



Article

# Insights into Regulation of C<sub>2</sub> and C<sub>4</sub> Photosynthesis in *Amaranthaceae/Chenopodiaceae* Using RNA-Seq

Christian Siadjeu <sup>1,\*</sup> , Maximilian Lauterbach <sup>2</sup> and Gudrun Kadereit <sup>1</sup>

<sup>1</sup> Systematics, Biodiversity and Evolution of Plants, Ludwig Maximilian University Munich, 80638 Munich, Germany; g.kadereit@biologie.uni-muenchen.de

<sup>2</sup> Martinstraße 20, 55116 Mainz, Germany; maxlauterbach0@gmail.com

\* Correspondence: christian.siadjeu@lmu.de

**Abstract:** *Amaranthaceae* (incl. *Chenopodiaceae*) shows an immense diversity of C<sub>4</sub> syndromes. More than 15 independent origins of C<sub>4</sub> photosynthesis, and the largest number of C<sub>4</sub> species in eudicots signify the importance of this angiosperm lineage in C<sub>4</sub> evolution. Here, we conduct RNA-Seq followed by comparative transcriptome analysis of three species from *Camphorosmeae* representing related clades with different photosynthetic types: *Threlkeldia diffusa* (C<sub>3</sub>), *Sedobassia sedoides* (C<sub>2</sub>), and *Bassia prostrata* (C<sub>4</sub>). Results show that *B. prostrata* belongs to the NADP-ME type and core genes encoding for C<sub>4</sub> cycle are significantly upregulated when compared with *Sed. sedoides* and *T. diffusa*. *Sedobassia sedoides* and *B. prostrata* share a number of upregulated C<sub>4</sub>-related genes; however, two C<sub>4</sub> transporters (DIT and TPT) are found significantly upregulated only in *Sed. sedoides*. Combined analysis of transcription factors (TFs) of the closely related lineages (*Camphorosmeae* and *Salsola*) revealed that no C<sub>3</sub>-specific TFs are higher in C<sub>2</sub> species compared with C<sub>4</sub> species; instead, the C<sub>2</sub> species show their own set of upregulated TFs. Taken together, our study indicates that the hypothesis of the C<sub>2</sub> photosynthesis as a proxy towards C<sub>4</sub> photosynthesis is questionable in *Sed. sedoides* and more in favour of an independent evolutionary stable state.

**Keywords:** *Amaranthaceae*; C<sub>4</sub> photosynthesis; Caryophyllales; *Chenopodiaceae*; complex trait evolution; gene regulation; *Salsola*; transcription factor



**Citation:** Siadjeu, C.; Lauterbach, M.; Kadereit, G. Insights into Regulation of C<sub>2</sub> and C<sub>4</sub> Photosynthesis in *Amaranthaceae/Chenopodiaceae* Using RNA-Seq. *Int. J. Mol. Sci.* **2021**, *22*, 12120. <https://doi.org/10.3390/ijms222212120>

Academic Editor: Alberto Mezzetti

Received: 12 October 2021

Accepted: 3 November 2021

Published: 9 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

C<sub>4</sub> photosynthesis is a carbon-concentration mechanism, enhancing CO<sub>2</sub> at the site of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). This mechanism leads to a decrease in the oxygenation reaction of RuBisCO, which in turn decreases photorespiration, because fewer toxic compounds resulting from the RuBisCO oxygenation reaction need to be recycled [1]. C<sub>4</sub> photosynthesis requires a series of biochemical, anatomical, and gene regulation changes compared with the C<sub>3</sub> photosynthesis ancestor [2,3]. C<sub>4</sub> photosynthesis has been a major subject in life science. Since the discovery of C<sub>4</sub> photosynthesis more than 50 years ago, its evolution is still under debate [4]. The current model of C<sub>4</sub> evolution relies heavily on the C<sub>3</sub>–C<sub>4</sub> intermediate (including C<sub>2</sub>) photosynthetic types as evolutionary stepping stones towards C<sub>4</sub> photosynthesis [5–7]. Most C<sub>3</sub>–C<sub>4</sub> intermediate species utilise C<sub>2</sub> photosynthesis, where a photorespiratory glycine shuttle and its decarboxylation by glycine decarboxylase (GDC) concentrate CO<sub>2</sub> in a bundle sheath-like compartment [5]. The establishment of this glycine-based CO<sub>2</sub> pump and the restriction of GDC activity in the bundle sheath cells (BSCs) is considered an important intermediate step in the evolution towards C<sub>4</sub> photosynthesis [8]. However, the absence of C<sub>4</sub> relatives in lineages with C<sub>3</sub>–C<sub>4</sub> intermediate phenotypes indicates that C<sub>2</sub> photosynthesis can be an evolutionarily stable state in their own right [9,10]. On the other hand, the hybrid origin of C<sub>2</sub> photosynthesis has been suggested because the anatomy of hybrids obtained from artificial crosses of *Atriplex prostrata* (C<sub>3</sub>) and *A. rosea* (C<sub>4</sub>) resemble the C<sub>3</sub>–C<sub>4</sub> intermediate using a glycine shuttle to concentrate CO<sub>2</sub> [11].

Despite its complexity, the  $C_4$  pathway independently evolved in at least 61 lineages of both monocot and eudicot lineages [12]. In eudicots, the *Amaranthaceae/Chenopodiaceae* alliance has the largest diversity of  $C_4$  syndromes with 15 independent origins of  $C_4$  identified, ten of which belong to *Chenopodiaceae sensu stricto* [12–18]. The closely related lineages *Camphorosmeae* and *Salsoleae*, belonging to the goosefoot family (*Chenopodiaceae*), are rich in  $C_4$  phenotypes [18] and contain a number of  $C_3$ – $C_4$  intermediate species, including  $C_2$ , proto-Kranz type and  $C_4$ -like type species. Both lineages are found in steppes, semi-deserts, salt marshes, and ruderal sites of Eurasia, South Africa, North America, and Australia [18,19]. *Camphorosmeae* comprise subshrubs and annuals, predominantly with moderately to strongly succulent leaves with a central aqueous tissue [16]. Evolutionary radiation was later in *Camphorosmeae* (early Miocene) than in *Salsoleae* (early to middle Oligocene) [18]. In *Camphorosmeae*,  $C_4$  photosynthesis likely evolved two times in the Miocene and different photosynthetic types are recognised on the basis of leaf anatomy with several  $C_4$  phenotypes [18,20].  $C_4$  photosynthesis in *Salsoleae* likely evolved multiple times and most species are  $C_4$  plants with terete leaves and *Salsoloid* Kranz anatomy in which a continuous dual layer of chlorenchyma cells encloses the vascular and water-storage tissue [18,21,22]. Therefore, these two sister groups constitute a central component allowing the investigation between and within each plant group for understanding the origin of  $C_2$  photosynthesis, the evolution of  $C_4$  photosynthesis and adjustments in gene regulation leading to different photosynthetic types. Indeed, new insight into  $C_4$  evolution were gained from studying *Salsoleae* lineage using high-throughput sequencing methods. A photosynthetic transition from  $C_3$  pathway in cotyledons to  $C_4$  pathway in leaves of the *Salsoleae* lineage (*Chenopodiaceae*, *Salsola soda* L.) has been identified [23]. This  $C_3$ -to- $C_4$  transition is thought to be a rather exceptional phenomenon since species conducting  $C_4$  pathway in all photosynthetically active tissues/organs are supposed to be the most abundant within this group. In addition, comparative transcriptomics revealed two proposed transporters associated with  $C_2$  and  $C_4$  photosynthesis [24]. However, for the *Camphorosmeae* lineage, transcriptome analysis and gene expression profiles of different photosynthesis types are still lacking. Moreover, for both lineages, differential gene expression of regulatory genes (e.g., transcription factors, TFs) involved in different photosynthetic pathways remains poorly understood.

The development of complex traits is controlled by the coordination of expression of many TFs and signaling pathways [25]. Thus, TFs play an important role in regulation of gene expression and are certainly responsible for the fine-tuning of the cell-specific expression patterns in  $C_4$  photosynthesis [26]. The characteristic expression pattern of PHOSPHOENOLPYRUVATE CARBOXYLASE (PEPC) in  $C_4$  plants (i.e., high abundance in mesophyll M and low abundance in BS cells), for example, could be controlled by a number of TFs from the ZINC FINGER HOMEODOMAIN (zf-HD) TF family [27]. Therefore, TFs are hot candidates for a stepwise evolutionary change of complex traits such as  $C_4$  photosynthesis. Reviewing nine studies on potential regulators of  $C_4$  photosynthesis in maize, Huang and Brutnell [28] found no TF consistently identified across these studies and suggested that consistent differential expression obtained between  $C_3$  and  $C_4$  sister lineages could be a more effective way to prioritise candidate TFs.

To fill this knowledge gap with new pieces of the puzzle, we (1) report transcriptome de novo assemblies and differential expression analysis between  $C_2$ ,  $C_3$ , and  $C_4$  species of *Camphorosmeae* (*Amaranthaceae*) using RNA-Seq, and (2) assess transcriptional regulator elements involved in  $C_3$ ,  $C_2$ , and  $C_4$  photosynthesis in the *Amaranthaceae/Chenopodiaceae* complex. In this latter objective, we merged transcriptome data generated in this study with the publicly available transcriptome data of  $C_2$ ,  $C_3$ , and  $C_4$  species of the sister lineage *Salsoleae*.

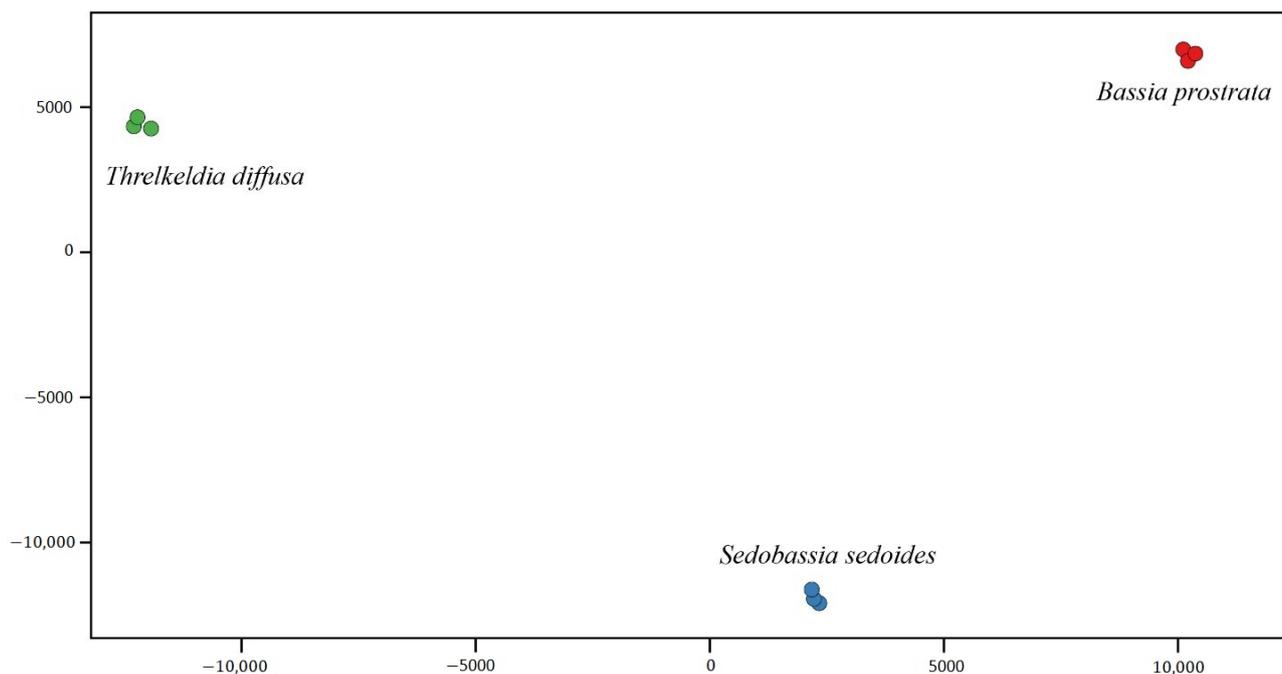
## 2. Results

### 2.1. Descriptive Statistics of RNA Data and De Novo Assembly

Between 27.4 and 37.9 million reads remained after quality filtering (98.68–99.06%) and were de novo assembled for each of the three species (Supplementary Dataset S1). Reduction in contigs by clustering resulted in 26,842 (*Sed. sedoides*, C<sub>2</sub>), 33,1653 (*T. diffusa*, C<sub>3</sub>), and 34,278 (*B. prostrata*, C<sub>4</sub>) contigs with an open reading frame. The number of BUSCOs genes recovered was 88.1, 88.6 and 90.9% for *T. diffusa*, *B. prostrata*, and *Sed. sedoides*, respectively (Supplementary Table S1).

### 2.2. Differential Expression Genes (DEGs) within *Camphorosmeae*

In total, 10,513 transcripts were expressed in at least six of the eight species and downstream analyses focused on this dataset (Supplementary Dataset S2). Principal component analysis of log<sub>2</sub> transformed read counts (normalised to TPM) showed that replicates of each species in *Camphorosmeae* were very similar, whereas different species were clearly distinct from each other (Figure 1). The first principal component explained 51.47% of the total variation and *Sed. sedoides* (C<sub>2</sub>) was positioned somewhere in between *B. prostrata* (C<sub>4</sub>) and *T. diffusa* (C<sub>3</sub>). This result was similar to what was found in *Salsoleae* [24] as the C<sub>3</sub>–C<sub>4</sub> intermediate species was positioned in between the C<sub>3</sub> and the C<sub>4</sub> species. The second component, explaining 42.51% of total variation, separated *Sed. sedoides* from the other two species.

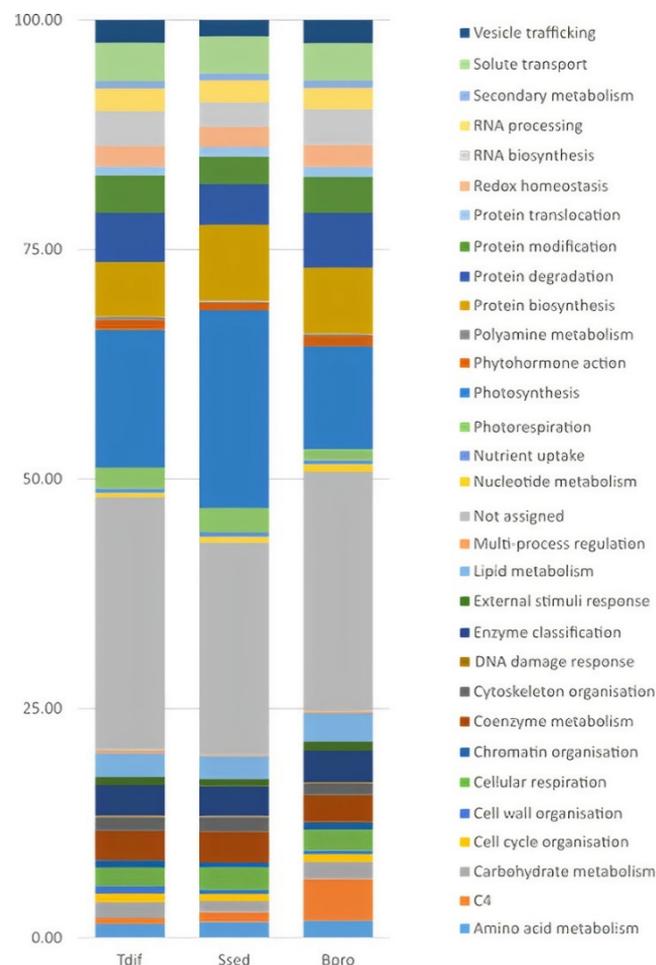


**Figure 1.** Principal component analysis of log<sub>2</sub>-transformed reads. The first (*x*-axis) and the second (*y*-axis) components are shown, which explain 51.47% and 42.51% of the total variation, respectively.

### 2.3. Functional Classification and Enrichment of DEGs within the *Camphorosmeae*

In all three species, transcriptional investment, defined as percentage of all read counts of transcripts (normalised to TPM) belonging to a particular MapMan category, was highest in the MapMan category ‘Not assigned’ (23.10–27.54%) (Figure 2, Supplementary Dataset S3). MapMan category ‘Photosynthesis’ was second highest in all three species; however, the amount differed between species from 11.24% in *B. prostrata* (C<sub>4</sub>), to 15.03% in *T. diffusa* (C<sub>3</sub>), up to 21.58% in *Sed. sedoides* (C<sub>2</sub>) (Figure 2, Supplementary Dataset S3). In general, high transcriptional investment in the category ‘Photosynthesis’ in *Sed. sedoides*

(C<sub>2</sub>) was caused by higher transcription of many genes of the sub-category ‘Calvin cycle’. However, transcription of a gene encoding RUBISCO ACTIVASE in *Arabidopsis* (AT2G39730) was the main driver of the difference among species (*Sed. sedoides* compared with *T. diffusa*: log<sub>2</sub> fold change of 2.09; *Sed. sedoides* compared with *B. prostrata*: log<sub>2</sub> fold change of 3.72; Supplementary Dataset S2). Transcripts belonging to the categories ‘Protein biosynthesis’ (5.93–8.27%), ‘Protein degradation’ (4.45–5.98%), and ‘Protein modification’ (3.03–4.09%) were also highly abundant in all three species. In *B. prostrata*, the category ‘C<sub>4</sub>’ was higher (transcriptional investment of 4.54%) compared with *Sed. sedoides* (1.09%) and *T. diffusa* (0.58%; Figure 2, Supplementary Dataset S3). ‘Photorespiration’ was about twice as high in *Sed. sedoides* and *T. diffusa* (2.59% and 2.21%, respectively) compared with *B. prostrata* (1.14%). This difference was caused by higher transcription of most genes of the category ‘photorespiration’ rather than a few genes (Supplementary Dataset S3). The categories with the lowest transcriptional investment in all three species were ‘DNA damage response’ (0.10–0.14%), ‘Polyamine metabolism’ (0.16–0.29%), and ‘Multi-process regulation’ (0.18–0.34%).



**Figure 2.** Distribution of transcriptional investment defined as the percentage of all transcripts belonging to a particular MapMan4 category [29] and the additional categories ‘C<sub>4</sub>’.

#### 2.4. Differential Expression of C<sub>4</sub>-Related Genes in C<sub>3</sub>, C<sub>2</sub> and C<sub>4</sub> Camphorosmeae Species

As observed in other C<sub>4</sub> species, most genes encoding for proteins involved in C<sub>4</sub> photosynthesis were significantly upregulated in the C<sub>4</sub> species *B. prostrata* (C<sub>4</sub>) when compared with *Sed. sedoides* (C<sub>2</sub>) and *T. diffusa* (C<sub>3</sub>) (Table 1, Supplementary Dataset S2–S4). Out of 18 C<sub>4</sub>-related transcripts, 16 and 13 transcripts encoding known C<sub>4</sub> cycle proteins were significantly upregulated ( $p < 0.001$ ) in *B. prostrata* (C<sub>4</sub>) compared with *T. diffusa* (C<sub>3</sub>) and

*Sed. sedoides* (C<sub>2</sub>), respectively. ALANINE AMINOTRANSFERASE (AlaAT, Log<sub>2</sub>FC = 5.17) was the most abundant followed by PYRUVATE ORTHOPHOSPHATE DIKINASE (PPdK, Log<sub>2</sub>FC = 4.63), PHOSPHOENOLPYRUVATE CARBOXYLASE (PEPC, Log<sub>2</sub>FC = 4.12), BILE ACID:SODIUM SYMPORTER FAMILY PROTEIN 2 (BASS2, Log<sub>2</sub>FC = 3.97), NADP-malic enzyme (NADP-ME Log<sub>2</sub>FC = 3.86), PEP/phosphate translocator (PPT, Log<sub>2</sub>FC = 3.61) in *B. prostrata* (C<sub>4</sub>) as compared with *T. diffusa* (C<sub>3</sub>). Conversely, BASS2 Log<sub>2</sub>FC = 3.51, PPdK Log<sub>2</sub>FC = 3.37, PHOSPHATE TRANSPORTER 1 (PHT1, Log<sub>2</sub>FC = 3.16), PEPC (Log<sub>2</sub>FC = 2.95), and AlaAT (Log<sub>2</sub>FC = 2.61) were in the top five of highly upregulated genes in *B. prostrata* (C<sub>4</sub>) when compared with *Sed. sedoides* (C<sub>2</sub>). However, not all C<sub>4</sub>-related genes were significantly upregulated in *B. prostrata* (C<sub>4</sub>) as compared with *Sed. sedoides* (C<sub>2</sub>). TRIOSE PHOSPHATE TRANSLOCATOR transporters [TPT (Bv8\_194450\_rkme.t1), a chloroplast dicarboxylate transporter isoform [DIT (Bv4\_072630\_xjai.t1)] and a CARBONIC ANHYDRASE isoform [CA (Bv8\_194450\_rkme.t1)] were significantly upregulated in *Sed. sedoides* (C<sub>2</sub>) when compared with *B. prostrata* (C<sub>4</sub>) (Supplementary Dataset S2). Interestingly, PHT1 (Bv3\_049110\_qgnh.t1) was significantly upregulated in *T. diffusa* (C<sub>3</sub>) as compared with *Sed. sedoides* (C<sub>2</sub>) (Supplementary Dataset S2).

**Table 1.** Differential expression of C<sub>4</sub>-related enzymes in leaves of *Camphorosmeae* species. *T. diffusa* (C<sub>3</sub>), *Sed. sedoides* (C<sub>2</sub>), *B. prostrata* (C<sub>4</sub>).

Transcripts	Genes	C <sub>2</sub> vs. C <sub>4</sub> *		C <sub>3</sub> vs. C <sub>4</sub> *		C <sub>3</sub> vs. C <sub>2</sub> *	
		Log <sub>2</sub> FC	Padj	Log <sub>2</sub> FC	P	Log <sub>2</sub> FC	Padj
Bv_006710_gkqg.t1	MDH	1.29	1.01 × 10 <sup>-10</sup>	2.84	9.92 × 10 <sup>-42</sup>	1.55	1.74 × 10 <sup>-14</sup>
Bv1_004490_tyfq.t1	PHT4	1.57	1.56 × 10 <sup>-11</sup>	2.89	1.72 × 10 <sup>-31</sup>	1.32	7.56 × 10 <sup>-08</sup>
Bv1_013550_fjqs.t1	PPdK	3.37	2.32 × 10 <sup>-54</sup>	4.63	2.78 × 10 <sup>-90</sup>	1.26	8.92 × 10 <sup>-10</sup>
Bv2_031080_twkf.t1	AspAt	1.32	2.84 × 10 <sup>-06</sup>	2.15	1.44 × 10 <sup>-13</sup>	0.82	0.003667
Bv3_049110_qgnh.t1	PHT1	3.16	8.05 × 10 <sup>-08</sup>	-	-	-	-
Bv4_072630_xjai.t1	DIT	-	-	1.30	4.88 × 10 <sup>-06</sup>	2.61	3.46 × 10 <sup>-19</sup>
Bv5_117240_yhsk.t1	PPT	1.94	7.13 × 10 <sup>-13</sup>	3.61	5.61 × 10 <sup>-35</sup>	1.67	2.18 × 10 <sup>-09</sup>
Bv6_135140_uyxu.t1	Asn Synthetase	2.56	5.60 × 10 <sup>-07</sup>	2.42	1.93 × 10 <sup>-06</sup>	-	-
Bv6_148840_uffy.t1	CA	3.80	6.66 × 10 <sup>-38</sup>	3.81	4.08 × 10 <sup>-38</sup>	-	-
Bv7_169130_kwer.t1	AlaAT	2.61	4.89 × 10 <sup>-17</sup>	5.17	4.74 × 10 <sup>-48</sup>	2.56	2.34 × 10 <sup>-15</sup>
Bv8_182550_kstq.t1	TPT	-	-	1.48	1.20 × 10 <sup>-10</sup>	2.29	2.71 × 10 <sup>-22</sup>
Bv8_194450_rkme.t1	CA	-	-	-	-	1.48	0.000263
Bv8_195530_sxjq.t1	BASS2	3.51	5.00 × 10 <sup>-68</sup>	3.97	6.26 × 10 <sup>-83</sup>	0.46	0.018625
Bv8_200290_ujgk.t1	TPT	-	-	3.25	1.25 × 10 <sup>-05</sup>	1.97	0.005513
Bv9_209750_xeaz.t1	PEPC	2.95	3.29 × 10 <sup>-17</sup>	4.12	6.98 × 10 <sup>-29</sup>	1.18	0.000429
Bv9_215520_prze.t1	DIT	1.48	1.23 × 10 <sup>-08</sup>	2.63	5.15 × 10 <sup>-22</sup>	1.16	2.27 × 10 <sup>-05</sup>
Bv9_224000_xpgi.t1	NHD	2.48	1.17 × 10 <sup>-21</sup>	2.68	7.04 × 10 <sup>-25</sup>	-	-
Bv9_224840_zmjw.t1	NADP-ME	1.29	8.44 × 10 <sup>-08</sup>	3.86	9.54 × 10 <sup>-47</sup>	2.56	1.34 × 10 <sup>-23</sup>

- not significantly expressed Padj > 0.05.

Twelve of the C<sub>4</sub>-related enzymes except ASPARAGINE SYNTHETASE (Asn Synthetase), SODIUM:HYDROGEN ANTIPORT (NHD) and a CARBONIC ANHYDRASE (CA) isoform (Bv6\_148840\_uffy.t1) were significantly upregulated in *Sed. sedoides* (C<sub>2</sub>) compared with *T. diffusa* (C<sub>3</sub>). These enzymes include typical C<sub>4</sub> enzymes such as PEPC, NADP-ME, PPdK, PHT4, Ala-AT, and ASPARTATE AMINOTRANSFERASE (Asp-AT), as well as C<sub>4</sub>-associated transporters such as BASS2 and DIT.

#### 2.5. Differential Expression of Key Photorespiration Genes in C<sub>3</sub>, C<sub>2</sub> and C<sub>4</sub> *Camphorosmeae* Species

Out of 14 transcripts associated with photorespiratory enzymes, 12 were annotated and assigned (Table 2). All 12 photorespiratory transcripts were significantly upregulated in *Sed. sedoides* (C<sub>2</sub>) as compared with *B. prostrata* (C<sub>4</sub>), including the core photorespiratory enzymes GLYCINE DECARBOXYLASE (GDC, T-, H-, P-, L-), GLUTAMATE:GLYOXYLATE

AMINOTRANSFERASE (GGT) and SERINE HYDROXYMETHYLTRANSFERASE (SHMT). In *Sed. sedoides* (C<sub>2</sub>), GGT, two SHMTs, GDC-T, GLYCOLATE OXIDASE (GOX), GDC-H, GLYCERATE 3-KINASE (GLYK), PHOSPHOGLYCOLATE PHOSPHATASE (PGP) were significantly upregulated when compared with *T. diffusa* (C<sub>3</sub>). Only one SHMT isoform (Bv6\_143730\_mggd.t1) was significantly upregulated in *B. prostrata* (C<sub>4</sub>) when compared with *Sed. sedoides* (C<sub>2</sub>) (Supplementary Dataset S2). In *T. diffusa* (C<sub>3</sub>), only one gene GDC-P was significantly upregulated as compared with *Sed. sedoides* (C<sub>2</sub>) (Supplementary Dataset S2). All photorespiratory genes except GLYK significantly expressed in *Sed. sedoides* (C<sub>2</sub>), compared with *B. prostrata* (C<sub>4</sub>), were significantly upregulated in *T. diffusa* (C<sub>3</sub>) when compared with *B. prostrata* (C<sub>4</sub>).

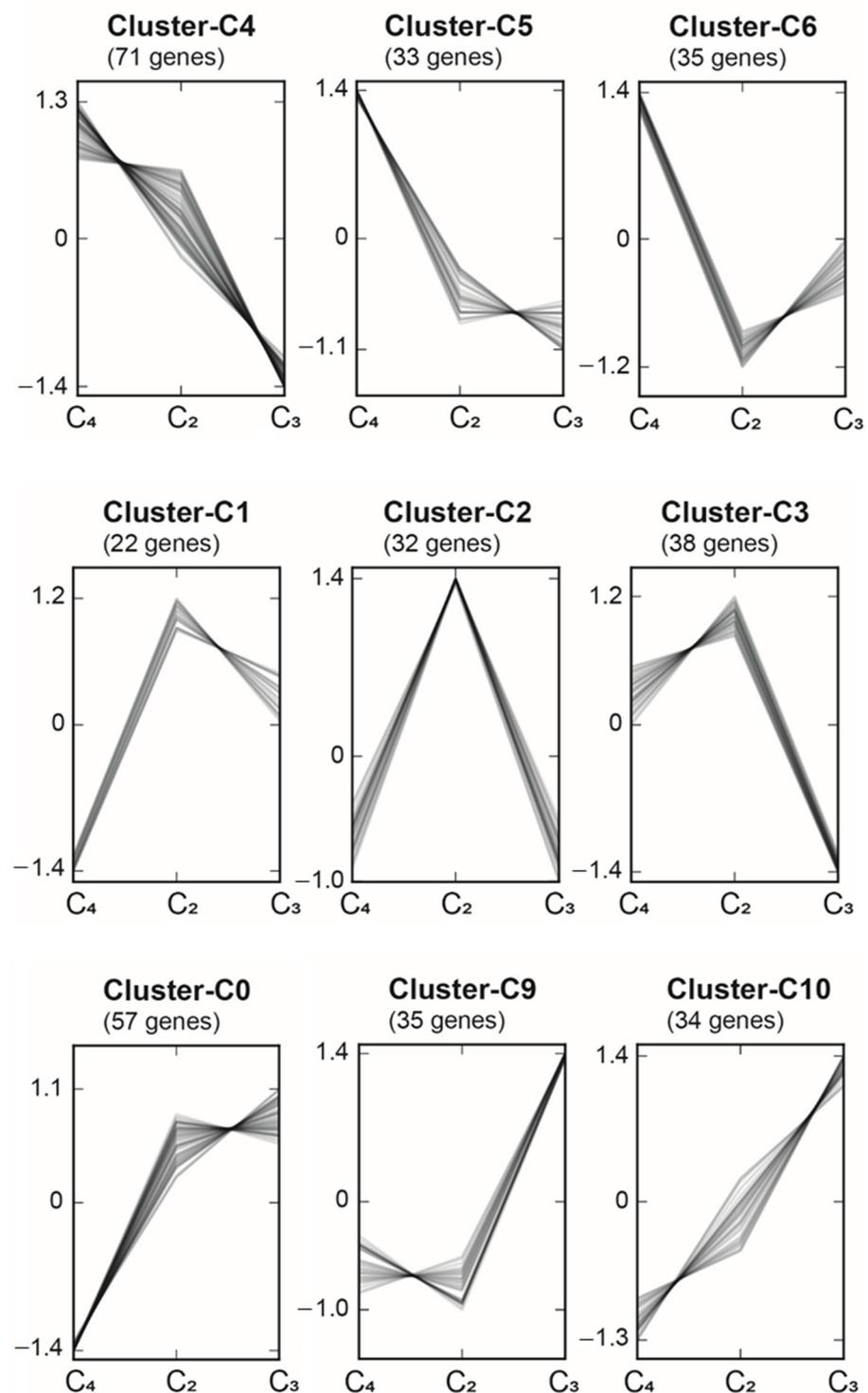
**Table 2.** Differential expression of photorespiratory transcripts in leaves between *Camphorosmeae* species. *T. diffusa* (C<sub>3</sub>), *Sed. sedoides* (C<sub>2</sub>), *B. prostrata* (C<sub>4</sub>).

Transcripts	Genes	C <sub>2</sub> * vs. C <sub>4</sub>		C <sub>3</sub> vs. C <sub>2</sub> *		C <sub>3</sub> * vs. C <sub>4</sub>	
		Log <sub>2</sub> FC	Padj	Log <sub>2</sub> FC	Padj	Log <sub>2</sub> FC	Padj
Bv5_106360_ipey.t1	GDC-H	1.862968	3.37 × 10 <sup>-15</sup>	0.990549	1.80 × 10 <sup>-05</sup>	0.872432	0.000167
Bv_012000_yknj.t1	GDC-P	1.253616	9.20 × 10 <sup>-08</sup>	-	-	2.068131	5.39 × 10 <sup>-18</sup>
Bv3_059720_tshd.t1	GDC-L	1.281024	1.13 × 10 <sup>-11</sup>	-	-	0.953812	3.95 × 10 <sup>-07</sup>
Bv3_065510_eeis.t1	SHMT	-	-	1.133566	0.000228	-	-
Bv4_073470_ismc.t1	AGT/SGT	1.254967	2.18 × 10 <sup>-07</sup>	-	-	1.130439	2.83 × 10 <sup>-06</sup>
Bv4_074740_miaa.t1	PGP	1.831751	1.99 × 10 <sup>-16</sup>	0.505181	0.019216	1.326602	1.77 × 10 <sup>-09</sup>
Bv4_094290_jgpp.t1	GOX	1.708751	1.51 × 10 <sup>-18</sup>	0.993515	1.94 × 10 <sup>-07</sup>	0.715241	0.000175
Bv5_107350_ydma.t1		2.435018	1.25 × 10 <sup>-13</sup>	-	-	2.32744	1.09 × 10 <sup>-12</sup>
Bv6_127540_qdph.t1	GDC-T	2.464812	2.21 × 10 <sup>-23</sup>	1.030248	1.32 × 10 <sup>-05</sup>	1.434602	2.66 × 10 <sup>-09</sup>
Bv6_148110_nuir.t1	GGT	2.380809	3.21 × 10 <sup>-21</sup>	1.223501	4.63 × 10 <sup>-07</sup>	1.157324	2.02 × 10 <sup>-06</sup>
Bv6_152820_wtfn.t1	SHMT	1.721992	2.64 × 10 <sup>-18</sup>	0.734934	0.000137	0.987067	3.75 × 10 <sup>-07</sup>
Bv8_184280_guso.t1		0.677745	0.006074	-	-	1.06935	3.85 × 10 <sup>-06</sup>
Bv9_213980_zwen.t1	HPR	0.96884	0.000282	-	-	0.531187	0.045313
Bv9_220360_xogt.t2	GLYK	0.689722	0.001207	0.708265	0.000885	-	-

- not significantly expressed Padj > 0.05.

## 2.6. Regulatory Elements in C<sub>3</sub>, C<sub>2</sub>, and C<sub>4</sub> Species of the Amaranthaceae/Chenopodiaceae Alliance

To identify regulatory genes putatively involved in the formation/regulation of C<sub>2</sub> and C<sub>4</sub> photosynthesis, expression patterns of TFs from The Plant Transcription Factor Database v.5.0 (PlantTFDB; [30,31]) using 1163 annotated TFs from *Beta vulgaris* (version 'BeetSet-2', [32]) were investigated. From the whole set of TFs included in PlantTFDB, 824 orthologous TFs were found, of which 494 TFs had orthologs in all eight species (three species from *Camphorosmeae*: *B. prostrata* (C<sub>4</sub>), *Sed. sedoides* (C<sub>2</sub>), *T. diffusa* (C<sub>3</sub>); five species from *Salsoleae*: *H. soparia* (C<sub>4</sub>), *Sal. divaricata* Pop-184 (C<sub>2</sub>), *Sal. divaricata* Pop-198 (C<sub>2</sub>), *Sal. oppositifolia* (C<sub>4</sub>), *Sal. soda* (C<sub>3</sub>/C<sub>4</sub>), *Sal. webbii* (C<sub>3</sub>)) and were further analysed using Clust. Based on the expression pattern, from the initial 494 TFs, 431 TFs were grouped into 11 clusters with between 22 and 71 genes per cluster (Cluster-C0 to Cluster-C10; Figure 3, Supplementary Figure S1). Nine of the 11 clusters were of particular interest, because these clusters included TFs that were highly expressed in each photosynthetic type when compared with others (Figure 3). Cluster-C4, Cluster-C5, and Cluster-C6 contained TFs that were highly expressed in C<sub>4</sub> species as compared with C<sub>3</sub> and C<sub>2</sub> species, whereas Cluster-C0, Cluster-C9, and Cluster-C10 comprised TFs that were highly expressed in C<sub>3</sub> species when compared with C<sub>2</sub> and C<sub>4</sub> species. Finally, Cluster-C1, Cluster-C2, Cluster-C3 encompassed TFs that were highly expressed in C<sub>2</sub> species as compared with C<sub>3</sub> and C<sub>4</sub> species.



**Figure 3.** Co-expressed gene clusters were generated using Clust v.1.10.7 [33]. The eight species were grouped into three different conditions ‘C<sub>4</sub> species’, ‘C<sub>2</sub> species’, and ‘C<sub>3</sub> species’. The 11 clusters contained between 22 and 71 genes.

Cluster-C4, Cluster-C5, and Cluster-C6 consisted of 71, 33 and 35 TFs, respectively, from 41 different TF families (Figure 3, Supplementary Dataset S5). Four TFs, all part of Cluster-C4, were present in all eight species, with the transcripts significantly (adjusted  $p$ -value  $\leq 0.01$ ) more abundant in all C<sub>4</sub> species compared with the two C<sub>3</sub> species (Table 3). These TFs comprised BBX15 (CO-like family, AT1G25440.1), SHR (TF family GRAS), SCZ (TF family HSF), and LBD41 (TF family LBD). Cluster-C0, Cluster-C9, and Cluster-C10,

respectively, comprised 57, 35, and 34 TFs (Figure 3, Supplementary Dataset S5). Here, two TFs of Cluster-10 (HSF, and NAC) and one of Cluster-9 (HD-ZIP) were significantly abundant in the studied C<sub>3</sub> species when compared with C<sub>4</sub> species (Table 4).

**Table 3.** Differentially expressed C<sub>4</sub>-related TFs in *Amaranthaceae/Chenopodiaceae*.

Lineage	Species (C <sub>3</sub> vs. C <sub>4</sub> *)	Cluster-4							
		BBX15 (CO-Like)		SHR (GRAS)		SCZ (HSF)		LBD41 (LBD)	
		LF <sub>2</sub> C	Padj						
Salsoleae	<i>Salweb</i> vs. <i>Hsco</i>	8.8	3.84 × 10 <sup>-24</sup>	2.23	1.55 × 10 <sup>-08</sup>	5.59	1.73 × 10 <sup>-05</sup>	3.71	6.05 × 10 <sup>-13</sup>
	<i>Salweb</i> vs. <i>Salopp</i>	11.86	6.56 × 10 <sup>-61</sup>	1.95	2.56 × 10 <sup>-06</sup>	5.09	0.000374	4.26	1.64 × 10 <sup>-17</sup>
	<i>Salweb</i> vs. <i>Salsod</i>	11.57	6.71 × 10 <sup>-62</sup>	2.63	1.85 × 10 <sup>-13</sup>	7.66	4.94 × 10 <sup>-20</sup>	2.84	2.06 × 10 <sup>-07</sup>
Camphorosmeae	<i>Tdif</i> vs. <i>Bpro</i>	10.99	3.77 × 10 <sup>-51</sup>	4.21	9.49 × 10 <sup>-12</sup>	3.4	2.71 × 10 <sup>-06</sup>	6.71	5.11 × 10 <sup>-10</sup>
	<i>Salweb</i> vs. <i>Bpro</i>	10.82	2.50 × 10 <sup>-47</sup>	1.45	0.0011	6.12	1.58 × 10 <sup>-07</sup>	2.22	0.00019
Salso. × Camph.	<i>Tdif</i> vs. <i>Hsco</i>	8.96	2.75 × 10 <sup>-27</sup>	4.99	2.89 × 10 <sup>-19</sup>	2.84	0.000272	8.23	1.45 × 10 <sup>-21</sup>
	<i>Tdif</i> vs. <i>Salsod</i>	11.72	1.85 × 10 <sup>-63</sup>	5.34	5.60 × 10 <sup>-25</sup>	5.25	3.76 × 10 <sup>-21</sup>	6.96	1.91 × 10 <sup>-12</sup>
	<i>Tdif</i> vs. <i>Salopp</i>	12.03	6.61 × 10 <sup>-65</sup>	4.7	1.56 × 10 <sup>-15</sup>	2.44	0.00602	8.68	3.84 × 10 <sup>-26</sup>

**Table 4.** Differentially expressed C<sub>3</sub>-related TFs in *Amaranthaceae/Chenopodiaceae*.

Lineage	Species (C <sub>4</sub> vs. C <sub>3</sub> *)	Cluster-9				Cluster-10		
		ATHB13 (HD-ZIP)		HSFA6B (HSF)		NAC083 (NAC)		
		LF <sub>2</sub> C	Padj	LF <sub>2</sub> C	Padj	LF <sub>2</sub> C	Padj	Padj
Salsoleae	<i>Hsco</i> vs. <i>Salweb</i>	9.02	5.45 × 10 <sup>-24</sup>	1.21	0.00017	0.73		0.00398
	<i>Salopp</i> vs. <i>Salweb</i>	1.71	2.77 × 10 <sup>-05</sup>	1.54	1.32 × 10 <sup>-05</sup>	1.42		4.44 × 10 <sup>-08</sup>
	<i>Salsod</i> vs. <i>Salweb</i>	1.14	0.00077	1.43	1.65 × 10 <sup>-06</sup>	0.76		0.0004
Camphorosmeae	<i>Bpro</i> vs. <i>Tdif</i>	8.89	2.80 × 10 <sup>-23</sup>	1.7	1.08 × 10 <sup>-06</sup>	11.74		3.28 × 10 <sup>-82</sup>
	<i>Bpro</i> vs. <i>Salweb</i>	8.99	8.45 × 10 <sup>-24</sup>	1.91	5.76 × 10 <sup>-08</sup>	11.28		1.31 × 10 <sup>-72</sup>
Salso. × Camph.	<i>Hsco</i> vs. <i>Tdif</i>	8.92	1.78 × 10 <sup>-23</sup>	1	0.00169	1.01		5.11 × 10 <sup>-05</sup>
	<i>Salsod</i> vs. <i>Tdif</i>	0.87	0.00956	1.23	4.09 × 10 <sup>-05</sup>	1.04		6.71 × 10 <sup>-07</sup>
	<i>Salopp</i> vs. <i>Tdif</i>	1.45	0.00038	1.34	0.00011	1.71		4.00 × 10 <sup>-11</sup>

Cluster-C1, Cluster-C2, and Cluster-C3 included 22, 32, and 38 TFs, respectively, of which one TF of Cluster-2 bHLH 106 (TF family bHLH) was significantly higher in the C<sub>2</sub> species when compared with the C<sub>3</sub> species and C<sub>4</sub> species (Table 5, Supplementary Dataset S2–S5). To assess the integration of C<sub>3</sub> and C<sub>4</sub> pathways into the intermediate C<sub>2</sub> pathway at the regulation level and vice versa, specific TFs of C<sub>4</sub> (Cluster-C4, Cluster-C5 and Cluster-C6) and C<sub>3</sub> (Cluster-C0, Cluster-C9, and Cluster-C10) pathways were assessed in the following comparisons: C<sub>2</sub> species vs. C<sub>3</sub> species and C<sub>2</sub> species vs. C<sub>4</sub> species. Then, specific TFs of the C<sub>2</sub> pathway (Cluster-C1, Cluster-C2 and Cluster-C3) were estimated in the pairwise comparison of C<sub>3</sub> species vs. C<sub>4</sub> species. Among the four TFs common to all C<sub>4</sub> species, only one TF (BBX15, TF family CO-like) was significantly upregulated in C<sub>2</sub> species when compared with C<sub>3</sub> species (Supplementary Dataset S2–S5). Conversely, no TF of C<sub>3</sub> species was highly expressed in C<sub>2</sub> species when compared with C<sub>4</sub> species.

**Table 5.** Differentially expressed C<sub>2</sub>-related TF in *Amaranthaceae/Chenopodiaceae*.

Lineage	Species (C <sub>3</sub> vs. C <sub>2</sub> *)	Cluster-2	
		LF <sub>2</sub> C	Padj
<i>Salsoleae</i>	<i>Sweb</i> vs. <i>Saldi1</i>	1.68	$8.90 \times 10^{-13}$
	<i>Sweb</i> vs. <i>Saldi2</i>	1.63	$4.07 \times 10^{-11}$
<i>Camphorosmeae</i>	<i>Tdif</i> vs. <i>Sedsed</i>	0.45	0.04616
<i>Salso.</i> × <i>Camph.</i>	<i>Tdif</i> vs. <i>Sdi1</i>	1.06	$3.33 \times 10^{-07}$
	<i>Tdif</i> vs. <i>Sdi2</i>	1.01	$2.79 \times 10^{-06}$
	<i>Sweb</i> vs. <i>Sedsed</i>	1.07	$4.19 \times 10^{-05}$
	Species (C <sub>4</sub> vs. C <sub>2</sub> *)	LF <sub>2</sub> C	Padj
<i>Salsoleae</i>	<i>Hsco</i> vs. <i>Saldi1</i>	1.37	$1.23 \times 10^{-10}$
	<i>Hsco</i> vs. <i>Saldi 2</i>	1.32	$3.58 \times 10^{-09}$
	<i>Salopp</i> vs. <i>Saldi1</i>	2.17	$7.37 \times 10^{-19}$
	<i>Salopp</i> vs. <i>Saldi2</i>	2.12	$1.03 \times 10^{-16}$
	<i>Salsod</i> vs. <i>Saldi1</i>	1.21	$3.24 \times 10^{-10}$
	<i>Salsod</i> vs. <i>Saldi2</i>	1.16	$7.30 \times 10^{-09}$
<i>Camphorosmeae</i>	<i>Bpro</i> vs. <i>Sedsed</i>	0.56	0.01398
<i>Salso.</i> × <i>Camph.</i>	<i>Bpro</i> vs. <i>Saldi1</i>	1.17	$2.14 \times 10^{-08}$
	<i>Bpro</i> vs. <i>Saldi2</i>	1.12	$2.21 \times 10^{-07}$
	<i>Hsco</i> vs. <i>Sedsed</i>	0.76	0.001176
	<i>Salopp</i> vs. <i>Sedsed</i>	1.56	$9.00 \times 10^{-09}$
	<i>Salsod</i> vs. <i>Sedsed</i>	0.6	0.00483

### 3. Discussion

#### 3.1. Transcriptome Analysis in *Camphorosmeae*

Gene expression analysis predominantly paved the way to understand the difference between derived photosynthetic types (C<sub>2</sub>, C<sub>4</sub>) and the ancestral C<sub>3</sub> photosynthesis [24,34,35]. Much progress in understanding C<sub>4</sub> and C<sub>2</sub> photosynthesis was achieved by comparing differentially expressed genes of closely related species in the genus *Flaveria* (*Asteraceae*) considered as a model organism to study the evolution of C<sub>4</sub> photosynthesis [6,7,34–37]. The goosefoot family (*Chenopodiaceae*) has a large number of C<sub>2</sub> and C<sub>4</sub> species that differ anatomically and ecologically from *Flaveria*. This family therefore represents a good supplementary alternative to decipher the convergent evolution of C<sub>4</sub> photosynthesis. With PCA based on gene expression, it was possible to clearly distinguish between *T. diffusa* (C<sub>3</sub>), representing the ancestral condition, and *Sed. sedoides* (C<sub>2</sub>) and *B. prostrata* (C<sub>4</sub>), representing derived conditions. The physiologically C<sub>3</sub>-C<sub>4</sub> intermediate *Sed. sedoides* (C<sub>2</sub>) was positioned in a triangle with *T. diffusa* (C<sub>3</sub>) and *B. prostrata* (C<sub>4</sub>) in terms of transcript variation. This result was similar to what was found in *Salsoleae* [24]. The first three components explained about 76% of the total variation, which was slightly higher than the 73% reported in *Salsoleae* [24]. Similar to *Salsoleae*, in *Camphorosmeae*, the three different photosynthesis types predominantly structure the gene expression pattern in assimilating tissue. Indeed, species with C<sub>3</sub>, C<sub>2</sub>, and C<sub>4</sub> photosynthesis differ in leaf anatomical structure. *T. diffusa* (C<sub>3</sub>) exhibits the *Neokochia* type characterised by an undifferentiated chlorenchyma of several layers. *Sed. sedoides* (C<sub>2</sub>) has the *Sedobassia* type consisting of kranz-like cells near peripheral vascular bundles. *B. prostrata* (C<sub>4</sub>) depicts the *Bassia prostrata* type with the chlorenchyma differentiated in an outer mesophyll and inner kranz-layer [20]. In contrast, the first three PCA components in a comparable study of *Flaveria* explained only 27% [35]. This difference could be due to the younger evolutionary age or other confounding factors affecting gene expression in the *Flaveria* study as suggested by Lauterbach et al. [24].

### 3.2. C<sub>4</sub> Key Enzymes in C<sub>4</sub> and C<sub>2</sub> *Camphorosmeae* Species

Analyses of differential gene expression between C<sub>3</sub> and C<sub>4</sub> species of *Camphorosmeae* showed that core C<sub>4</sub> cycle proteins were highly abundant in *B. prostrata* (C<sub>4</sub>). Similar results were found in *Cleome* [38], *Flaveria* [34,35] and *Salsoleae* [24]. Traditionally, three biochemical subtypes of C<sub>4</sub> photosynthesis are classified according to the predominant type of decarboxylation releasing CO<sub>2</sub> around RUBisCo in the BSCs: NAD-ME, NADP-ME, and PEP-CK. However, PEPCK should be considered as a supplemental subtype to either NAD-ME or NADP-ME [39]. Significant expression of NADP-ME indicates that *B. prostrata* (C<sub>4</sub>) uses a NADP-ME type C<sub>4</sub> cycle. Asparagine synthetase (ASN) and NHD were found significantly expressed and upregulated only in *B. prostrata* (C<sub>4</sub>) as compared with *T. diffusa* (C<sub>3</sub>) and *Sed. sedoides* (C<sub>2</sub>). ASN was reported upregulated in C<sub>4</sub> species *Gynandropsis gynandra* when compared with closely related C<sub>3</sub> species *Tarenaya hassleriana* (*Cleomaceae*), as well as in C<sub>4</sub> leaves of *Sal. soda* when compared with its C<sub>3</sub> cotyledones [23]. On the other hand, NHD was found upregulated in C<sub>4</sub> species compared with C<sub>3</sub> and C<sub>2</sub> species of *Flaveria* [35]. Moreover, the top three highly expressed C<sub>4</sub> enzymes in *B. prostrata* (C<sub>4</sub>) as compared with *T. diffusa* (C<sub>3</sub>) were Ala-AT, PPDK, and BASS2. ASN is involved in ammonium metabolism and asparagine in nitrogen transport [24]. Achievement of the C<sub>4</sub> cycle requires the transport of pyruvate to the mesophyll cell (MC) for regeneration of PEP. While Ala-AT plays an important role in pyruvate generation, PPDK intervenes in the regeneration of PEP. Pyruvate transport is mediated by the BASS2/NHD transport system [40]. Taken all together, this indicates not only a possible functional connection between nitrogen metabolism and the switch from C<sub>3</sub> to C<sub>4</sub> pathway as suggested by Lauterbach et al. [24] and Mallmann et al. [35], but also the capacity to shuttle pyruvate from the BS plastid. In this regard, the pyruvate shuttle ensures the regeneration of the CO<sub>2</sub> acceptor (PEP), and therefore maintains the C<sub>4</sub> pathway.

Eleven C<sub>4</sub>-related genes were found significantly upregulated in *Sed. sedoides* (C<sub>2</sub>) compared with *T. diffusa* (C<sub>3</sub>), including, for example, PEPC, NADP-ME, PPdK, and PHT4. Upregulation of C<sub>4</sub> typical enzymes such as PEPC, NADP-ME, PPdK, PPT was also reported in the C<sub>2</sub> species when compared with the C<sub>3</sub> species in studies of *Flaveria* and *Salsoleae* [24,35]. This result suggests that genes associated with the C<sub>4</sub> cycle are present in *Sed. sedoides* (C<sub>2</sub>) and play an important role in C<sub>2</sub> metabolism. A DIT isoform (Bv4\_072630\_xjai.t1), NADP-ME and Ala-AT were the three most upregulated C<sub>4</sub> enzymes in *Sed. sedoides* (C<sub>2</sub>) as compared with *T. diffusa* (C<sub>3</sub>). Moreover, we found two transporters (TPT and DIT) upregulated in *Sed. sedoides* (C<sub>2</sub>) when compared with *B. prostrata* (C<sub>4</sub>). These transporters were found highly expressed in some C<sub>2</sub> species when compared with C<sub>4</sub> species in *Flaveria* [35]. DIT is a putative plastidial dicarboxylate transporter and TPT is the chloroplast envelope triose-phosphate/phosphate translocator (TPT) [41]. Based on simulated data, it was shown that a high TPT capacity is required to obtain high assimilation rates and to decrease the CO<sub>2</sub> leakage from BSCs to MCs [39]. The most likely reason for upregulation of these genes is their involvement in decreasing the CO<sub>2</sub> leakage from the Kranz-like cells back to the MCs due to the presence of RuBisCo, which is not the case for C<sub>4</sub> plants. This explains the low CO<sub>2</sub> compensation observed in C<sub>2</sub> species [42]. Thus, C<sub>2</sub> plants upregulate a distinct set of C<sub>4</sub> enzymes to handle constraints related to the C<sub>2</sub> pathway and not an entirely congruent set. This does not support their interpretation as an intermediate state towards C<sub>4</sub> photosynthesis, but is more in line with their interpretation as an independent evolutionarily stable state ([43] and refs. therein).

### 3.3. C<sub>4</sub> Key Enzymes in C<sub>4</sub> and C<sub>2</sub> *Camphorosmeae* Species

Transcripts associated with photorespiration were about twice as abundant in C<sub>3</sub>-C<sub>4</sub> intermediate (*Sed. sedoides*) and *T. diffusa* (C<sub>3</sub>) compared with *B. prostrata* (C<sub>4</sub>). Likewise, we found key photorespiration enzymes were differentially expressed and upregulated in the C<sub>2</sub> species (*Sed. sedoides*) when compared with *T. diffusa* (C<sub>3</sub>) and *B. prostrata* (C<sub>4</sub>). This corroborates the expression patterns reported in the C<sub>2</sub> species of *Flaveria* [34,35] and *Salsoleae* [24], implying a successful integration of C<sub>2</sub> photosynthesis in *Sed. sedoides* (C<sub>2</sub>).

Our transcript data showed that GDC-P and a SHMT isoform (Bv6\_143730\_mggd.t1) were downregulated in *Sed. sedoides* (C<sub>2</sub>) when compared with *T. diffusa* (C<sub>3</sub>) and *B. prostrata* (C<sub>4</sub>), respectively. Similar results were obtained in the C<sub>2</sub> species of the genus *Flaveria* [35]. Schulze et al. [36] showed that downregulation of GDC-P was closely linked to the establishment of the C<sub>2</sub> pathway in *Flaveria*. Since GDC-P and a SHMT isoform are known to be involved in glycine decarboxylation, their downregulation in *Sed. sedoides* (C<sub>2</sub>) might have similar consequences. It is worth noticing that the number of significantly upregulated photorespiratory genes in *Sed. sedoides* (C<sub>2</sub>) was equal to *T. diffusa* (C<sub>3</sub>) when compared with *B. prostrata* (C<sub>4</sub>).

A significant reduction in almost all photorespiratory genes was observed in *B. prostrata* (C<sub>4</sub>). All photorespiratory genes except GLYK were downregulated in *B. prostrata* (C<sub>4</sub>) as compared with *T. diffusa* (C<sub>3</sub>). Mallmann et al. [35] reported significant downregulation of all photorespiratory genes in C<sub>4</sub> *Flaveria* except the transport proteins DIT1 and DIT2 and one isoform of GLDH. On the other hand, GLYK was expressed in the M of C<sub>4</sub> *Sorghum bicolor* [44]. GLYK catalyses the regeneration of 3-phosphoglycerate (3-PG). The localisation of GLYK within the leaf cells of *B. prostrata* (C<sub>4</sub>) could clarify its high expression and role.

### 3.4. Regulation of C<sub>3</sub>, C<sub>2</sub> and C<sub>4</sub> Photosynthesis in Amaranthaceae/Chenopodiaceae

Transcription factors are proteins that bind to the DNA promoter or enhancer regions of specific genes and regulate their expression. They have a crucial role on plant growth, development and adaptation under various stress conditions, and therefore are excellent candidates for modifying complex traits in plants [45]. C<sub>3</sub>, C<sub>2</sub> and C<sub>4</sub> species of *Salsola* and *Camphorosmeae* are widely spread in desert, semi-desert, saline, and arid regions [18,19]. In former *Chenopodiaceae*, C<sub>4</sub> photosynthesis evolved as an adaptation to hot, dry, or saline areas from the C<sub>3</sub> ancestor which was already preadapted to grow in these harsh environments [15]. We focused on TFs that were differentially expressed between C<sub>3</sub>, C<sub>2</sub>, and C<sub>4</sub> species/states irrespective of the lineage, to further reduce the amount of differentially expressed TFs to a small subset of actually C<sub>4</sub>-, C<sub>2</sub>-, and C<sub>3</sub>-related changes. Indeed, a small number of TFs were found differentially expressed between C<sub>3</sub>, C<sub>2</sub>, and C<sub>4</sub> species/states.

Cluster analysis showed that BBX15, SHR, SCZ, and LBD41 were co-regulated and significantly more abundant in all C<sub>4</sub> species irrespective of the lineage when compared with C<sub>3</sub> species. The families to which these TF families belong play an important role in regulatory networks controlling plant growth and development, and plant adaptive responses to various environmental stress conditions [46–50]. Except the LBD TF family, the SHR, HSF, and CO-like families have been shown to be involved in the development of C<sub>4</sub> Kranz anatomy in *Zea mays* L. and potentially involved in the establishment of C<sub>4</sub> M and Kranz cell identities [51–55]. However, members of the LBD TF family are key regulators of plant organ development, leaf development, pollen development, plant regeneration, stress response, and anthocyanin and nitrogen metabolisms [50,56]. Since mRNA of all the four TFs was highly abundant in C<sub>4</sub> species and co-regulated, our data suggest a critical role of these TFs in the development of any C<sub>4</sub> Kranz anatomy in the *Amaranthaceae/Chenopodiaceae* complex.

We found that C<sub>3</sub> species enhanced different TFs compared with C<sub>4</sub> species. Three TFs (ATHB13, HD-ZIP family), HSF6B (HSF TF family), and NAC083 (NAC TF family), in which two TFs (HSF6B and NAC083) are co-regulated, were significantly higher in all C<sub>3</sub> species when compared with C<sub>4</sub> species. As C<sub>4</sub> TFs, they are involved in plant growth, development, and stress tolerance. The NAC TF family was shown to contribute to root and shoot apical meristems formation in *Arabidopsis* [57,58], organogenesis [59], salt and drought tolerance in *Arabidopsis* [60], leaf senescence in tobacco [61], and secondary cell wall formation in cotton [62]. The HD-Zip TF family was reported to regulate plant growth adaptation to abiotic stress such as salt and drought in apple and *Arabidopsis* [63,64]. Interestingly, HD-ZIP, HSF, and NAC TF families were suggested to control the C<sub>4</sub> photosynthesis in maize and rice [55,65]. However, in these studies, these TF families were

detected using development gradient transcriptome comparison only on C<sub>4</sub> maize and rice plants. This may imply higher activity of these TF families in C<sub>3</sub> species. Nevertheless, different transcripts of these TFs families were involved when compared with the present study. Thus, a significant expression of these TFs in C<sub>3</sub> species could indicate a potential function of these TFs in the C<sub>3</sub> pathway.

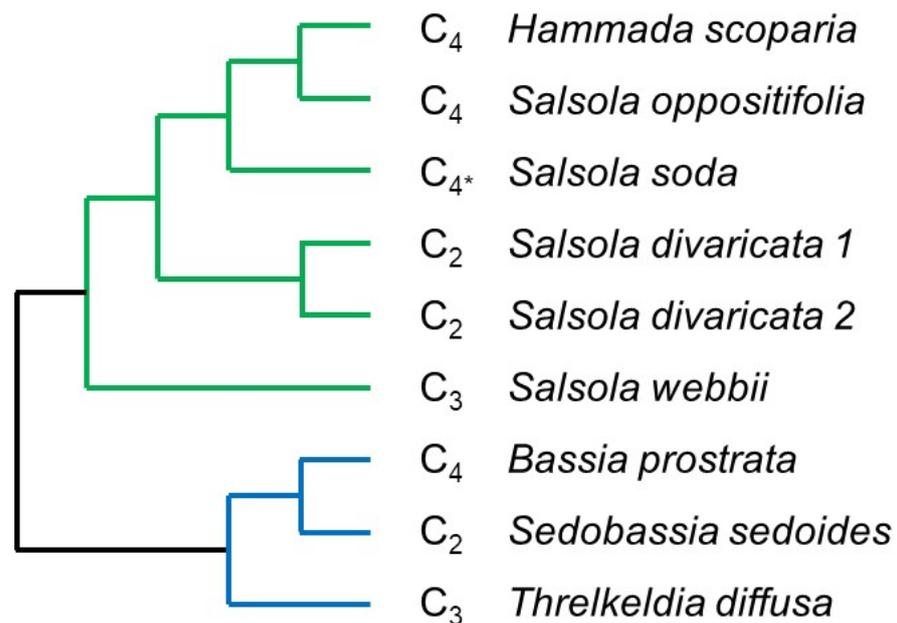
In C<sub>2</sub> species, one transcript of the BASIC HELIX-LOOP-HELIX (bHLH106) protein from the bHLH TF family was found to be upregulated compared with C<sub>4</sub> and C<sub>3</sub> species. Two TFs of the bHLH TF family were shown to regulate a C<sub>4</sub> photosynthesis gene in maize [66]. This upregulation of bHLH106 in all C<sub>2</sub> species may suggest its possible role in the development and establishment of the C<sub>2</sub> photosynthesis specificities relative to other photosynthesis types. Interestingly, one C<sub>4</sub>-specific TF (BBX15, TF family CO-like) was significantly higher in C<sub>2</sub> species when compared with C<sub>3</sub> species. Thus, this TF could be responsible for similarities of C<sub>4</sub> photosynthesis found in C<sub>2</sub> species such as the Kranz-like anatomy. Surprisingly, no C<sub>3</sub>-specific TF was significantly expressed in C<sub>2</sub> species when compared with C<sub>4</sub> species. This indicates that C<sub>2</sub> and C<sub>4</sub> photosynthesis represent more derived types of photosynthesis compared with C<sub>3</sub> photosynthesis. Nonetheless, this seems to be inconsistent with the current model of C<sub>4</sub> evolution which relies heavily on the interpretation of the physiological intermediacy of C<sub>2</sub> photosynthesis as an evolutionary stepping stone to C<sub>4</sub> [8]. One would expect C<sub>3</sub>-specific TFs to be higher in C<sub>2</sub> species when compared with C<sub>4</sub> species if the C<sub>2</sub> photosynthetic type represents an intermediate step along the evolution of C<sub>3</sub>-to-C<sub>4</sub> photosynthesis as revealed by differential expression analysis of core photorespiratory genes in C<sub>2</sub> and C<sub>3</sub> species of *Camphorosmeae* (this study) and *Salsoleae* [24].

Taking the results of this study together, the unique derived TF profile of the C<sub>2</sub> intermediate species suggests an evolutionarily stable state in its own right. Similarities with C<sub>4</sub> relatives might result from a hybrid origin involving C<sub>3</sub> and C<sub>4</sub> parental lineages, parallel recruitment of a number of TFs in C<sub>4</sub> and C<sub>2</sub> lineages or common ancestry, and later divergent evolution. The position of *Sedobassia* as sister to *Bassia* (all C<sub>4</sub>) allows all of these three scenarios [16]. For C<sub>2</sub> species in *Salsoleae*, however, phylogenomic evidence points to a hybrid origin of the *Sal. divaricata* agg. (Tefarikis et al., in prep.). Further phylogenomic analyses are needed to discern if an early hybridisation event of a C<sub>4</sub> (or ancestral preadapted C<sub>4</sub>) lineage and a C<sub>3</sub> lineage led to the origin of the *Sedobassia* lineage which then evolved towards stable C<sub>2</sub> photosynthesis.

## 4. Materials and Methods

### 4.1. Plant Material

Plants of three *Camphorosmeae* species (*Bassia prostrata* (L.) Beck (C<sub>4</sub>), *Sedobassia sedoides* (Schrad.) Freitag and G. Kadereit (C<sub>2</sub>), and *Threlkeldia diffusa* R.Br. (C<sub>3</sub>) (Figure 4, for voucher information see Supplementary Table S2) were grown from seeds in custom mixed potting soil in a glasshouse at the Botanic Garden, Johannes Gutenberg University Mainz, Germany at a minimum temperature of 18 °C in the night. Daytime temperatures varied from 25 to 35 °C in the summer and from 20 to 25 °C in the winter. Plants were watered once a week in the winter and twice a week in the summer and kept at 16 h light/ 10 h dark with natural light and an additional light intensity of ca. 300 μmol m<sup>-2</sup> s<sup>-1</sup>. Leaf samples of the three species were harvested between 16th April and 16th May 2014 between 10:30 a.m. and 13:00 p.m., immediately frozen in liquid nitrogen, and stored at -80 °C for RNA extraction.



**Figure 4.** Phylogenetic relationships between species of the current study. The photosynthetic type is indicated. C<sub>4</sub>\*, species with C<sub>4</sub> photosynthesis in leaves/assimilating shoots but C<sub>3</sub> in cotyledons. Green colour represents *Salsoleae*; blue represents *Camphorosmeae*.

#### 4.2. RNA Isolation and Sequencing

Total RNA extraction, library preparation, and mRNA sequencing were performed as described by Lauterbach et al. [23,24]. Total RNA was extracted from 16–55 mg leaf tissue of *B. prostrata*, *Sed. sedoides*, and *T. diffusa*. Sequencing of 101 bp single-end reads was performed on an Illumina HiSeq2000 platform. For each species, three individuals were sequenced (i.e., biological triplicates). Sequencing reads of these three species are available under study accession PRJEB36559.

#### 4.3. Data Access

RNA-Seq data of the five *Salsoleae* species were retrieved from Lauterbach et al. (2017 a, b; study accession numbers PRJNA321979 and PRJEB22023) (Figure 4, Supplementary Table S2). These data comprise: cotyledons, and first and second leaf pair of *Salsola soda* (C<sub>3</sub>/C<sub>4</sub>), cotyledons and leaves of the *Salsola divaricata* population 184 (C<sub>2</sub>, Pop-184), *Salsola divaricata* population 198 (C<sub>2</sub>, Pop-198), and *Salsola oppositifolia* (C<sub>4</sub>); leaves of *Salsola webbii* (C<sub>3</sub>); and assimilating shoots of *Hammada scoparia* (C<sub>4</sub>). For all of these samples, triplicates per species and organ were available [23,24].

#### 4.4. RNA-Seq Data Processing

Single-end sequencing reads were checked for quality using the FASTQC tool ([www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), accessed on 15 March 2021), and filtered and trimmed using Trimmomatic v.0.38 [67]. For each species, de novo assembly was conducted using quality-filtered reads of all replicates of leaves and, where present, cotyledons of the respective species with default parameters in Trinity v.2.1.1 [68]. Quality of assemblies were assessed with BUSCO v.3.0 (Benchmarking Universal Single-Copy Orthologs, [69]) using the ‘Eudicotyledons odb10’ dataset [70]. Number of contigs of de novo assemblies were reduced by clustering via CD-HIT-EST v.4.7 [71,72] and only contigs with an open reading frame were included in the downstream analysis, which was conducted with TransDecoder v.5.3.0 (github.com/TransDecoder/TransDecoder accessed on 29 February 2020) followed by another round of CD-HIT-EST. Orthology assignment between the nine de novo assemblies was carried out by conditional reciprocal best (crb) BLAST v.0.6.9 [26] run locally using protein-coding sequences of *Beta vulgaris* (version

'BeetSet-2', [32]) as a reference. Only contigs were included in downstream analyses that had ortholog assignments between at least six of the eight species. Besides 'BeetSet-2' from *Beta vulgaris*, contigs were annotated using *Arabidopsis* (TAIR10). Reads of each of the replicates were separately mapped against these reduced data sets via bowtie2 v.2.3.4.1 [73]. Re-formatting and final extraction of read counts (excluding supplementary alignments) were carried out in Samtools v.1.3 [74].

#### 4.5. Differential Gene Expression Analysis

Read counts were normalised into transcripts per million (TPM) and used for differential gene expression analysis. Here, pairwise comparison between all eight species was statistically evaluated using edgeR [75] in R (R Core Team, 2018). Hierarchical clustering using Pearson's correlation and principal component analysis of  $\log_2$  transformed read counts were carried out with Multiexperiment Viewer (MeV) v.4.9 (<http://mev.tm4.org/> accessed on 5 February 2020). Co-expressed gene clusters of (1) all expressed transcripts and (2) transcripts annotated as transcription factors were carried out with Clust v.1.10.7 [33]. Pathways were defined in MapMan4 categories [29] with the additional category 'C<sub>4</sub>'. To identify TFs putatively involved in the formation/regulation of C<sub>2</sub> and C<sub>4</sub> photosynthesis, the 1163 annotated TFs from *Beta vulgaris* (version 'BeetSet-2', [32]) from The Plant Transcription Factor Database v.5.0 (PlantTFDB; [30,31]) were used. Here, two different datasets were combined (1: leaf transcriptome data of the three *Camphorosmeae* species *T. diffusa* (C<sub>3</sub>), *Sed. sedoides* (C<sub>2</sub>), and *B. prostrata* (C<sub>4</sub>); 2: leaf transcriptome data of the five *Salsoleae* species *Sal. webbii* (C<sub>3</sub>), *Sal. divaricata* Pop-184 (C<sub>2</sub>), *Sal. divaricata* Pop-198 (C<sub>2</sub>), *H. scoparia* (C<sub>4</sub>), *Sal. oppositifolia* (C<sub>4</sub>), and *Sal. soda* (C<sub>4</sub>); for study accession numbers see above) and transcribed TFs grouped using Clust v.1.10.7 [33] and grouping all samples based on the photosynthetic type into the three conditions C<sub>3</sub>, C<sub>2</sub>, and C<sub>4</sub>. VENNY v. 2.1 (<https://bioinfogp.cnb.csic.es/tools/venny/> accessed on 12 July 2021) were deployed to find intersected TFs across all pairwise comparisons.

## 5. Conclusions

The transcriptome data of the *Chenopodiaceae* family provided new insight into C<sub>4</sub> evolution. Proteins encoding for C<sub>4</sub> transporters (DIT and TPT) were found significantly upregulated in *Sed. sedoides* (C<sub>2</sub>) when compared with *B. prostrata* (C<sub>4</sub>). Upregulation of those transporters reduces CO<sub>2</sub> leakage from BSCs to MC, which could otherwise be detrimental to C<sub>2</sub> photosynthesis due to the presence of RuBisCo in the MC. This suggests evolution of a stable C<sub>2</sub> photosynthesis independent of C<sub>4</sub> photosynthesis. Combined analysis of TFs of the sister lineages provides further support of this result. Indeed, while one C<sub>4</sub>-specific TF (BBX15) was significantly higher in C<sub>2</sub> species when compared with C<sub>3</sub> species, no C<sub>3</sub>-specific TFs were higher in C<sub>2</sub> species compared with C<sub>4</sub> species. Finally, apart from well-known TFs involved in the development of C<sub>4</sub> Kranz anatomy such as SHR, BBX15, SCZ, and LBD41 may also be associated with its development and physiology. Furthermore, bHLH106 could be related to specific C<sub>2</sub> anatomy and BBX15 to a characteristic C<sub>4</sub>-like expression pattern found in species with C<sub>2</sub> photosynthesis. This study sheds light on the differentiated regulation and evolution of transcription factors in C<sub>2</sub> and C<sub>4</sub> photosynthesis.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijms222212120/s1>, Table S1. Quality assessment of transcriptome de novo assemblies using the Eudicotyledons odb10 dataset in BUSCO v.3.0. Table S2. Voucher information of species used in this study, Dataset S1. Read Mapping statistics of *Camphorosmeae* species, Dataset S2. Differential expression analysis of pairwise comparisons between *Camphorosmeae* species, Dataset S3. Transcriptional investment of *Camphorosmeae* species, Dataset S4. Annotation and normalised transcript count of C<sub>4</sub>-related and photorespiratory genes, Dataset S5. Clustered TFs, Figure S1. All clusters of TFs per photosynthesis types.

**Author Contributions:** Conceptualisation, G.K.; methodology, G.K., M.L. and C.S.; software, M.L. and C.S.; validation, M.L. and C.S.; formal analysis, C.S. and M.L.; investigation, M.L.; resources, G.K.; data curation, C.S. and M.L.; writing—original draft preparation, C.S.; writing—review and editing, G.K., M.L. and C.S.; visualisation, C.S.; supervision, G.K.; project administration, G.K.; funding acquisition, G.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the Deutsche Forschungsgemeinschaft (DFG) with grants to GK (KA1816/7-3).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available upon request.

**Acknowledgments:** We would like to thank Kumari Billakurthi and the late Udo Gowik for their useful advice and support during this study. We also thank the Millenium Seed Bank (MSB Kew), M. Höhn (Budapest) and G. Somogyi (Hungary) for the contribution of seed samples. We thank the “Genomics and Transcriptomics laboratory” of the “Biologisch-Medizinisches Forschungszentrum” (BMFZ) at the Heinrich-Heine-University Duesseldorf (Germany) for technical support and conducting the Illumina sequencing. Parts of this research were performed using the supercomputer Mogon and/or advisory services offered by Johannes Gutenberg University Mainz (hpc.uni-mainz.de), which is a member of the AHRP and the Gauss Alliance e.V. We thank C. Wild (Botanical Garden Univ. Mainz) for cultivating the plants.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Bauwe, H.; Hagemann, M.; Fernie, A.R. Photorespiration: Players, partners and origin. *Trends Plant Sci.* **2010**, *15*, 330–336. [[CrossRef](#)]
- Gowik, U.; Westhoff, P. The Path from C3 to C4 photosynthesis. *Plant Physiol.* **2011**, *155*, 56–63. [[CrossRef](#)] [[PubMed](#)]
- Christin, P.A.; Osborne, C.P.; Chatelet, D.S.; Columbus, J.T.; Besnard, G.; Hodkinson, T.R.; Garrison, L.M.; Vorontsova, M.S.; Edwards, E.J. Anatomical enablers and the evolution of C4 photosynthesis in grasses. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1381–1386. [[CrossRef](#)]
- Kadereit, G.; Bohley, K.; Lauterbach, M.; Tefarikis, D.T.; Kadereit, J.W. C3–C4 intermediates may be of hybrid origin—A reminder. *New Phytol.* **2017**, *215*, 70–76. [[CrossRef](#)] [[PubMed](#)]
- Sage, R.F.; Khoshravesh, R.; Sage, T.L. From proto-Kranz to C4 Kranz: Building the bridge to C4 photosynthesis. *J. Exp. Bot.* **2014**, *65*, 3341–3356. [[CrossRef](#)]
- Bräutigam, A.; Gowik, U. Photorespiration connects C3 and C4 photosynthesis. *J. Exp. Bot.* **2016**, *67*, 2953–2962. [[CrossRef](#)]
- Schlüter, U.; Weber, A.P.M. Regulation and Evolution of C4 Photosynthesis. *Annu. Rev. Plant Biol.* **2020**, *71*, 183–215. [[CrossRef](#)]
- Sage, R.F.; Monson, R.K.; Ehleringer, J.R.; Adachi, S.; Pearcy, R.W. Some like it hot: The physiological ecology of C4 plant evolution. *Oecologia* **2018**, *187*, 941–966. [[CrossRef](#)]
- Monson, R.K.; Edwards, G.E.; Ku, M.S.B. C3–C4 Intermediate Photosynthetic Plants. *Bioscience* **1984**, *34*, 563–566. [[CrossRef](#)]
- Edwards, G.E.; Ku, M.S.B. Biochemistry of C3–C4 Intermediates. In *The Biochemistry of Plants*; Hatch, M.D., Boardman, N.K., Eds.; Academic Press: London, UK, 1987; Volume 10, pp. 275–325.
- Oakley, J.C.; Sultmanis, S.; Stinson, C.R.; Sage, T.L.; Sage, R.F. Comparative studies of C3 and C4 Atriplex hybrids in the genomics era: Physiological assessments. *J. Exp. Bot.* **2014**, *65*, 3637–3647. [[CrossRef](#)]
- Sage, R.F. A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: Species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.* **2016**, *67*, 4039–4056. [[CrossRef](#)]
- Pyankov, V.I.; Artyusheva, E.G.; Edwards, G.E.; Black, C.C.; Soltis, P.S. Phylogenetic analysis of tribe *Salsola* (Chenopodiaceae) based on ribosomal ITS sequences: Implications for the evolution of photosynthesis types. *Am. J. Bot.* **2001**, *88*, 1189–1198. [[CrossRef](#)]
- Kadereit, G.T.; Borsch, K.; Weising, H.F. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C4 photosynthesis. *Int. J. Plant Sci.* **2003**, *164*, 959–986. [[CrossRef](#)]
- Kadereit, G.; Ackerly, D.; Pirie, M.D. A broader model for C4 photosynthesis evolution in plants inferred from the goosefoot family (*Chenopodiaceae* s.s.). *Proc. R. Soc. B Biol. Sci.* **2012**, *279*, 3304–3311. [[CrossRef](#)]
- Kadereit, G.; Lauterbach, M.; Pirie, M.D.; Arafah, R.; Freitag, H. When do different C4 leaf anatomies indicate independent C4 origins? Parallel evolution of C4 leaf types in *Camphorosmeae* (Chenopodiaceae). *J. Exp. Bot.* **2014**, *65*, 3499–3511. [[CrossRef](#)]
- Schütze, P.; Freitag, H.; Weising, K. An integrated molecular and morphological study of the subfamily *Suaedoideae* ulbr. (Chenopodiaceae). *Plant Syst. Evol.* **2003**, *239*, 257–286. [[CrossRef](#)]
- Kadereit, G.; Freitag, H. Molecular phylogeny of *Camphorosmeae* (Camphorosmoideae, Chenopodiaceae): Implications for biogeography, evolution of C4-photosynthesis and taxonomy. *Taxon* **2011**, *60*, 51–78. [[CrossRef](#)]

19. Akhani, H.; Edwards, G.; Roalson, E.H. Diversification of the Old World *Salsoleae* s.l. (Chenopodiaceae): Molecular Phylogenetic Analysis of Nuclear and Chloroplast Data Sets and a Revised Classification Author (s): Hossein Akhani, Gerald Edwards, and Eric H. Roalson Diversification O. *Int. J. Plant Sci.* **2007**, *168*, 931–956. [[CrossRef](#)]
20. Freitag, H.; Kadereit, G. C3 and C4 leaf anatomy types in *Camphorosmeae* (Camphorosmoideae, Chenopodiaceae). *Plant Syst. Evol.* **2014**, *300*, 665–687. [[CrossRef](#)]
21. Voznesenskaya, E.V.; Koteyeva, N.K.; Akhani, H.; Roalson, E.H.; Edwards, G.E. Structural and physiological analyses in *Salsoleae* (Chenopodiaceae) indicate multiple transitions among C3, intermediate, and C4 photosynthesis. *J. Exp. Bot.* **2013**, *64*, 3583–3604. [[CrossRef](#)]
22. Schüssler, C.; Freitag, H.; Koteyeva, N.; Schmidt, D.; Edwards, G.; Voznesenskaya, E.; Kadereit, G. Molecular phylogeny and forms of photosynthesis in tribe *Salsoleae* (Chenopodiaceae). *J. Exp. Bot.* **2017**, *68*, 207–223. [[CrossRef](#)]
23. Lauterbach, M.; Billakurthi, K.; Kadereit, G.; Ludwig, M.; Westhoff, P.; Gowik, U. C3 cotyledons are followed by C4 leaves: Intra-individual transcriptome analysis of *Salsola soda* (Chenopodiaceae). *J. Exp. Bot.* **2017**, *68*, 161–176. [[CrossRef](#)]
24. Lauterbach, M.; Schmidt, H.; Billakurthi, K.; Hankeln, T.; Westhoff, P.; Gowik, U.; Kadereit, G. De novo transcriptome assembly and comparison of C3, C3-C4, and C4 species of tribe *Salsoleae* (Chenopodiaceae). *Front. Plant Sci.* **2017**, *8*, 1939. [[CrossRef](#)]
25. Monteiro, A.; Podlaha, O. Wings, horns, and butterfly eyespots: How do complex traits evolve? *PLoS Biol.* **2009**, *7*, 0209–0216. [[CrossRef](#)] [[PubMed](#)]
26. Aubry, S.; Kelly, S.; Kümpers, B.M.C.; Smith-Unna, R.D.; Hibberd, J.M. Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C4 Photosynthesis. *PLoS Genet.* **2016**, *10*, e1006087. [[CrossRef](#)]
27. Windhövel, A.; Hein, I.; Dabrowa, R.; Stockhaus, J. Characterization of a novel class of plant homeodomain proteins that bind to the C4 phosphoenolpyruvate carboxylase gene of *Flaveria trinervia*. *Plant Mol. Biol.* **2001**, *45*, 201–214. [[CrossRef](#)] [[PubMed](#)]
28. Huang, P.; Brutnell, T.P. A synthesis of transcriptomic surveys to dissect the genetic basis of C4 photosynthesis. *Curr. Opin. Plant Biol.* **2016**, *31*, 91–99. [[CrossRef](#)] [[PubMed](#)]
29. Schwacke, R.; Ponce-Soto, G.Y.; Krause, K.; Bolger, A.M.; Arsova, B.; Hallab, A.; Gruden, K.; Stitt, M.; Bolger, M.E.; Usadel, B. MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Mol. Plant* **2019**, *12*, 879–892. [[CrossRef](#)]
30. Jin, J.; Zhang, H.; Kong, L.; Gao, G.; Luo, J. PlantTFDB 3.0: A portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* **2014**, *42*, 1182–1187. [[CrossRef](#)]
31. Jin, J.; Tian, F.; Yang, D.C.; Meng, Y.Q.; Kong, L.; Luo, J.; Gao, G. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **2017**, *45*, D1040–D1045. [[CrossRef](#)]
32. Dohm, J.C.; Minoche, A.E.; Holtgräwe, D.; Capella-Gutiérrez, S.; Zakrzewski, F.; Tafer, H.; Rupp, O.; Sörensen, T.R.; Stracke, R.; Reinhardt, R.; et al. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* **2014**, *505*, 546–549. [[CrossRef](#)]
33. Abu-Jamous, B.; Kelly, S. *Clust*: Automatic extraction of optimal co-expressed gene clusters from gene expression data. *Genome Biol.* **2018**, *19*, 172. [[CrossRef](#)]
34. Gowik, U.; Bräutigam, A.; Weber, K.L.; Weber, A.P.M.; Westhoff, P. Evolution of C4 photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C4? *Plant Cell* **2011**, *23*, 2087–2105. [[CrossRef](#)] [[PubMed](#)]
35. Mallmann, J.; Heckmann, D.; Bräutigam, A.; Lercher, M.J.; Weber, A.P.M.; Westhoff, P.; Gowik, U. The role of photorespiration during the evolution of C4 photosynthesis in the genus *Flaveria*. *eLife* **2014**, *2014*, e02478. [[CrossRef](#)]
36. Schulze, S.; Mallmann, J.; Burscheidt, J.; Koczor, M.; Streubel, M.; Bauwe, H.; Gowik, U.; Westhoff, P. Evolution of C4 photosynthesis in the genus *Flaveria*: Establishment of a photorespiratory CO<sub>2</sub> pump. *Plant Cell* **2013**, *25*, 2522–2535. [[CrossRef](#)] [[PubMed](#)]
37. Sage, R.F. Russ Monson and the evolution of C4 photosynthesis. *Oecologia* **2021**, *21*, 1–18. [[CrossRef](#)]
38. Bräutigam, A.; Kajala, K.; Wullenweber, J.; Sommer, M.; Gagneul, D.; Weber, K.L.; Carr, K.M.; Gowik, U.; Maß, J.; Lercher, M.J.; et al. An mRNA blueprint for C4 photosynthesis derived from comparative transcriptomics of closely related C3 and C4 species. *Plant Physiol.* **2011**, *155*, 142–156. [[CrossRef](#)] [[PubMed](#)]
39. Wang, Y.; Bräutigam, A.; Weber, A.P.M.; Zhu, X.G. Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. *J. Exp. Bot.* **2014**, *65*, 3567–3578. [[CrossRef](#)] [[PubMed](#)]
40. Furumoto, T.; Yamaguchi, T.; Ohshima-Ichie, Y.; Nakamura, M.; Tsuchida-Iwata, Y.; Shimamura, M.; Ohnishi, J.; Hata, S.; Gowik, U.; Westhoff, P.; et al. A plastidial sodium-dependent pyruvate transporter. *Nature* **2011**, *476*, 472–476. [[CrossRef](#)]
41. Walters, R.G.; Ibrahim, D.G.; Horton, P.; Kruger, N.J. A mutant of *Arabidopsis* lacking the triose-phosphate/phosphate translocator reveals metabolic regulation of starch breakdown in the light. *Plant Physiol.* **2004**, *135*, 891–906. [[CrossRef](#)]
42. Sage, R.F. Photorespiratory compensation: A driver for biological diversity. *Plant Biol.* **2013**, *15*, 624–638. [[CrossRef](#)] [[PubMed](#)]
43. Lundgren, M.R. C2 photosynthesis: A promising route towards crop improvement? *New Phytol.* **2020**, *228*, 1734–1740. [[CrossRef](#)] [[PubMed](#)]
44. Döring, F.; Streubel, M.; Bräutigam, A.; Gowik, U. Most photorespiratory genes are preferentially expressed in the bundle sheath cells of the C4 grass *Sorghum bicolor*. *J. Exp. Bot.* **2016**, *67*, 3053–3064. [[CrossRef](#)]
45. Ambawat, S.; Sharma, P.; Yadav, N.R.; Yadav, R.C. MYB transcription factor genes as regulators for plant responses: An overview. *Physiol. Mol. Biol. Plants* **2013**, *19*, 307–321. [[CrossRef](#)]

46. Kotak, S.; Port, M.; Ganguli, A.; Bicker, F.; Von Koskull-Döring, P. Characterization of C-terminal domains of Arabidopsis heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES motifs essential for activator function and intracellular localization. *Plant J.* **2004**, *39*, 98–112. [[CrossRef](#)]
47. Pernas, M.; Ryan, E.; Dolan, L. Schizoriza Controls Tissue System Complexity in Plants. *Curr. Biol.* **2010**, *20*, 818–823. [[CrossRef](#)] [[PubMed](#)]
48. Schmidt, R.; Schippers, J.H.M.; Welker, A.; Mieulet, D.; Guiderdoni, E.; Mueller-Roeber, B. Transcription factor oshsf1b regulates salt tolerance and development in oryza sativa ssp. japonica. *AoB Plants* **2012**, *12*, pls011. [[CrossRef](#)]
49. Gangappa, S.N.; Botto, J.F. The BBX family of plant transcription factors. *Trends Plant Sci.* **2014**, *19*, 460–470. [[CrossRef](#)]
50. Grimplet, J.; Pimentel, D.; Agudelo-Romero, P.; Martinez-Zapater, J.M.; Fortes, A.M. The Lateral Organ Boundaries Domain gene family in grapevine: Genome-wide characterization and expression analyses during developmental processes and stress responses. *Sci. Rep.* **2017**, *7*, 15968. [[CrossRef](#)]
51. Slewinski, T.L.; Anderson, A.A.; Zhang, C.; Turgeon, R. Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant Cell Physiol.* **2012**, *53*, 2030–2037. [[CrossRef](#)]
52. Slewinski, T.L. Using evolution as a guide to engineer Kranz-type C4 photosynthesis. *Front. Plant Sci.* **2013**, *4*, 212. [[CrossRef](#)]
53. Wang, P.; Kelly, S.; Fouracre, J.P.; Langdale, J.A. Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. *Plant J.* **2013**, *75*, 656–670. [[CrossRef](#)] [[PubMed](#)]
54. Fouracre, J.P.; Ando, S.; Langdale, J.A. Cracking the Kranz enigma with systems biology. *J. Exp. Bot.* **2014**, *65*, 3327–3339. [[CrossRef](#)]
55. Lori Tausta, S.; Li, P.; Si, Y.; Gandotra, N.; Liu, P.; Sun, Q.; Brutnell, T.P.; Nelson, T. Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C4-related processes. *J. Exp. Bot.* **2014**, *65*, 3543–3555. [[CrossRef](#)]
56. Semiarti, E.; Ueno, Y.; Tsukaya, H.; Iwakawa, H.; Machida, C.; Machida, Y. The asymmetric leaves 2 gene of Arabidopsis thaliana regulates formation of a symmetric lamina, establishment of venation and repression of meristem-related homeobox genes in leaves. *Development* **2001**, *128*, 1771–1783. [[CrossRef](#)] [[PubMed](#)]
57. Xie, Q.; Frugis, G.; Colgan, D.; Chua, N.H. Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes Dev.* **2000**, *14*, 3024–3036. [[CrossRef](#)]
58. Vroemen, C.W.; Mordhorst, A.P.; Albrecht, C.; Kwaaitaal, M.A.C.J.; De Vries, S.C. The Cup-Shaped Cotyledon 3 gene is required for boundary and shoot meristem formation in Arabidopsis. *Plant Cell* **2003**, *15*, 1563–1577. [[CrossRef](#)]
59. Yamaguchi, M.; Ohtani, M.; Mitsuda, N.; Kubo, M.; Ohme-Takagi, M.; Fukuda, H.; Demura, T. Vnd-Interacting 2, a NAC domain transcription factor, negatively regulates xylem vessel formation in Arabidopsis. *Plant Cell* **2010**, *22*, 1249–1263. [[CrossRef](#)]
60. Huang, Q.; Wang, Y.; Li, B.; Chang, J.; Chen, M.; Li, K.; Yang, G.; He, G. TaNAC29, a NAC transcription factor from wheat, enhances salt and drought tolerance in transgenic Arabidopsis. *BMC Plant Biol.* **2015**, *15*, 268. [[CrossRef](#)] [[PubMed](#)]
61. Li, W.; Li, X.; Chao, J.; Zhang, Z.; Wang, W.; Guo, Y. NAC family transcription factors in tobacco and their potential role in regulating leaf senescence. *Front. Plant Sci.* **2018**, *871*, 1900. [[CrossRef](#)] [[PubMed](#)]
62. Zhang, J.; Huang, G.Q.; Zou, D.; Yan, J.Q.; Li, Y.; Hu, S.; Li, X.B. The cotton (*Gossypium hirsutum*) NAC transcription factor (FSN1) as a positive regulator participates in controlling secondary cell wall biosynthesis and modification of fibers. *New Phytol.* **2018**, *217*, 625–640. [[CrossRef](#)]
63. Ebrahimian-Motlagh, S.; Ribone, P.A.; Thirumalaikumar, V.P.; Allu, A.D.; Chan, R.L.; Mueller-Roeber, B.; Balazadeh, S. Jungbrunnen 1 confers drought tolerance downstream of the HD-Zip I Transcription factor AtHB13. *Front. Plant Sci.* **2017**, *8*, 2118. [[CrossRef](#)] [[PubMed](#)]
64. Zhang, Q.; Chen, T.; Wang, X.; Wang, J.; Gu, K.; Yu, J.; Hu, D.; Hao, Y. Genome-wide Identification and Expression Analyses of Homeodomain-leucine Zipper Family Genes Reveal Their Involvement in Stress Response in Apple (*Malus × domestica*). *Hortic. Plant J.* **2021**, in press. [[CrossRef](#)]
65. Li, P.; Ponnala, L.; Gandotra, N.; Wang, L.; Si, Y.; Tausta, S.L.; Kebrom, T.H.; Provart, N.; Patel, R.; Myers, C.R.; et al. The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* **2010**, *42*, 1060–1067. [[CrossRef](#)]
66. Borba, A.R.; Serra, T.S.; Górska, A.; Gouveia, P.; Cordeiro, A.M.; Reyna-Llorens, I.; Kneřová, J.; Barros, P.M.; Abreu, I.A.; Oliveira, M.M.; et al. Synergistic binding of bHLH transcription factors to the promoter of the maize NADP-ME gene used in C4 photosynthesis is based on an ancient code found in the ancestral C3 state. *Mol. Biol. Evol.* **2018**, *35*, 1690–1705. [[CrossRef](#)] [[PubMed](#)]
67. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
68. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **2013**, *29*, 644–652. [[CrossRef](#)]
69. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
70. Kriventseva, E.V.; Kuznetsov, D.; Tegenfeldt, F.; Manni, M.; Dias, R.; Simão, F.A.; Zdobnov, E.M. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **2019**, *47*, D807–D811. [[CrossRef](#)] [[PubMed](#)]

71. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [[CrossRef](#)] [[PubMed](#)]
72. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [[CrossRef](#)] [[PubMed](#)]
73. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
74. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
75. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]