differentially distributed between these tissues in *Miscanthus* in July, the same month our study was conducted [18, 22]. Specifically, the abundance of starch in stems was up to 6x more concentrated in the leaf than stem, the below-ground biomass preferentially accumulated starch, and soluble sugars tended to be in greater concentrations in the stems compared to leaves [22]. Our transcriptional observations therefore largely reflect the distribution of carbohydrates; starch metabolism transcripts were DE in the leaf where starch is the most abundant carbohydrate, and sucrose metabolising enzymes were DE in the predominantly sucrose accumulating stem [18]. Fewer

356 DEGs were observed in roots. Seasonal carbohydrate profiling of rhizomes in four
357 genotypes showed that the soluble sugar contents were similar between genotypes
358 and across two sites located 340km apart [18].
359
360 We observed that multiple genes involved in the synthesis (AGP, SS2, SS3, BE2)
361 and degradation of starch in the chloroplast (AMY3, ISA3, SEX4, BAM1) were down-
362 regulated in high biomass yielding genotypes. We also observed down-regulation of
363 genes involved in the synthesis (SPS5) and degradation (SUS) of sucrose in high
364 biomass-yielding genotypes. Genes involved in the starch metabolic pathway are up-
365 regulated by a high sugar status [39-41], as there was a negative relationship
366 between yield and soluble sugar (i.e. high yielders had lower sugar), it is consistent
367 that the expression of sugar stimulated genes would be lower in high yielding
368 genotypes.
369
370 Contrary to this, we noticed the up-regulation with a high fold-change in high
371 biomass yield genotypes of triosephosphate isomerase TIM/PDTPI, which encodes a
372 plastidic triose phosphate isomerase [42], and Phosphofructokinase 2 (PFK2). PFK2
373 catalyses the penultimate step before usable energy is extracted from the
374 phosphorylated products of photosynthesis. This enzyme is, therefore, a main
375 control point of glycolysis. The observation that high biomass plants have low
376 carbohydrates can seem counter-intuitive, but the rationale is highly logical; high
377 biomass plants maximise growth at the expense of their carbon reserves [43],
378 whereas slow-growing types accumulate their reserves. The upregulation of the
379 PFK2 gene encoding a major glycolytic enzyme is suggestive of a more rapid
380 metabolism of photosynthate to fuel growth in the high yielding types. In summary,

starch and sucrose synthesis was down-regulated in high yielding *Miscanthus* hybrids, while glycolysis and export of triose phosphates was up-regulated in high yielding *Miscanthus* hybrids.

These results support that high yielding *Miscanthus* genotypes were more rapidly accumulating structural mass, likely cellulose via sucrose metabolism [44-46], at the expense of starch [17, 18, 47]. The latter is further supported by the significant differences in the fructose-to-starch (but not glucose-to-starch) ratio between high and low yielding hybrids [17], which was also observed between the sequenced hybrids. Fructose is an indicator of sucrose metabolism, because it is produced exclusively from the metabolism of sucrose by the action of sucrose synthases (SUS), while glucose is produced by the metabolism of both sucrose and starch [48, 49]. Furthermore, in a C13 labelling experiment, it was observed that a greater proportion of the labelled carbon was observed in the insoluble fraction (mainly comprising cellulose) of a rapidly growing *Miscanthus* genotype, whereas a greater proportion was partitioned into starch in a slower-growing type [17]. Our results, therefore, add to these previous observations with the addition of transcriptomic evidence of the relationship between carbon metabolism, partitioning and growth.

We observed a significant enrichment of “response to stress” genes in stems. However, a further analysis did not reveal more details, only 32 of the 72 DEG in this category had a homologous protein in *A. thaliana*, and 23 of these were annotated as "response to stimulus", i.e. several types of environmental stress. Changes in starch metabolism are linked to changes in source-sink carbon allocation for protection against environmental stresses [50], and may expose differences between

groups with phenotypic differences. On the other hand, several of the genes involved in the starch and sucrose metabolism are confirmed redox-regulated enzymes (e.g. AGP, SS3, BE2, AMY3, ISA3, SEX4, and BAM1), which partially explains the enrichment of in the "oxidoreductase activity" (GO:55114). The homologous in *A. thaliana* of *Miscanthus* genes annotated in "oxidoreductase activity" were usually involved in metabolic processes (GO:44699).

Our results evidence a direct relationship between high expression of essential enzymatic genes in the starch and sucrose synthesis pathway, high starch concentrations, and lower biomass production. The strong interconnectivity between genotype, chemotype and agronomic traits opens the door to use the expression of well-characterised genes in the starch and sucrose pathway for the early selection of high biomass yielding genotypes from large *Miscanthus* populations.

ACKNOWLEDGEMENTS

We would like to thank Dr Anne Maddison for her technical support in generating the carbohydrate data. This work was funded by the Biotechnology and Biological Sciences Council (BBSRC) in projects BBS/E/T/000PR9818 (JDV), BBS/E/W/10963A01A (KF) and BB/CSP1730/1 (KF). Sequencing was funded through the “Capacity and Capability Challenge” funding awarded to Earlham Institute’s *Genomic Pipelines* group by BBSRC.

DATA AVAILABILITY

Raw reads are deposited in the Short Reads Archive (SRA) under Bioproject ID PRJNA639832. The R code used in the analysis is deposited in Zenodo

(http://doi.org/10.5281/zenodo.3834007) and Github (https://github.com/joseja/miscanthus_starch_rnaseq).

REFERENCES

1. Somerville C, Youngs H, Taylor C, Davis SC, Long SP: **Feedstocks for lignocellulosic biofuels**. *Science* 2010, **329**(5993):790-792.

2. Donnison IS, Fraser MD: **Diversification and use of bioenergy to maintain future grasslands**. *Food and Energy Security* 2016, **5**(2):67-75.

3. Heaton EA, Dohleman FG, Miguez AF, Juvik JA, Lozovaya V, Widholm J, Zabotina OA, McIsaac GF, David MB, Voigt TB: ***Miscanthus*: a promising biomass crop**. In: *Advances in Botanical Research.* vol. 56: Elsevier; 2010: 75-137.

4. Hodkinson TR, Chase MW, Lledó DM, Salamin N, Renvoize SA: **Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid trnL intron and trnL-F intergenic spacers**. *Journal of Plant Research* 2002, **115**(5):381-392.

5. Kam J, Thomas D, Pierre S, Ashman C, McCalmont JP, Purdy SJ: **A new carbohydrate retaining variety of *Miscanthus* increases biogas methane yields compared to *M x giganteus* and narrows the yield advantage of maize**. *Food and Energy Security* 2020:e224.

6. Kiesel A, Wagner M, Lewandowski I: **Environmental performance of *Miscanthus*, switchgrass and maize: Can C4 perennials increase the sustainability of biogas production?** *Sustainability* 2017, **9**(1):5.

7. Wagner M, Mangold A, Lask J, Kiesel A, Lewandowski I: **Economic and environmental performance of *Miscanthus* cultivated on marginal land for biogas production. GCB Bioenergy 11: 34–49**. In.; 2018.

458 8. McCalmont JP, Hastings A, McNamara NP, Richter GM, Robson P, Donnison
459 IS, Clifton-Brown J: **Environmental costs and benefits of growing**
460 ***Miscanthus* for bioenergy in the UK**. *GCB Bioenergy* 2017, **9**(3):489-507.
461 9. Hodkinson T, Klaas M, Jones M, Prickett R, Barth S: ***Miscanthus*: a case**
462 **study for the utilization of natural genetic variation**. *Plant Genetic*
463 *Resources* 2015, **13**(3):219-237.
464 10. Zub HW, Brancourt-Hulmel M: **Agronomic and physiological performances**
465 **of different species of *Miscanthus*, a major energy crop. A review**.
466 *Agronomy for Sustainable Development* 2010, **30**(2):201-214.
467 11. Christian D, Riche A, Yates N: **Growth, yield and mineral content of**
468 ***Miscanthus x giganteus* grown as a biofuel for 14 successive harvests**.
469 *Industrial Crops and Products* 2008, **28**(3):320-327.
470 12. Davey CL, Jones LE, Squance M, Purdy SJ, Maddison AL, Cunniff J,
471 Donnison I, Clifton-Brown J: **Radiation capture and conversion efficiencies**
472 **of *Miscanthus sacchariflorus*, *M. sinensis* and their naturally occurring**
473 **hybrid *M. x giganteus***. *GCB Bioenergy* 2017, **9**(2):385-399.
474 13. Hastings A, Mos M, Yesufu JA, McCalmont J, Schwarz K, Shafei R, Ashman
475 C, Nunn C, Schuele H, Cosentino S: **Economic and environmental**
476 **assessment of seed and rhizome propagated *Miscanthus* in the UK**.
477 *Frontiers in Plant Science* 2017, **8**:1058.
478 14. Arnoult S, Brancourt-Hulmel M: **A review on *Miscanthus* biomass**
479 **production and composition for bioenergy use: genotypic and**
480 **environmental variability and implications for breeding**. *BioEnergy*
481 *Research* 2015, **8**(2):502-526.

15. Robson P, Jensen E, Hawkins S, White SR, Kenobi K, Clifton-Brown J, Donnison I, Farrar K: **Accelerating the domestication of a bioenergy crop: identifying and modelling morphological targets for sustainable yield increase in *Miscanthus***. *Journal of Experimental Botany* 2013, **64**(14):4143-4155.

16. Clifton-Brown J, Harfouche A, Casler MD, Dylan Jones H, Macalpine WJ, Murphy-Bokern D, Smart LB, Adler A, Ashman C, Awty-Carroll D: **Breeding progress and preparedness for mass-scale deployment of perennial lignocellulosic biomass crops switchgrass, *Miscanthus*, willow and poplar**. *GCB Bioenergy* 2019, **11**(1):118-151.

17. Maddison AL, Camargo-Rodriguez A, Scott IM, Jones CM, Elias DMO, Hawkins S, Massey A, Clifton-Brown J, McNamara NP, Donnison IS *et al*: **Predicting future biomass yield in *Miscanthus* using the carbohydrate metabolic profile as a biomarker**. *Global Change Biology Bioenergy* 2017, **9**(7):1264-1278.

18. Purdy SJ, Cunniff J, Maddison AL, Jones LE, Barraclough T, Castle M, Davey CL, Jones CM, Shield I, Gallagher J: **Seasonal carbohydrate dynamics and climatic regulation of senescence in the perennial grass, *Miscanthus***. *BioEnergy Research* 2015, **8**(1):28-41.

19. Slavov GT, Nipper R, Robson P, Farrar K, Allison GG, Bosch M, Clifton-Brown JC, Donnison IS, Jensen E: **Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus* sinensis**. *New Phytologist* 2014, **201**(4):1227-1239.

20. Van der Weijde T, Kiesel A, Iqbal Y, Muylle H, Dolstra O, Visser RG, Lewandowski I, Trindade LM: **Evaluation of *Miscanthus sinensis* biomass quality as feedstock for conversion into different bioenergy products**. *GCB Bioenergy* 2017, **9**(1):176-190.

21. da Costa RM, Pattathil S, Avci U, Winters A, Hahn MG, Bosch M: **Desirable plant cell wall traits for higher-quality *Miscanthus* lignocellulosic biomass**. *Biotechnology for Biofuels* 2019, **12**(1):1-18.

22. Purdy SJ, Maddison AL, Cunniff J, Donnison I, Clifton-Brown J: **Non-structural carbohydrate profiles and ratios between soluble sugars and starch serve as indicators of productivity for a bioenergy grass**. *AoB Plants* 2015, **7**.

23. de Souza AP, Arundale RA, Dohleman FG, Long SP, Buckeridge MS: **Will the exceptional productivity of *Miscanthus x giganteus* increase further under rising atmospheric CO2?** *Agricultural and Forest Meteorology* 2013, **171**:82-92.

24. Miyake H: **Starch accumulation in the bundle sheaths of C3 plants: a possible pre-condition for C4 photosynthesis**. *Plant and Cell Physiology* 2016, **57**(5):890-896.

25. Purdy SJ, Maddison AL, Nunn CP, Winters A, Timms-Taravella E, Jones CM, Clifton-Brown JC, Donnison IS, Gallagher JA: **Could *Miscanthus* replace maize as the preferred substrate for anaerobic digestion in the United Kingdom? Future breeding strategies**. *GCB Bioenergy* 2017, **9**(6):1122-1139.

26. Keurentjes JJ: **Genetical metabolomics: closing in on phenotypes**. *Current Opinion in Plant Biology* 2009, **12**(2):223-230.

27. Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE: **Genomic and metabolic prediction of complex heterotic traits in hybrid maize**. *Nature Genetics* 2012, **44**(2):217.

28. Krueger F: **Trim Galore**. *A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.* 2015. https://github.com/FelixKrueger/TrimGalore

29. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner**. *Bioinformatics* 2013, **29**(1):15-21.

30. Bray NL, Pimentel H, Melsted P, Pachter L: **Near-optimal probabilistic RNA-seq quantification**. *Nature Biotechnology* 2016, **34**(5):525-527.

31. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research**. *Bioinformatics* 2005, **21**(18):3674-3676.

32. Soneson C, Love MI, Robinson MD: **Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences**. *F1000Research* 2015, **4**.

33. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome biology* 2014, **15**(12):550.

34. Alexa A, Rahnenfuhrer J: **topGO: enrichment analysis for gene ontology**. *R package version 2.40.0.* 2010.

35. Wickham H: **ggplot2: elegant graphics for data analysis**: Springer-Verlag, New York, 2016.

556 36. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T,
557 Dirks W, Hellsten U, Putnam N: **Phytozome: a comparative platform for**
558 **green plant genomics**. *Nucleic Acids Research* 2012, **40**(D1):D1178-D1186.
559 37. Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, Kim
560 M, Rosen BD, Cheng C-Y, Moreira W, Mock SA: **Araport: the Arabidopsis**
561 **information portal**. *Nucleic Acids Research* 2015, **43**(D1):D1003-D1009.
562 38. Kanehisa M, Sato Y: **KEGG Mapper for inferring cellular functions from**
563 **protein sequences**. *Protein Science* 2020, **29**(1):28-35.
564 39. Mita S, Suzuki-Fujii K, Nakamura K: **Sugar-Inducible Expression of a Gene**
565 **for [beta]-Amylase in *Arabidopsis thaliana***. *Plant Physiology* 1995,
566 **107**(3):895-904.
567 40. Nakamura K, Ohto M-a, Yoshida N, Nakamura K: **Sucrose-induced**
568 **accumulation of β-amylase occurs concomitant with the accumulation**
569 **of starch and sporamin in leaf-petiole cuttings of sweet potato**. *Plant*
570 *Physiology* 1991, **96**(3):902-909.
571 41. Price J, Laxmi A, Martin SKS, Jang J-C: **Global transcription profiling**
572 **reveals multiple sugar signal transduction mechanisms in *Arabidopsis***.
573 *The Plant Cell* 2004, **16**(8):2128-2150.
574 42. Chen M, Thelen JJ: **The plastid isoform of triose phosphate isomerase is**
575 **required for the postgerminative transition from heterotrophic to**
576 **autotrophic growth in *Arabidopsis***. *The Plant Cell* 2010, **22**(1):77-90.
577 43. Sulpice R, Pyl E-T, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H,
578 Gibon Y, Usadel B, Poree F, Piques MC: **Starch as a major integrator in the**
579 **regulation of plant growth**. *Proceedings of the National Academy of*
580 *Sciences* 2009, **106**(25):10348-10353.

581 44. Coleman HD, Yan J, Mansfield SD: **Sucrose synthase affects carbon**
582 **partitioning to increase cellulose production and altered cell wall**
583 **ultrastructure**. *Proceedings of the National Academy of Sciences* 2009,
584 **106**(31):13118-13123.
585 45. Mood SH, Golfeshan AH, Tabatabaei M, Jouzani GS, Najafi GH, Gholami M,
586 Ardjmand M: **Lignocellulosic biomass to bioethanol, a comprehensive**
587 **review with a focus on pretreatment**. *Renewable and Sustainable Energy*
588 *Reviews* 2013, **27**:77-93.
589 46. McFarlane HE, Döring A, Persson S: **The cell biology of cellulose**
590 **synthesis**. *Annual Review of Plant Biology* 2014, **65**:69-94.
591 47. Lloyd JR, Kossmann J: **Starch trek: The search for yield**. *Frontiers in Plant*
592 *Science* 2019, **9**:1930.
593 48. Smith AM, Zeeman SC, Smith SM: **Starch degradation**. *Annual Review Plant*
594 *Biology* 2005, **56**:73-98.
595 49. Koch K: **Sucrose metabolism: regulatory mechanisms and pivotal roles**
596 **in sugar sensing and plant development**. *Current Opinion in Plant Biology*
597 2004, **7**(3):235-246.
598 50. Dong S, Beckles DM: **Dynamic changes in the starch-sugar**
599 **interconversion within plant source and sink tissues promote a better**
600 **abiotic stress response**. *Journal of Plant Physiology* 2019, **234**:80-93.
601

602 Figure 1: Concentrations of non-structural carbohydrates, number of tillers, and
603 biomass yield in a mapping population comprised of 102 *M. sinensis* X *M.*
604 *sacchariflorus* hybrids. Values for four hybrids with contrasting phenotypes ("high"
605 and "low"), which were selected for RNA sequencing, are highlighted (Triangles).
606 Significant differences (T-test) between the hybrids are annotated under each
607 phenotype. The two parents of the family were also sequenced and phenotyped
608 (diamonds). Boxplots summarise the distribution of values for the whole family for
609 each phenotype.
610
611 Figure 2: Principal component analysis of the normalised gene counts from 23 RNA-
612 seq libraries generated from leaves (diamonds), stems (squares) and roots (circles)
613 obtained from four *M. sinensis* X *M. sacchariflorus* hybrids (green shapes) with
614 contrasting phenotypes and their parents (red and blue shapes). Gene counts were
615 obtained from Kallisto alignments and normalised using DESeq2 for the top 1,000
616 most variable genes.
617
618 Figure 3: Number of differentially expressed genes shared between root, leaf and
619 stem tissues among the "High NCS" and "Low NCS" *Miscanthus* hybrids at FDR <
620 0.05 (A), and between the hybrids and their progenitors (B). A gene only was
621 considered DE between hybrids and parents when it was DE against both parents.
622
623 Figure 4: GO SLIM terms (rows) that were significantly enriched ($p < 0.05$) in each
624 tissue (columns) among differentially expressed genes (DEG) from the "High NCS"
625 and "Low NCS" *Miscanthus* hybrids DE analysis. The size of a bubble is proportional
626 to the number of DEG annotated with that GO term. Rows are sorted by descending

627 p-value (F-fisher test) and the bubble colour is representative to the obtained p-
628 value, from lower (dark green) to higher (light green). Yellow ($p > 0.05$) and white
629 ($p > 0.1$) bubbles were not enriched. All the enriched GO SLIM terms for the
630 “biological process” (top 8 rows) and “molecular function” (bottom 5 rows) GO
631 categories were included.
632
633 Figure 5: Schema of the starch and sucrose metabolism in plants, highlighting critical
634 differentially expressed (DE) proteins between “High NSC” and “Low NSC”
635 Enzymatic codes are shown between brackets. DE *Miscanthus* genes are included
636 under their respective protein (The prefix “Misin_” is not included in the gene name).
637 Genes were differentially expressed in leaves (coloured in green), stems (orange) or
638 both tissues (blue).
639

640 **Table 1: RNA-seq libraries used in this study.**

Genotype	tissue	Library	Group
112 (hybrid)	root	LIB2338	High NSC / Low yield / Fewer tillers
	stem	LIB2339	
	leaf	LIB2340	
90 (hybrid)	root	LIB2341	
	stem	LIB2342	
	leaf	LIB2343	
120 (hybrid)	root	LIB2344	Low NSC / High yield / Many tillers
	stem	LIB2345	
	leaf	LIB2346	
18 (hybrid)	root	LIB2347	
	stem	LIB2348	
	leaf	LIB2349	
M. sacch 297	stem	LIB2352	Progenitors
	leaf	LIB2353	
	stem	SAM1158	
	root	SAM1159	
	leaf	SAM1160	
	root*	SAM1161	
M. sinensis 102	stem	LIB2350	
	leaf	LIB2351	
	stem	SAM1162	
	root	SAM1163	
	leaf	SAM1164	

641 *M. sacch = M. sacchariflorus; *root tip*

642 **Table 2: Carbohydrate and secondary metabolic enzymatic reactions**
643 **differentially expressed between "high NSC" and "low NSC" *Miscanthus***
644 **hybrids.**

645

646 *Leaf FC or Stem FC = Log2 fold-change expression between "high NSC" / "Low*
647 *NSC" hybrids in either leaf or stem tissues.*

ENZYME NAME	ENZIME CODE	GENE	LEAF FC	STEM FC
Malate dehydrogenase (NADP(+))	1.1.1.82	Misin07G271500	-3.12	-0.85
Laccase	1.10.3.2	Misin06G334400	-2.40	
Indole-2-monooxygenase	1.14.14.153 1.14.13.138	Misin19G207900	1.52	
		MisinT014900	11.40	5.99
Indolin-2-one monooxygenase	1.14.13.138 1.14.14.157	MisinT219600	2.22	
		Misin07G204200	5.73	
		Misin01G349900	-4.68	
		Misin09G192700	3.53	4.73
3-hydroxyindolin-2-one monooxygenase	1.14.13.139 1.14.14.109	MisinT014600	10.87	8.05
Ent-isokaurene C2/C3-hydroxylase	1.14.13.143 1.14.14.76	Misin01G158200		1.83
Ent-cassa-12,15-diene 11-hydroxylase	1.14.13.145 1.14.14.112	Misin15G165600	5.04	5.67
Trans-cinnamate 4-monooxygenase	1.14.14.91	Misin05G312600	-2.11	
Camalexin synthase	1.14.19.52	Misin04G105800	1.58	
Aldehyde dehydrogenase (NAD(+))	1.2.1.3	Misin04G200300	0.81	
L-glutamate gamma-semialdehyde dehydrogenase	1.2.1.88	Misin17G216100	4.19	4.17
Pyruvate dehydrogenase	1.2.4.1	Misin18G109800		4.21
Delta(24)-sterol reductase	1.3.1.72	MisinT029700	2.52	1.50
Nicotinamide N-methyltransferase	2.2.1.1	Misin11G031500		2.33
Sinapoylglucose--malate O-sinapoyltransferase	2.3.1.92	Misin12G095900	2.55	3.49
Malate synthase	2.3.3.9	Misin11G121200	5.07	5.71
Glycogen phosphorylase	2.4.1.1	Misin05G335800		1.07
Sucrose synthase	2.4.1.13	Misin01G358800		1.98
Sucrose-phosphate synthase	2.4.1.14	Misin10G070300		1.10
1,4-alpha-glucan branching enzyme	2.4.1.18	Misin07G352300		1.92
		Misin18G276400		1.77
Starch synthase	2.4.1.21	MisinT393000		1.63
		Misin19G100900		2.44
Dimethylallyltranstransferase	2.5.1.1	Misin04G333300	-1.40	
Glutathione transferase	2.5.1.18	Misin02G293100		4.15
		Misin02G286600		2.01
		MisinT258000	2.57	1.55
		MisinT404400	2.51	3.37
6-phosphofructokinase	2.7.1.11	Misin12G113600	-6.94	-6.65
Pyruvate kinase	2.7.1.40	Misin06G200500	0.64	
Mitogen-activated protein kinase	2.7.11.24	Misin01G390800		1.77
Glucose-1-phosphate adenylyltransferase	2.7.7.27	Misin17G255500		1.72
Fucose-1-phosphate guanylyltransferase	2.7.7.30	Misin02G490600	1.84	2.90

6-phosphogluconolactonase	3.1.1.31	Misin07G251100	-2.48	
Sugar-phosphatase	3.1.3.23	Misin10G086500	-1.67	
2-carboxy-D-arabinitol-1-phosphatase	3.1.3.63	Misin10G020200	-3.42	-2.49
Glycerophosphodiester phosphodiesterase	3.1.4.46	Misin08G144100	0.82	
Alpha-amylase	3.2.1.1	Misin04G207500	1.67	
Beta-amylase	3.2.1.2	Misin02G205400	2.36	2.37
		MisinT552400		3.03
		Misin04G312400		5.04
		Misin15G034600		1.09
Beta-glucosidase	3.2.1.21	Misin11G141900	-1.58	
		Misin11G111200	-1.26	
		Misin06G358300	1.27	
		Misin12G147300		1.12
		Misin11G142000		1.21
Alpha-galactosidase	3.2.1.22	MisinT167900		0.81
Beta-galactosidase	3.2.1.23	Misin03G233500	-1.03	
Beta-fructofuranosidase	3.2.1.26	Misin11G067200		3.63
Glucan endo-1,3-beta-D-glucosidase	3.2.1.39	Misin17G123500	-2.79	
		Misin16G118700	-3.26	
		Misin02G326400	2.53	
		Misin01G337100	2.41	
		MisinT226600	2.45	4.22
		Misin01G145100		2.22
		Misin02G115300		1.99
Beta-N-acetylhexosaminidase	3.2.1.52	Misin03G316100	3.91	
		Misin17G142700		-0.57
Non-reducing end alpha-L-arabinofuranosidase	3.2.1.55	Misin10G067800	-1.95	
Isoamylase	3.2.1.68	Misin07G322000	-1.02	
		Misin17G131000		1.04
		Misin03G195400		1.36
		Misin04G215400		1.07
Endo-1,4-beta-xylanase	3.2.1.8	Misin05G078900	-2.78	
Glucose-6-phosphate 1-epimerase	5.1.3.15	Misin06G202700	-1.67	
Triose-phosphate isomerase	5.3.1.1	Misin03G235900		-0.59
Phosphoglycerate mutase	5.4.2.11/.12	Misin02G341300	2.45	4.35
Ent-copalyl diphosphate synthase	5.5.1.13	Misin01G047600	2.51	

648

649 **Table 3: Thirty-nine differentially expressed genes were involved in twelve**
650 **reactions in the starch and sucrose metabolism and three of the glycolysis**
651 **reactions were highlighted in our analysis.**

652

653 *Leaf/stem = Log2 fold-change expression "high NSC" / "Low NSC" hybrids in either*
654 *lead or stem tissues; Ath/Rice = Homologous protein in Arabidopsis thaliana and rice*
655 *(The prefix "LOC_" is not included in the name).*

GENE	LEAF	STEM	EC	Protein name	Ath	Rice
Misin01G145100		2.22	3.2.1.39	BG8	AT1G64760	Os03g45390
Misin01G337100	2.41		3.2.1.39	*Beta-1,3-glucanase*		Os03g25790
Misin01G358800		1.98	2.4.1.13	SUS3	AT4G02280	Os03g22120
Misin02G115300		1.99	3.2.1.39	T11I18	AT3G04010	Os03g45390
Misin02G205400	2.36	2.37	3.2.1.2	BAM1	AT3G23920	Os10g32810
Misin02G326400	2.53		3.2.1.39	*Beta-1,3-glucanase*		Os03g25790
Misin02G341300	2.45	4.35	5.4.2.11	*Phosphoglycerate mutase*		Os03g21260
Misin03G195400		1.36	3.2.1.68	ISA3	AT4G09020	Os09g29404
Misin03G235900		-0.59	5.3.1.1	TIM	AT2G21170	Os09g36450
Misin03G316100	3.91		3.2.1.52	HEXO2	AT1G05590	Os07g38790
Misin04G207500	1.67		3.2.1.1	AMY1	AT4G25000	
Misin04G215400		1.07	3.2.1.68	ISA3	AT4G09020	Os09g29404
Misin04G312400		5.04	3.2.1.2	*Beta-amylase*		
Misin05G335800		1.07	2.4.1.1	PHS2	AT3G46970	Os01g63270
Misin06G202700	-1.67		5.1.3.15	F15G16.1/SF10	AT3G61610	Os01g46950
Misin06G358300	1.27		3.2.1.21	BGLU42/4	AT5G36890	Os01g67220
Misin07G322000	-1.02		3.2.1.68	LSF1/SEX4	AT3G01510	Os08g29160
Misin07G352300		1.92	2.4.1.18	SBE2.2	AT5G03650	Os02g32660
Misin10G070300		1.10	2.4.1.14	SPS5		Os11g12810
Misin11G067200		3.63	3.2.1.26	cwINV4/OsCIN2	AT2G36190	Os04g33740
Misin11G111200	-1.26		3.2.1.21	BGLU14	AT2G25630	
Misin11G121200		5.71	2.3.3.9	MLS	AT5G03860	Os04g40990
Misin11G141900	-1.58		3.2.1.21	BGLU45/18	AT1G61810	Os04g43410
Misin11G142000		1.21	3.2.1.21	BGLU18		Os04g43410
Misin12G113600		-6.65	2.7.1.11	PFK2	AT5G47810	Os09g30240
Misin12G147300		1.12	3.2.1.21	BGLU46	AT1G61820	Os04g43390
Misin15G034600		1.09	3.2.1.2	*Beta-amylase*		
Misin16G118700	-3.26		3.2.1.39	BG1	AT3G57270	
Misin17G123500	-2.79		3.2.1.39	BG3	AT3G57240	
Misin17G131000		1.04	3.2.1.68	DBE1	AT1G03310	Os05g32710
Misin17G142700		-0.57	3.2.1.52	HEXO3	AT1G65590	Os05g34320
Misin17G216100		4.17	1.2.1.88	ALDH12A1	AT5G62530	
Misin17G255500		1.72	2.7.7.27	AGPL3/APL3		Os05g50380
Misin18G276400		1.77	2.4.1.18	*Glycogen branching*		
Misin19G100700		5.08	3.4.24.-	FTSH6	AT5G15250	Os06g12370
Misin19G100900		2.44	2.4.1.21	SS2		Os06g12450
MisinT226600	2.45	4.22	3.2.1.39	BGL2	AT3G57260	
MisinT393000		1.63	2.4.1.21	SS3	AT1G11720	

MisinT552400		3.03	3.2.1.2	BAM1	AT3G23920	Os10g32810

656

SUPPLEMENTARY MATERIALS

Suppl. Figure S1: Number of differentially expressed genes shared between root, leaf and stem tissues between the hybrids and the *M. sinensis* progenitor.

Suppl. Figure S2: Number of differentially expressed genes shared between root, leaf and stem tissues between the hybrids and the *M. sacchariflorus* progenitor.

Suppl. Figure S3: GO SLIM terms (rows) that were significantly enriched ($p < 0.05$) in each tissue (columns) among differentially expressed genes (DEG) from the expression analysis between the hybrids and both progenitors. The size of a bubble is proportional to the number of DEG annotated with that GO term. Rows are sorted by descending p-value (F-fisher test) and the bubble colour is representative to the obtained p-value, from lower (dark green) to higher (light green). Yellow ($p > 0.05$) and white ($p > 0.1$) bubbles were not enriched. All the enriched GO SLIM terms for the “biological process” (top 8 rows) and “molecular function” (bottom 5 rows) GO categories were included.

Suppl. Figure S4: Down-regulated enzymatic reactions in the “starch and sucrose metabolism” pathway from KEGG (KEGG pathway ath00500) that were down-regulated in “high NSC” hybrids, which had higher concentrations of starch and sucrose.

Suppl. Figure S5: Enzymatic reactions in the “glycolysis/gluconeogenesys” pathway from KEGG (KEGG pathway ath00010) that were down-regulated (red boxes) or up-

682 regulated (green boxes) in “high NSC” hybrids, which had higher concentrations of
683 starch and sucrose.
684
685 Suppl. Table S1: Individual trait scores and Person correlation between traits.
686
687 Suppl. Table S2: Traits significantly different (T-test) between the sequenced
688 samples.
689
690 Suppl. Table S3: Normalised counts, expression fold-change and P-values for all the
691 genes in roots, stem and leaf tissue between groups of hybrids.
692
693 Suppl. Table S4: Normalised counts, expression fold-change and P-values for all the
694 genes in roots, stem and leaf tissue between hybrids and parents.
695
696 Suppl. Table S5: Functional annotation, GO and enzyme codes for all the genes in
697 the reference genome.
698
699 Suppl. Table S6: Enriched GO terms among DEG between groups of hybrids.
700
701 Suppl. Table S7: Enriched GO terms among DEG between hybrids and parents.
702
703 Suppl. Table S8: Detailed functional annotation of 88 DEG within the enriched
704 “carbohydrate metabolism” GO term.
705

706 Suppl. Table S9: Detailed functional annotation of 29 DEG within the enriched

707 “generation of precursor metabolites and energy” GO term.

708

709 Suppl. Table S10: Detailed functional annotation of 72 DEG within the enriched

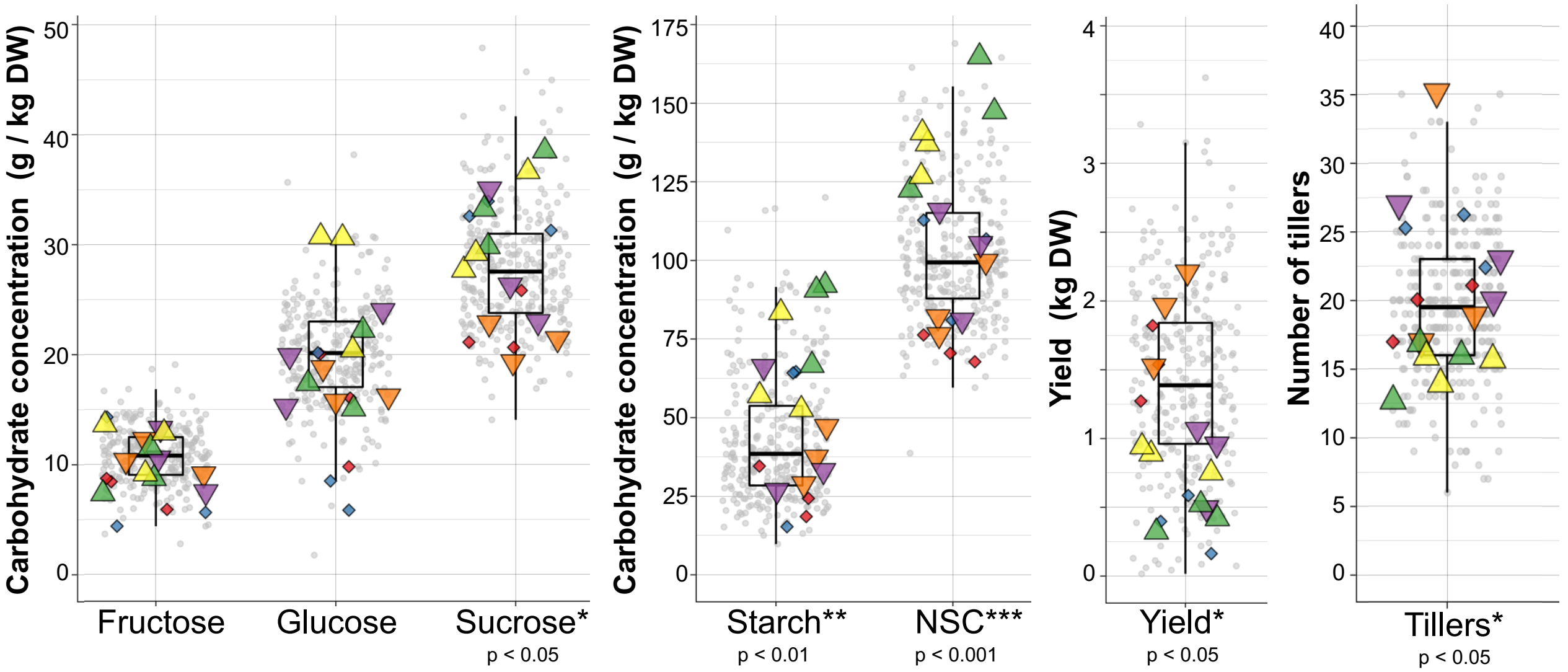
710 “response to stress” GO term.

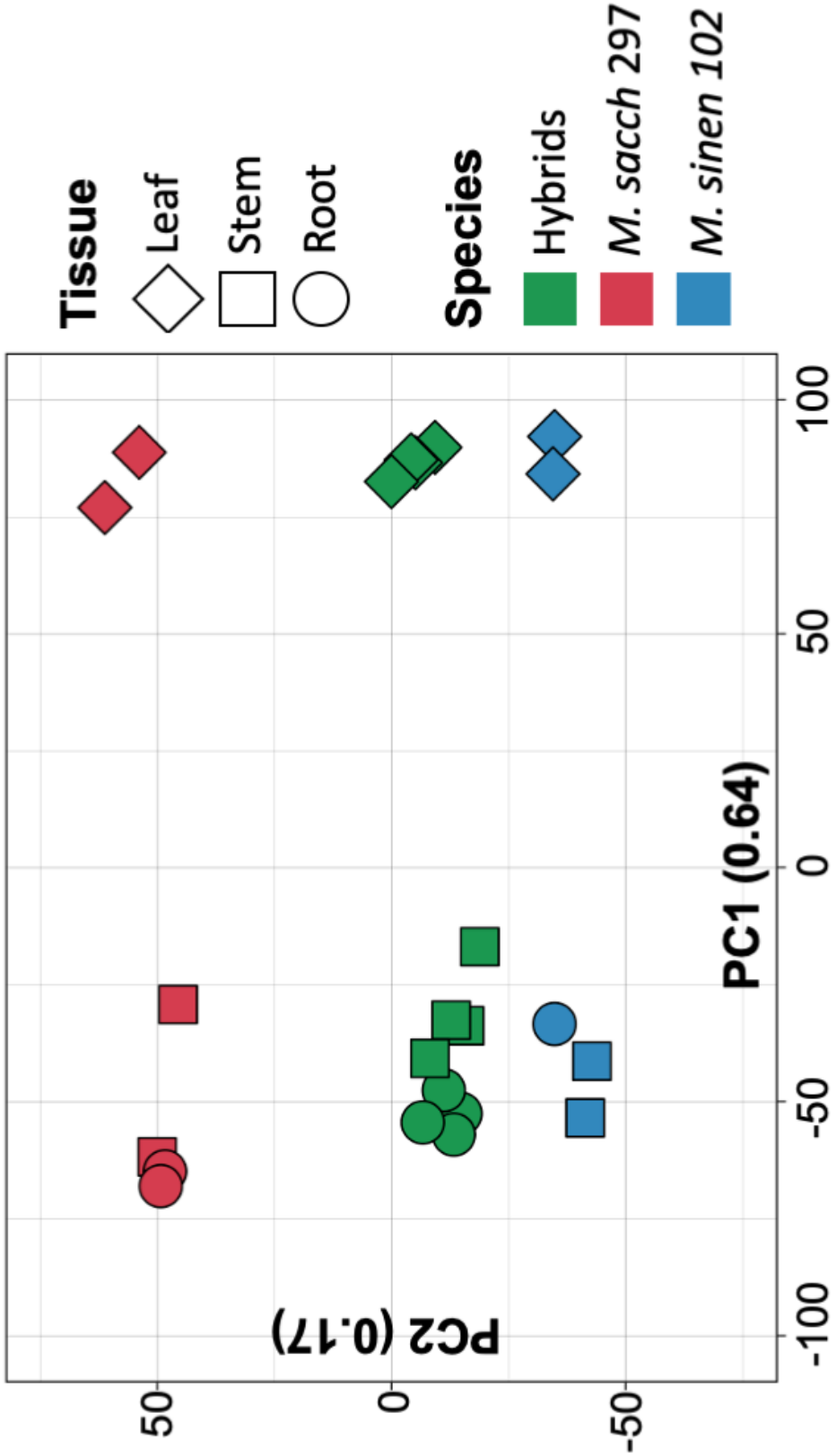
711

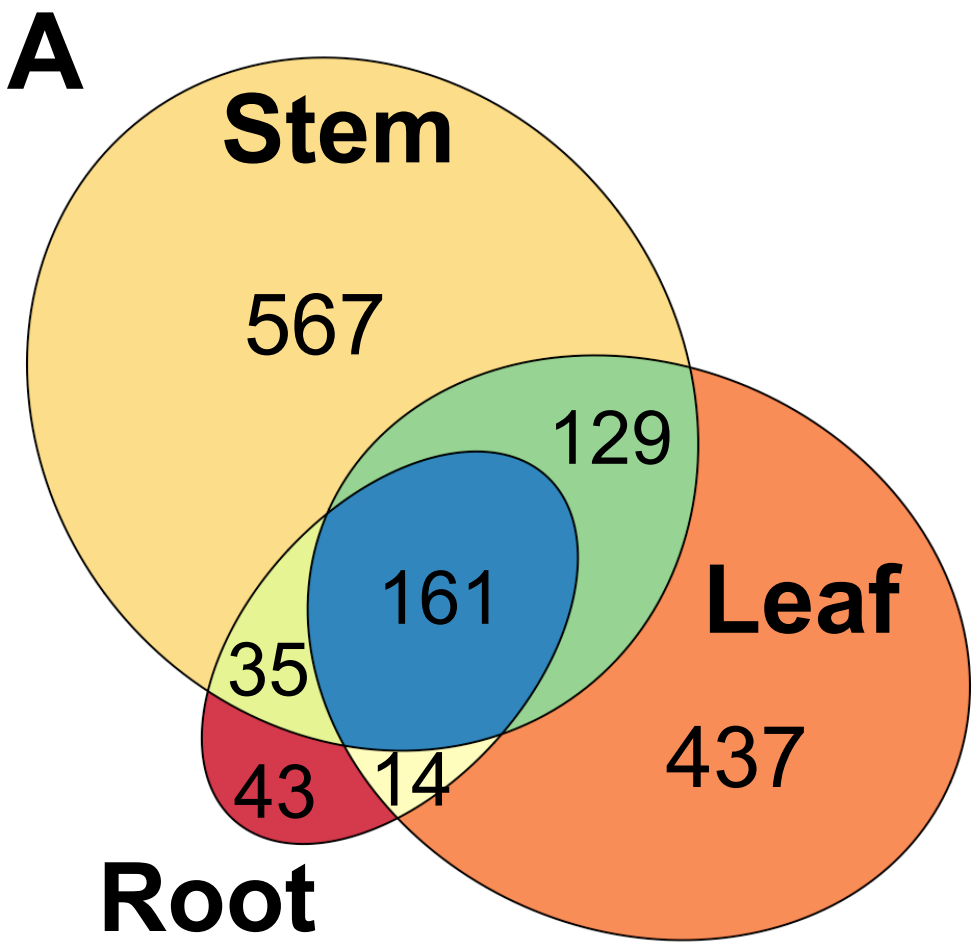
712 Suppl. Table S11: Detailed functional annotation of 31 DEG within the enriched

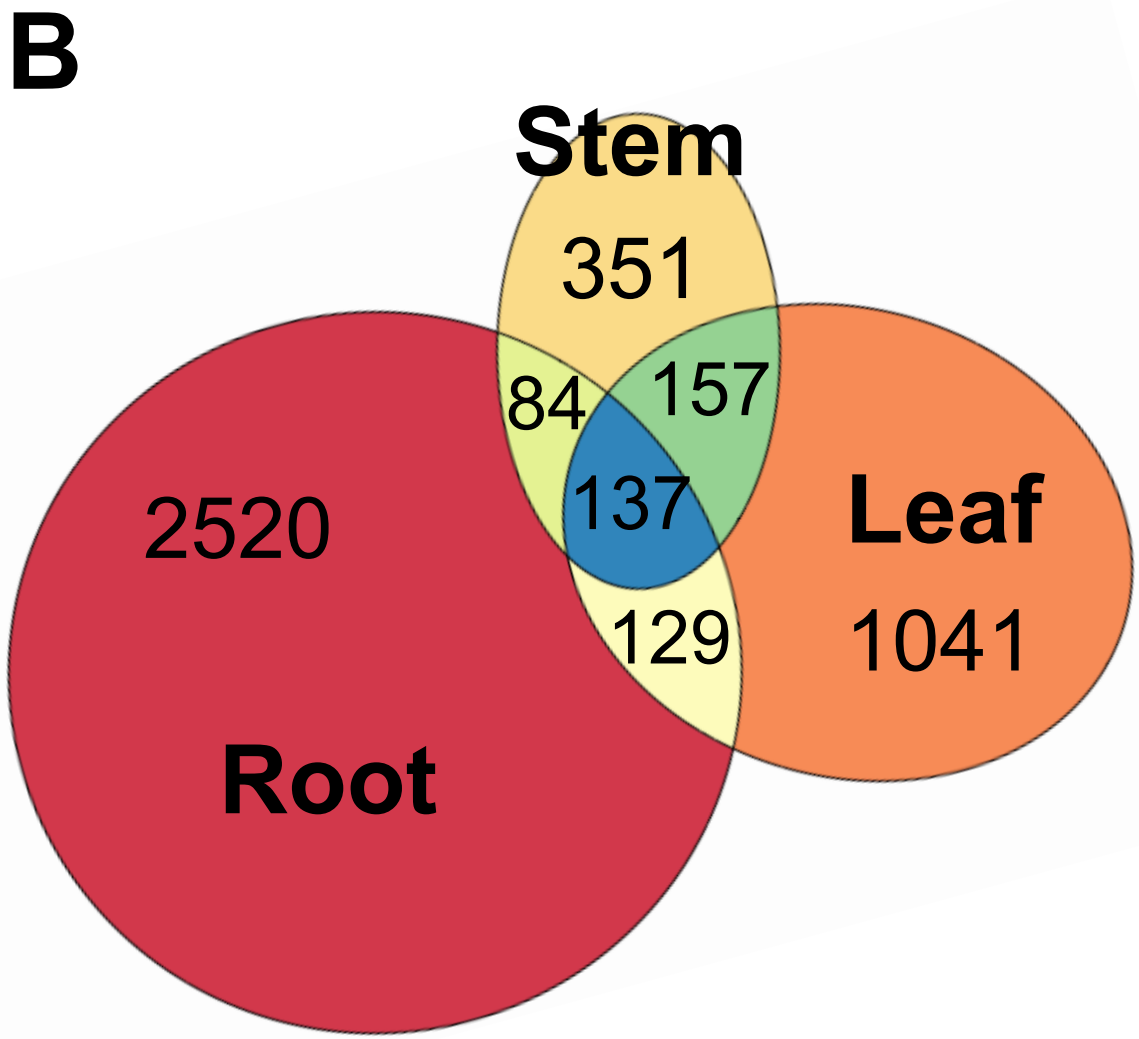
713 “secondary metabolism” GO term.

714


Carbohydrate concentration (g / kg DW)
0
10
20
30
40
50
Fructose
Glucose
Sucrose*
p < 0.05
Carbohydrate concentration (g / kg DW)
0
25
50
75
100
125
150
175
Starch**
p < 0.01
NSC***
p < 0.001
Yield (kg DW)
0
1
2
3
4
Yield*
p < 0.05
Number of tillers
0
5
10
15
20
25
30
35
40
Tillers*
p < 0.05




A
Stem
567
129
161
Leaf
437
35
43
14
Root


B
Stem
351
84
157
137
2520
Leaf
1041
129
Root