



OPEN

## Morphological and metabolomics profiling of intraspecific *Arabidopsis* hybrids in relation to biomass heterosis

Quynh Thi Ngoc Le<sup>1,2,8</sup>, Naoya Sugi<sup>1,3,8</sup>, Masaaki Yamaguchi<sup>4</sup>, Touko Hirayama<sup>4</sup>, Makoto Kobayashi<sup>5</sup>, Yutaka Suzuki<sup>6</sup>, Miyako Kusano<sup>4,5</sup> & Hiroshi Shiba<sup>4,7</sup>✉

Heterosis contributes greatly to the worldwide agricultural yield. However, the molecular mechanism underlying heterosis remains unclear. This study took advantage of *Arabidopsis* intraspecific hybrids to identify heterosis-related metabolites. Forty-six intraspecific hybrids were used to examine parental effects on seed area and germination time. The degree of heterosis was evaluated based on biomass: combinations showing high heterosis of F<sub>1</sub> hybrids exhibited a biomass increase from 6.1 to 44% over the better parent value (BPV), whereas that of the low- and no-heterosis hybrids ranged from –19.8 to 9.8% over the BPV. Metabolomics analyses of F<sub>1</sub> hybrids with high heterosis and those with low one suggested that changes in TCA cycle intermediates are key factors that control growth. Notably, higher fumarate/malate ratios were observed in the high heterosis F<sub>1</sub> hybrids, suggesting they provide metabolic support associated with the increased biomass. These hybrids may produce more energy-intensive biomass by speeding up the efficiency of TCA fluxes. However, the expression levels of TCA-process-related genes in F<sub>1</sub> hybrids were not associated with the intensity of heterosis, suggesting that the post-transcriptional or post-translational regulation of these genes may affect the productivity of the intermediates in the TCA cycle.

The growth of the world population has heightened the need to improve agricultural productivity to ensure food security. Hybrids with optimal combinations have superior performance regarding biomass, yield, or development rate and can significantly contribute to solving this problem<sup>1</sup>. Heterosis (or hybrid vigor) is an important phenomenon in agriculture that has been researched in various crops, including rice, wheat, maize, sorghum, sugar beet, cotton, and canola<sup>2–6</sup>.

Although many studies have used natural variation in intraspecific hybrids to understand heterotic mechanisms<sup>7–11</sup>, the correlation between genetic distance and heterosis remains a subject of debate. In maize, crosses with a greater genetic distance between parents have improved heterosis<sup>12</sup>. Conversely, in *Arabidopsis*, the lack of a relationship between genetic distance and the heterosis level of the parents concerning the dry weight of the shoots<sup>13</sup> or the fresh weight recorded during the vegetative growth phase has been reported<sup>7</sup>. A variation in heterosis levels in terms of biomass was detected within reciprocal hybrids, even if the parental accessions were closely related, as in the case of the Columbia (Col), C24, Landsberg *erecta* (*Ler*), and Wassilewskija (Ws)<sup>7,9</sup> *Arabidopsis* ecotypes.

The analytical techniques used in metabolomics, including gas chromatography time-of-flight mass spectrometry (GC-TOF-MS), have resulted in a relatively broad coverage of compound classes that can be rapidly detected in tissues. This provides opportunities for profiling different metabolite levels between hybrids and their parents, and it may afford useful information regarding heterosis in plants<sup>4,14,15</sup>. A diallel cross population

<sup>1</sup>Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Ten-Nodai, Tsukuba, Ibaraki, Japan. <sup>2</sup>Thuyloi University, 175 Tay Son, Dong Da, Hanoi, Viet Nam. <sup>3</sup>Kihara Institute for Biological Research, Yokohama City University, Yokohama, Kanagawa, Japan. <sup>4</sup>Degree Programs in Life and Earth Sciences, Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Ten-Nodai, Tsukuba, Ibaraki, Japan. <sup>5</sup>RIKEN Center for Sustainable Resource Science, Suehiro 1-7-22, Tsurumi, Yokohama, Japan. <sup>6</sup>Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Japan. <sup>7</sup>Tsukuba-Plant Innovation Research Center, University of Tsukuba, Ten-Nodai 1-1-1, Tsukuba, Ibaraki, Japan. <sup>8</sup>These authors contributed equally: Quynh Thi Ngoc Le and Naoya Sugi. ✉email: shiba.hiroshi.gm@u.tsukuba.ac.jp

in rice was generated by crossing 18 inbred lines, with the metabolic profiles of the parents being used to determine three predicted agronomic traits, namely, heading date, grain yield, and plant height, using the parental metabolic profiles<sup>4</sup>. In *Arabidopsis*, a recombinant inbred line population generated from a cross between Col and C24 exhibited transgressive segregating biomass. The use of such a population in a metabolomics approach was based on the assumption of a high correlation between the growth rate and metabolic profiles<sup>14</sup>.

Several known metabolites in the central metabolic pathway are negatively correlated with biomass. These include sucrose-, glucose-, and fructose-6-phosphate; the intermediates of the tricarboxylic acid (TCA) cycle (such as citrate, succinate, malate, and 2-oxoglutarate (2-OG)); and amino acids (such as glutamine and phenylalanine)<sup>14,16</sup>. The TCA cycle has emerged as a central mitochondrial hub that drives the production of the ATP used in photosynthesis optimization<sup>17</sup>. Riewe et al. and Le et al.<sup>18,19</sup> reported an association between increased fumarate/malate ratios and advantageous growth development. However, the relationship between hybrid vigor levels and changes in central metabolites in multiple intraspecific hybrids of *Arabidopsis* remains unknown.

We investigated the physiological characteristics of seed area, germination time, and vegetative growth phenotypes in multiple intraspecific lines of *Arabidopsis thaliana* at matched germination times. The establishment of a new cultivation system was an attempt to detect differences in heterosis levels in parental combinations with minimal effects on the seed size and germination time on growth at the seedling stage. Four intraspecific combinations, i.e., Col × C24, Col × Rld-1, Col × Ws-0, and Col × Ler-1, with different levels of hybrid vigor were selected for comparative metabolomics analysis between the intraspecific hybrids and their parental lines to identify changes in metabolites associated with heterosis. Gene expression variations related to changes in metabolites were also investigated. This study is important for elucidating the relationship between TCA intermediates and heterosis levels regarding biomass and clarifying the metabolic information of intraspecific *Arabidopsis* hybrids.

## Results

**Physiological characteristics of intraspecific *Arabidopsis* hybrids.** *Selection of intraspecific hybrid combinations that were expected to have varying heterosis levels.* Heterosis can be calculated as the percentage increase in the trait value of the F<sub>1</sub> hybrid over the average trait value of the two parents (mid-parent value, MPV) or the trait value of the better-performance parent (better-parent value, BPV). As a result, it is frequently ideal to compare a hybrid's performance to both BPV and MPV. For a hybrid to be considered commercially advantageous, it typically needs to outperform the BPV for agronomically significant characteristics like vegetative biomass or seed yield.

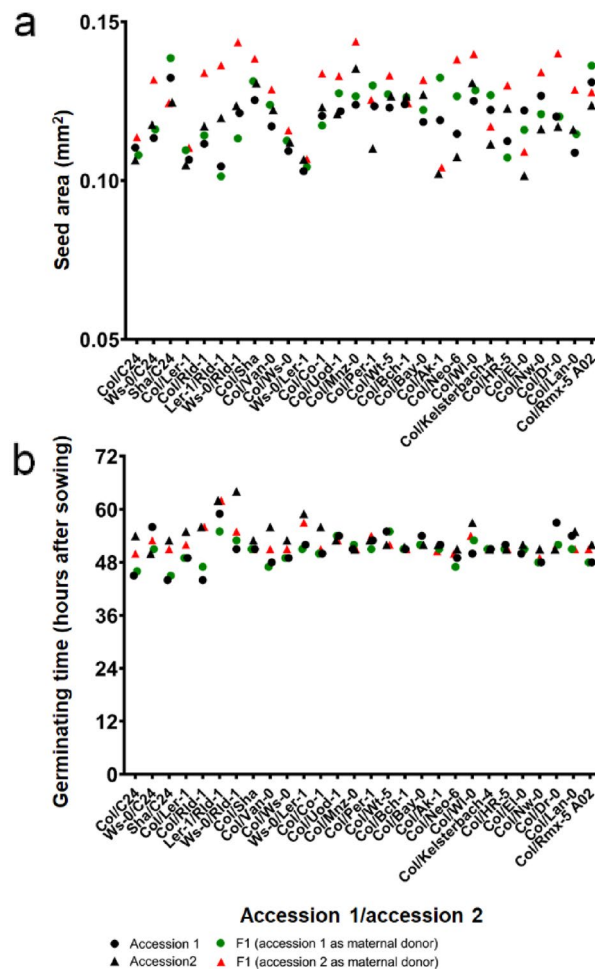
The 202 intraspecific lines (Dataset S1) used in this experiment were self-fertilized via hydroponics under uniform conditions to reduce interindividual variation during germination and growth<sup>19</sup>. Differences in flowering time among intraspecific lines affect biomass during the vegetative growth phase<sup>9</sup>. Therefore, of the 202 accessions, 80 with delayed flowering were excluded from the analysis.

It is also known that differences in seed size among intraspecific hybrids affect the changes in plant biomass<sup>13</sup>. For example, the hybrid seeds of the C24 and Col lines exhibited almost the same seed size as that of their parents, whereas the seed size of the Est line was increased compared with that of the Col line. Thus, although both the Col × C24 and Est × Col combinations showed increased biomass in their F<sub>1</sub> hybrids (13.9% and 35.6% increase from the mid parental value (MPV) for Col × C24 and C24 × Col, respectively, and 37% increase from the MPV for Est × Col), the heterosis in the Est × Col hybrid can be attributed to seed size (Fig. S1). Therefore, we measured the seed areas in the remaining 122 accessions and compared them to the seed area of the control line (Col) ( $P < 0.01$ , Student's *t*-test) (Fig. S2). We found that 48 accessions had a larger seed area than Col. These accessions were excluded from further analysis. After considering the genetic background and plant growth process, we finally selected 28 accessions for intraspecific hybrid evaluation (Dataset S2).

**Seed area and germination time of intraspecific F<sub>1</sub> hybrids.** To investigate physiological traits and quantify biomass, seed size was first measured and compared between the hybrids and their parents. Measurements of hybrid seeds from the 28 combinations of reciprocal crosses exhibited a variation in seed area in several intraspecific hybrid combinations, suggesting that it could be attributed to maternal contribution (Figs. 1A and S3). Rld-1 × Ws-0, Rld-1 × Ler-1, and Rld-1 × Col had a significantly increased seed area compared with the BPV ( $P < 0.01$ ), whereas Rld-1, as the paternal donor, had the opposite effect on seed size. Similarly, the seeds of C24 × Ws-0 hybrids were notably larger than the BPV, whereas C24 × Col and C24 × Sha, with C24 as a maternal donor, were not statistically different from the BPV. Conversely, Col, Ws-0, and Ler-1 had little effect on F<sub>1</sub> seed size. Therefore, to minimize the variation in seed size between the intraspecific hybrids and their parents as much as possible, crosses with other intraspecific lines using Col as one of the parents were used in subsequent experiments.

The timing of germination, as a common physiological trait, is considered the first process of plant growth<sup>20</sup>. To assess differences in the timing of germination among the parental lines and the hybrids, the seed germination time was tracked (Figs. 1B and S4), and the germination timing of the F<sub>1</sub> hybrids was compared with that of their parents. The results indicated that the maternal environment has a significant effect on seed dormancy and germination and that both seed size and seed germination largely reflect the effect of the maternal plants.

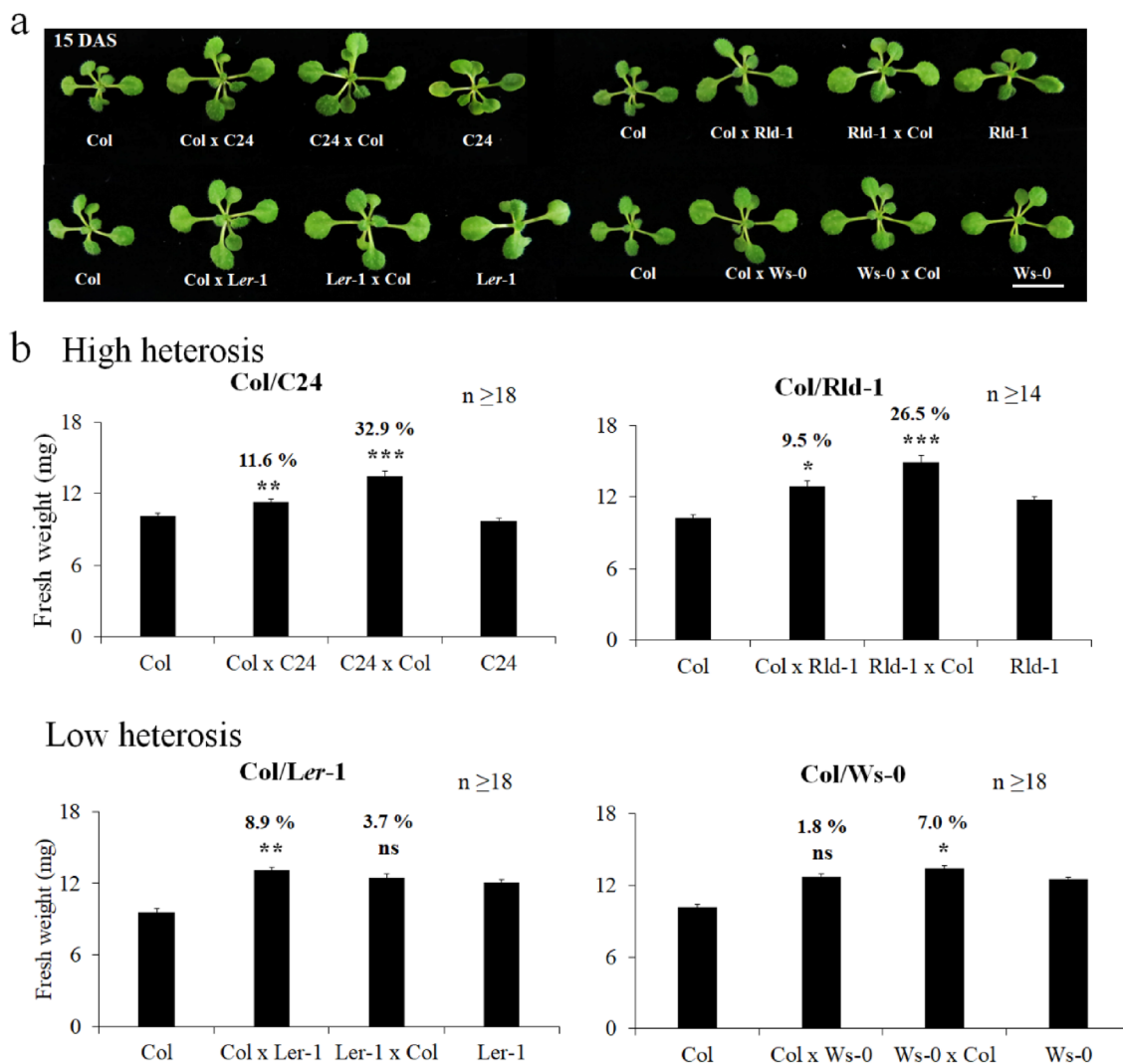
**Changes in heterosis levels according to a hybrid combination.** To minimize the effects of the environmental differences and interindividual variability between the hybrids and the parental lines on biomass and developmental stages, the plants were grown under short-day and highly controlled conditions<sup>16,19</sup>. The parents and hybrids were adjusted to the same germination time by sowing seeds at different times based on germination time data (Figs. 2 and S5). From the agronomical point of view, hybrids need to exceed the BPV regarding the total weight to contribute to yield improvement. In addition, various time points in the develop-



**Figure 1.** Effect of parent-origin on seed area and germinating time. **(a)** Seed area of differential parental combinations.  $n \geq 100$  seeds collected from at least three individual plants. **(b)** Germination time of intraspecific hybrids and their parents. Germinating time was calculated on the time at which 50% of cotyledons fully opened.

mental life cycle can modulate the differences in hybrid vigor<sup>9</sup>. This study focused on the early developmental stage to avoid the impact of flowering on biomass<sup>21</sup>. We classified heterosis according to the percentage of the fresh weight of F<sub>1</sub> hybrids at 15 DAS compared with the BPV<sup>9,13</sup>: (1) high heterosis: two reciprocal F<sub>1</sub> hybrids had significantly higher values than the BPV ( $P < 0.05$ ), with at least one reciprocal hybrid having a 10% higher value than the BPV; (2) weak heterosis: a difference from the BPV ranging from 0 to 10%, with at least one reciprocal hybrid having a significantly higher value than the BPV; (3) no heterosis: neither of the hybrids exhibited significantly higher values than the BPV; and (4) unclassifiable combinations: hybrids obtained by reciprocal crosses that showed opposite trends.

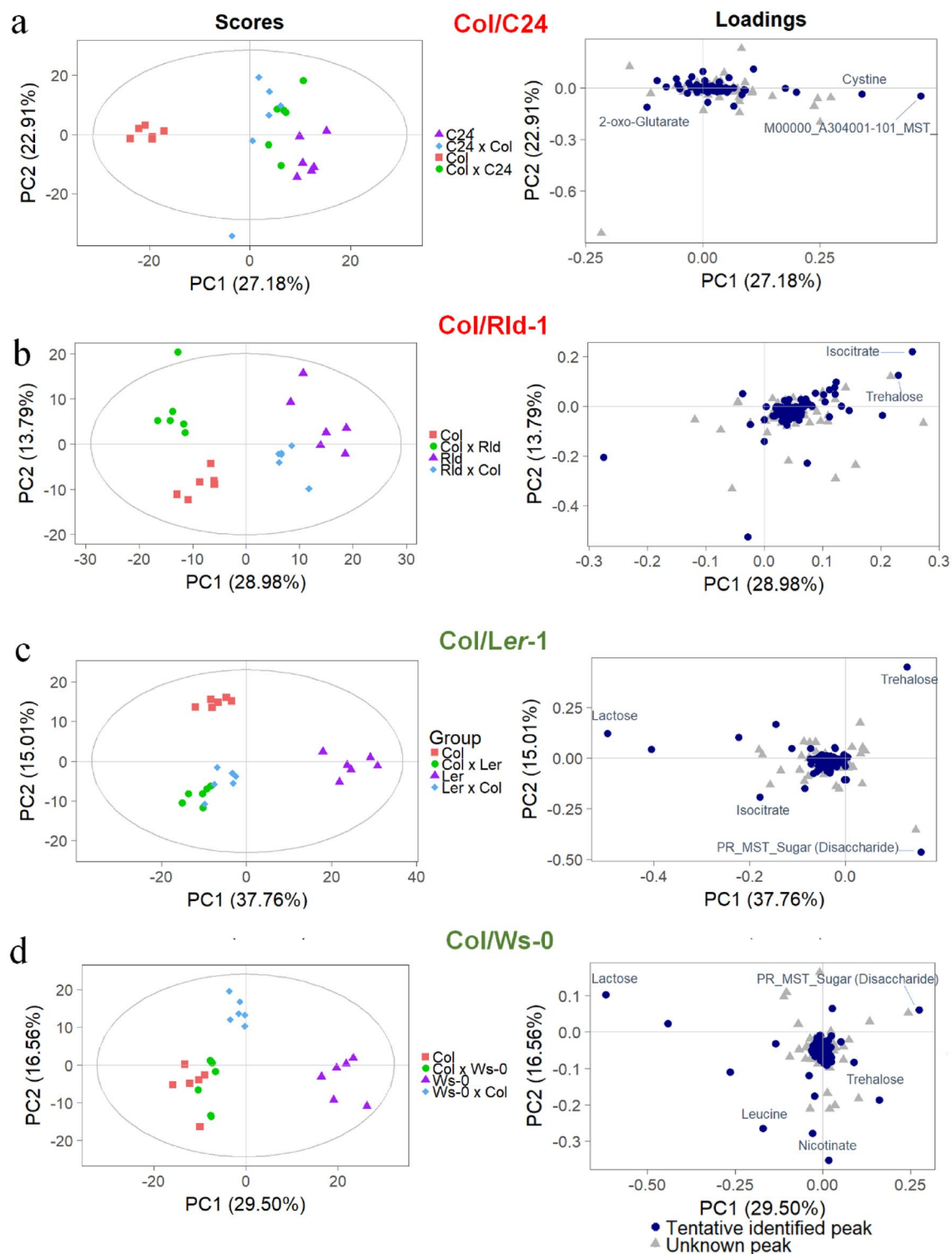
The intraspecific hybrids generally were at the 10-leaf stage at almost 15 DAS under our experimental growth conditions (data not shown). The hybrids obtained by crossbreeding the 28 lines exhibited different levels of growth regarding fresh weight, i.e., from -21 to 44% compared with the BPV (Fig. S5). High heterosis was detected in 9 of the 28 combinations (32% of the total hybrids). The most effective combinations were Col × Neo-6 (44% of Neo-6 × Col and 17.2% of Col × Neo-0), Col × C24 (32.9% of C24 × Col and 11.6% of Col × C24), and Col × Co-1 (36.7% of Co-1 × Col and 24.8% of Col × Co-1). Two combinations, namely, Col × Ler-1 and Col × Ws-0, were classified into the low-heterosis class (2), with BPVs ranging from 1.8 to 8.9%. 11 of the 28 combinations did not exhibit heterosis, and the remaining six combinations showed different results in reciprocal crosses. Similarly, data from the leaf area were gathered at 5–15 DAS to support the fresh weight data (Fig. S6). It is important to note that the F<sub>1</sub> hybrids obtained from the two reciprocal crosses of each combination exhibited comparable trends regarding both fresh weight and leaf area, which implies that the heterosis mechanism could also be detected by other methods. Hence, all classes, i.e., (1), (2), and (3), were considered potential candidate combinations for the extensive testing of heterosis models.



**Figure 2.** The phenotype of representative hybrids with different levels of fresh weight heterosis. **(a)** Col/C24 and Col/Rld-1 represent high heterosis levels; Col/Ler-1 and Col/Ws-0 show low heterosis levels in 15 DAS. White bar represents 1 cm. **(b)** Levels of fresh weight heterosis of representative four hybrid combinations. Numbers above each hybrid's bar represent a level from BPV. Black asterisks show above BPV; ns: not significant, \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ , Student's *t*-test. *n* represents in detail the number of plants collected from each combination. Data indicate the average and error bar (standard error).

**Metabolite profiling of intraspecific *Arabidopsis* hybrids.** *Principal component analysis (PCA) of the various hybrid combinations.* To identify potential metabolomics differences in biomass heterosis, hybrids of C24 and Rld-1 (high heterosis group) and Ler-1 and Ws-0 (low-heterosis group) with Col as the common parent were analyzed for primary metabolites using gas chromatography time-of-flight mass spectrometry (GC-TOF-MS) (Fig. 2). Of the 188 peaks detected, 103 were known metabolites (Dataset S3).

A global analysis using PCA score plots was performed after the  $\log_2$ -transformation of the data to detect differences or similarities among the reciprocal hybrids and their parents based on non-targeted metabolite profiling (Fig. 3). Non-targeted metabolite profiling was used in this investigation to uncover all metabolites including well-known metabolites changes and unexpected metabolic responses. The score plots revealed total variability values of 50.09% (PC1: 27.17%; PC2: 22.92%) for Col  $\times$  C24, 42.77% (PC1: 28.98%; PC2: 13.79%) for Col  $\times$  Rld-1, 52.77% (PC1: 37.76%; PC2: 15.01%) for Col  $\times$  Ler-1, and 46.06% (PC1: 29.50%; PC2: 16.56%) for Col  $\times$  Ws-0. Furthermore, the Col  $\times$  C24 and C24  $\times$  Col groups were lumped together and segregated from the Col and C24 parents (Fig. 3A), similar to the Col  $\times$  Ler-1 and Ler-1  $\times$  Col hybrids, which showed slight differences in metabolic profiles but were distinct from their parents (Fig. 3C). The Col  $\times$  Rld-1 hybrids were affected by the maternal donor, and score plots tended to be closer to the maternal one (Fig. 3B). The Ws-0  $\times$  Col group was divided by both parental lines, but Col  $\times$  Ws-0 and Col were close according to the PCA scores and loading plots (Fig. 3D). These results suggest that many of the candidate metabolites that may affect heterosis obtained from comprehensive metabolite profiling of different hybrid combinations were not correlated with heterosis levels and might be dependent on ecotype background. Therefore, it was necessary to focus on individual metabolites that showed differences between parents and F<sub>1</sub> hybrids according to heterosis level.

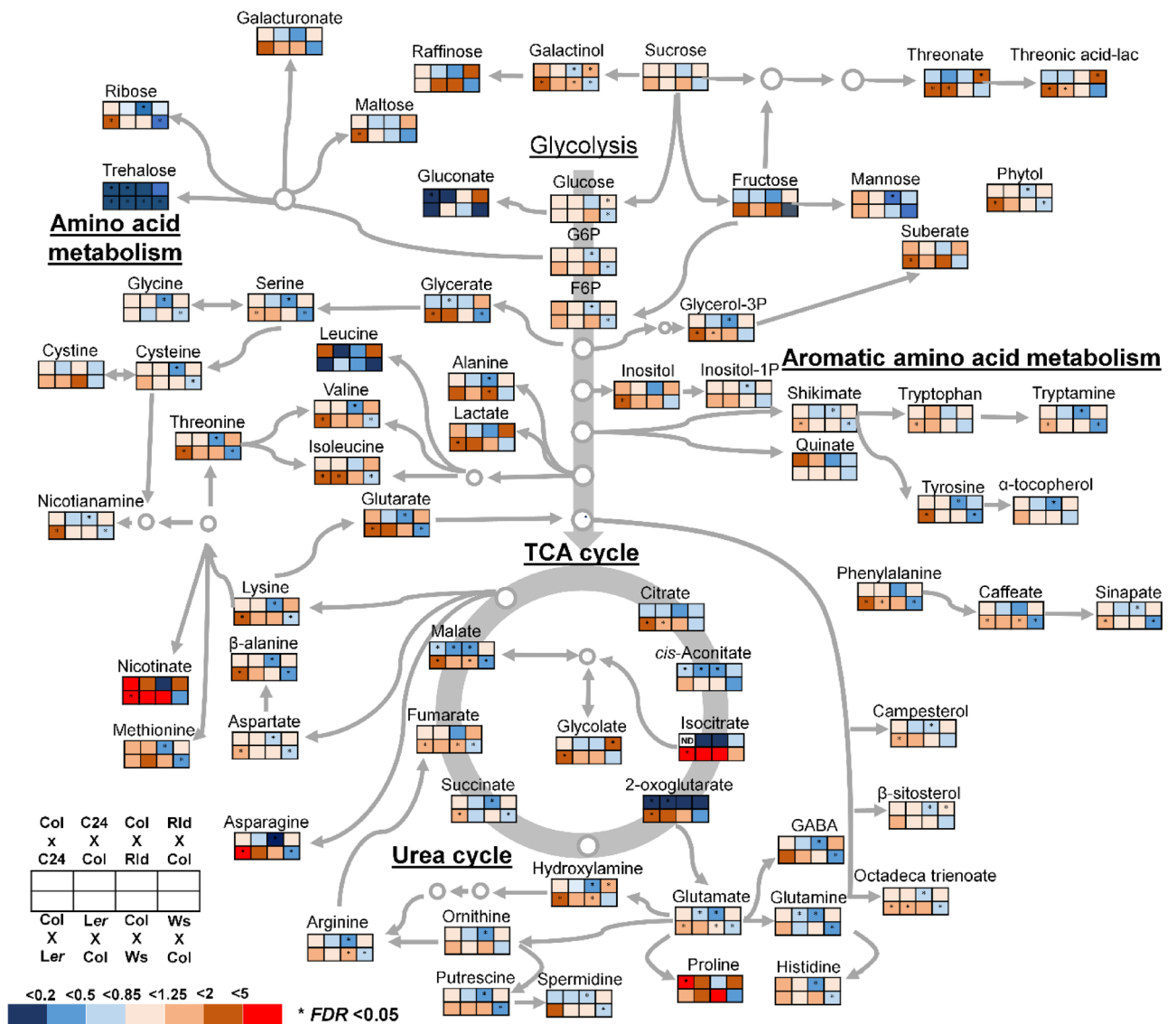


**Figure 3.** PCA plots for metabolites obtained from *Arabidopsis* combinations. Score and loading plots of (a) Col/C24, (b) Col/Rld-1, (c) Col/Ler-1, and (d) Col/Ws-0. The  $\log_2$  transformation of the normalized data matrix (188 peak areas  $\times$  24 leaf samples/one combination) from GC-TOF-MS was used. Red words: high biomass heterosis, green words: low-biomass heterosis.

**Metabolite differences among hybrids with high and low biomass heterosis.** The differences in metabolite profiling between the hybrids and their parents were analyzed using multiple *t*-tests ( $FDR < 0.05$ ). Metabolites that were significantly increased or decreased in the comparison of  $F_1$  individuals with their par-



ents are shown in the heat map (Fig. S7). The full dataset used in this heat map is presented in Dataset S4. The pattern of the significant metabolite changes between the hybrid and the parental lines for the Col×Rld-1 and Col×Ws-0 hybrids showed an opposite trend compared with the hybrids obtained by reciprocal crosses. Conversely, the pattern of the significant metabolite changes between the hybrids and the parental lines for Col×C24, as well as Col×Ler-1, showed a similar trend compared with the hybrids obtained by reciprocal crosses (Fig. S7, Dataset S3). Notably, the lowest number of metabolites showing significant changes between the hybrids and the parental lines was lowest in crosses between Col and C24, which showed high heterosis (Fig. S7, Dataset S3). Figure 4 presents the metabolic map of the central metabolic pathways of F<sub>1</sub> hybrids with their mid-parent value (MPV). The common sugars and sugar derivatives were mostly unaltered in the F<sub>1</sub> hybrids, and the levels of sucrose, maltose, raffinose, glucose, and ribose showed no clear trend between F<sub>1</sub> hybrids with high and/or low heterosis. Previously, Groszmann et al. reported that sugar transporter-related genes, including *STP1*, *STP14*, *ERD6*, and *PLT6*, are upregulated in all three hybrids (i.e., C24×Ler, C24×Col, and Col×Ler)<sup>22</sup>. In our transcriptome data, no difference was found in the expression levels of *STP1*, *STP14*, *ERD6*, and *PLT6* in both Col×C24 and C24×Col hybrids (Fig. S8). Starch content is also compared in hybrids and their parents to identify effects on hybrid vigor. Hybrids of Col×C24, Col×Ws-0, and Ler-1×Ws-0 were selected as representatives of high heterosis, low heterosis, and no heterosis, respectively, to analyze the amounts of starch. Results showed that reciprocal hybrids had starch levels comparable to their parental lines at all heterosis levels during the mid-light phase (ZT6) and the late-light phase (ZT11) (Fig. S9). In addition, our metabolome data suggested that amino acid levels exhibited unclear trends, whereas the main metabolic changes detected between the F<sub>1</sub>

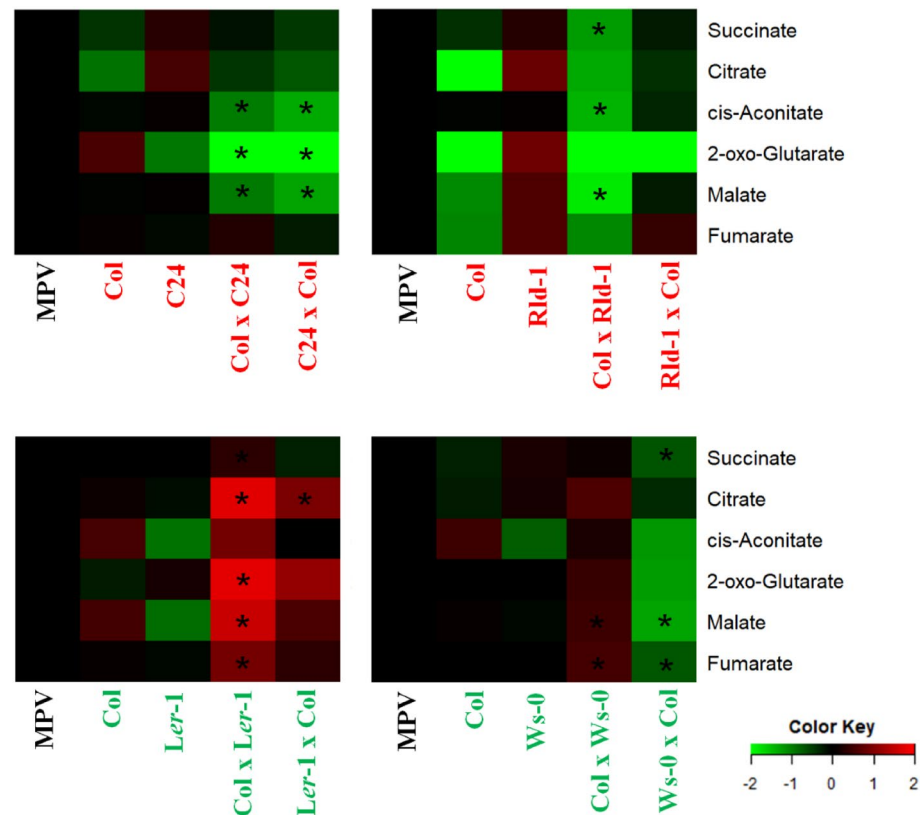


**Figure 4.** Metabolic change of F<sub>1</sub> hybrid to their parents. The fold change of the F<sub>1</sub> hybrid's metabolic change is normalized concerning the mid-parent value. Fold change is visualized by color; red, upregulated in F<sub>1</sub> hybrid; blue, downregulated in F<sub>1</sub> hybrid. ND not detected. \*  $FDR < 0.05$ , significant increase/decrease of F<sub>1</sub> hybrid as compared with each parent, based on multiple *t*-tests.

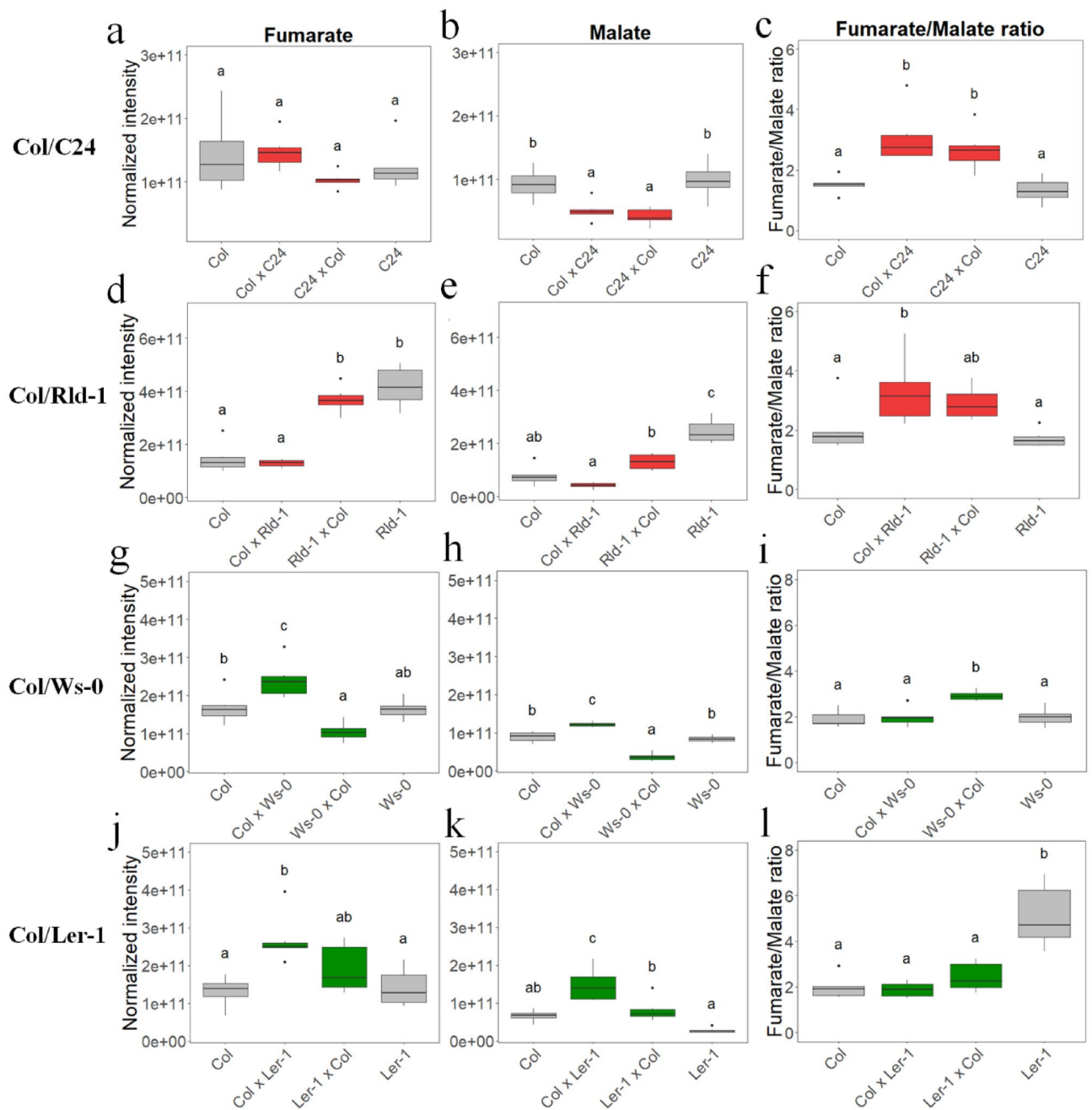
hybrids and their parents pertained to TCA cycle intermediates, citrate, cis-aconitate, isocitrate, and 2-oxoglutarate, which were downregulated in the  $F_1$  hybrids with high heterosis (Fig. 4). Therefore, we decided to focus on TCA cycle intermediates to examine their involvement in heterosis.

TCA cycle intermediates are the principal components responsible for driving biomass, as reported previously<sup>14,17</sup>. Therefore, we focused on the metabolites in the TCA cycle in several hybrid combinations. We found a decreasing trend in Col  $\times$  C24 and Col  $\times$  Rld-1, with higher levels of heterosis (Figs. 4 and 5). The three intermediates of the TCA cycle, namely, malate, 2-oxoglutarate, and aconitate, were detected at lower levels in high heterosis hybrids, i.e., Col  $\times$  C24, Col  $\times$  Rld, and Ws-0  $\times$  Col, compared with the MPV (Figs. 4 and 5). By contrast, the low-heterosis Col  $\times$  Ler- and Col  $\times$  Ws-0 hybrids showed higher levels than the MPV (Figs. 4 and 5). The fumarate/malate ratios of the Col  $\times$  C24, C24  $\times$  Col, Col  $\times$  Rld, Rld  $\times$  Col, and Ws  $\times$  Col hybrids were increased by 2.2-, 1.9-, 1.8-, 1.5-, and 1.5-fold, respectively, compared with the MPV (Fig. 6). Malate has a greater pKa value than fumarate, and a higher fumarate/malate ratio offers a beneficial mechanism for the pH-mediated modulation of intracellular flow and biomass<sup>23</sup>. These results suggest that the components of the TCA cycle play an important role in the mechanism of heterosis.

To understand the expression levels of TCA cycle-related genes, we performed RNA sequencing in the hybrids and parental lines for Col  $\times$  C24. One-way ANOVA revealed that only two genes (i.e., *FUM2* and *ATACA2*) showed a significantly differential expression pattern among the hybrids and parents among 78 TCA cycle-related genes based on MapMan annotation<sup>24</sup> (Fig. S10a and Dataset S5). The differential expression pattern of the two genes was mainly based on differences between the two parents rather than on  $F_1$ -hybrid-specific regulation (Fig. S10b). Although the RNA sequencing data of Col  $\times$  C24 did not reveal any  $F_1$ -hybrid-specific expression pattern for the TCA cycle-related genes, we also performed a quantitative PCR analysis of *FUM1* and *FUM2* in the Col  $\times$  C24, Col  $\times$  Rld-1, Col  $\times$  Ws-0, and Col  $\times$  Ler-1 hybrids because it was previously reported that fumarase determines the fumarate/malate ratios<sup>18</sup>. As shown in Fig. S11, the expression levels of *FUM1* and *FUM2* were not significantly different between the two hybrids and their BPVs. This indicates that the high fumarate/malate ratio observed in high heterosis hybrids was not caused by changes in transcriptional regulation. Post-transcriptional or post-translational regulation may correlate with changes in the metabolites of the TCA cycle. Changes in the fumarate/malate ratio could also be caused by increased metabolic activity in the  $F_1$  hybrids.



**Figure 5.** Levels of TCA cycle intermediates in hybrid combinations. Asterisks mean a significant difference by multiple  $t$ -tests ( $FDR < 0.05$ ), increased/decreased metabolites compared with both parents. Differences in red/green colors exhibit the  $\log_2$  fold change (up/down) normalized to MPV (MPV: control = 0 in  $\log_2$  fold change). Red/green words mean the name of high/low-biomass heterosis.



**Figure 6.** Analysis of fumarate, malate, and fumarate/malate ratio in Col/C24, Col/Rld-1, Col/Ws-0, and Col/Ler-1. Col/C24: fumarate (a), malate (b), and fumarate/malate ratio (c). Col/Rld-1: fumarate (d), malate (e), and fumarate/malate ratio (f). Col/Ws-0: fumarate (g), malate (h), and fumarate/malate ratio (i). Col/Ler-1: fumarate (j), malate (k), and fumarate/malate ratio (l). Box plots present median value and upper and lower quartiles. Difference letters over box plots are displayed significant value according Tukey's test,  $P < 0.05$ .

## Discussion

The effect of parental origin on seed size has been recognized<sup>9,25</sup>. This study examined several F<sub>1</sub> hybrids that develop larger mature embryos compared with their parental lines. Hybrids of C24 and Ler-1 have been suggested to promote or suppress embryo heterosis<sup>9</sup>. We found that the use of the Rld-1 and C24 accessions as maternal donors increased F<sub>1</sub> seed size (Figs. 1 and S3). Thus, for some of the F<sub>1</sub> hybrid combinations examined here, it is possible that F<sub>1</sub>-hybrid-specific modifications based on maternal lineage traits caused the increase in seed size. A possible explanation for this observation is that the embryos of F<sub>1</sub> hybrid seeds obtained by reciprocal crosses are epigenetically different even though they are genetically identical. In addition, the composition of the seed coat is affected by the maternal origin, suggesting that the contribution from the two parents is unequal. It is also possible that the maternal-derived organelle genome (chloroplast and mitochondrial DNA), which is important



for seed size heterosis, exhibited different interactions with the nuclear genome in the two parents compared with the hybrid lines<sup>26,27</sup>.

C24 × *Ler* hybrids exhibit a high level of heterosis regarding biomass production in *Arabidopsis*<sup>28–30</sup>, with upregulation of photosynthesis-related genes and early emergence of new leaves<sup>28</sup>. The seeds of reciprocal C24 × *Ler* hybrids germinated 7–12 h earlier than did their parents<sup>29</sup>. However, when the germination times of the parental and F<sub>1</sub> hybrids were matched, the chloroplast-related genes of F<sub>1</sub> hybrids were not downregulated or changed compared with their parents<sup>31</sup>, suggesting that the changes detected in C24 × *Ler* hybrids may be attributed to differences in developmental stages between the hybrids and their parents. This is a potential explanation for the differences observed in the developmental stages of the hybrids and their parents. Therefore, in this study, the germination times were identical among the lines to eliminate the effect of the differences in developmental stages. In addition, early germination is considered an indicator of a heterotic phenotype<sup>29</sup>. However, in our data from multiple intraspecific hybrids, none of the nine combinations that exhibited high heterosis in seedlings showed early germination in hybrids (Figs. 1 and S4). Thus, although an early germination time may contribute to a higher biomass<sup>28,29</sup>, it does not necessarily seem to be an indicator of heterosis in *Arabidopsis*.

Carbohydrates or sugars are a primary source of carbon and energy and play a fundamental role in plant growth regulation<sup>32–34</sup>. *Arabidopsis* allotetraploids were reported to contain more starch and sugars, as well as increased biomass, with almost 12% more chlorophyll and 10% more starch than the BPV<sup>35</sup>. Using a set of recombinant inbred lines from a cross of Col-0 and C24 ecotype, carbohydrates or sugars-related metabolites such as glucose-6-phosphate, fructose-6-phosphate, or sucrose have significantly correlated to biomass<sup>14</sup>. Therefore, we assessed whether sugar derivatives were associated with the heterotic characteristics in several F<sub>1</sub> hybrids. In this study, the hybrids were grown under short-day conditions without sucrose added to the medium, thus limiting the carbon source required for growth. We hypothesized that if sucrose is involved in the heterosis mechanism, then differences in sugar-induced carbohydrate metabolic pathways in hybrids should emerge. Sucrose may play a role in the induction of faster growth rates, especially during early development. In our study, trehalose tended to be lower in all F<sub>1</sub> hybrids compared with inbred lines under short-day conditions without sucrose (Dataset S3). However, in our study, there were no clear differences in the overall carbohydrate metabolic pathway and no changes in metabolites at different levels of heterosis. In addition, the sugar-transport-responsive *STP1*, *STP14*, *ERD6*, and *PLT6* genes were upregulated in the C24 × *Ler*, C24 × Col, and Col × *Ler* hybrids<sup>22</sup>. Nevertheless, our transcriptome data showed no significant upregulation of these genes in Col × C24 and C24 × Col compared with their parents (Fig. S8). Moreover, starch metabolism is essential for growth<sup>36</sup>. Therefore, we also examined the starch accumulation in the hybrids of Col × C24 (high heterosis), Col × *Ws-0* (low heterosis), and *Ws-0* × *Ler-1* (no heterosis) to determine whether starch contributed to heterosis. Under both ZT6 and ZT11 tested, the reciprocal hybrids at all heterosis classes had starch levels no greater than the parental lines (Fig. S9). These results suggest that the carbohydrate metabolite profiling and starch are not altered when biomass heterosis occurs in the F<sub>1</sub> hybrids of *Arabidopsis* at least 15DAS young growth stage under short-day growing conditions.

In oxygen metabolism, the TCA cycle is essential for large energy intermediates<sup>17</sup>. Also, regarding the close association between biomass increase and metabolic composition, it has been reported that central metabolic pathways, i.e., the TCA cycle and amino acid pathway, are negatively correlated with biomass gain<sup>14</sup>. A QTL analysis in recombinant inbred lines (RILs) and the closely related species Col × C24 and C24 × Col revealed that *FUM2* was a candidate gene for heterosis<sup>37</sup>. Recently, associations between heterotic growth traits and fumarate and malate levels, as well as high fumarate/malate ratios, were detected in large populations of *Arabidopsis* accessions, RILs, near-isogenic lines, hybrids, and different photoperiods<sup>18,19</sup>. In this study, several intraspecific hybrids were used to identify metabolites associated with growth. As a result, we observed higher fumarate/malate ratios in hybrids showing high heterosis levels, similar to that reported previously (Fig. 6). Thus, the TCA cycle appears to be a core factor associated with vigorous growth. In oxygen metabolism, the TCA cycle is essential for large energy intermediates<sup>17</sup>. Fumaric acid has a lower pK<sub>a</sub> value than that malic acid, and a higher fumarate/malate ratio provides an advantageous mechanism for the pH-mediated regulation of intracellular flux and biomass<sup>23</sup>. In *Asparagus sprengeri*, a decrease in cytosolic pH is associated with  $\gamma$ -aminobutyrate (GABA) synthesis (glutamate + H<sup>+</sup> ≥ GABA + CO<sub>2</sub>)<sup>38</sup>. The *fum2* knock-down mutant exhibited a higher malate content, whereas the fumarate content remained low, and a decrease in biomass was detected under high-nitrogen conditions. This indicates that nitrate is not a major source of ammonium and that fumarate and malate help in maintaining the pH as nitrate is converted to ammonium<sup>23</sup>. The final core function of the TCA process is the production of energy<sup>39</sup>. Moreover, evidence has emerged in support of higher TCA cycle fluxes, photosynthetic capacity, and crop yields afforded by stomatal opening and closing movements<sup>40–42</sup>. At the protein level, enzymes that participate in the TCA and Calvin–Benson cycles were broadly upregulated, indicating an increase in the flux to provide more energy for enhancing growth at longer photoperiods<sup>43</sup>. The details of the upregulation of photosynthesis by controlling fumarate and malate concentrations remain unclear<sup>40,42</sup>. In this report, the TCA cycle may have been regulated in the parental accessions, thus limiting the energy flux necessary for growth. In the case of high heterosis, hybrids can accelerate the TCA flux and can produce more energy as biomass. Previously, reduced levels of TCA cycle intermediates were shown to result in enhanced rates of photosynthesis and aerial growth throughout carbon flux<sup>14,44</sup>. This negative correlation seems to indicate that strong growth is triggered to diminish the levels of central metabolites rather than an increase in the supply of the components used for cellular synthesis<sup>14</sup>. These examples reveal the importance of mitochondrial reactions for cellular function and, hence, for higher biomass in plants. Therefore, we hypothesized that photosynthesis and/or nitrogen metabolism occur when heterosis has a high TCA cycle flux. Because TCA cycle-related genes do not exhibit any significant changes in their transcriptional levels, we assumed that the reduced levels observed in TCA cycle intermediates occurred at the post-transcriptional level or were caused by the action of specific regulatory factors. The detailed effects of these regulatory mechanisms involved in such interactions remain to be studied.

In conclusion, we report here that general physiological traits, such as seed area and germination time, were affected by the maternal donor. In addition, the minimization of environmental differences and interindividual variability, which affect biomass and developmental stages, led to various levels of heterosis and provided a taxonomy of combinations for further study. The metabolomics analysis of intraspecific *Arabidopsis thaliana* revealed that TCA fluxes seem to be involved in the acquisition of heterotic phenotypes. This will provide insight into the phenomena of heterosis and higher plant growth.

## Materials and methods

**Plant material and growth conditions.** A total of 202 accessions obtained from the *Arabidopsis* Biological Resource Center (the Ohio State University, U.S.) used in this study (Dataset S1) were planted at Tsukuba-Plant Innovation Research Center, University of Tsukuba, Tsukuba (36° 11' N and 140° 10' E) in 2017–2019. The experimental study of the plants, including the collection of plant material, complied with relevant institutional, national, and international guidelines and legislation. Seeds of both F<sub>1</sub> hybrids and parental controls were generated by hand pollination as reference<sup>13</sup> and were collected at a comparable time.

Seeds were sterilized, stored at 4 °C for three days, and then sown on square dishes containing 1/2 Murashige and Skoog (MS) salts with vitamins, pH 5.7, 1% (w/v) agar without sucrose. Plates were divided into four compartments comprising of two parental lines and their hybrids and then replaced in developmental chambers under short-day conditions (12 h in light/12 h in the dark) at 22 °C, fluxes of 120 μmol photons m<sup>2</sup> s<sup>-1</sup>. 5 DAS seedlings were transferred to 150 mm-diameter plates and developed under the same conditions. The dishes were rotated around the development chamber daily to ensure uniform lighting conditions<sup>16</sup>.

For seed germinating time, the complete opening of the cotyledons was recorded every 12 h after sowing (HAS) until 72 HAS. Data were collected every 4 h from 48 to 60 HAS as green cotyledons appeared within this period. The time of germination was defined as the time when 50% of the seed had fully opened cotyledons. Based on these data, the seed sowing times of the phenotypic experiments were matched to eliminate differences in germination time between the parents and their reciprocal hybrids.

**Measurement of seed area, leaf area, and shoot fresh weight.** After harvest, seed images were recorded by VHX-700F digital microscopy (KEYENCE, Japan), and their areas were measured with ImageJ software<sup>45</sup>. At least five individual plants were measured in each line. Leaf area and fresh weight data were collected as described<sup>19</sup>. Using ImageJ software, the leaf area of seedlings from 5 to 15 DAS was calculated<sup>44</sup>. Data on leaf numbers were collected by counting all visible leaves. Each plant's seedlings were weighed when they had 10 leaves and 15 DAS on them, which is known as their fresh weight.

The mid-parent value (MPV) was calculated as the average of the trait value of the two parents. The better-parent value (BPV) was defined as the trait value of the better-performing parent in the cross.

**Metabolite data with GC-TOF-MS.** Samples for GC-TOF-MS analysis were collected at zeitgeber time (ZT) = 6, along with documentation of shoot fresh weight at 15 DAS. The entire rosettes were harvested in 6 replicates per line at the same 10-leaf number stage, each replicate comprising two bulked plants; then, they were promptly frozen in liquid nitrogen and kept at -80 °C until extraction preparation.

Extraction of each sample was done by mixing each sample with a ratio of 25 mg of tissues (fresh weight) per 01 μl of extraction medium (methanol/chloroform/water [3:1:1 v/v/v]) containing 10 stable isotope reference compounds (Cambridge Isotope Laboratories, USA; Spectra Stable Isotopes, USA and C/D/N ISOTOPES, Canada) using a Retsch mixer mill MM310 at a frequency of 30 Hz for 3 min at 4 °C. Each isotope was adjusted to the final concentration of 22.5 ng/μl/injection. After centrifuged for 5 min at 15,100×g, 400 μl of the supernatant was used for further analysis. The extracts were dried in a Savant SPD2010 SpeedVac Concentrator (Thermo Electron Corporation, Waltham, MA, USA).

Each sample was then added with 20 μl of methoxyamine hydrochloride (20 mg/ml in pyridine) at room temperature for methoximation. After 30 h of derivatization, the sample was trimethylsilylated by shaking with 20 μl of *N*-methyl-*N*-trimethylsilyltrifluoroacetamide with 1% trimethylchlorosilane (Pierce, USA) at 37 °C in 1 h, then added with 20 μl of *n*-heptane. In the vacuum glove box, VSC-100 (Sanplatec, Japan) was packed with 99.9995% (G3 grade) dry nitrogen, and all derivatization steps were carried out. The analysis of metabolites by GC-TOF/MS has been described previously<sup>46</sup>.

**Quantification of starch.** Starch quantification was performed as previously described<sup>47</sup>, with slight modifications. Seedlings were grown for 15 days in short-day conditions, and then the whole rosettes were harvested at ZT6 and ZT11 and weighed. Harvested rosettes were immediately frozen in liquid nitrogen, and two or three plants bulked to generate three replicates. Firstly, plant materials were ground and suspended in 1 ml 0.7 M ice-cold perchloric acid. Then, the soluble fractions were separated and removed by centrifugation (15,300 g, 15 min, 4 °C). The insoluble fractions were rinsed twice with water and washed out the chlorophyll with 80% (v/v) ethanol heated to 50 °C five times at least. The pellet containing starch was rinsed twice with water and resuspended in 1.2 ml of water. 500 μl of resuspension was dispensed to every two tubes, and one tube was used as blank. Reaction buffer and thermostable α-amylase were added and heated to 100 °C for 15 min, and 10 μl of α-amylglucosidase solution was added when the mixture cooled to 50 °C using a heat block for 5 min and incubated for 30 min. Finally, the starch was quantified by measuring the product of D-glucose using the glucose oxidase/peroxidase (GOPOD) method.

**Statistical analysis.** Student's *t*-test ( $P < 0.05$ ) was used to compare phenotypic data. For the metabolic analysis, all variables were transformed (log<sub>e</sub>); then, multiple *t*-tests were performed with a false discovery rate

(FDR) approach and a desired FDR of <math><0.05^{48}</math>. Box plot visualization and statistical analysis above box plots (Tukey's test,  $P < 0.05$ ) were performed using R software. Principal component analysis (PCA) was used in R software along with `pcaMethods` and `ggplot2` to display changes in the levels of 188 metabolites discovered by GC-TOF-MS. For the transcriptome analysis, one way-ANOVA was performed with an FDR approach and a desired FDR of <math><0.05</math>. Bar plot visualization and statistical analysis above bar plots (Tukey's test,  $P < 0.05$ ) were performed using R software.

**Expression analysis.** 15 DAS *Arabidopsis* seedlings from Col, C24, Rld-1, Ws-0, *Ler-1*, and their reciprocal  $F_1$  hybrids harvested at ZT6 were used as samples for RNA expression profiling. The RNA extraction, cDNA synthesis, and quantitative real-time polymerase chain reaction were performed as previously described<sup>49</sup>. Gene-specific primer sequences were as follows: *FUM1* (AT2G47510) (5'-TGCGTTATGCCACCTCTCTG-3' and 5'-TTGCTTTCCCAATCGTCGGA-3') and *FUM2* (AT5G50950) (5'-GTCGCATGTACTCTACCCCG-3' and 5'-GTGCAGCCAGAGCTTCAAAC-3'). Expressions of these target genes were normalized using *AT5G12240* as a control gene<sup>49</sup>. RNA-seq was performed for the  $F_1$  progeny of reciprocal crosses between the Col-0, C24, and their reciprocal  $F_1$  hybrids of 15 DAS seedlings. Sample preparation, sequencing, and data analyses by MinION (Oxford Nanopore Technologies, UK) were performed as previously<sup>50</sup>.

## Data availability

The raw sequence reads we generated in this study have been deposited at DDBJ Sequenced Read Archive under accession numbers DRX409928—DRX409939 of PRJDB14952.

Received: 16 December 2022; Accepted: 7 June 2023

Published online: 12 June 2023

## References

- Long-ping, Y. Development of hybrid rice to ensure food security. *Rice Sci.* **21**, 1–2 (2014).
- Hochholding, F. & Baldauf, J. A. Heterosis in plants. *Curr. Biol.* **28**, R1089–R1092 (2018).
- Cheng, S.-H. *et al.* Progress in research and development on hybrid rice: a super-domesticated in China. *Ann. Bot.* **100**, 959–966 (2007).
- Dan, Z. *et al.* Metabolic prediction of important agronomic traits in hybrid rice (*Oryza sativa* L.). *Sci. Rep.* **6**, 21732 (2016).
- Fujimoto, R. *et al.* Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. *Breed. Sci.* **68**, 145–158 (2018).
- de Abreu e Lima, F. *et al.* Metabolic robustness in young roots underpins a predictive model of maize hybrid performance in the field. *Plant J.* **90**, 319–329. <https://doi.org/10.1111/tpj.13495> (2017).
- Yang, M. *et al.* Genomic architecture of biomass heterosis in *Arabidopsis*. *PNAS* **114**, 8101–8106 (2017).
- Barth, S., Busimi, A., Utz, H. F. & Melchinger, A. Heterosis for biomass yield and related traits in five hybrids of *Arabidopsis thaliana* L. *Heynh. Heredit.* **91**, 36–42 (2003).
- Groszmann, M. *et al.* Intraspecific *Arabidopsis* hybrids show different patterns of heterosis despite the close relatedness of the parental genomes. *Plant Physiol.* **166**, 265–280 (2014).
- Moore, S. & Lukens, L. An evaluation of *Arabidopsis thaliana* hybrid traits and their genetic control. *G-Genes Genomes Genet.* **1**, 571–579 (2011).
- Miller, M. *et al.* Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of *Arabidopsis*. *Nat. Commun.* **6**, 7453 (2015).
- Moll, R., Lonnquist, J., Fortuno, J. V. & Johnson, E. The relationship of heterosis and genetic divergence in maize. *Genetics* **52**, 139–144 (1965).
- Meyer, R. C., Törjék, O., Becher, M. & Altmann, T. Heterosis of biomass production in *Arabidopsis*, Establishment during early development. *Plant Physiol.* **134**, 1813–1823 (2004).
- Meyer, R. C. *et al.* The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *PNAS* **104**, 4759–4764 (2007).
- Meyer, R. C. *et al.* Heterosis manifestation during early *Arabidopsis* seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. *Plant J.* **71**, 669–683. <https://doi.org/10.1111/j.1365-313X.2012.05021.x> (2012).
- Sugi, N. *et al.* Integrated transcript and metabolite profiling reveals coordination between biomass size and nitrogen metabolism in *Arabidopsis*  $F_1$  hybrids. *Plant Biotechnol.* **38**, 67–75 (2021).
- Sweetlove, L. J. *et al.* Not just a circle: Flux modes in the plant TCA cycle. *Trends Plant Sci.* **15**, 462–470 (2010).
- Riewe, D. *et al.* A naturally occurring promoter polymorphism of the *Arabidopsis FUM2* gene causes expression variation, and is associated with metabolic and growth traits. *Plant J.* **88**, 826–838 (2016).
- Le, Q. T. N. *et al.* Association analysis of phenotypic and metabolomic changes in *Arabidopsis* accessions and their  $F_1$  hybrids affected by different photoperiod and sucrose supply. *Plant Biotechnol. (Tokyo)* **36**, 155–165. <https://doi.org/10.5511/plantbiotechnology.19.0604a> (2019).
- Rajjou, L. *et al.* Seed germination and vigor. *Annu. Rev. Plant Biol.* **63**, 507–533 (2012).
- Fujimoto, R. *et al.* Heterosis of *Arabidopsis* hybrids between C24 and Col is associated with increased photosynthesis capacity. *PNAS* **109**, 7109–7114 (2012).
- Groszmann, M. *et al.* Hormone-regulated defense and stress response networks contribute to heterosis in *Arabidopsis*  $F_1$  hybrids. *PNAS* **112**, E6397–E6406 (2015).
- Pracharoenwattana, I. *et al.* *Arabidopsis* has a cytosolic fumarase required for the massive allocation of photosynthate into fumaric acid and for rapid plant growth on high nitrogen. *Plant J.* **62**, 785–795 (2010).
- Thimm, O. *et al.* MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939 (2004).
- Wang, L. *et al.* *Arabidopsis* Col/Ler and Ws/Ler hybrids and Hybrid Mimics produce seed yield heterosis through increased height, inflorescence branch and silique number. *Planta* **252**, 1–13 (2020).
- Castillo-Bravo, R. *et al.* Parent-of-origin effects on seed size modify heterosis responses in *Arabidopsis thaliana*. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2022.835219> (2022).
- Flood, P. J. *et al.* Reciprocal cybrids reveal how organellar genomes affect plant phenotypes. *Nat. Plants* **6**, 13–21 (2020).
- Wang, L. *et al.* Cotyledons contribute to plant growth and hybrid vigor in *Arabidopsis*. *Planta* **249**, 1107–1118 (2019).
- Wang, L. *et al.* PIF4-controlled auxin pathway contributes to hybrid vigor in *Arabidopsis thaliana*. *PNAS* **114**, E3555–E3562 (2017).
- Wang, L. *et al.* Hybrid mimics and hybrid vigor in *Arabidopsis*. *PNAS* **112**, E4959–E4967 (2015).

31. Zhu, A. *et al.* Early changes of gene activity in developing seedlings of Arabidopsis hybrids relative to parents may contribute to hybrid vigor. *Plant J.* **88**, 597–607 (2016).
32. Eveland, A. L. & Jackson, D. P. Sugars, signaling, and plant development. *J. Exp. Bot.* **63**, 3367–3377 (2011).
33. Patrick, W. J., Botha, C. F. & Birch, G. R. Metabolic engineering of sugars and simple sugar derivatives in plants. *Plant Biotechnol. J.* **11**, 142–156 (2013).
34. Lastdrager, J., Hanson, J. & Smeekens, S. Sugar signals and the control of plant growth and development. *J. Exp. Bot.* **65**, 799–807 (2014).
35. Ni, Z. *et al.* Altered circadian rhythms regulate growth vigor in hybrids and allopolyploids. *Nature* **457**, 327–331 (2009).
36. Sulpice, R. *et al.* Starch as a major integrator in the regulation of plant growth. *PNAS* **106**, 10348–10353 (2009).
37. Liseč, J. *et al.* Identification of heterotic metabolite QTL in *Arabidopsis thaliana* RIL and IL populations. *Plant J.* **59**, 777–788 (2009).
38. Crawford, L. A., Bown, A. W., Breitzkreuz, K. E. & Guinel, F. C. The synthesis of [gamma]-aminobutyric acid in response to treatments reducing cytosolic pH. *Plant Physiol.* **104**, 865–871 (1994).
39. Broddrick, J. T. *et al.* Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E8344–E8353 (2016).
40. Nunes-Nesi, A., Araújo, W. L. & Fernie, A. R. Targeting mitochondrial metabolism and machinery as a means to enhance photosynthesis. *Plant Physiol.* **155**, 101–107. <https://doi.org/10.1104/pp.110.163816> (2011).
41. Nunes-Nesi, A. *et al.* Deficiency of mitochondrial fumarate activity in tomato plants impairs photosynthesis via an effect on stomatal function. *Plant J.* **50**, 1093–1106. <https://doi.org/10.1111/j.1365-313X.2007.03115.x> (2007).
42. Araújo, W. L. *et al.* Metabolic control and regulation of the tricarboxylic acid cycle in photosynthetic and heterotrophic plant tissues. *Plant Cell Environ.* **35**, 1–21. <https://doi.org/10.1111/j.1365-3040.2011.02332.x> (2012).
43. Seaton, D. D. *et al.* Photoperiodic control of the *Arabidopsis* proteome reveals a translational coincidence mechanism. *Mol. Syst. Biol.* **14**, e7962 (2018).
44. Timm, S. *et al.* Mitochondrial dihydrolipoyl dehydrogenase activity shapes photosynthesis and photorespiration of *Arabidopsis thaliana*. *Plant Cell* **27**, 1968–1984 (2015).
45. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH image to imageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
46. Kusano, M. *et al.* Unbiased characterization of genotype-dependent metabolic regulations by metabolomic approach in *Arabidopsis thaliana*. *BMC Syst. Biol.* **1**, 53 (2007).
47. Liu, K. *et al.* Young seedlings adapt to stress by retaining starch and retarding growth through ABA-dependent and-independent pathways in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* **515**, 699–705 (2019).
48. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.* **57**, 289–300 (1995).
49. Czechowski, T. *et al.* Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* **139**, 5–17 (2005).
50. Seki, M. *et al.* Evaluation and application of RNA-Seq by MinION. *DNA Res.* **26**, 55–65 (2019).

## Acknowledgements

We thank Prof. Kazuki Saito (RIKEN Center for Sustainable Resource Science) for helpful discussions regarding the experiments. We are grateful to Ms. Kazuko Ito and Ms. Masumi Nakagawara (Tsukuba-Plant Innovation Research Center, University of Tsukuba), and Ms. Tomoko Nishizawa (RIKEN Center for Sustainable Resource Science) for their excellent technical assistance. Using of experimental equipment was supported by the Tsukuba-Plant Innovation Research Center, the University of Tsukuba, and the RIKEN Center for Sustainable Resource Science.

## Author contributions

H.S. conceived the research plans. Q.T.N.L., N.S., M.K., and H.S. designed all the experiments and conducted data. Q.T.N.L., N.S., M.Y., and T.H. performed research and analyzed data. M.K. and M.K. worked on the metabolic analysis. N.S., M.Y., and Y.S. worked on transcriptome analyses. Q.T.N.L., N.S., M.K., and H.S. wrote the manuscript. All of the authors critically read and approved the final manuscript.

## Funding

This work was supported in part by the Tsukuba Innovation Arena (TIA) collaborative research program “Kakehashi” (TK18-07 and TK19-062) to H.S. from TIA, and by the Grant-in-Aid for Scientific Research (C) (22K05928) and Funding Program for Next Generation World-Leading Researchers (NEXT Program) (GS018) to H.S. from JSPS. This research was also supported by the “Sustainable Food Security Research Project” in the form of an operational grant from the National University Corporation, Japan. Q.T.N. Le is supported by a doctoral scholarship from The Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-36618-y>.

**Correspondence** and requests for materials should be addressed to H.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023