Received: 18 December 2023 | Accepted: 5 February 2024

DOI: 10.1002/ajb2.16310

SPECIAL ISSUE ARTICLE

Botany

Halophytes and heavy metals: A multi-omics approach to understand the role of gene and genome duplication in the abiotic stress tolerance of *Cakile maritima*

Shawn K. Thomas[1,2,3] | **Kathryn Vanden Hoek**[4] | **Tasha Ogoti**[5] | **Ha Duong**[3,4] | **Ruthie Angelovici**[1,3] | **J. Chris Pires**[6] | **David Mendoza-Cozatl**[3,7] | **Jacob Washburn**[3,8] | **Craig A. Schenck**[3,4]

[1]Division of Biological Sciences, University of Missouri, Columbia, MO 65211, USA

[2]Bioinformatics and Analytics Core, University of Missouri, Columbia, MO 65211, USA

[3]Interdisciplinary Plant Group, University of Missouri, Columbia, MO 65211, USA

[4]Department of Biochemistry, University of Missouri, Columbia, MO 65211, USA

[5]Department of Computer Science, University of Missouri, Columbia, MO 65211, USA

[6]Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523-1170, USA

[7]Division of Plant Sciences and Technology, University of Missouri, Columbia, MO 65211, USA

[8]Plant Genetics Research Unit, USDA-ARS, Columbia, MO 65211, USA

Correspondence

Shawn K. Thomas, Division of Biological Sciences, University of Missouri, Columbia, MO 65211 USA.
Email: shawnkt4@gmail.com

Craig A. Schenck, Department of Biochemistry, University of Missouri, Columbia, MO 65211 USA.
Email: caschenck@missouri.edu

This article is part of joint special issues of the *American Journal of Botany* and *Applications in Plant Sciences*: "Twice as Nice: New Techniques and Discoveries in Polyploid Biology."

Abstract

Premise: The origin of diversity is a fundamental biological question. Gene duplications are one mechanism that provides raw material for the emergence of novel traits, but evolutionary outcomes depend on which genes are retained and how they become functionalized. Yet, following different duplication types (polyploidy and tandem duplication), the events driving gene retention and functionalization remain poorly understood. Here we used *Cakile maritima*, a species that is tolerant to salt and heavy metals and shares an ancient whole-genome triplication with closely related salt-sensitive mustard crops (*Brassica*), as a model to explore the evolution of abiotic stress tolerance following polyploidy.

Methods: Using a combination of ionomics, free amino acid profiling, and comparative genomics, we characterize aspects of salt stress response in *C. maritima* and identify retained duplicate genes that have likely enabled adaptation to salt and mild levels of cadmium.

Results: *Cakile maritima* is tolerant to both cadmium and salt treatments through uptake of cadmium in the roots. Proline constitutes greater than 30% of the free amino acid pool in *C. maritima* and likely contributes to abiotic stress tolerance. We find duplicated gene families are enriched in metabolic and transport processes and identify key transport genes that may be involved in *C. maritima* abiotic stress tolerance.

Conclusions: These findings identify pathways and genes that could be used to enhance plant resilience and provide a putative understanding of the roles of duplication types and retention on the evolution of abiotic stress response.

KEYWORDS

abiotic stress, amino acid, gene duplication, ionomics, polyploidy

Gene duplications are a source of evolutionary novelty. In plants, gene duplication events are abundant and have spurred the evolution of diverse adaptive responses to extreme environments (Vanneste et al., 2014; Panchy et al., 2016). However, evolutionary outcomes depend on which genes are retained, with most duplicated genes being lost. Although rare, retention of duplicates appears to be a non-random process (Rody et al., 2017). Retention and functional divergence depend on the type of duplication event and the original function of the gene that was

duplicated (Rody et al., 2017; Birchler and Yang, 2022). The two major mechanisms for duplicating genes are polyploidy events resulting in whole-genome duplications and triplications (WGD, WGT) and small-scale duplications (SSD) (Edger and Pires, 2009; Freeling, 2009; Conant et al., 2014; Conant, 2014; Vance and McLysaght, 2023). Retention of duplicate genes and functional divergence drive gene family expansions, which can lead to novel traits (Vanneste et al., 2014; Panchy et al., 2016; DeGiorgio and Assis, 2021). After polyploidy, redundancy, increased dosage, and rewiring of gene networks can lead to novel metabolic, regulatory, or developmental pathways key in stress biology (Van de Peer et al., 2021; Tossi et al., 2022).

Two major abiotic stresses impacting agriculture are salinization of soils and accumulation of heavy metals (Minhas et al., 2017; Gul et al., 2022). Halophytes (plants tolerant to salt) and heavy metal accumulators employ diverse strategies to mitigate stresses and thrive in inhospitable soils (Hasanuzzaman et al., 2014; Kazachkova et al., 2018). Some of these strategies include exclusion, uptake, and sequestration of salt and/or heavy metals, salt glands to exude excess ions, and alteration of metabolite profiles to cope with stress (Mendoza-Cózatl et al., 2011; Hasanuzzaman et al., 2014; Oh et al., 2015; Kazachkova et al., 2018; Yan et al., 2020). However, most crop species are glycophytes; that is, they are plants susceptible to salt stress (Hasanuzzaman et al., 2014). Since crop wild relatives can be locally adapted to diverse environments, some of which are impacted by salt and heavy metals (Lopes et al., 2015; Milla et al., 2018; Melino and Tester, 2023), they provide a source of genetic variation that may be useful in crop improvement strategies. Metabolically stress-adapted plants often accumulate osmoprotectants, such as the amino acid proline and sugars, before stress or in response to stress (Slama et al., 2015). Halophytes display a wide range of salt concentration tolerances, and depending on the threshold to define a halophyte, there are estimated to be anywhere from around 350–2600 halophytes distributed across the green plant phylogeny (Flowers and Colmer, 2008; Flowers et al., 2010; Bromham, 2015) Often, strategies for abiotic stress tolerance have evolved following gene duplication and neofunctionalization or subfunctionalization of duplicated genes (Schranz et al., 2012). Natural variation within crop wild relatives can be used to identify mechanisms by which these plants cope with abiotic stress and provide genetic tools to enhance crop resilience.

Cakile maritima Scop. (Brassicaceae) is an facultative halophyte (Arbelet-Bonnin et al., 2018, 2019) native to coastal dunes in the Mediterranean and has invaded coastlines across the globe. It is a crop wild relative that is closely related to canola and other *Brassica* crops (Arias and Pires, 2012; Walden et al., 2020; Hendriks et al., 2023). *Cakile maritima* exhibits typical halophytic behavior in response to salt stress such as exclusion, uptake, and sequestration (Hamed-Louati et al., 2016; Arbelet-Bonnin et al., 2018, 2019) and has been shown to be tolerant to heavy metals (Taamalli et al., 2014). Yet how *C. maritima* has evolved to tolerate abiotic stress is not well known. *Cakile maritima* shares the recent whole-genome triplication (WGT) with *Brassica* crops called Br-α WGT (Lysak et al., 2005, 2007; Tang et al., 2012; Emery et al., 2018; Qi et al., 2021) and is closely related to other salt-tolerant Brassicaceae including *Eutrema salsugineum* and *Schrenkiella parvula* that do not share the more recent Br-α WGT (Figure 1A) (Walden et al., 2020; Hendriks et al., 2023). Thus, *C. maritima* is well positioned phylogenetically to investigate the role of duplication events (SSD and WGT) on evolution of stress tolerance.

In this study, we investigated the role of retained duplicate genes in stress adaptation from the Br-α WGT and SSD events in *C. maritima* using a combination of ionomics, free amino acid profiling, and comparative genomics. We found that *C. maritima* can take up ions and overaccumulate stress metabolites. We identified expanded gene families, some of which show evidence of functional divergence, and hypothesized that these may contribute to the abiotic stress adaptation of *C. maritima*. These findings provide insight into the roles of duplication on abiotic stress tolerance.

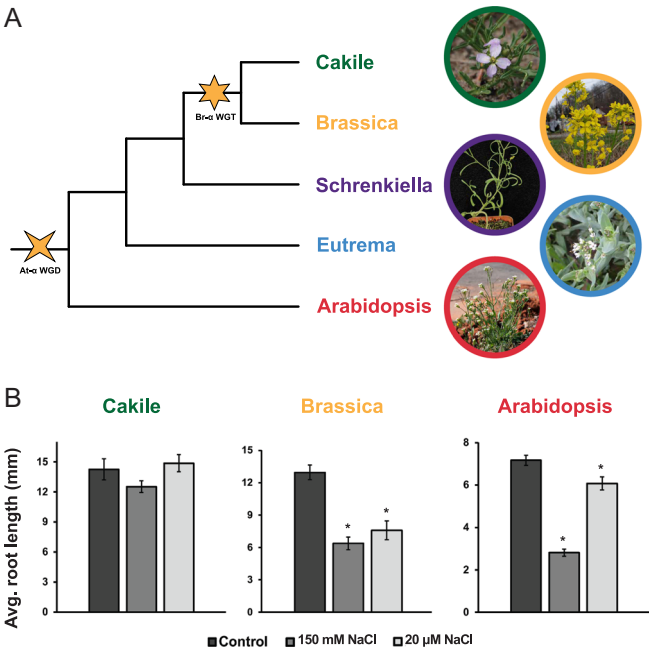

FIGURE 1 *Cakile maritima* is tolerant to salt and heavy metals and shares the Br-α WGT with *Brassica rapa*. (A) Cladogram depicting relationships between species used in this study with representative images. Ancient polyploidy events are indicated by the four- and six-pointed stars. *C. maritma* and *B. rapa* images are from the public domain (licensed under https://creativecommons.org/publicdomain/zero/1.0/). *Schrenkiella parvula* image provided by Dr. Pramod Pantha. *Eutrema salsugineum* image observed in Russia by Cepreй (licensed under https://creativecommons.org/licenses/by-nc/4.0/). *Arabidopsis thaliana* image observed in Germany by Wolfgang Jauch (licensed under https://creativecommons.org/licenses/by/4.0/) (B) *A. thaliana, B. rapa*, and *C. maritima* were grown for 6 days on ½ MS plates and moved to plates containing 150 mM NaCl (salt stress), 20 µM cadmium (Cd^{2+} heavy metal stress), or control plates, then grown for 5 days. Root length was measured using ImageJ. Bars represent mean root length ± SEM of N ≥ 6 biological replicates. $^{*}P < 0.01$ in unpaired *t*-test.

MATERIALS AND METHODS

Plant growth on plates

Seeds of *Arabidopsis thaliana* (L.) Heyn. Col-0 (Brassicaceae), *Brassica rapa* L. R500 (Brassicaceae), and *Cakile maritima* (henceforth referred to as Arabidopsis, Brassica, and Cakile; accession information is available in Appendix S1) were surface-sterilized with 90% v/v ethanol for 1 min, 15% v/v bleach for 12 min, and rinsed five times for 6 min each with sterilized water. Sterilized seeds were placed on plates containing half-strength Murashige and Skoog (½x MS) agar plus 1% w/v sucrose for germination. After 6 days, seedlings were moved to plates to induce different stresses: ½x MS agar + 0.5% sucrose supplemented with 150 mM NaCl for salt stress, ½x MS agar + 0.5% sucrose supplemented with 20 μM $CdCl_2$ for heavy metal stress, or ½x MS agar + 0.5% sucrose for the control (Tran et al., 2022). The NaCl and Cd^{2+} concentrations were chosen because they would impose a stress, but would not kill sensitive plants within the time frames of the experiments (Tran et al., 2022). Plants were grown at 23°C with a photosynthetic photon flux density of ~140 mM m^{-2} s^{-1} and 16 h light/8 h dark. After 5 days, plants on the plates were imaged with a camera (Canon PowerShot SX540 HS), and root length was measured using ImageJ software of $N \geq 6$ biological replicates (Figure 1; Appendix S2).

Hydroponics and ionomics

Cakile seeds were surface-sterilized and germinated on ½x Hoagland modified basal salt agar in plastic cups (236.5 mL [8 oz]) with dome lids to increase humidity. Once the seedlings were approximately 2.5 cm tall, the lids were removed, and the cups were placed in a flat tray filled with water. When seedlings had adapted to lower humidity levels (after approximately 6 days), they were transferred to hydroponics containers.

To house the hydroponics solution, we used black plastic cans (3.79 L [1 gallon]) with lids. In the lid, three large holes were cut to house a foam sponge with a slit down the middle. Cakile seedlings were placed inside the foam sponge, and the containers were filled with ½x Hoagland solution that had been aerated and adjusted to pH 5.6 as described by Nguyen et al. (2016). The solution was changed every 2 weeks. Hydroponics containers with plants were kept in a growth chamber set to 22–23°C, 55% humidity, photosynthetic photon flux density of 120–140 mM m^{-2} s^{-1}, and a 16 h light/8 h dark. Cakile plants were grown for 4 weeks before stress treatments.

Cakile plants were transferred to new hydroponics solutions containing 150 mM NaCl, 20 μM $CdCl_2$, and 150 mM NaCl + 20 μM $CdCl_2$ or ½x MS+0.5% sucrose control solution. Leaf and root tissue were collected at 0, 24, and 48 h of stress treatment to identify elemental changes occurring rapidly during stress treatment. Tissue samples were washed, processed, and analyzed using inductively coupled plasma optical emission spectroscopy (ICP-OES) following methods from Chen et al. (2006) for data on K^+, Na^+, Fe^{2+}, Zn^{2+}, and Cd^{2+} concentrations (μg/mg dry mass) (Figure 2). We monitored changes in these metals because K^+ and Zn^{2+} are essential nutrients and their levels can be affected by Na^+ (during salt stress) and Cd^{2+}, respectively (Hauser and Horie, 2010; Mendoza-Cózatl et al., 2011). Maintenance of high K^+/Na^+ ratios is an important phenotype in salt-tolerant plants (Hauser and Horie, 2010), and Cd^{2+} can disrupt Zn^{2+} transporter activity (Mendoza-Cózatl et al., 2011). Fe^{2+} was also measured because it is linked to stress-induced reactive oxygen species (ROS) accumulation, which leads to iron-sulfur cluster degradation (Zandalinas et al., 2020).

Amino acid extraction, detection, and quantification

Tissue was collected from plants grown in a growth chamber for analysis of amino acids without stress (Figure 3A; Appendices S3–S5). Whole seedlings were used from salt treatments on plates (Figure 3B; Appendix S6). All plants were grown with 16 h light/8 h dark at 22°C with a photosynthetic photon flux density at ~140 mM m^{-2} s^{-1}. Tissue was harvested and flash frozen in liquid nitrogen and lyophilized for 72 h. Lyophilized tissue was pulverized with 3-mm glass beads using a Spex SamplePrep 2010 Geno/Grinder (Cole-Parmer, Metuchen, NJ, USA), and free amino acids were extracted, detected, and quantified by liquid chromatography-tandem mass spectrometry (LC-MS/MS) as previously described (Angelovici et al., 2013; Yobi et al., 2020).

Identification of expanding and contracting gene families

The primary coding sequence (CDS) and amino acid sequences from Arabidopsis (Lamesch et al., 2012), Brassica (Lou et al., 2020), Cakile (Brassicales Map Alignment Project, DOE-JGI, http://bmap.jgi.doe.gov/), *Eutrema salsugineum* (Pall.) Al-Shehbaz & Warwick (Brassicaceae) (Yang et al., 2013) (henceforth referred to as Eutrema) and *Schrenkiella parvula* (Shrenk) D.A. German & Al-Shehbaz (Brassicaceae) (Dassanayake et al., 2011; Oh et al., 2014) (henceforth referred to as Schrenkiella) were obtained from Phytozome and the Brassicales Map Alignment Project (BMAP) project (with permission from the project's principal collaborators; data sources can be found in Appendix S1). Orthologous gene families (orthogroups) were inferred using amino acid sequences in OrthoFinder version 2.5.4 (Emms and Kelly, 2019) with parameters -t 20 and -a 20 to use 20 parallel analysis threads for sequence search, MSA and tree inference (-t 20), and internal, RAM intensive tasks (-a 20). All other parameters were left as default. Using the inferred orthogroups from Orthofinder, we examined gene family

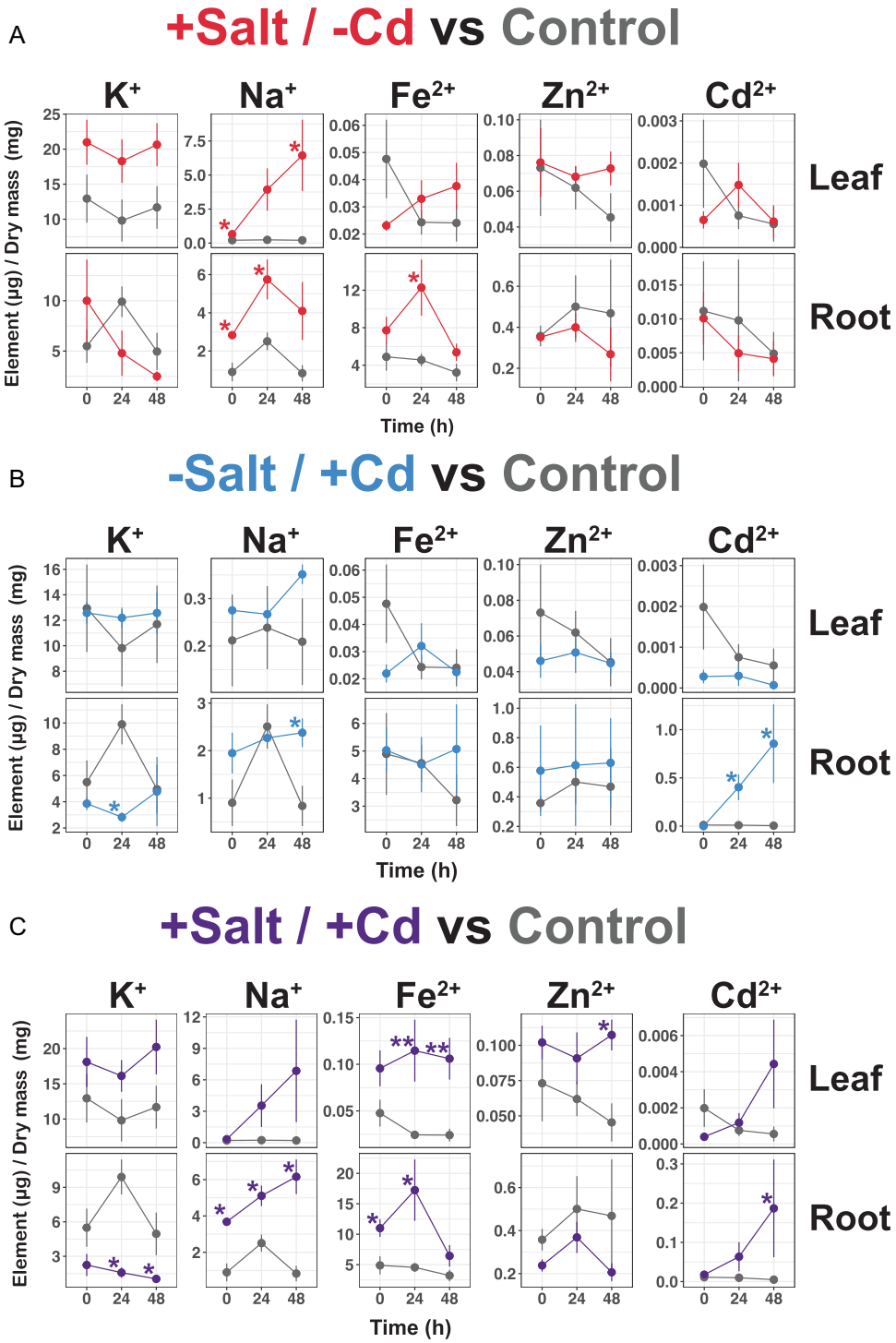

FIGURE 2 Elemental profile for *Cakile maritima* after abiotic stress. Plants grown hydroponically for 4 weeks, then grown under control conditions (-Salt/-Cd), (A) +Salt/-Cd (150 mM NaCl), (B) -Salt/+Cd (20 uM Cd^{2+}) or (C) +Salt/+Cd (150 mM NaCl + 20 uM Cd^{2+}). Following stress treatment, root and shoot tissue samples were harvested at 0, 24, and 48 h. Tissue samples were washed, processed, and elemental concentrations K^+, Na^+, Fe^{2+}, Zn^{2+}, and Cd^{2+} were measured using ICP-OES. Each graph shows measurements over time per each tissue and ion combination in stress treatments; red = +Salt/-Cd, blue = -Salt/+Cd, and purple = +Salt/+Cd) vs the control (grey = -Salt/-Cd). Points represent the mean ($N = 3$); error bars show standard error. Significant differences in the experimental treatments by time compared to the control (-Salt/-Cd) are denoted by an asterisk (*$P < 0.05$; **$P < 0.01$, Kruskal–Wallis test).

expansion and contraction of orthogroups using CAFE5 (Mendes et al., 2021). Gene counts per species for all orthogroups were extracted from the Orthofinder output, and orthogroups with more than 100 gene copies in one or more species were removed. Two hypotheses were tested to account for variation in evolutionary rates (λ) across the phylogeny: First, λ is constant across the tree and, second, λ changes after the Br-α WGT event. In the first hypothesis,

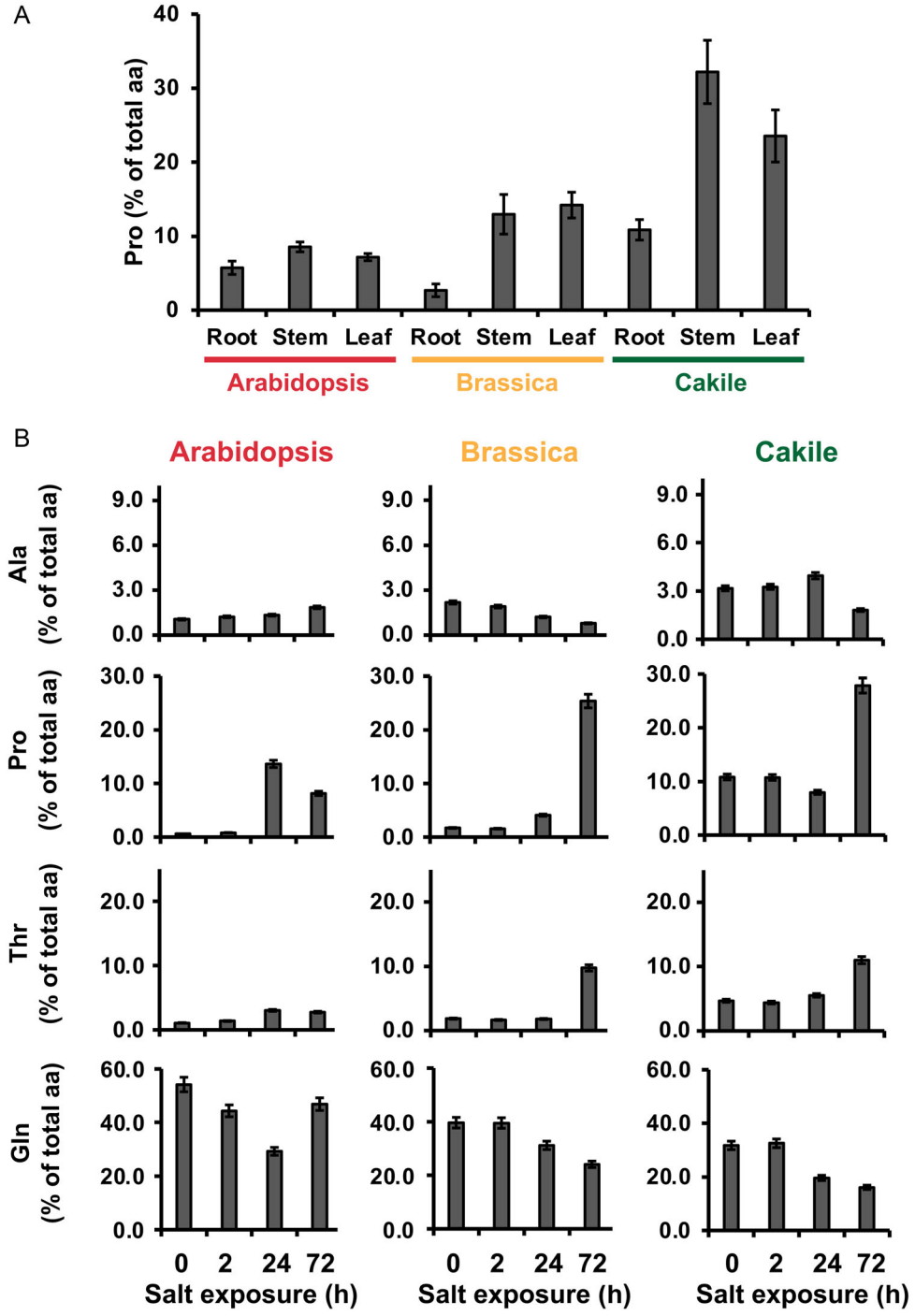

FIGURE 3 Proline accumulates in *Cakile maritima* before salt stress. (A) Proline composition of *Arabidopsis thaliana*, *Brassica rapa*, and *C. maritima* tissues grown under standard conditions. Pro levels were higher in *C. maritima* before stress in all tissues analyzed compared to *A. thaliana* and *B. rapa*. Bars are means ± SEM of $N \geq 4$ biological replicates. (B) Amino acid (aa) changes during salt stress. *A. thaliana*, *B. rapa*, and *C. maritima* were grown for 6 days on ½ MS plates and moved to plates containing 150 mM NaCl; roots and leaves from whole seedlings were analyzed for amino acids after 0, 2, 24, and 72 h of salt stress. Select amino acids are shown (full amino acid profiles: Appendix S6). Bars represent means ± SEM of $N \geq 4$ biological replicates.

one λ parameter was inferred for the whole tree, and in the second hypothesis, two separate λ parameters were inferred, one for the Br-α WGT clade (including Cakile and Brassica branches; Figures 1A, 4A) and another for the rest of the tree. The model with a separate λ parameter for the Br-α clade had a higher likelihood (final likelihood, –lnL: 234,507) than the model with constant λ parameter (final likelihood, –lnL: 244,248) and was used for downstream analyses. The λ parameters estimated for the Br-α WGT clade and the rest of the tree were ~0.022 and ~0.004, respectively. The cutoff for significant gene family expansions or contractions was set as $P < 0.05$. Orthogroups with significant

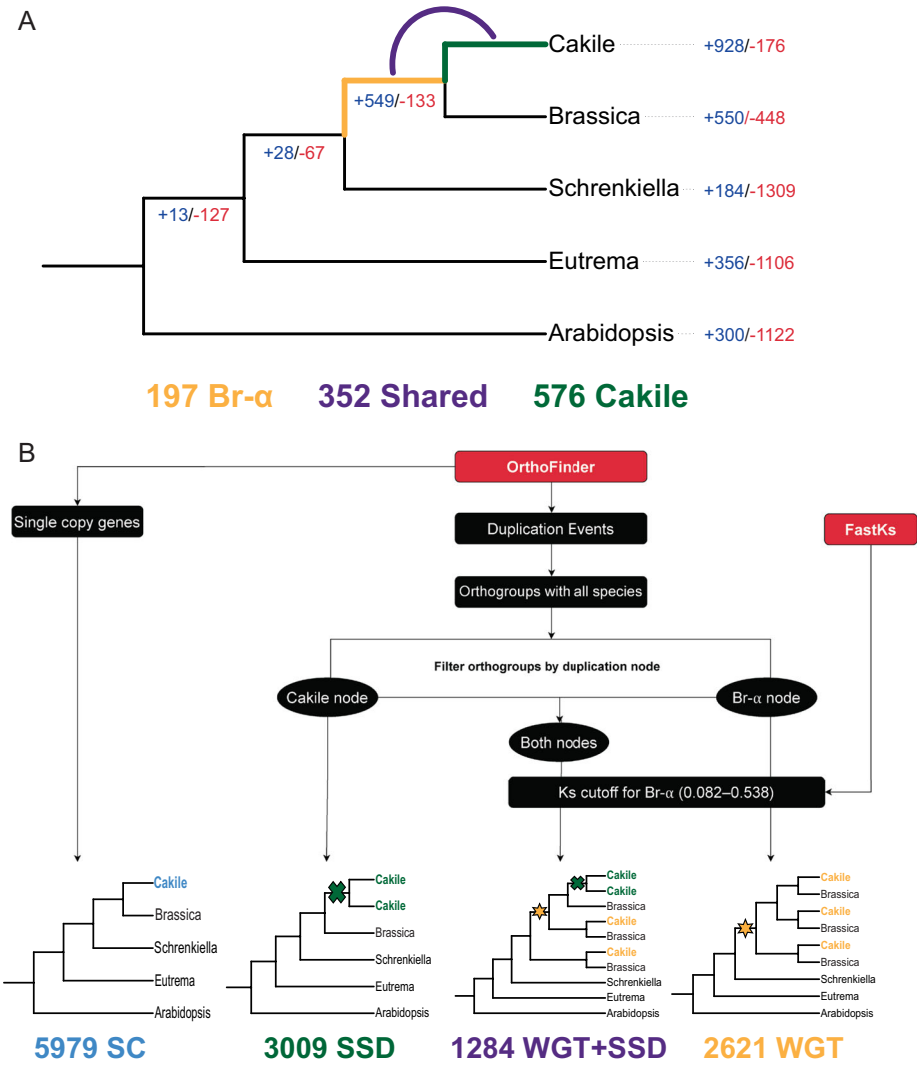

FIGURE 4 Expanded gene families and orthogroup classification for *Cakile maritima*. (A) Summary of CAFE results for significantly expanding/contracting orthogroups. The blue and red numbers respectively indicate the number of significant ($P < 0.05$) gene family expansions and contractions per branch as inferred by CAFE. Significantly expanding orthogroups at the Br-α and *C. maritima* branches were extracted and grouped as Br-α-specific, *C. maritima*-specific, and shared orthogroups. Counts for these groups are in gold, green, and purple, respectively. (B) Duplication classification workflow of *C. maritima* genes. Inferred orthogroups were classified as single copy (SC), small-scale duplication (SSD), whole-genome triplication (WGT), or WGT+SSD based on their duplication status in relation to the *C. maritima* and Br-α nodes. Phylogenetic tree under each classification show which branches are enriched for duplications and the numbers next to them represent the number of orthogroups in each class. Green X: SSD duplication event; yellow star: Br-α WGT.

expansions at the Cakile branch and the branch shared by Brassica and Cakile (Br-α branch) were binned into three sets: Cakile specific expansions, Br-α specific expansions, and expansions shared at both branches.

Identification and classification of duplicate gene classes

Orthogroups in which all five species (Cakile, Brassica, Arabidopsis, Eutrema, Schrenkiella) retained sequences were classified into duplication classes: single copy (SC), small-scale duplication (SSD), whole-genome triplication (WGT), and whole-genome triplication and small-scale duplication (WGT +SSD). This filtering was based on Cakile duplications enriched at branches of ortholog gene trees (Figure 4). The SC group represents genes maintained in single copies in all five species. The SSD group represents orthogroups enriched for duplications only at the Cakile branch in gene trees. The WGT group represents orthogroups enriched for duplications only at the branch shared by Brassica and Cakile (Br-α branch). The WGT+SSD group represents orthogroups that are enriched for duplications at both the Cakile and the Br-α branch. This approach for classifying duplicate genes in Cakile has some limitations as differential retention of duplicates in various lineages could skew the expected patterns described above. If gene loss is not consistent across species, differential retention of genes could make WGT genes appear as SSD as

genes in other species sharing the polyploidy were lost and not retained. Further it is possible for duplicates showing a WGT pattern to be a retained SSD in the ancestral lineage. For example, an SSD in the ancestor of Cakile and Brassica would be inferred as being derived from WGT as it would appear as a duplication at the Br-α branch.

The WGT groups (WGT and WGT+SSD) were subjected to additional filtering based on the distribution of the synonymous divergence (Ks) between paralogs in Cakile. The FASTKs pipeline (McKain et al., 2016) was used to obtain the Ks distribution for Cakile and assess pairwise ortholog divergence between Cakile and Brassica to Arabidopsis, and Cakile to Brassica. In this pipeline, CDS sequences for each sample were aligned against themselves with an e-value cutoff of 1e-40 using BLAST (Altschul et al., 1990). Putative pairs were identified as BLAST hits when they met the following criteria: (1) greater than 40% identity in alignment, (2) not an exact match (100% identity), and (3) longer than 300 bp of alignment length. Amino acid sequences for the putative pairs were aligned with MUSCLE (Edgar, 2004) and converted to CDS using PAL2NAL (Suyama et al., 2006). Ks, Ka, and Ka/Ks were estimated for the aligned pairs using codeml in PAML v.4.8 (Yang, 2007) with the same parameters used by McKain et al. (2012). The previously identified WGT groups were filtered to only include orthogroups containing Cakile genes within the Ks range of 0.082 to 0.538 indicative of the Br-α WGT (Qi et al., 2021). Despite some inferential limitations, as stated above, we expect that large-scale patterns of gene duplication classes described in this study are accurate.

Functional enrichment of orthogroup classes

The previously identified significantly expanding orthogroups and duplication classes were assessed for Gene Ontology (GO) term enrichment. Orthologous Arabidopsis genes in these classes were used as the input to gain a putative understanding of functional binning of all biological processes and specific processes related to salt, metal, amino acid metabolism, and ROS. For each class, GO enrichment analysis was performed using ShinyGO 0.77 (Ge et al., 2020). Arabidopsis was used as the reference database and the parameters FDR cutoff = 0.05 and Pathway size min-max = 20–2000 were used. After enrichment, the resulting FDR enrichment values and GO Biological Process terms were extracted for the duplication classes. These were used as inputs for REVIGO (Supek et al., 2011) to cluster terms based on semantic similarity.

Associated loci to the GO terms response to salt stress (GO:0009651) and response to cadmium ion (GO:0071276) from The Arabidopsis Information Resource (TAIR) database were extracted and mapped to duplication classes to gain a specific understanding of salt and cadmium functional binning per duplication class. For each duplication class, GO enrichment analysis was performed using ShinyGO version 0.77 (Ge et al., 2020). Arabidopsis was used as the reference database and all other parameters were default (FDR cutoff = 0.05; pathway size min-max = 2–2000). After enrichment, the resulting FDR enrichment values, GO molecular function, and KEGG terms were extracted.

RESULTS

Cakile is tolerant to salt and cadmium stress

To quantify the degree of abiotic stress tolerance, we challenged Cakile along with other Brassicaceae lineages Arabidopsis and Brassica, both known to be salt and cadmium sensitive, with salt and cadmium stress. Arabidopsis and Brassica showed significantly shorter roots and photobleaching on cadmium and salt (Figure 1B; Appendix S2), whereas Cakile was not affected by these treatments and maintained root growth (Figure 1B; Appendix S2).

Cakile uptakes salt and cadmium in a tissue-specific manner

To investigate the ability of Cakile to accumulate elements in a tissue-specific manner, we performed elemental profiling (ionomics) from various tissues of Cakile grown in hydroponic media during stress treatments. Upon the addition of salt in the hydroponic media, Na^+ levels increased slightly in both tissues over the 48 h of salt treatment (Figure 2). K^+ levels were consistently higher in leaves compared to roots, with a slight decrease in the roots over the 48-h period (Figure 2). Fe^{2+} levels showed a transient spike in the roots after 24 h and then returned to baseline levels, whereas they did not change in the leaf tissue (Figure 2). Zn^{2+} levels did not change substantially upon salt stress and as expected, Cd^{2+} was not detectable (Figure 2). When cadmium was applied in the hydroponic media, Cd^{2+} levels increased in the roots, but not the leaves, indicating that Cakile does not exclude Cd^{2+}, but does, at least in this context, take up Cd^{2+} in its roots (Figure 2). Cakile does not appear to move Cd^{2+} into its leaves within 48 h (Figure 2). Zn^{2+}, Fe^{2+}, K^+, and Na^+ levels were unaltered following cadmium stress (Figure 2). When salt and cadmium were applied together, the response was similar to the salt-only treatment, including a transient spike in Fe^{2+} at 24 h in the roots, but not leaves (Figure 2). Finally, there was a very minor increase in Cd^{2+} in the roots after 48 h, but not as dramatic a change as observed when Cd^{2+} stress was applied alone (Figure 2).

Cakile metabolic response during salt stresses

To better understand how Cakile metabolically responds to salt stress, we extracted and quantified total free amino acids from various tissues in the presence and absence of stress. In general, roots contained the least amount of total free amino acids and leaves the most across the species (Appendix S3).

Cakile had the lowest amount of total free amino acids compared to Arabidopsis and Brassica (Appendix S3), making comparisons of absolute levels of amino acids between species challenging. Additionally, changes in low-abundance amino acids can be masked by the most-abundant amino acids, for example, glutamine, glutamate, and aspartate (Appendix S4). Relative amino acid composition can capture important, but subtle, metabolic changes across the species; thus, percentage compositions (the amount of an individual amino acid over the sum of all free amino acids in a single extraction) were compared across species (Appendix S5). The percentage compositions of most amino acids were consistent across species and tissues, for example, serine, valine, and tyrosine (Appendix S5). One major exception was the relative proline (Pro) composition. As a percentage of the total free amino acids, Pro was dramatically higher in Cakile compared with Arabidopsis and Brassica (Figure 3A). In Cakile, Pro contributed to 10.8, 32.2, and 23.5% of the amino acid composition in the roots, stems, and leaves, respectively (Figure 3A), compared to 5.7, 8.5, and 7.2% Pro in Arabidopsis roots, stems, and leaves, respectively, and 2.7, 12.9, and 14.2% Pro in Brassica roots, stems, and leaves (Figure 3A).

Cakile, Arabidopsis, and Brassica were also investigated for amino acid composition changes during salt treatment. Whole seedlings were harvested at 0, 2, 24, and 72 h during salt treatment to analyze amino acids. Amino acid changes could be classified into three general categories during salt stress: no change, increase, or decrease. Amino acids such as alanine had no substantial change during salt treatment (Figure 3B; Appendix S6). Amino acids such as threonine and Pro increased during salt treatment; however, Pro increased substantially more in Arabidopsis and Brassica (13.5- and 15-fold, respectively) compared with Cakile (2.6-fold) during salt stress (Figure 3B; Appendix S6). Amino acids such as glutamine tended to decrease during salt treatment (Figure 3B; Appendix S6). A principal component analysis of the amino acid content during salt stress revealed that the three species could be distinctly grouped (Appendix S7A). The major amino acids contributing to the model were the branched chain amino acids (leucine, valine, and isoleucine), and amino acids involved in nitrogen metabolism including aspartate, asparagine, glutamate, and glutamine (Appendix S7B), whereas Pro was only a minor contributor to the model (Appendix S7B). Nitrogen metabolism amino acids were inversely correlated with Pro, suggesting that when Pro levels increased, generally nitrogen metabolism amino acids decreased (Figure 3B; Appendices S6, S7B).

Expanding and contracting Cakile gene families

To identify pathways and genes contributing to Cakile abiotic stress response, we first identified expanding and contracting gene families at the Br-α WGT branch and the Cakile specific branch in relation to the other Brassicaceae species used in the phylogenetic analyses. In total, 24,813 orthogroups were identified and after filtering served as the input for CAFE5 analysis (Mendes et al., 2021). Testing multiple evolutionary rate hypotheses identified a model where the Br-α clade had a separate evolutionary rate from the rest of the tree as the optimal model with the highest likelihood (final likelihood –lnL: 234,507). This model identified 549/708 significantly expanding/contracting orthogroups at the Br-α branch and 928/176 significantly expanding and contracting orthgroups at the Cakile specific branch (Figure 4A). GO enrichment of expanding orthogroups identified biological processes related to specialized metabolism, photosynthesis, response to amino acid and defense response enriched at Br-α; protein ubiquitination, proteolysis, RNA dependent DNA biosynthesis at the Cakile branch; RNA-dependent DNA biosynthesis, fungal defense response, pollen recognition, unsaturated fatty acid metabolism shared at both branches (Appendix S8).

Classification of duplicates in Cakile

We hypothesized that duplication and divergence has contributed to Cakile abiotic stress adaptation; therefore, expanded gene families in Cakile could arise from polyploidy events, including the Br-α WGT and/or Cakile-specific SSDs. Orthogroups were classified based on their duplication status: single copy (SC), small-scale duplication (SSD), whole-genome triplication (WGT), and whole-genome triplication plus small-scale duplication (WGT+SSD). These classes were inferred based on the phylogenetic context of orthogroup gene trees and the distribution of the synonymous divergence (Ks) between Cakile paralogs (Figure 4B). For example, when a single copy was phylogenetically distributed across all five species in a gene tree, these were classified as SC. Gene trees with duplications localized to a Cakile specific branch were classified as SSD, and gene trees classified as WGT only contained duplications at branches shared by Cakile and Brassica (Br-α). Finally, WGT+SSD represent cases where gene trees contained duplications at both Cakile-specific branches and Br-α branches. Of the 24,813 orthogroups, 17,152 were present across all five species (Cakile, Brassica, Schrenkiella, Eutrema, and Arabidopsis); 5979 of these were present as single-copy genes across all the species analyzed (Figure 4B); 3009 orthogroups were classified as resulting from SSD only and showed enrichment only in Cakile (Figure 4B); 2621 orthogroups were classified as resulting from WGT only; and 1284 were classified as resulting from both WGT and SSD (WGT+SSD) (Figure 4B). The remaining orthogroups were enriched for duplications only at nodes outside of Br-α and Cakile and were not included in downstream analyses.

Partitioning of biological process terms across Cakile duplication classes

Using GO term enrichment of biological processes, we discerned the partitioning of biological processes across

Cakile duplication classes and identified where terms related to salt and cadmium tolerance occurred. The Arabidopsis orthologs in orthogroups identified as different Cakile duplication classes and were subject to gene enrichment analyses. All significantly enriched terms per duplication class can be found in Appendix S9. In summary, 107, 65, 681, and 304 GO biological process terms were enriched (FDR < 0.05) in the SC, SSD, WGT and WGT+SSD duplication classes, respectively. The SC category contained GO terms associated with core cellular functions such as DNA repair, RNA modification, macromolecule modification, and plastid organization (Figure 5A). Transport and metabolic processes such as amide biosynthetic process, peptide biosynthetic process and cellular amide metabolic process were enriched in the SSD category (Figure 5B). The WGT category had many functionally enriched categories associated with regulation/regulatory elements, abiotic responses, and core developmental functions, such as regulation of metabolic

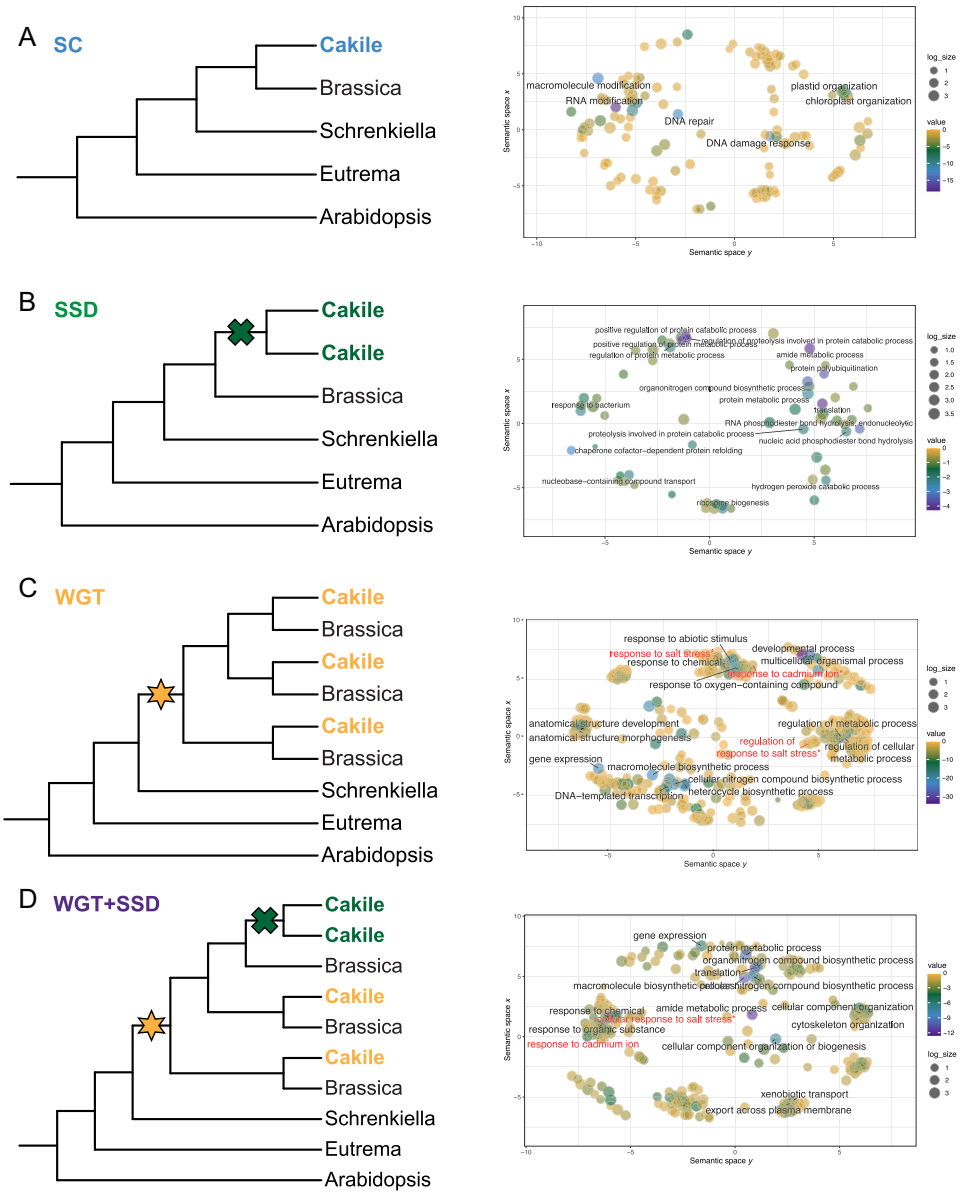

FIGURE 5 Gene ontology enrichment of *Cakile maritima* duplication classes. *Arabidopsis thaliana* genes from orthogroups were used as input for GO Enrichment using ShinyGO (Ge et al., 2020). Enriched GO biological process terms and their FDR enrichment values were used as inputs for REVIGO (Supek et al., 2011) to cluster semantically similar terms. The *x*- and *y*-axes represent semantic space used to cluster semantically similar GO terms (Supek et al., 2011). Displayed terms include the most significant terms based on EnrichmentFDR value (SC: value <−10; SSD: value <−2; WGT: value <−20; WGT +SSD: value <−4.59). Salt- or cadmium-specific terms are shown in red; terms with an asterisk did not meet the cutoff for the displayed term. The full list of significantly enriched categories can be found in Appendix S9. Color of circles is based on EnrichmentFDR value; size of circles indicates the number of terms clustered. (A) Single-copy (SC) orthogroups. (B) Small-scale duplication (SSD) orthogroups. (C) Whole-genome triplication orthogroups. (D) Whole-genome triplication + small-scale duplication (WGT+SSD) orthogroups. Phylogenetic trees show which branches are enriched for duplications. Green X: SSD duplication event; yellow star: Br-α WGT.

process, developmental process, and DNA-templated transcription (Figure 5C). Finally, the WGT+SSD category was enriched for response to cadmium ion, xenobiotic transport, amino acid metabolism, and terms related to core cellular process processes (Figure 5D).

Given the metabolic and ionomic response of Cakile following abiotic stress, we expected to identify biological processes related to salt, metal, amino acid metabolism, and ROS. Consistent with our hypothesis, salt-specific terms were found exclusively in the WGT (response to salt stress and regulation of response to salt stress) and WGT+SSD (response to salt stress and cellular response to salt stress) categories (Figure 5). Metal-related terms were identified across all duplication classes except SC, but cadmium-specific terms were identified only in the WGT and WGT+SSD categories (Figure 5). In relation to the high potassium abundance in salt-treated Cakile (Figure 2), potassium ion transport was found in the WGT category (Figure 5C). Amino acid metabolism terms were found in all categories except SC. Specifically, amide biosynthetic process and peptide biosynthetic process were among the top enriched terms in the SSD and WGT+SSD categories (Figure 5). This term was also enriched in the WGT category; however, it was the 523rd ranked term by FDR enrichment (Figure 5). Enrichment of amino-acid-related metabolic processes tend to be retained more commonly following Cakile-specific duplications and may be involved in metabolic pre-adaptation to abiotic stress. Finally, ROS-specific terms were found in all categories except for SC. Response to oxidative stress was found in the SSD and WGT categories, reactive oxygen metabolic process was found in SSD, and response to oxygen-containing compound was found in the WGT and WGT+SSD categories (Figure 5). Based on the trends observed, duplicate retention post polyploidy is enriched for regulation and stress terms, whereas Cakile-specific duplications are enriched in metabolic processes, highlighting the relevance of duplication type in various aspects of abiotic stress response.

Salt- and cadmium-responsive genes show evidence of duplication and divergence

Next, we investigated the duplication status and divergence of known salt- and cadmium-responsive genes to understand what pathways may be contributing to Cakile stress tolerance. Enriched GO molecular function and KEGG pathways were identified from known salt- (Figure 6) and cadmium-responsive genes (Appendix S10). There were 501 salt-responsive genes identified across the duplication classes and of those that were classified in SSD, WGT, and SSD+WGT, terms related to transport, cofactor, amino acid, and sugar metabolism were highly enriched (Figure 6A, B). Additionally, Pro metabolism was enriched in the SSD only category (Figure 6B).

To identify whether any of these duplicates show evidence of divergence, we determined Ka/Ks ratios between Cakile and Arabidopsis orthologs for salt-responsive genes (Figure 6C). Salt-responsive genes falling within the SC and WGT categories had low Ka/Ks ratios, indicating sequence conservation (Figure 6C), whereas salt-responsive genes within the SSD and WGT+SSD categories had a bimodal distribution and higher mean Ka/Ks ratios, which were significantly different from the distributions observed in the SC category (Figure 6C; Appendix S11). Ka/Ks ratios between Brassica and Arabidopsis, and Cakile and Brassica were also investigated (Appendix S12). Brassica and Arabidopsis had similar Ka/Ks distribution patterns to those of Cakile and Arabidopsis (Appendix S12A). However, Ka/Ks distribution patterns between Cakile and Brassica had higher means across all categories, and SSD, WGT, and WGT+SSD distributions of Cakile and Brassica all significantly differed from the SC distribution (Appendix S12B). Cakile and Brassica also shared more total ortholog pairs compared to Arabidopsis, which may influence Ka/Ks ratio distributions. Overall, higher Ka/Ks ratios between orthologs in the SSD and WGT+SSD categories indicate that some genes are diverging in sequence, perhaps indicators of sub- or neofunctionalization and abiotic stress response. However, in the absence of expression data, we cannot determine whether Cakile genes with high Ka/Ks ratios are expressed, which would be crucial for functionalization.

A similar analysis was performed using all the cadmium-responsive genes, which consists of a much smaller list than salt-responsive genes (42 cadmium-responsive genes). Some overlapping categories were identified in the SSD, WGT, and WGT+SSD categories including cofactor, amino acid, sugar, and specialized metabolism, but protein degradation categories were unique to cadmium-responsive genes (Appendix S10A, S10B). Ka/Ks ratios were calculated for all the cadmium-responsive genes, and we saw no significant differences between groups. (Appendix S10C). The low number of cadmium-responsive genes limits further interpretation of these data.

DISCUSSION

Duplications are important for stress tolerance

Duplications are a source of genetic material for the evolution of novel traits. Although gene loss is often the outcome following duplication, maintenance and functional divergence can lead to new functions (DeGiorgio and Assis, 2021). Expansions and contractions in gene families post polyploidy have contributed to the evolution of functional networks enabling crop domestication and environmental adaptation (Salman-Minkov et al., 2016; Zhang et al., 2019; Van de Peer et al., 2021; Xu et al., 2021; Tossi et al., 2022; Thomas et al., 2023). Abiotic stress tolerance is a multifaceted response encompassing morphological, transport, metabolic, and regulatory changes (Mickelbart et al., 2015; Slama

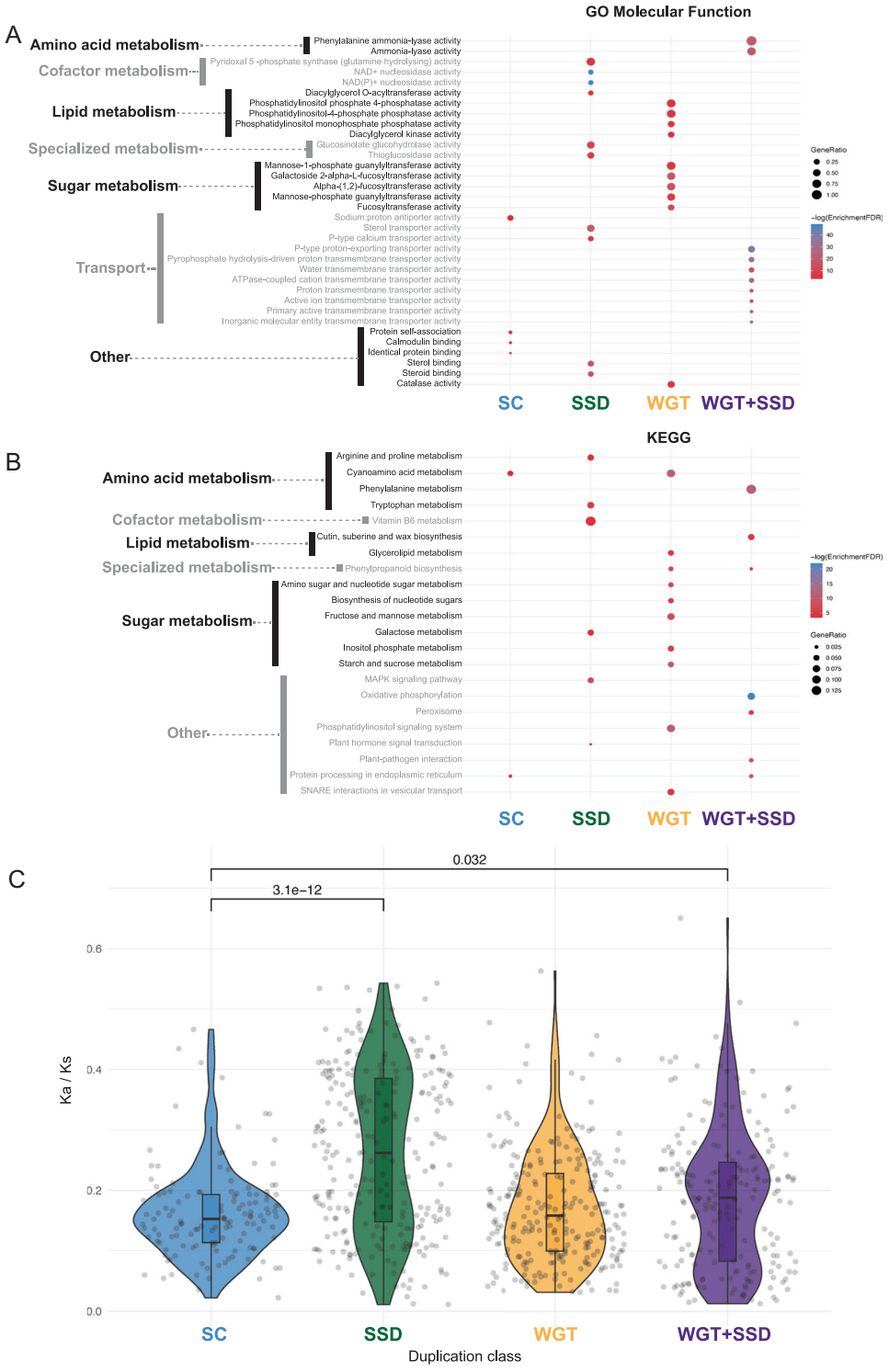

FIGURE 6 Top 10 GO terms/pathways per duplication class enriched with genes associated with response to salt stress. TAIR loci associated with response to salt stress (GO:0009651) found in duplication classes were used as inputs for functional enrichment using ShinyGO (Ge et al., 2020). The top 10 terms per duplication class for (A) GO molecular function and (B) KEGG. The size of circles indicates the proportion of genes in enrichment/total genes associated with the term/pathway; circle color indicates –log (Enrichment FDR). (C) Ka/Ks distributions per duplication class for loci associated with salt stress. Violin plots with nested boxplots of Ka/Ks ratios for loci associated with response to salt stress gene ontology term (GO:0009651). The *t*-test comparisons with $P < 0.05$ are displayed above distributions.

et al., 2015; Gul et al., 2022). In mangrove species, polyploid-derived genes have enabled environmental adaptations to fluctuating salinity (Xu et al., 2021; Thomas et al., 2023). Small-scale duplications (SSDs) also contribute to novel phenotypes; for example, genomes of salt-tolerant Brassicaceae species Eutrema and Schrenkiella are enriched for SSDs linked to salt tolerance (Kazachkova et al., 2018). In the present study, duplicated Cakile genes were classified as resulting from SSD, whole-genome triplication (WGT) or retained after both (WGT+SSD). The duplicated Cakile genes associated with putative salt and heavy metal responses are more likely retained in SSD, WGT, and WGT+SSD, suggesting that different duplication classes contribute to various aspects of abiotic stress tolerance in Cakile.

Dosage sensitivity explains patterns of duplicate retention

The ability of gene(s) and pathways to tolerate changes in copy number have dramatic impacts on the evolution of novel traits. The gene balance hypothesis proposes that the stoichiometric balance of genes whose products act in multi-subunit complexes (gene networks) is important for a complete and functional network (Birchler and Newton, 1981; Birchler et al., 2001; Edger and Pires, 2009; Freeling, 2009; Makino and McLysaght, 2010; Birchler and Veitia, 2012; Conant et al., 2014; Conant, 2014). Dosage sensitive genes are genes in which a change in copy number alters expression and protein abundance and may interrupt the stoichiometric balance of a functional gene network (Bird et al., 2023). Duplications in dosage sensitive genes are tolerated when the entire network is duplicated (i.e., via polyploidy) and gene balance is not perturbed; however, duplications such as SSD that only duplicate one or a few members of a network at a time are not tolerated because these disrupt network/gene balance. In contrast, dosage insensitive genes are more amenable to SSDs because they are not as highly connected and reliant on the stoichiometry of a network. This pattern of dosage sensitive genes tending to be over-retained post polyploidy and dosage insensitive genes tending to be over-retained after SSD is called reciprocal retention and has been shown in many studies (Edger and Pires, 2009; Freeling, 2009; Makino and McLysaght, 2010; De Smet et al., 2013; Conant et al., 2014; Conant, 2014; Tasdighian et al., 2017; Emery et al., 2018; Gout et al., 2023).

Overall, our results are consistent with patterns of reciprocal retention in Cakile after gene duplication. GO terms were classified as dosage sensitive, dosage insensitive, or neither (Appendix S13A) based on reciprocal retention of Arabidopsis duplicates (Song et al., 2020). In Cakile, the WGT class contained a higher proportion of dosage sensitive GO terms (Appendix S13A), including highly connected functions such as macromolecular complexes, transcriptional regulation, and signal transduction (Figure 5). On the other hand, the SSD class had a higher proportion of dosage insensitive compared to dosage sensitive genes, which may indicate that Cakile genes retained after SSD are involved in metabolism and biosynthetic processes (Figure 5). However, across all duplication classes, most GO terms were not classified as dosage sensitive or insensitive. Response to salt stress (GO:0009651) and response to cadmium ion (GO:0046686) GO terms were both found in the WGT and WGT+SSD categories, but were not found to be dosage sensitive or insensitive because they had a higher than average retention after both polyploidy and SSD (Appendix S13B). These data suggest that functions associated with abiotic stress adaptation have either weak or no dosage sensitivity and can diversify and contribute to adaptation through combinations of polyploidy and SSD.

Salt-tolerance functions are partitioned into various duplication classes

Based on patterns of reciprocal retention, we hypothesized that Cakile duplicated genes involved in different aspects of abiotic stress response would be partitioned into the SSD, WGT, and WGT+SSD classes. Within salt-response GO terms, we found enrichment in molecular function and metabolic pathways generally associated with amino acid metabolism, cofactor metabolism, lipid metabolism, specialized metabolism, and sugar metabolism and transport (Figure 6A, B). Amino acid metabolism terms were found across all duplication classes, and specifically, Pro metabolism was enriched in the SSD class (Figure 6B). Cofactor and specialized metabolism were mostly enriched in the SSD class (Figure 6A, B). Lipid and sugar metabolism terms were mainly partitioned into the WGT class and transport was mainly found in the WGT+SSD class (Figure 6). Our results are consistent with previous studies that show specialized metabolic genes are over-retained post SSD but not polyploidy, and metabolism and transport are often retained in multiple copies and are seen as drivers of environmental stress tolerance (Dunham et al., 2002; Gresham et al., 2008; Selmecki et al., 2006, 2015; Sunshine et al., 2015; Moore et al., 2019). Genes for core metabolic processes are more conserved compared with evolutionarily agile genes for specialized metabolism (Schenck and Last, 2020; Schenck and Busta, 2022), and we find that genes for lipid and sugar metabolism may be more constrained by dosage balance than those involved in specialized metabolism. Noncoding RNAs (ncRNA) play roles in the regulation of gene expression in response to stress (Di et al., 2014; Wang et al., 2017). Although ncRNA were not identified in this study, there are likely many in the Cakile genome that could contribute to stress tolerance and occur through different duplication types.

Ionomic response to abiotic stress

Plants that are salt and/or heavy metal tolerant have evolved diverse adaptive strategies, including exclusion, uptake, and

sequestration (Mendoza-Cózatl et al., 2011; Hasanuzzaman et al., 2014; Kazachkova et al., 2018; Yan et al., 2020). The elemental profiling suggests that Cakile can tolerate salt by accumulating Na^+ in leaf tissue and cadmium through Cd^{2+} sequestration in the root tissues with limited translocation to the leaves (Figure 2). The addition of salt in the hydroponic media tended to show a pattern of higher K^+ in leaf tissue; however, this result was not statistically significant (Figure 2). During salt stress, Na^+ ions can compete with K^+ ions for binding to K^+ transporters due to ionic similarities (Munns, 2002). Maintenance of high K^+/Na^+ ratios can mitigate ionic competition, and the pattern of high K^+ in Cakile leaf tissue could suggest that Cakile maintains high K^+/Na^+ ratios to mitigate salt stress. During cadmium treatments, there was a significant reduction of K^+ in roots at 24 h (cadmium only) and at 24 and 48 h (both stressors applied). Cadmium stress has been shown to negatively impact nutrient uptake (Qin et al., 2020; Naciri et al., 2021). With Zn^{2+}, there were no obvious patterns that emerged from the data; however, there was a significant increase of Zn^{2+} at 48 h in leaf tissue when salt and cadmium were applied. Interestingly, salt stress induced a transient spike in Fe^{2+} levels in the roots after 24 h, but Cd^{2+} stress alone had little effect on Fe^{2+} levels (Figure 2). Salt stress induces ROS, which can degrade Fe-S clusters, so perhaps Cakile increases Fe^{2+} uptake to compensate for this degradation (Zandalinas et al., 2020). We found ROS-related terms were enriched across all duplication classes and may be related to this pattern. When Cakile was grown on cadmium, an increase in Cd^{2+} was observed in the roots, but not the leaves, suggesting uptake in the roots but not translocation, at least within the 48-h window monitored (Figure 2). However, when salt and cadmium treatments were applied together, an increase in Cd^{2+} was observed in the roots; however, the increase was substantially lower than when cadmium was applied alone (Figure 2). Cakile may respond differently when multiple abiotic stresses are supplied, which could have led to the decreased Cd^{2+} uptake observed (Figure 2). These data provide initial insights into the ionomic changes during abiotic stress response in Cakile.

Genes for transporters have been reported as dosage sensitive (Blanc and Wolfe, 2004; Maere et al., 2005); therefore, we expected to find transporters enrichment in the WGT category. However, we found that transporters tolerated duplications post WGT and SSD because they were mainly found in the WGT+SSD category (Figure 6). HKT1 is a sodium transporter in Arabidopsis and other plants (Kazachkova et al., 2018). Natural variation of *AtHKT1* is associated with salt tolerance in some coastal Arabidopsis accessions (Busoms et al., 2018). Cakile possesses four *HKT1* orthologs in the SSD category, two of which have greater than mean Ka/Ks values (Appendix S11A, S14) when compared to Arabidopsis, and further comparing Ka/Ks ratios of Cakile and Brassica *HKT1* orthologs, we see similar levels of divergence. These data suggest that *HKT1* orthologs may have diverged in Cakile. Multiple *HKT1* orthologs are found in other salt-tolerant plants. In closely related Brassicaceae species, *HKT1* expansion was the result of SSD (Kazachkova et al., 2018), whereas in mangrove *HKT1* expansion is the result of polyploidy (Xu et al., 2021; Thomas et al., 2023). Another sodium transporter, AtNHX1, is also associated with salt tolerance in Arabidopsis (Pehlivan et al., 2016). Cakile has three *AtNHX1* homologs; however, none had high divergence rates (Appendix S11A). It is possible that expansion of *AtNHX1* in Cakile contributes to abiotic stress tolerance; however, this divergence could be through transcriptional or subcellular localization changes, which were not captured in this study. Thus, evolutionary context and environmental pressures influence the fates of duplicated genes.

Metabolic response to abiotic stress

Metabolically stress-adapted plants often accumulate osmoprotectants, such as Pro, before stress or in response to stress (Verslues and Sharma, 2010; Slama et al., 2015). Amino acid profiling of Cakile supports that it is metabolically prepared for stress (Figure 3). Before salt stress, Pro overaccumulates in Cakile compared to Arabidopsis and Brassica, and only minor changes were observed in Pro levels following salt stress in Cakile (Figure 3; Appendices S5, S6). Over 30% of the amino acid composition in Cakile stems consists of Pro before stress treatment (Figure 3A). Other Brassicaceae species including Schrenkiella and Eutrema show metabolic stress adaptation via increases in Pro content (Tran et al., 2022). Downregulation of Pro catabolic genes appear to enable Pro overaccumulation in these species and to lead to stress tolerance (Gong et al., 2005; Tran et al., 2022). Amino-acid-related processes, including Pro metabolism, were enriched in SSD categories and show evidence of sequence divergence (Figure 6), suggesting that Pro metabolism in Cakile may have diverged to support Pro accumulation. However, not all Pro biosynthetic or catabolic genes have greater than mean Ka/Ks (Appendices S11, S14), indicating that Cakile Pro metabolism may have diverged at a transcriptional level, like in other salt-tolerant species (Tran et al., 2022). Although Pro levels increased in Brassica and Arabidopsis following stress treatment (Figure 3B; Appendix S6), this increase did not translate into salt tolerance (Figure 1B), suggesting that Pro accumulation by itself is insufficient for some abiotic stresses and additional metabolic processes likely contribute to Cakile stress tolerance. Other metabolic processes including sugar and lipid metabolism were mostly enriched in WGT (Figure 6). Cytochrome P450 (CYP) genes play key roles in metabolic pathways related to plant development and stress response (Xu et al., 2015). Many CYPs were enriched in both the SSD (CYP71B families) and WGT +SSD (CYP72A and CYP709B families) classes, and several showed higher than mean Ka/Ks ratios (Appendices S11, S12, S14), which may suggest functional divergence. In particular, Cakile possesses two *CYP709B3* orthologs

with high sequence divergence, and this gene has been characterized in Arabidopsis to play a role in the regulation of salt stress response (Mao et al., 2013). These CYP duplicate genes may contribute to novel specialized metabolites in Cakile. Although we did not capture sugars, lipids, or specialized metabolites in our metabolic profiling, alteration in these pathways may also contribute to abiotic stress tolerance in Cakile. Large-scale metabolomic characterization could give further insights into patterns seen by the comparative genomics analyses.

CONCLUSIONS

This study identifies duplicated genes in Cakile, a mustard crop wild relative, which may mediate tolerance to salt and cadmium stress. Abiotic-stress-related genes in Cakile grouped into different duplication classes, and sequence divergence of some of these genes suggest potential sub-/neofunctionalization. Validation of these putative stress-related Cakile genes can provide experimental evidence into their specific functions in abiotic stress tolerance. This study demonstrates the power of using crop wild relatives because they can provide insights into the evolution of environmental adaptation. Further they provide novel genetic resources that can aid in the development of stress-resilient crops to meet global agricultural challenges.

AUTHOR CONTRIBUTIONS

The conception and design of the experiments were performed by S.K.T., C.A.S., D.M.C., J.C.P., R.A., and J.W. Data acquisition and analyses were performed by S.K.T., C.A.S., T.O., K.V.H., H.N.D., and D.M.C. The first draft of the manuscript was written by S.K.T., K.V.H., and C.A.S. All authors read, revised, and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The computational aspect of this work was done using the high-performance computing infrastructure provided by Research Computing Support Services and funded in part by the National Science Foundation under grant number CNS-1429294 at the University of Missouri, Columbia MO (MU). We thank Armando Moreno Geraldes and Kathryn Hodgins for providing Cakile seeds. We thank Emily Walter for help with hydroponics and Norma Castro-Guerrero for help with processing samples for elemental profiling. We thank Brian Mooney of the Gehrke Proteomics Center at MU for help with amino acid profiling. We thank Jim Birchler, Kevin Bird, and Makenzie Mabry for providing useful comments. We thank the Bioinformatics in Plant Science (BIPS) program at MU for providing funding for K.V.H. and T.O. (NSF MCB – 2224839). We thank the Bioinformatics and Analytics Core at MU for providing funding for SKT. We thank the Department of Energy Joint Genome Institute and collaborators for prepublication access to the *Cakile maritima* genome sequence. The work (proposal: 10.46936/10.25585/60000980) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. We thank the reviewers for careful reading and comments that have improved the manuscript.

DATA AVAILABILITY STATEMENT

Genome data used in this study were obtained from DOE-JGI Phytozome (https://phytozome-next.jgi.doe.gov/). Specific data sources can be found in Appendix S1. Scripts, raw and intermediate data available at https://github.com/shawnkt/HalophytesAndHeavyMetals.

ORCID

Shawn K. Thomas http://orcid.org/0000-0002-6198-8847
Kathryn Vanden Hoek http://orcid.org/0000-0001-6419-3364
Tasha Ogoti http://orcid.org/0000-0002-7813-4282
Ha Duong http://orcid.org/0000-0001-6087-746X
Ruthie Angelovici http://orcid.org/0000-0001-5150-0695
J. Chris Pires http://orcid.org/0000-0001-9682-2639
David Mendoza-Cozatl http://orcid.org/0000-0002-9616-0791
Jacob Washburn http://orcid.org/0000-0003-0185-7105
Craig A. Schenck http://orcid.org/0000-0002-5711-7213

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Angelovici, R., A. E. Lipka, N. Deason, S. Gonzalez-Jorge, H. Lin, J. Cepela, R. Buell, et al. 2013. Genome-wide analysis of branched-chain amino acid levels in *Arabidopsis* seeds. *Plant Cell* 25: 4827–4843.

Arbelet-Bonnin, D., I. Ben Hamed-Laouti, P. Laurenti, C. Abdelly, K. Ben Hamed, and F. Bouteau. 2018. Cellular mechanisms to survive salt in the halophyte *Cakile maritima*. *Plant Science* 272: 173–178.

Arbelet-Bonnin, D., I. Ben-Hamed-Louati, P. Laurenti, C. Abdelly, K. Ben-Hamed, and F. Bouteau. 2019. *Cakile maritima*, a promising model for halophyte studies and a putative cash crop for saline agriculture. *Advances in Agronomy*, 155: 45–78.

Arias, T., and J. C. Pires. 2012. A fully resolved chloroplast phylogeny of the brassica crops and wild relatives (Brassicaceae: Brassiceae): novel clades and potential taxonomic implications. *Taxon* 61: 980–988.

Birchler, J. A., U. Bhadra, M. P. Bhadra, and D. L. Auger. 2001. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Developmental Biology* 234: 275–288.

Birchler, J. A., and K. J. Newton. 1981. Modulation of protein levels in chromosomal dosage series of maize: the biochemical basis of aneuploid syndromes. *Genetics* 99: 247–266.

Birchler, J. A., and R. A. Veitia. 2012. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences, USA* 109: 14746–14753.

Birchler, J. A., and H. Yang. 2022. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell* 34: 2466–2474.

Bird, K. A., J. C. Pires, R. VanBuren, Z. Xiong, and P. P. Edger. 2023. Dosage-sensitivity shapes how genes transcriptionally respond to

allopolyploidy and homoeologous exchange in resynthesized *Brassica napus*. *Genetics* 225: iyad114.

Blanc, G., and K. H. Wolfe. 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16: 1679–1691.

Bromham, L. 2015. Macroevolutionary patterns of salt tolerance in angiosperms. *Annals of Botany* 115: 333–341.

Busoms, S., P. Paajanen, S. Marburger, S. Bray, X.-Y. Huang, C. Poschenrieder, L. Yant, and D. E. Salt. 2018. Fluctuating selection on migrant adaptive sodium transporter alleles in coastal *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* 115: E12443–E12452.

Chen, A., E. A. Komives, and J. I. Schroeder. 2006. An improved grafting technique for mature *Arabidopsis* plants demonstrates long-distance shoot-to-root transport of phytochelatins in Arabidopsis. *Plant Physiology* 141: 108–120.

Conant, G. C. 2014. Comparative genomics as a time machine: How relative gene dosage and metabolic requirements shaped the time-dependent resolution of yeast polyploidy. *Molecular Biology and Evolution* 31: 3184–3193.

Conant, G. C., J. A. Birchler, and J.C. Pires. 2014. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Current Opinion in Plant Biology* 19: 91–98.

Dassanayake, M., D.-H. Oh, J. S. Haas, A. Hernandez, H. Hong, S. Ali, D.-J. Yun, et al. 2011. The genome of the extremophile crucifer *Thellungiella parvula*. *Nature Genetics* 43: 913–918.

DeGiorgio, M., and R. Assis. 2021. Learning retention mechanisms and evolutionary parameters of duplicate genes from their expression data. *Molecular Biology and Evolution* 38: 1209–1224.

De Smet, R., K. L. Adams, K. Vandepoele, M. C. E. Van Montagu, S. Maere, and Y. Van de Peer. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences, USA* 110: 2898–2903.

Di, C., J. Yuan, Y. Wu, J. Li, H. Lin, L. Hu, T. Zhang, et al. 2014. Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant Journal* 80: 848–861.

Dunham, M. J., H. Badrane, T. Ferea, J. Adams, P. O. Brown, F. Rosenzweig, and D. Botstein. 2002. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences, USA* 99: 16144–16149.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Edger, P. P., and J. C. Pires. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research* 17: 699–717.

Emery, M., M. M. S. Willis, Y. Hao, K. Barry, K. Oakgrove, Y. Peng, J. Schmutz, et al. 2018. Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. *PLoS Genetics* 14: e1007267.

Emms, D. M., and S. Kelly. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20: 238.

Flowers, T. J., and T. D. Colmer. 2008. Salinity tolerance in halophytes. *New Phytologist* 179: 945–963.

Flowers, T. J., H. K. Galal, and L. Bromham. 2010. Evolution of halophytes: multiple origins of salt tolerance in land plants. *Functional Plant Biology* 37: 604–612.

Freeling, M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology* 60: 433–453.

Ge, S. X., D. Jung, and R. Yao. 2020. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36: 2628–2629.

Gong, Q., P. Li, S. Ma, S. Indu Rupassara, and H. J. Bohnert. 2005. Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. *Plant Journal* 44: 826–839.

Gout, J.-F., Y. Hao, P. Johri, O. Arnaiz, T. G. Doak, S. Bhullar, A. Couloux, et al. 2023. Dynamics of gene loss following ancient whole-genome duplication in the cryptic *Paramecium* complex. *Molecular Biology and Evolution* 40: msad107.

Gresham, D., M. M. Desai, C. M. Tucker, H. T. Jenq, D. A. Pai, A. Ward, C. G. DeSevo, et al. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genetics* 4: e1000303.

Gul, Z., Z.-H. Tang, M. Arif, and Z. Ye. 2022. An insight into abiotic stress and influx tolerance mechanisms in plants to cope in saline environments. *Biology* 11: 597.

Hamed-Louati, I. B., F. Bouteau, C. Abdelly, and K. B. Hamed. 2016. Impact of repetitive salt shocks on seedlings of the halophyte *Cakile maritima*. *Environmental Control in Biology* 54: 23–30.

Hasanuzzaman, M., K. Nahar, M. M. Alam, P. C. Bhowmik, M. A. Hossain, M. M. Rahman, M. N. V. Prasad, et al. 2014. Potential use of halophytes to remediate saline soils. *BioMed Research International* 2014: 589341.

Hauser, F., and T. Horie. 2010. A conserved primary salt tolerance mechanism mediated by HKT transporters: a mechanism for sodium exclusion and maintenance of high K^+/Na^+ ratio in leaves during salinity stress. *Plant, Cell & Environment* 33: 552–565.

Hendriks, K. P., C. Kiefer, I. A. Al-Shehbaz, C. D. Bailey, A. Hooft van Huysduynen, L. A. Nikolov, L. Nauheimer, et al. 2023. Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. *Current Biology* 33: 4052–4068.e6.

Kazachkova, Y., G. Eshel, P. Pantha, J. M. Cheeseman, M. Dassanayake, and S. Barak. 2018. Halophytism: What have we learnt from *Arabidopsis thaliana* relative model systems? *Plant Physiology* 178: 972–988.

Lamesch, P., T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, et al. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40: D1202–D1210.

Lopes, M. S., I. El-Basyoni, P. S. Baenziger, S. Singh, C. Royo, K. Ozbek, H. Aktas, et al. 2015. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *Journal of Experimental Botany* 66: 3477–3486.

Lou, P., S. Woody, K. Greenham, R. VanBuren, M. Colle, P. P. Edger, R. Sartor, et al. 2020. Genetic and genomic resources to study natural variation in *Brassica rapa*. *Plant Direct* 4: e00285.

Lysak, M. A., K. Cheung, M. Kitschke, and P. Bures. 2007. Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiology* 145: 402–410.

Lysak, M. A., M. A. Koch, A. Pecinka, and I. Schubert. 2005. Chromosome triplication found across the tribe *Brassiceae*. *Genome Research* 15: 516–525.

Maere, S., S. De Bodt, J. Raes, T. Casneuf, M. Van Montagu, M. Kuiper, and Y. Van de Peer. 2005. Modeling gene and genome duplications in eukaryotes. *Proceedings of the National Academy of Sciences, USA* 102: 5454–5459.

Makino, T., and A. McLysaght. 2010. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proceedings of the National Academy of Sciences, USA* 107: 9270–9274.

Mao, G., T. Seebeck, D. Schrenker, and O. Yu. 2013. CYP709B3, a cytochrome P450 monooxygenase gene involved in salt tolerance in *Arabidopsis thaliana*. *BMC Plant Biology* 13: 169.

McKain, M. R., H. Tang, J. R. McNeal, S. Ayyampalayam, J. I. Davis, C. W. dePamphilis, T. J. Givnish, et al. 2016. A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. *Genome Biology and Evolution* 8: 1150–1164.

McKain, M. R., N. Wickett, Y. Zhang, S. Ayyampalayam, W. R. McCombie, M. W. Chase, J. C. Pires, et al. 2012. Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). *American Journal of Botany* 99: 397–406.

Melino, V., and M. Tester. 2023. Salt-tolerant crops: time to deliver. *Annual Review of Plant Biology* 74: 671–696.

Mendes, F. K., D. Vanderpool, B. Fulton, and M. W. Hahn. 2021. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36: 5516–5518.

Mendoza-Cózatl, D. G., T. O. Jobe, F. Hauser, and J. I. Schroeder. 2011. Long-distance transport, vacuolar sequestration, tolerance, and transcriptional responses induced by cadmium and arsenic. *Current Opinion in Plant Biology* 14: 554–562.

Mickelbart, M. V., P. M. Hasegawa, and J. Bailey-Serres. 2015. Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nature Reviews Genetics* 16: 237–251.

Milla, R., J. M. Bastida, M. M. Turcotte, G. Jones, C. Violle, C. P. Osborne, J. Chacón-Labella, et al. 2018. Phylogenetic patterns and phenotypic profiles of the species of plants and mammals farmed for food. *Nature Ecology & Evolution* 2: 1808–1817.

Minhas, P. S., J. Rane, and R. K. Pasala. 2017. Abiotic stress management for resilient agriculture. Springer, Singapore.

Moore, B. M., P. Wang, P. Fan, B. Leong, C. A. Schenck, J. P. Lloyd, M. D. Lehti-Shiu, et al. 2019. Robust predictions of specialized metabolism genes through machine learning. *Proceedings of the National Academy of Sciences, USA* 116: 2344–2353.

Munns, R., 2002. Comparative physiology of salt and water stress. *Plant, Cell & Environment*. 25: 239–250. https://doi.org/10.1046/j.0016-8025.2001.00808.x

Naciri, R., M. Lahrir, C. Benadis, M. Chtouki, and A. Oukarroum. 2021. Interactive effect of potassium and cadmium on growth, root morphology and chlorophyll a fluorescence in tomato plant. *Scientific Reports* 11: 5384.

Nguyen, N. T., S. A. McInturf, and D. G. Mendoza-Cózatl. 2016. Hydroponics: a versatile system to study nutrient allocation and plant responses to nutrient availability and exposure to toxic elements. *Journal of Visualized Experiments* 113: e54317.

Oh, D.-H., B. J. Barkla, R. Vera-Estrella, O. Pantoja, S.-Y. Lee, H. J. Bohnert, and M. Dassanayake. 2015. Cell type-specific responses to salinity – the epidermal bladder cell transcriptome of *Mesembryanthemum crystallinum*. *New Phytologist* 207: 627–644.

Oh, D.-H., H. Hong, S. Y. Lee, D.-J. Yun, H. J. Bohnert, and M. Dassanayake. 2014. Genome structures and transcriptomes signify niche adaptation for the multiple-ion-tolerant extremophyte *Schrenkiella parvula*. *Plant Physiology* 164: 2123–2138.

Panchy, N., M. Lehti-Shiu, and S.-H. Shiu. 2016. Evolution of gene duplication in plants. *Plant Physiology* 171: 2294–2316.

Pehlivan, N., L. Sun, P. Jarrett, X. Yang, N. Mishra, L. Chen, A. Kadioglu, et al. 2016. Co-overexpressing a plasma membrane and a vacuolar membrane sodium/proton antiporter significantly improves salt tolerance in transgenic Arabidopsis plants. *Plant & Cell Physiology* 57: 1069–1084.

Qi, X., H. An, T. E. Hall, C. Di, P. D. Blischak, M. T. W. McKibben, Y. Hao, et al. 2021. Genes derived from ancient polyploidy have higher genetic diversity and are associated with domestication in *Brassica rapa*. *New Phytologist* 230: 372–386.

Qin, S., H. Liu, Z. Nie, Z. Rengel, W. Gao, C. Li, and P. Zhao. 2020. Toxicity of cadmium and its competition with mineral nutrients for uptake by plants: a review. *Pedosphere* 30: 168–180.

Rody, H. V. S., G. J. Baute, L. H. Rieseberg, and L. O. Oliveira. 2017. Both mechanism and age of duplications contribute to biased gene retention patterns in plants. *BMC Genomics* 18: 46.

Salman-Minkov, A., N. Sabath, and I. Mayrose. 2016. Whole-genome duplication as a key factor in crop domestication. *Nature Plants* 2: 16115.

Schenck, C. A., and R. L. Last. 2020. Location, location! Cellular relocalization primes specialized metabolic diversification. *FEBS Journal* 287: 1359–1368.

Schenck, C. A., and L. Busta. 2022. Using interdisciplinary, phylogeny-guided approaches to understand the evolution of metabolism. *Plant Molecular Biology* 109: 355–357.

Schranz, M. E., S. Mohammadin, and P. P. Edger. 2012. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Current Opinion in Plant Biology* 15: 147–153.

Selmecki, A., A. Forche, and J. Berman. 2006. Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. *Science* 313: 367–370.

Selmecki, A. M., Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shoresh, A. L. Sorenson, S. De, et al. 2015. Polyploidy can drive rapid adaptation in yeast. *Nature* 519: 349–352.

Slama, I., C. Abdelly, A. Bouchereau, T. Flowers, and A. Savouré. 2015. Diversity, distribution and roles of osmoprotective compounds accumulated in halophytes under abiotic stress. *Annals of Botany* 115: 433–447.

Song, M. J., B. I. Potter, J. J. Doyle, and J. E. Coate. 2020. Gene balance predicts transcriptional responses immediately following ploidy change in *Arabidopsis thaliana*. *Plant Cell* 32: 1434–1448.

Sunshine, A. B., C. Payen, G. T. Ong, I. Liachko, K. M. Tan, and M. J. Dunham. 2015. The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLoS Biology* 13: e1002155.

Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6: e21800.

Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34: W609–W612.

Taamalli, M., R. Ghabriche, T. Amari, M. Mnasri, L. Zolla, S. Lutts, C. Abdely, and T. Ghnaya. 2014. Comparative study of Cd tolerance and accumulation potential between *Cakile maritima* L. (halophyte) and *Brassica juncea* L. *Ecological Engineering* 71: 623–627.

Tang, H., M. R. Woodhouse, F. Cheng, J. C. Schnable, B. S. Pedersen, G. Conant, X. Wang, et al. 2012. Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* 190: 1563–1574.

Tasdighian, S., M. Van Bel, Z. Li, Y. Van de Peer, L. Carretero-Paulet, and S. Maere. 2017. Reciprocally retained genes in the angiosperm lineage show the hallmarks of dosage balance sensitivity. *Plant Cell* 29: 2766–2785.

Thomas, S. K., H. An, and J. C. Pires. 2023. Mangroves and multiplications: influence of genome duplications on salt tolerance. *Molecular Ecology* 32: 275–277.

Tossi, V. E., L. J. Martínez Tosar, L. E. Laino, J. Iannicelli, J. J. Regalado, A. S. Escandón, I. Baroli, et al. 2022. Impact of polyploidy on plant tolerance to abiotic and biotic stresses. *Frontiers in Plant Science* 13: 869423.

Tran, K.-N., G. Wang, D.-H. Oh, J. C. Larkin, A. P. Smith, and M. Dassanayake. 2022. Multiple paths lead to salt tolerance - pre-adaptation vs dynamic responses from two closely related extremophytes. *bioRxiv* https://doi.org/10.1101/2021.10.23.465591 [preprint].

Van de Peer, Y., T.-L. Ashman, P. S. Soltis, and D. E. Soltis. 2021. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* 33: 11-26.

Vance, Z., and A. McLysaght. 2023. Ohnologs and SSD paralogs differ in genomic and expression features related to dosage constraints. *Genome Biology and Evolution* 15: evad174.

Vanneste, K., G. Baele, S. Maere, and Y. Van de Peer. 2014. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Research* 24: 1334–1347.

Verslues, P. E., and S. Sharma. 2010. Proline metabolism and its implications for plant-environment interaction. *Arabidopsis Book* 8: e0140.

Walden, N., D. A. German, E. M. Wolf, M. Kiefer, P. Rigault, X.-C. Huang, C. Kiefer, et al. 2020. Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. *Nature Communications* 11: 3795.

Wang, J., X. Meng, O. B. Dobrovolskaya, Y. L. Orlov, and M. Chen. 2017. Non-coding RNAs and their roles in stress response in plants. *Genomics, Proteomics & Bioinformatics* 15: 301–312.

Xu, J., X. Y. Wang, and W. Z. Guo. 2015. The cytochrome P450 superfamily: key players in plant development and defense. *Journal of Integrative Agriculture* 14: 1673–1686.

Xu, S., Z. Guo, X. Feng, S. Shao, Y. Yang, J. Li, C. Zhong, et al. 2021. Where whole-genome duplication is most beneficial: adaptation of mangroves to a wide salinity range between land and sea. *Molecular Ecology* 32: 460–475.

Yan, A., Y. Wang, S. N. Tan, M. L. Mohd Yusof, S. Ghosh, and Z. Chen. 2020. Phytoremediation: a promising approach for revegetation of heavy metal-polluted land. *Frontiers in Plant Science* 11: 359.

Yang, R., D. E. Jarvis, H. Chen, M. A. Beilstein, J. Grimwood, J. Jenkins, S. Shu, et al. 2013. The reference genome of the halophytic plant *Eutrema salsugineum*. *Frontiers in Plant Science* 4: 46.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591.

Yobi, A., C. Bagaza, A. Batushansky, V. Shrestha, M. L. Emery, S. Holden, S. Turner-Hissong, et al. 2020. The complex response of free and bound amino acids to water stress during the seed setting stage in *Arabidopsis*. *Plant Journal* 102: 838–855.

Zandalinas, S. I., L. Song, S. Sengupta, S. A. McInturf, D. G. Grant, H.-B. Marjault, N. A. Castro-Guerrero, et al. 2020. Expression of a dominant-negative AtNEET-H89C protein disrupts iron-sulfur metabolism and iron homeostasis in *Arabidopsis*. *Plant Journal* 101: 1152–1169.

Zhang, K., X. Wang, and F. Cheng. 2019. Plant polyploidy: origin, evolution, and its influence on crop domestication. *Horticultural Plant Journal* 5: 231–239.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Taxon sampling and genome data sources.

Appendix S2. Representative images following salt and cadmium stress. Arabidopsis, Brassica, and Cakile were grown for 6 days on ½ MS plates and moved to plates containing 150 mM NaCl (salt stress), 20 μM cadmium (Cd^{2+} heavy metal stress) or control plates and grown for an additional 5 days.

Appendix S3. Total absolute free amino acid levels in Arabidopsis, Brassica, and Cakile tissues grown under standard conditions. Amino acids were extracted from root, stem, and leaf tissue under standard growth conditions in the absence of exogenous abiotic stress. Amino acids are shown as the sum of the absolute amount of each individual amino acid. Data are the mean of $N \geq 4$ biological replicates ± standard deviation.

Appendix S4. Absolute levels of individual amino acids in root, stem, and leaf tissue of Arabidopsis, Brassica, and Cakile grown under standard conditions in the absence of exogenous abiotic stress. The three-letter abbreviations are used for each amino acid. Data are the means of $N \geq 4$ biological replicates ± standard deviation.

Appendix S5. Free amino acid composition in leaf, root, and stems of Arabidopsis, Brassica, and Cakile grown under standard conditions in the absence of exogenous abiotic stress. Amino acids are expressed as a percentage of the total amino acid complement in different tissues. The three-letter abbreviations are used for each amino acid. Data are the mean of $N \geq 4$ biological replicates.

Appendix S6. Free amino acid composition of Cakile, Brassica, and Arabidopsis during salt stress. Amino acids were extracted from whole seedlings of Arabidopsis, Brassica, and Cakile under salt stress for different times. Plants were grown for 6 days on ½ MS plates and moved to plates containing 150 mM NaCl. Whole seedlings were collected and analyzed for amino acids after 0, 2, 24, and 72 h of salt stress. Amino acids are expressed as a percentage of the total amino acid complement at each timepoint. The three-letter abbreviations are used for each amino acid. Data are the mean of $N \geq 4$ biological replicates.

Appendix S7. PCA of amino acid changes during salt stress. (A) PCA of amino acid content from whole seedlings of Arabidopsis, Brassica, and Cakile during salt stress for different times. Amino acid content was normalized by subtracting means and dividing by the standard deviations. PC1 and PC2 accounted for 79.59% of variance. Symbol size is based on the squared cosine values (cos2). (B) Loadings plot shows relative contributions and correlations of each amino acid to changes observed during salt stress. The standard three-letter code is used to abbreviate each amino acid.

Appendix S8. GO biological processes enriched at significantly expanding gene families. TAIR loci found in significantly expanding gene families that are (A) Br-α branch specific, (B) shared at the Br-α and Cakile branches, and (C) Cakile branch specific were used as inputs for functional enrichment using ShinyGO (Ge; Jung and Yao, 2020). The top enriched GO biological process terms per group are shown. The size of circles indicates the number of genes, circle color indicates –log (Enrichment FDR), and fold enrichment is indicated by the length of the lines in the chart.

Appendix S9. Partitioning GO biological process terms across Cakile duplication classes. (A) Single-copy (SC) global enrichment. (B) Small-scale duplication (SSD) global enrichment. (C) Whole-genome triplication (WGT) global enrichment. (D) Whole-genome triplication+small-scale duplication (WGT+SSD) global enrichment.

Appendix S10. Top 10 GO terms/pathways enriched with loci associated with response to cadmium ion. TAIR loci associated with response to cadmium ion (GO:0046686) and found in duplication classes were used as inputs for functional enrichment using ShinyGO (Ge et al., 2020). The top 10 terms per duplication class for (A) GO molecular function and (B) KEGG per duplication class. The size of circles indicates the proportion of genes in enrichment/total genes associated with the term/pathway, and circle color indicates –log (Enrichment FDR). (C) Ka/Ks distributions per duplication class for response to cadmium ion loci. Violin plots with nested boxplots of Ka/Ks ratios for loci associated with response to cadmium ion (GO:0046686). The *t*-test comparisons with $P < 0.05$ are displayed above distributions.

Appendix S11. Ka, Ks, Ka/Ks values between (A) Cakile and Arabidopsis genes in the salt-specific enrichment and

their duplication class; (B) Cakile and Arabidopsis genes in the cadmium-specific enrichment and their duplication class; (C) Brassica and Arabidopsis genes in the salt-specific enrichment and their duplication class; (D) Cakile and Brassica genes in the salt-specific enrichment and their duplication class.

Appendix S12. Ka/Ks distributions per duplication class for response to salt stress loci. Violin plots with nested boxplots of Ka/Ks ratios for loci associated with response to salt stress (GO:0009651). The *t*-test comparisons with $P < 0.05$ are displayed above distributions. (A) Brassica vs. Arabidopsis. (B) Cakile vs. Brassica.

Appendix S13. Response to salt stress and response to cadmium ion GO Terms were not classifiable as dosage sensitive or insensitive. (A) Log10-transformed counts of enriched GO biological process terms with FDR < 0.05 per duplication class classified as dosage sensitive, insensitive, and NA per Song et al. (2020) assignments of GO terms. (B) Adaptation of Figure 1a from Song et al. (2020). Yellow data points indicate class I GO terms (putatively dosage balance insensitive; unbalanced > 0.38 and $\alpha < 0.3$), and blue data points indicate class II GO terms (putatively dosage balance sensitive; unbalanced < 0.38 and $\alpha > 0.3$). Response to salt stress (GO:0009651) and response to cadmium ion (GO:0046686) did not fall into either group and had unbalanced = 0.44, $\alpha = 0.35$ and unbalanced = 0.41, $\alpha = 0.37$, respectively.

Appendix S14. Arabidopsis gene annotations of Ka/Ks values above means between Cakile and Arabidopsis genes. (A) small-scale duplication (SSD) class. (B) whole-genome triplication+SSD (WGT+SSD) class.

How to cite this article: Thomas, S. K., K. Vanden Hoek, T. Ogoti, H. Duong, R. Angelovici, J. C. Pires, D. Mendoza-Cozatl, J. Washburn, and C. A. Schenck. 2024. Halophytes and heavy metals: A multi-omics approach to understand the role of gene and genome duplication in the abiotic stress tolerance of *Cakile maritima*. *American Journal of Botany* 111: e16310. https://doi.org/10.1002/ajb2.16310