CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Apollo

- Quantification of the degree to which transcript abundance differs between C_3 and C_4 leaves
- Identification of novel components of C_4 metabolism
- Intersection with mathematical models to explain evolution of the complex C_4 phenotype
- Indication that C_4 photosynthesis is underpinned by both convergent and parallel evolution of structural genes and also regulators

*Manuscript
Click here to view linked References

1 **Insights into C_4 metabolism from comparative deep sequencing**

2

3 Steven J. Burgess and Julian M. Hibberd

4

5 Department of Plant Sciences, Downing Street, University of Cambridge, Cambridge, CB2 3EA, UK

6 Email: SJB: sjb287@cam.ac.uk, JMH: jmh65@cam.ac.uk

7

8 For correspondence; email - jmh65@cam.ac.uk

9

10 **Abstract**

11 C_4 photosynthesis suppresses the oxygenation activity of Ribulose Bisphosphate Carboxylase Oxygenase
12 and so limits photorespiration. Although highly complex, it is estimated to have evolved in sixty-six plant
13 lineages, with the vast majority lacking sequenced genomes. Transcriptomics has recently initiated
14 assessments of the degree to which transcript abundance differs between C_3 and C_4 leaves, identified novel
15 components of C_4 metabolism, and also led to mathematical models explaining the repeated evolution of
16 this complex phenotype. Evidence is accumulating that this complex and convergent phenotype is partly
17 underpinned by parallel evolution of structural genes, but also regulatory elements in both *cis* and *trans*.
18 Furthermore, it appears that initial events associated with acquisition of C_4 traits likely represent
19 evolutionary exaptations related to non-photosynthetic processes.

20 **Introduction**

21 C_3 plants inherited a carbon fixation system developed by the photosynthetic bacteria, with primary
22 carbon fixation being catalysed by the enzyme Ribulose Bisphosphate Carboxylase Oxygenase (RuBisCO).
23 The oxygenase activity of RuBisCO generates the toxic intermediate phosphogylcollate, and although this
24 can be detoxified and carbon partially recovered by the photorespiratory pathway, energy is expended in
25 the process. As the oxygenase function of RuBisCO increases with ambient temperature, it is thought that
26 in tropical and sub-tropical habitats, significant selection pressure led to the convergent evolution of
27 carbon concentrating mechanisms [1]. Phylogeny indicates that land plants have repeatedly evolved either
28 temporal (Crassulacean Acid Metabolism) or spatial carbon concentrating mechanisms (C_4 photosynthesis)
29 [2].

30 Although highly complex, the C_4 pathway is estimated to have evolved in at least sixty-six lineages of
31 plants [3]. Initial analysis of clades that contain C_3 and C_4 species but also 'C_3-C_4' intermediates identified
32 the most common early traits likely associated C_4 photosynthesis, and this led to the development of
33 models that depict the evolution of this complex phenotype along a relatively linear path of trait acquisition
34 [4]. More recently, probabalistic modelling within a Bayesian framework identified flexibility in when C_4
35 component traits evolve, but also found four major paths likely associated with acquisition of these traits
36 [5]. Despite this flexibility in the acquisition of C_4 component traits, the core C_4 metabolic machinery has
37 converged upon a similar architecture in all C_4 lineages. For example, in all C_4 species, HCO_3^- is initially fixed
38 by phospho*enol*pyruvate carboxylase (PEPC) (Figure 1), which has a higher affinity for HCO_3^- than RuBisCO
39 does for CO_2 [6]. C_4 acids then diffuse down a concentration gradient into insulated cellular, or sub-cellular
40 [7] compartments where C_4 acid decarboxylases increase the local concentration of CO_2 around RuBisCO,
41 thereby reducing its oxygenation activity. In most C_4 species, an altered arrangement of cells within the leaf
42 known as Kranz anatomy facilitates the compartmentation of carboxylation and decarboxylation (Figure
43 1A). There are three basic biochemical pathways defined by the predominant C_4 acid decarboxylase that
44 releases CO_2 around RuBisCO, but there are also at least 25 forms of Kranz anatomy documented (Figure 1A
45 and 1B).

46 Progress in understanding C_4 leaf anatomy has recently been critically assessed [8]. Here we focus on
47 how deep sequencing is influencing our understanding of C_4 biochemistry and argue that combined with
48 allied technologies it is opening up a new era of C_4 research. These approaches are helpful for at least three
49 reasons. First, many years of mutant screens, biochemistry and molecular biology have so far failed to
50 unlock many of the molecular components that regulate or induce the C_4 system [9,10], sequencing offers
51 the opportunity to identify candidate genes for these traits. Second, the C_4 pathway should correctly be
52 viewed as a system. Deep sequencing now makes it possible to move from analysis of individual genes and
53 their gene products, to assessing the simultaneous behaviours of both the system and its components.
54 Third, computational advances that have been driven by deep sequencing datasets provide the opportunity
55 to study the natural diversity of all C_4 lineages, rather than being limited to well-studied 'model' species for

56 which genome sequence is available. With this as background, we now assess how deep sequencing has
57 influenced the understanding of core components of the C_4 cycle, the *trans*-factors likely responsible for
58 their compartmentation between mesophyll (M) and bundle sheath (BS) cells of the C_4 leaf, and also the
59 evolutionary processes that have governed the transition from the ancestral C_3 photosynthetic system to
60 the derived C_4 metabolic pathway.

61 **Defining mRNAs associated with C_4 photosynthesis**

62 Approximately forty genes encoding core C_4 cycle enzymes and components of the Calvin-Benson-
63 Bassham cycle (CBB) have long been known to be involved in C_4 metabolism. RNA-seq has been used to
64 report mRNA signatures associated with the 'NAD-ME', 'NADP-ME' or 'PEPCK' biochemical sub-types [11–
65 13], and along with theoretical and modelling approaches, has provided clear evidence that often two of
66 the decarboxylases operate in parallel, with their relative contributions varying depending on conditions
67 [14–17]. However, our understanding of what changes leaf anatomy such that contact between tissues
68 involved in carbon assimilation and reduction is increased (Figure 1A), and also what sets up and then
69 maintains the patterns of gene expression required for the C_4 cycle are rudimentary. These factors are
70 important, as an understanding of C_4 genetics has implications for strategies being adopted to engineer the
71 pathway into C_3 crop species, dictating whether efforts should be focused on alterations to individual
72 genes, transcriptional regulators or hormone metabolism and signalling. Deep sequencing has allowed
73 estimates of the extent to which global patterns of mRNA abundance differ between C_3 and C_4 leaves. This
74 approach was initiated in the *Cleomaceae,* which in addition to containing C_3 and C_4 species, is
75 phylogenetically the closest-C_4-containing clade to C_3 *Arabidopsis thaliana* [18]. 603 genes showed
76 differential mRNA abundance in C_4 compared with C_3 leaves [12]. Furthermore, in addition to confirmation
77 that mRNAs encoding core C_4 and CBB cycles were up and down-regulated respectively, previously
78 unidentified characteristics of the C_4 leaf as well as new components of the C_4 cycle were reported. For
79 example, reduced abundance of mRNAs encoding ribosomal sub-units in C_4 compared with C_3 leaves was
80 reported [12], while *BASS2,* which was subsequently shown to encode the long-sought-after pyruvate
81 transporter associated with C_4 photosynthesis was up-regulated [19]. Subsequent analysis has led to
82 increased numbers of genes being linked to the C_4 cycle [13] and Table 1. The highest reported differences
83 in transcript abundance between C_3 and C_4 tissues are derived from *Eleocharis*, a species that is able to
84 switch from C_3 to C_4 depending on whether it is aquatic or terrestrial (Table 1). However, a proportion of
85 the mRNAs reported to be differentially abundant in C_4 compared with C_3 *Eleocharis* are likely associated
86 with the different light and temperature conditions caused by the aquatic to terrestrial switch [20].

87 Comparison of estimates of the number of changes associated with each of the three biochemical sub-
88 types (Figure 1) led to suggestions that establishment of the PEPCK C_4 sub-type requires the fewest
89 changes, in part because of reduced requirements for alterations in photosystem accumulation between
90 mesophyll and bundle sheath cells [11]. An overview of statistics from these studies (Table 1) shows that as
91 sequencing depths have increased there has been an increase in the predicted number of differentially
92 expressed genes, likely due to better quantification of low abundance transcripts. However, as no
93 annotated genomes were available for these species, the data are based either on cross-species mapping of
94 reads, or gene models created by *de novo* transcriptome assembly [21–23]. Both of these approaches
95 introduce inaccuracy compared with direct read mapping to a well-annotated genome. It is important to
96 note that the absolute number of differentially expressed genes detected through congeneric comparisons

97 is clearly dependent on the phylogenetic distance, statistical cut-offs, quality of transcriptome assemblies
98 and number of species sampled (Table 1). As the number of independent C_4 lineages that are assessed with
99 RNA-seq increase, estimates of the conserved alterations to mRNA abundance will become more reliable.
100 However, it is clear from the current estimates which range from hundreds to thousands of genes showing
101 differential expression in C_4 compared with C_3 leaves, research needs to focus on identification of key
102 transcription factors and signalling events that underlie these patterns of gene expression.
103
104 **Compartmentation of gene expression between cell-types of the C_4 leaf**
105 As with analysis of any organ or tissue, the C_4 leaf is composed of multiple distinct cell types, and the
106 specialisation of M and BS cells in C_4 leaves (Figure 1) is considered a hallmark of the C_4 pathway. The first
107 publications on global mRNA populations of M and BS cells of C_4 leaves were conducted on maize and
108 supported existing knowledge of genes known to be differentially expressed between these cell types
109 [24,25]. Analysis of two independent C_4 lineages from the grasses indicated that the absolute abundance of
110 mRNAs in M and BS cells of grasses that evolved C_4 photosynthesis independently was statistically more
111 convergent than other differentially expressed genes [26]. This implies that strong selection pressures
112 acted on genes associated with the C_4 pathway to generate very similar expression in separate C_4 lineages.
113 As the M and BS transcriptomes of more C_4 species become available this quantitative convergence could
114 be used to generate a predictive framework that allows unknown components of C_4 photosynthesis to be
115 identified. Although it has long been clear that transcriptional, post-transcriptional and post-translational
116 processes all play a part in generating the C_4 metabolic system [9], omics approaches are now initiating
117 non-biased and systems level quantification of their importance. For example, quantitative proteomics and
118 transcriptomics indicated that the ratio of each cognate protein to its mRNA varies during C_4 leaf
119 development, and that the ratio is often highest where protein function is most relevant [27]. Taken
120 together, these findings start to provide an oversight of the extent of post-transcriptional and post-
121 translational regulation in the C_4 leaf.
122 Transcriptomic datasets derived from M and BS cells of C_4 leaves highlight an area of ignorance, namely
123 the mRNA populations associated with these two cell types in leaves of ancestral C_3 plants. Without this
124 information it has not been possible to define how much patterns of gene expression have altered in M and
125 BS cells of C_4 compared with those cells in C_3 leaves. A major hurdle was our inability to isolate M and BS
126 cells from C_3 leaves, however immunopurification of ribosomes from specific cell types [28] has initiated
127 our understanding of the BS in C_3 *Arabidopsis thaliana*. Although it was previously known that veinal cells of
128 C_3 plants possessed characteristics of C_4 photosynthesis [30,31], ribosome tagging and deep sequencing of
129 associated mRNAs indicated that components of the C_4 cycle are also preferentially expressed in the C_3 BS
130 [29]. This work also highlighted a role for the C_3 BS in sulphur metabolism, a characteristic that had
131 previously been reported of the C_4 BS [32]. Thus, as more C_3 lineages are sampled, we will develop a much
132 clearer understanding of the extent to which metabolic characteristics currently associated with C_4

133 photosynthesis are actually ancestral and present in either M or BS cells of C_3 leaves. We therefore
134 conclude that technologies are in place to significantly improve our understanding of M and BS cells in both
135 C_3 and C_4 plants. Data from these approaches are being used to formulate models that relate to the
136 molecular drivers associated with the repeated evolution of this complex trait, and it is this that will be
137 explored in the next section.

138

139 **Insights into the molecular drivers of C_4 metabolism**

140 It has been clear for some time that prior to their recruitment into C_4 photosynthesis, the major proteins
141 of C_4 photosynthesis typically accumulate at relatively low levels in a constitutive manner in C_3 leaves [33].
142 Through comparison with a gene expression atlas of closely related species, it is now proposed that
143 expression of orthologues to C_4 genes show a variety of expression patterns, and peak in various tissues, in
144 the C_3 ancestral system [34]. Deep sequencing data has also now provided the insight into the extent to
145 which genes of the C_4 cycle become co-regulated with photosynthesis genes in leaves of both C_4
146 monotyledons and dicotyledons [23,35]. Overall, these data imply that during the evolution of C_4
147 photosynthesis, genes of the C_4 cycle are co-opted into the gene regulatory networks that govern
148 photosynthesis gene expression in the ancestral C_3 state [23,34].

149 The identification of transcription factors responsible for these alterations in expression of genes
150 encoding components of the C_4 cycle is an area where significant progress still needs to be made. However,
151 comparative transcriptomics has now identified candidate regulators for the C_4 cycle in maize [24,25,35–
152 37], *Setaria* [26,38], *Flaveria* [13] and *Gynandropsis gynandra* (formerly known as *Cleome gynandra*)
153 [23,34]. Interestingly, independent lineages of C_4 plants appear to have up-regulated homologous
154 transcriptional regulators in either M or BS cells. This has been reported for two independent lineages of C_4
155 grasses [26] but also for the C_4 dicotyledon *G. gynandropsis* and the C_4 monocotyledon maize [23]. These
156 data indicated that M or BS preferential expression is not only associated with parallel evolution of
157 regulatory DNA [39] and histone marks [40], but also the recruitment of transcription factors [23,26].

158 Another striking finding facilitated by deep sequencing has been quantification of the extent to which
159 specific members of multi-gene families are recruited into the C_4 pathway. This was initially reported after
160 phylogenetic reconstructions of individual genes such as *PEPC* [41], but the extent of this process was not
161 clear. Transcriptomics has now quantified this phenomenon in *Alloteropsis*, which contains C_3 and C_4
162 subspecies [42]. In maize and *Setaria*, which represent two independent lineages of C_4 grass, 87% of C_4
163 cycle proteins that are up-regulated in C_4 leaves are syntenic orthologues, indicating that the same
164 ancestral gene has repeatedly been recruited into the pathway [26]. Again, the mechanism behind this
165 phenomenon is not clear, but it is possible that these orthologues are repeatedly used into the C_4 pathway
166 because they are part of pre-existing gene regulatory networks that are recruited into C_4 photosynthesis.
167 These data further emphasize that the highly complex C_4 photosynthesis trait is underpinned by a mixture
168 of both convergent and parallel evolution [39,42].

The combination of deep sequencing and metabolic flux modelling has demonstrated the power of an integrated approach, and lead to an enticing hypothesis concerning the repeated evolution of C_4 photosynthesis. Comparing C_3, C_3-C_4 and C_4 species in *Flaveria*, RNA-seq data coupled to metabolic modelling predicted that loss of the full photorespiratory pathway in the M cells of C_3 plants, which is the most common biochemical alteration thought to initiate C_4 evolution [2], leads to a nitrogen imbalance between M and BS cells [43] (Figure 2). The most parsimonious alterations to central metabolism that corrects this imbalance in the leaf is to induce, and compartment, the key components of the C_4 cycle into either M or BS cells (Figure 2). These data strongly imply that the metabolic remodelling during these early stages of C_4 evolution represent an evolutionary exaptation that was initially not related to photosynthetic efficiency *per se*. Thus, it now appears that metabolic and also morphological alterations to C_3 leaves were both unrelated to photosynthesis [5,42,44]. Later in the evolutionary process it is thought that each alteration to the C_4 cycle leads to a steady increase in photosynthetic performance [45], and this is then followed by evolutionary fine-tuning mediated by amino acid substitutions that modify allosteric regulation of these proteins for the C_4 leaf [46]. In the future, deep sequencing will also allow us to determine whether parallel changes to amino acids are associated with parallel or convergent evolution to the nucleotides encoding them. Moving ahead, perhaps a similar combined modelling, sequencing and hormone approach is required to make progress in understanding the molecular basis of Kranz anatomy.

Summary

The use of deep sequencing in C_4 research is in its infancy, and so far is mostly limited to RNA-seq. It is also true that the initial phase has identified many genes that could be important for C_4 photosynthesis, but for which functional analysis has not yet been undertaken. However, it is clear that use of deep sequencing has initiated an unbiased and objective study of C_4 photosynthesis in species that previously lacked any transcriptomic or genomic resources. As outlined above, deep sequencing and improved computational pipelines for data analyses have started to provide significant new insight. This includes defining core components of the C_4 cycle, identifying variations in C_4 metabolism both within and between species, and also providing inference into evolutionary mechanisms associated with the polyphyletic appearance of this highly complex system.

197 **Acknowledgements**

198 We thank the BBSRC for grant BB/1002243/1 and the EU *3to4* program for financial support.

199 1. Christin P-A, Osborne CP: **The evolutionary ecology of C_4 plants**. *New Phytol.* 2014, **204**:765–781.

200 2. Sage RF: **Photosynthetic efficiency and carbon concentration in terrestrial plants: the C_4 and CAM**
201 **solutions**. *J. Exp. Bot.* 2014, **65**:3323–3325.

202 3. Sage RF, Christin P-A, Edwards EJ: **The C_4 plant lineages of planet Earth**. *J. Exp. Bot.* 2011, **62**:3171–
203 3181.

204 4. Sage R: **The evolution of C_4 photosynthesis**. *New Phytol* 2004, **161**:341–370.

205 5. Williams BP, Johnston IG, Covshoff S, Hibberd JM: **Phenotypic landscape inference reveals multiple**
206 **evolutionary paths to C_4 photosynthesis**. *eLife* 2013, **2**.

207 6. Hatch MD: **C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and**
208 **ultrastructure**. *Biochim. Biophys. Acta* 1987, **895**:81–106.

209 7. Voznesenskaya EV, Franceschi VR, Kiirats O, Artyusheva EG, Freitag H, Edwards GE: **Proof of C_4**
210 **photosynthesis without Kranz anatomy in *Bienertia cycloptera* (Chenopodiaceae)**. *Plant J.* 2002,
211 **31**:649–662.

212 8. Fouracre JP, Ando S, Langdale J a: **Cracking the Kranz enigma with systems biology.** *J. Exp. Bot.*
213 2014, **65**:3327–39.

214 9. Hibberd JM, Covshoff S: **The regulation of gene expression required for C_4 photosynthesis**. *Annu*
215 *Rev Plant Biol* 2010, **61**:181–207.

216 10. Langdale JA: **C_4 cycles: past, present, and future research on C_4 photosynthesis.** *Plant Cell* 2011,
217 **23**:3879–92.

218 11. Bräutigam A, Schliesky S, Külahoglu C, Osborne CP, Weber APM: **Towards an integrative model of C_4**
219 **photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-**
220 **ME, and PEP-CK C4 species.** *J. Exp. Bot.* 2014, **65**:3579–93.

221 12. Brautigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D, Weber KL, Carr KM, Gowik U, Mass J,
222 Lercher MJ, et al.: **An mRNA blueprint for C_4 photosynthesis derived from comparative**
223 **transcriptomics of closely related C_3 and C_4 species**. *Plant Physiol* 2010, **155**:142–156.

224 ++ The initial use of deep sequencing of closely related C_3 and C_4 plants to provide quantitative insight into
225 the degree that their leaf transcriptomes differ, and identifying new candidate proteins important for the
226 C_4 cycle.

227

228 13. Gowik U, Brautigam A, Weber KL, Weber AP, Westhoff P: **Evolution of C_4 photosynthesis in the**
229 **genus Flaveria: how many and which genes does it take to make C_4?** *Plant Cell* 2011, **23**:2087–
230 2105.

231 14. Furbank RT: **Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid**
232 **decarboxylation types?** *J. Exp. Bot.* 2011, doi:10.1093/jxb/err080.

233 15. Bellasio C, Griffiths H: **Acclimation of C4 metabolism to low light in mature maize leaves could limit**
234 **energetic losses during progressive shading in a crop canopy.** *J. Exp. Bot.* 2014, **65**:3725–36.

235 16. Wang Y, Bräutigam A, Weber APM, Zhu X-G: **Three distinct biochemical subtypes of C_4**
236 **photosynthesis? A modelling analysis.** *J. Exp. Bot.* 2014, **65**:3567–78.

237 17. Sommer M, Bräutigam A, Weber APM: **The dicotyledonous NAD malic enzyme C_4 plant *Cleome***
238 ***gynandra* displays age-dependent plasticity of C_4 decarboxylation biochemistry.** *Plant Biol.* 2012,
239 **14**:621–9.

240 18. Brown NJ, Parsley K, Hibberd JM: **The future of C_4 research - Maize, *Flaveria* or *Cleome*?** *Trends*
241 *Plant Sci.* 2005, **10**:215–221.

242 19. Furumoto T, Yamaguchi T, Ohshima-Ichie Y, Nakamura M, Tsuchida-Iwata Y, Shimamura M, Ohnishi
243 J, Hata S, Gowik U, Westhoff P, et al.: **A plastidial sodium-dependent pyruvate transporter**. *Nature*
244 2011, **476**:472–475.

245 ++Functional analysis showing that a gene identified after deep sequencing of closely related C_3 and C_4
246 species encoded an elusive pyruvate transporter.

247

248 20. Chen T, Guang X, Yongjun Z: **Major alterations in transcript profiles between C_3 – C_4 and C_4**
249 **photosynthesis of an amphibious species *Eleocharis baldwinii***. *Plant Mol Biol* 2014, **86**:93–110.

250 21. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury
251 R, Zeng Q, et al.: **Full-length transcriptome assembly from RNA-Seq data without a reference**
252 **genome**. *Nat Biotech* 2011, **29**:644–652.

253 22. Schulz MH, Zerbino DR, Vingron M, Birney E: **Oases: robust *de novo* RNA-seq assembly across the**
254 **dynamic range of expression levels**. *Bioinformatics* 2012, **28** :1086–1092.

255 23. Aubry S, Kelly S, Kümpers BMC, Smith-Unna RD, Hibberd JM: **Deep evolutionary comparison of gene**
256 **expression identifies parallel recruitment of trans-factors in two independent origins of C_4**
257 **photosynthesis**. *PLoS Genet* 2014, **10**:e1004365.

258 24. Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, et al.:
259 **The developmental dynamics of the maize leaf transcriptome**. *Nat Genet* 2010, **42**:1060–1067.

260 25. Chang YM, Liu WY, Shih AC, Shen MN, Lu CH, Lu MY, Yang HW, Wang TY, Chen SC, Chen SM, et al.:
261 **Characterizing regulatory and functional differentiation between maize mesophyll and bundle**
262 **sheath cells by transcriptomic analysis**. *Plant Physiol* 2012, **160**:165–177.

263 26. John CR, Smith-Unna RD, Woodfield H, Covshoff S, Hibberd JM: **Evolutionary Convergence of Cell-**
264 **Specific Gene Expression in Independent Lineages of C_4 Grasses**. *Plant Physiol.* 2014, **165** :62–75.

265 27. Ponnala L, Wang Y, Sun Q, van Wijk KJ: **Correlation of mRNA and protein abundance in the**
266 **developing maize leaf**. *Plant J.* 2014, **78**:424–440.

267 28. Zanetti ME, Chang I-F, Gong F, Galbraith DW, Bailey-Serres J: **Immunopurification of Polyribosomal**
268 **Complexes of Arabidopsis for Global Analysis of Gene Expression**. *Plant Physiol.* 2005, **138** :624–
269 635.

270 29. Aubry S, Smith-Unna RD, Boursnell CM, Kopriva S, Hibberd JM: **Transcript residency on ribosomes**
271 **reveals a key role for the *Arabidopsis thaliana* bundle sheath in sulfur and glucosinolate**
272 **metabolism**. *Plant J.* 2014, **78**:659–673.

273 30. Hibberd JM, Quick WP: **Characteristics of C_4 photosynthesis in stems and petioles of C_3 flowering**
274 **plants**. *Nature* 2002, **415**:451–454.

275 31. Brown NJ, Palmer BG, Stanley S, Hajaji H, Janacek SH, Astley HM, Parsley K, Kajala K, Quick WP,
276 Trenkamp S, et al.: **C_4 acid decarboxylases required for C_4 photosynthesis are active in the mid-vein**
277 **of the C_3 species *Arabidopsis thaliana*, and are important in sugar and amino acid metabolism**.
278 *Plant J.* 2010, **61**:122–133.

279 32. Leegood RC: **Roles of the bundle sheath cells in leaves of C_3 plants**. *J. Exp. Bot.* 2008, **59**:1663–1673.

280 33. Aubry S, Brown NJ, Hibberd JM: **The role of proteins in C_3 plants prior to their recruitment into the**
281 **C_4 pathway**. *J. Exp. Bot.* 2011, doi:10.1093/jxb/err012.

282 34. Kulahoglu C, Denton a. K, Sommer M, Mass J, Schliesky S, Wrobel TJ, Berckmans B, Gongora-Castillo
283 E, Buell CR, Simon R, et al.: **Comparative transcriptome atlases reveal altered gene expression**
284 **modules between two Cleomaceae C_3 and C_4 plant species**. *Plant Cell* 2014, **26**:3243–3260.

285 ++ Detailed analysis of samples from closely related C_3 and C_4 plants providing key insights into genes
286 recruited into C_4 photosynthesis.

287

288 35. Tausta SL, Li P, Si Y, Gandotra N, Liu P, Sun Q, Brutnell TP, Nelson T: **Developmental dynamics of**
289 **Kranz cell transcriptional specificity in maize leaf reveals early onset of C_4-related processes**. *J. Exp.*
290 *Bot.* 2014, **65**:3543–3555.

291 36. Pick TR, Bräutigam A, Schlüter U, Denton AK, Colmsee C, Scholz U, Fahnenstich H, Pieruschka R,
292 Rascher U, Sonnewald U, et al.: **Systems analysis of a maize leaf developmental gradient redefines**
293 **the current C_4 model and provides candidates for regulation.** *Plant Cell* 2011, **23**:4208–20.

294 37. Wang L, Czedik-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, Arrivault S, Dedow LK, Bryant
295 DW, Zhou W, et al.: **Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of**
296 **maize and rice**. *Nat Biotech* 2014, **32**:1158–1165.

297 ++ Deep sequencing of C_3 and C_4 grasses combined with an algorthim to normalise between leaves of
298 different growth habits.

299

300 38. Lin J-J, Yu C-P, Chang Y-M, Chen SC-C, Li W-H: **Maize and millet transcription factors annotated**
301 **using comparative genomic and transcriptomic data**. *BMC Genomics* 2014, **15**:818.

302 39. Brown NJ, Newell CA, Stanley S, Chen JE, Perrin AJ, Kajala K, Hibberd JM: **Independent and parallel**
303 **recruitment of preexisting mechanisms underlying C4 photosynthesis**. *Science.* 2011, **331**:1436–
304 1439.

305 40. Heimann L, Horst I, Perduns R, Dreesen B, Offermann S, Peterhansel C: **A Common histone**
306 **modification code on C_4 genes in maize and its conservation in *Sorghum* and *Setaria italica*.** *Plant*
307 *Physiol.* 2013, **162**:456–69.

308 41. Christin P-A, Petitpierre B, Salamin N, Büchi L, Besnard G: **Evolution of C_4 Phospho*enol*pyruvate**
309 **Carboxykinase in Grasses, from Genotype to Phenotype**. *Mol. Biol. Evol.* 2009, **26**:357–365.

310 42. Christin P-A, Boxall SF, Gregory R, Edwards EJ, Hartwell J, Osborne CP: **Parallel recruitment of**
311 **multiple genes into C_4 photosynthesis.** *Genome Biol. Evol.* 2013, **5**:2174–87.

312 43. Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik, U: **The role of**
313 **photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria***. *eLife* 2014,
314 10.7554/eLife.02478.

315 ++RNAseq used in conjunction with with flux modeling to provide amazing insight into the evolution of C_4
316 photosynthesis.

317 44. Griffiths H, Weller G, Toy LF, Dennis RJ: **You're so vein: bundle sheath physiology, phylogeny and**
318 **evolution in C_3 and C_4 plants**. *Plant Cell Env.* 2013, **36**:249–261.

319 45. Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber AP, Lercher MJ: **Predicting C_4**
320 **photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape**.
321 *Cell* 2013, **153**:1579–1588.

322 46. Blasing OE, Ernst K, Streubel M, Westhoff P, Svensson P: **The non-photosynthetic**
323 **phospho*enol*pyruvate carboxylases of the C_4 dicot *Flaveria trinervia* - implications for the**
324 **evolution of C_4 photosynthesis**. *Planta* 2002, **215**:448–56.

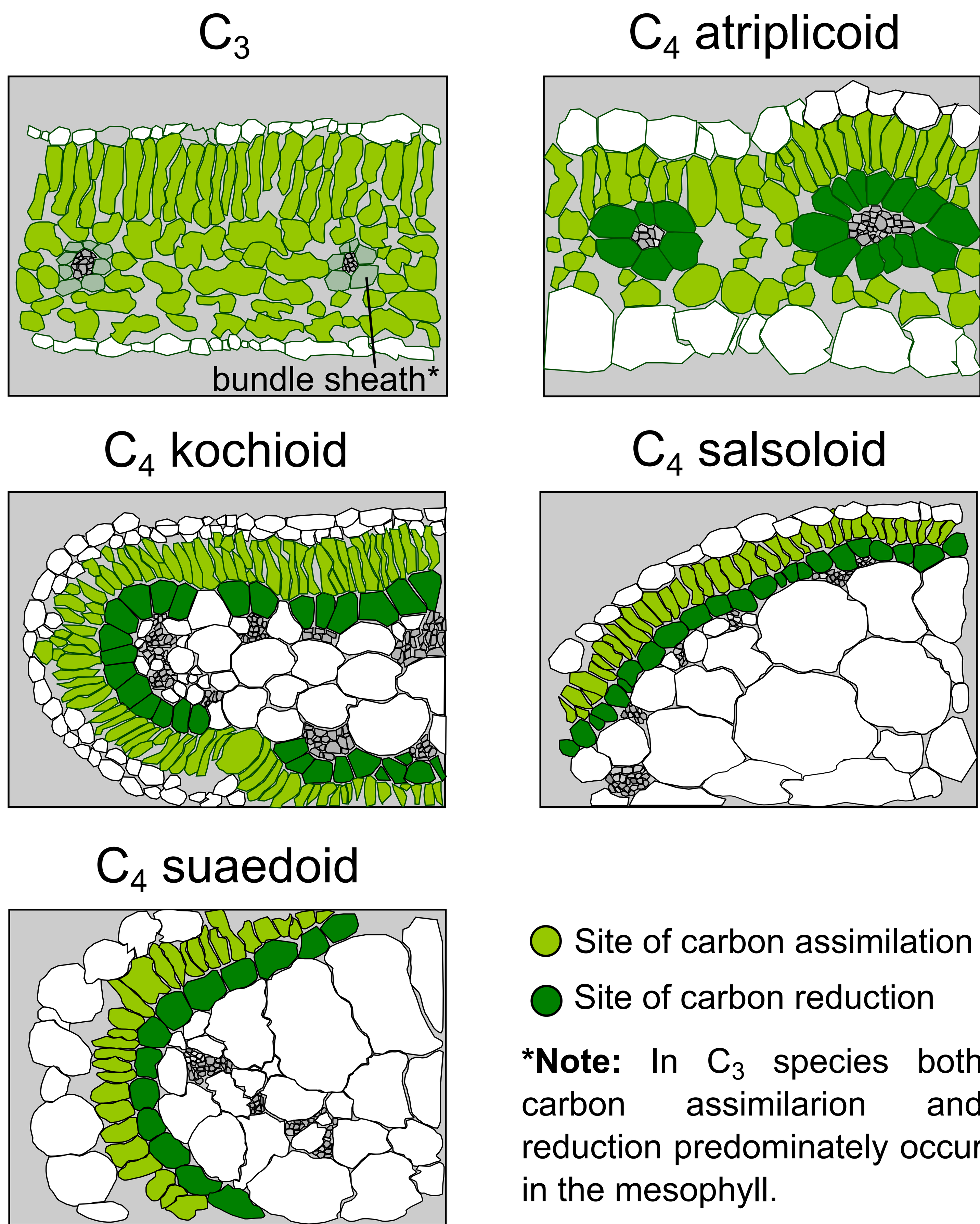
325 **Figure Legends**

326 **Figure 1:** Schematics illustrating variation in leaf anatomy and C_4 biochemical cycles of C_4 leaves. **A.**
327 Diagrams representing transverse sections through a C_3 leaf, and four anatomical variations in Kranz
328 anatomy. Images are based on those reported by [47]. B. The three main cycles that have classically been
329 used to define the three biochemcial sub-types of C_4 photosynthesis. AlaAT = Alanine aminotransferase,
330 AspAT = Aspartate aminotransferase, CA= Carbonic anhydrase, PEPC = Phospho*enol*pyruvate carboxylase,
331 PEPCK = Phospho*enol*pyruvate carboxykinase, NADP-MDH = NADP-dependent malate dehydrogenase,
332 NADP-ME = NADP-dependent malic enzyme, NAD-ME = NAD-dependent malic enzyme, PPDK =
333 Pyruvate,orthophosphate dikinase, CBB = Calvin Benson Bassham cycle, Ala = alanine, Asp = aspartate, Mal
334 = malate, OAA = oxaloacetic acid, Pyr = Pyruvate, PEP = phospho*enol*pyruvate.

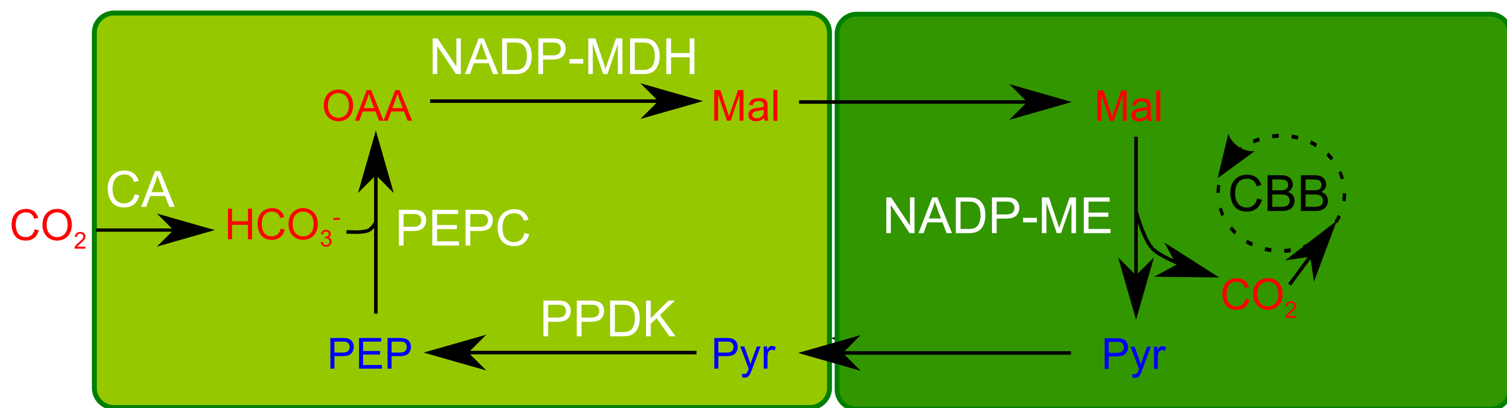
335

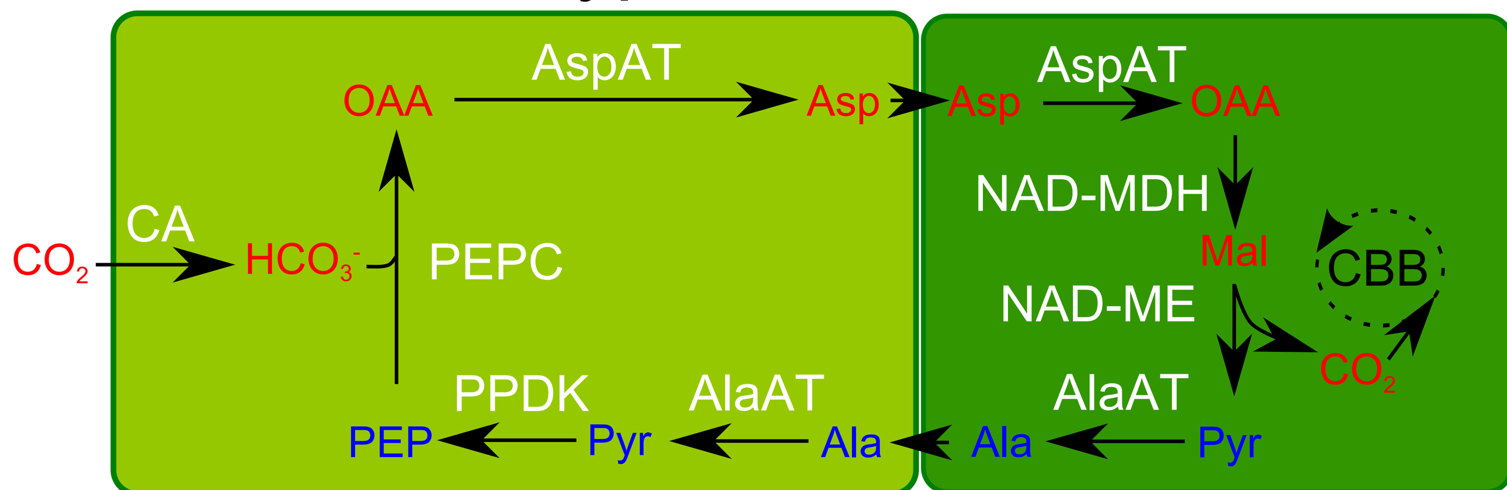
336 **Figure 2: Impacts of deep sequencing on understanding C_4 metabolism**. Representation of model
337 predicting initial events associated with the evolution of C_3-C_4 intermediacy (based on [43]). Loss of
338 photorespiration in the mesophyll cells would lead to lead to an imbalance in nitrogen metabolism
339 between mesophyll and bundle sheath cells, and accumulation of ammonia (yellow circle) in the bundle
340 sheath. Upregulation of a C_4-like pathway rebalances this nitrogen imbalance. The three panels represent
341 photorespiration (C_2 cycle) operating in both mesophyll and bundle sheath cells of a C_3 leaf (A), the C_2 cycle
342 being lost in the mesophyll cells of C_3-C_4 intermediate species, and the subsequent development of a C_4-like
343 cycle (B), and finally complete implementation of the C_4 cycle (C). Abbreviations as in Figure 1, as well as
344 Glu = glutamate, Gly= glycine, 2-OG = 2-oxoglutarate, Ser = serine. Dashed lines indicate low metabolic flux.
345 Red circles represent carbon atoms while yellow circles represent amine groups.

Figure in original format
Click here to download Figure in original format: Figure_1.pdf



B





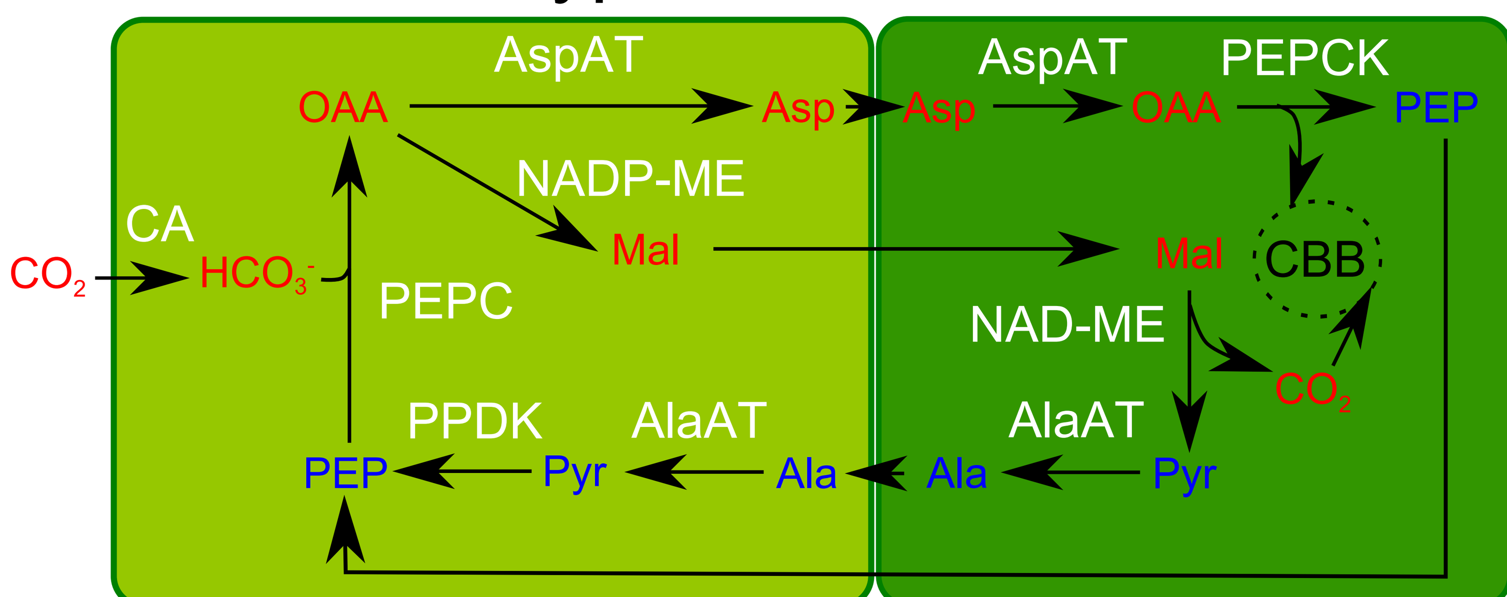
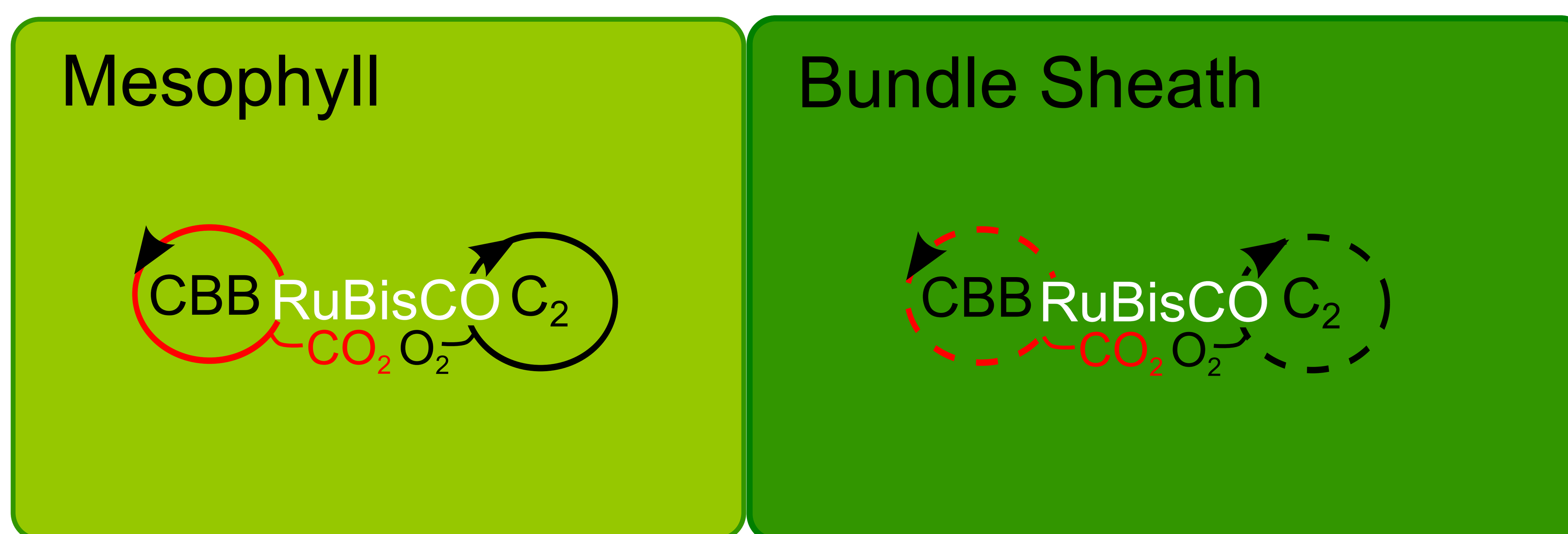
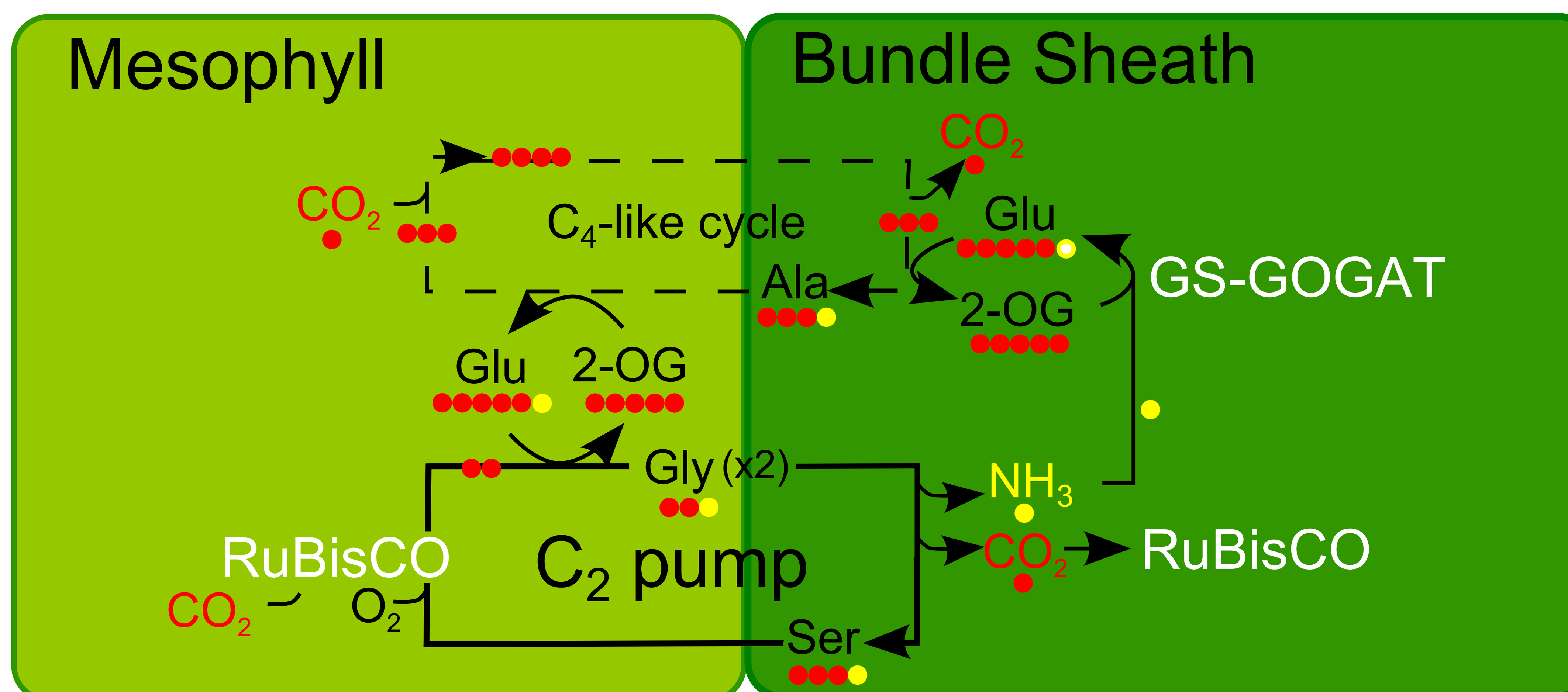


Figure in original format
Click here to download Figure in original format: figure_2_final.pdf

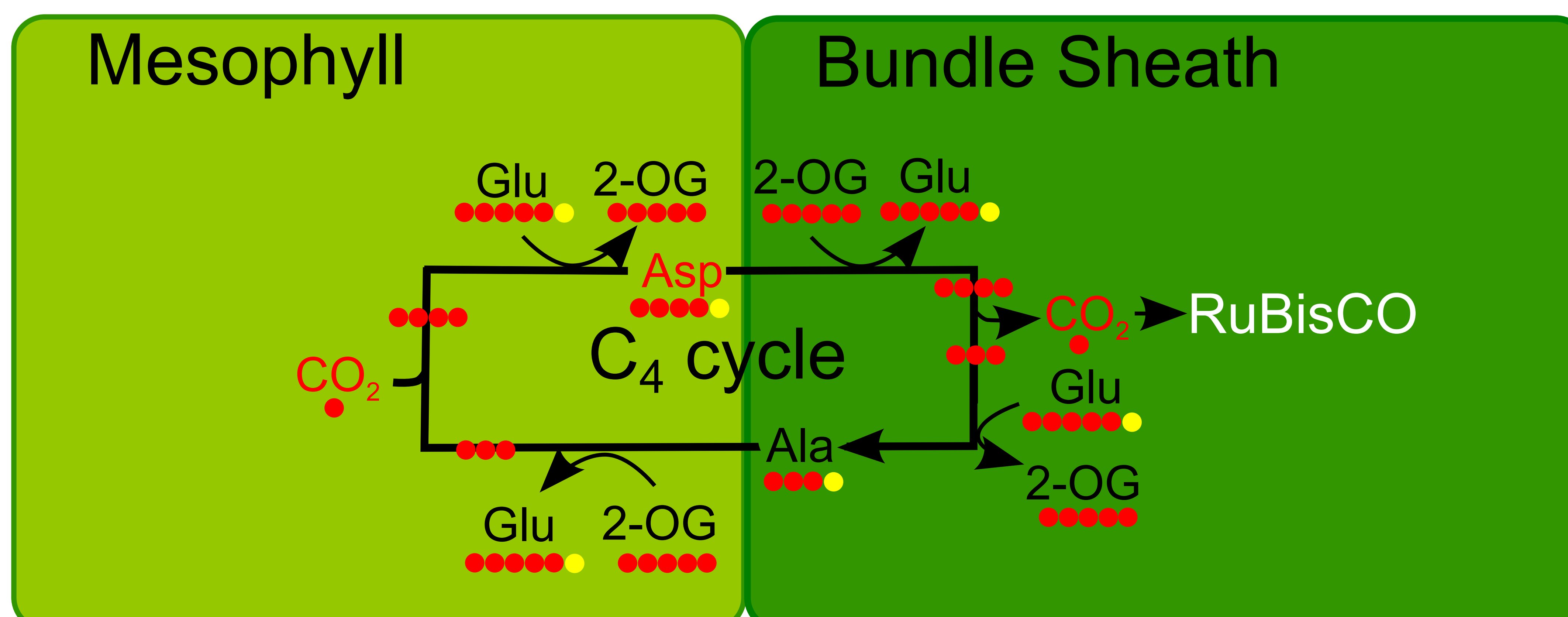
A



B



C



Table

	Bräutigam et al. (2011) [12]	Gowik et al. (2011) [13]	Bräutigam et al. (2014) [11]	Chen et al. (2014) [20]
Total number of DE transcripts	603	3582	1168	8848
Transcripts more abundant in C_3	258	1418	792	4184
Transcripts more abundant in C_4	345	2164	376	4664
% Transcriptome DE	1.4	NA*	6.1	13.5

Table 1: Comparisons of transcript abundance in closely related C_3 versus C_4 photosynthetic tissues. The total number of transcripts annotated as being differentially expressed (DE) in each study is listed, along with the numbers up or down regulated. Data expressed as percent of the total transcriptome are also reported for each study. Bräutigam *et al.* 2011 assessed C_4 *Gynandropsis gynandra* versus C_3 *Tareneya hassleriana.* Gowik *et al.* 2011 assessed C_4 *Flaveria bidentis* and *Flaveria trinervia* as well as C_3-C_4 *Flaveria ramosissima* and C_3 *Flaveria pringlei* and *Flaveria robusta.* Bräutigam *et al.* 2014 assessed *Panicum maximum* and *Dicanthelium clandestinum*. Chen *et al.* 2014 assessed C_4 and C_3 culms of *Eleocharis baldwinii.* *NA: the values for DE transcripts were based on multispecies comparisons which prohibits expressing the number of DE transcripts as a percentage of transcriptome.