Elucidation of arsenic detoxification mechanism in *Marchantia polymorpha*: the role of ACR3

Mingai Li[1,2#], Aurélien Boisson-Dernier[3], Daniela Bertoldi[4], Francisco Ardini[5], Roberto Larcher[4], Marco Grotti[5], Claudio Varotto[1,2#]

[1]Biodiversity, Ecology and Environment Area, Research and Innovation Centre, Fondazione Edmund Mach, via Mach 1, 38098, San Michele all'Adige, Trento, Italy

[2]NBFC, National Biodiversity Future Center, Palermo 90133, Italy

[3]Université Côte d'Azur, INRAE, CNRS, Institut Sophia Agrobiotech, 400 Route des Chappes, BP167, 06903 Sophia Antipolis Cedex, France

[4]Department of Food and Transformation, Technology Transfer Centre of Fondazione Edmund Mach, E. Mach 1, 38098 San Michele all'Adige (TN), Italy

[5]Department of Chemistry and Industrial Chemistry, University of Genoa, Via Dodecaneso 31, Genoa, Italy

Figure S1. Scheme of the genome-editing targets sites in the Mp*ACR3* gene and summary of the mutations. (A) Black rectangles represent exons and light grey lines introns. All features are drawn to scale. The purple and pink arrowheads indicate the target sites 1 (first exon) and 2 (second exon), respectively. (B) For each target site (underlined), the wild-type Mp*ACR3* genomic sequence is shown with the PAM sequence highlighted in red. Numbers above the DNA sequence indicate the base positions with respect to the start of the gene's CDS. The encoded amino acid sequence is reported under each DNA sequence. The symbol "//" indicates portions of the gene or protein sequences not shown to allow the display of the first stop codon in the mutant sequence. Two independent mutant sequences are summarized for each target site, Mp*acr3-1-5^{ge}* and Mp*acr3-2-1^{ge}* for target 1 and 2, respectively. Dashes indicate deletions, bases in blue mutations, and bases in green insertions.

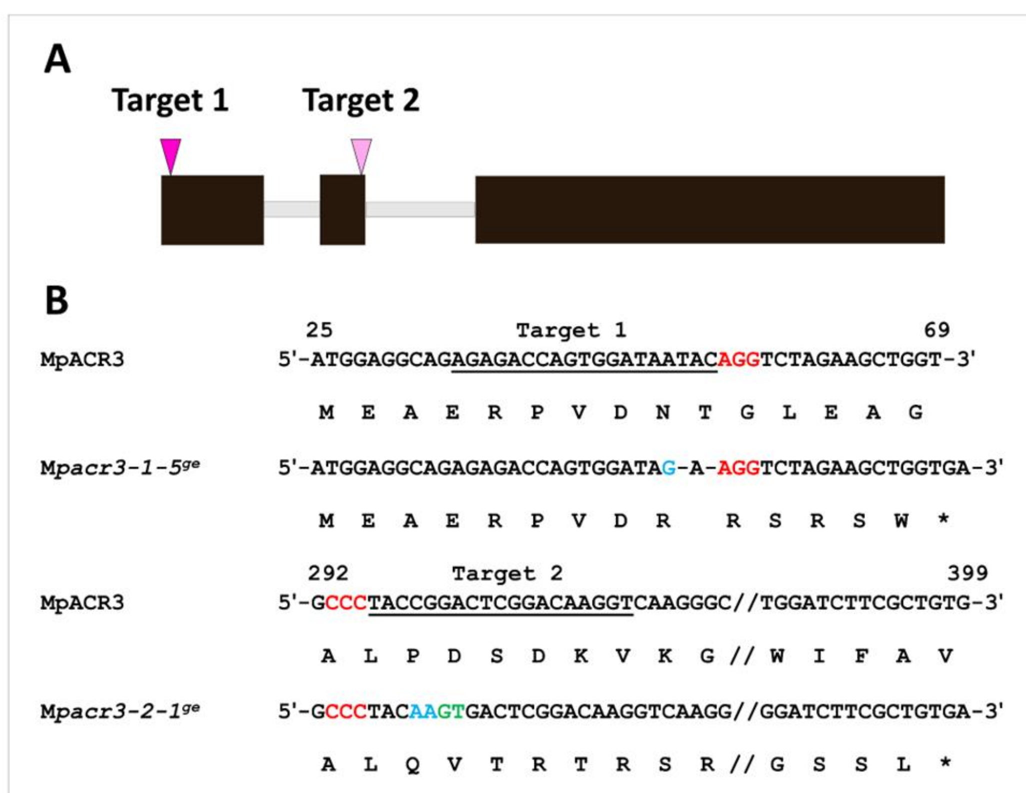

Fig. S2. Expression level analysis of Mp*ACR3* in Cam2 and 13 independent transgenic lines overexpressing Mp*ACR3* in *M. polymorpha* by semi-quantitative RT-PCR.

Figure S3. Total arsenic content of transgenic *M. polymorpha* lines with gain- and loss-of-function of *MpACR3*. The WT *M. polymorpha* ecotype (Cam2), two independent transgenic lines overexpressing *MpACR3* (Mp*ACR3-ox-9* and Mp*ACR3-ox-10*) and two independent *Mpacr3* knockout mutants (Mp*acr3-1-5^{ge}* and Mp*acr3-2-1^{ge}*) were treated with the indicated concentrations of AsIII or AsV for 72 hours and then the total amounts of As present in the plants were measured. (A) Total As content of Cam2 WT and Mp*acr3* knockout mutants expressed in mg of As per Kg of dry plant tissue. (B) Total As content of Cam2 WT and Mp*ACR3* overexpressing lines expressed in mg of As per Kg of dry plant tissue. Colored and error bars in the graphs represent average and standard deviation, respectively (n = 4 biological replicates). The number of stars corresponds to the level of statistical significance of a Student's *t*-test between the As content for each genotype compared to that of WT Cam2 plants. *: p <= 0.05; **: p <= 0.01; ***: p <= 0.001.

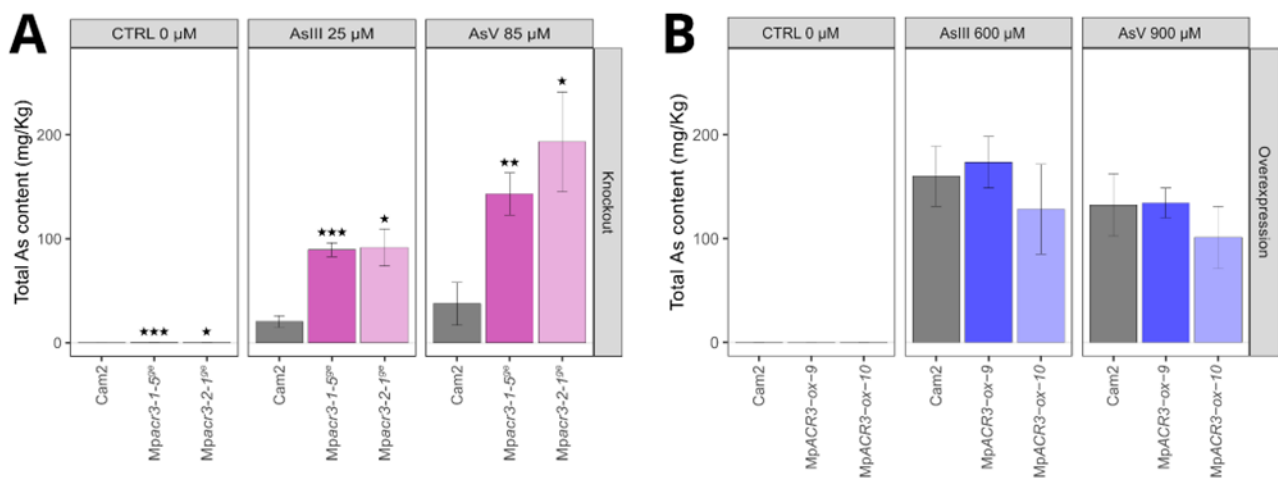

Fig. S4. Agarose electrophoresis gels for *M. polymorpha* complementation lines of a Mp*acr3* knockout mutant with the Mp*ACR3* genomic locus tagged with a 3xCitrine C-terminal fusion. A) Loading control of genomic DNA for the lines tested. B) PCR-specific amplification of the Mp*ACR3* transgene to verify its integration in the genome. WT: Cam2 wild-type *M. polymorpha*. Lines C6 and C8 are transgenic escapes lacking the *ACR3* transgene carried on the T-DNA.

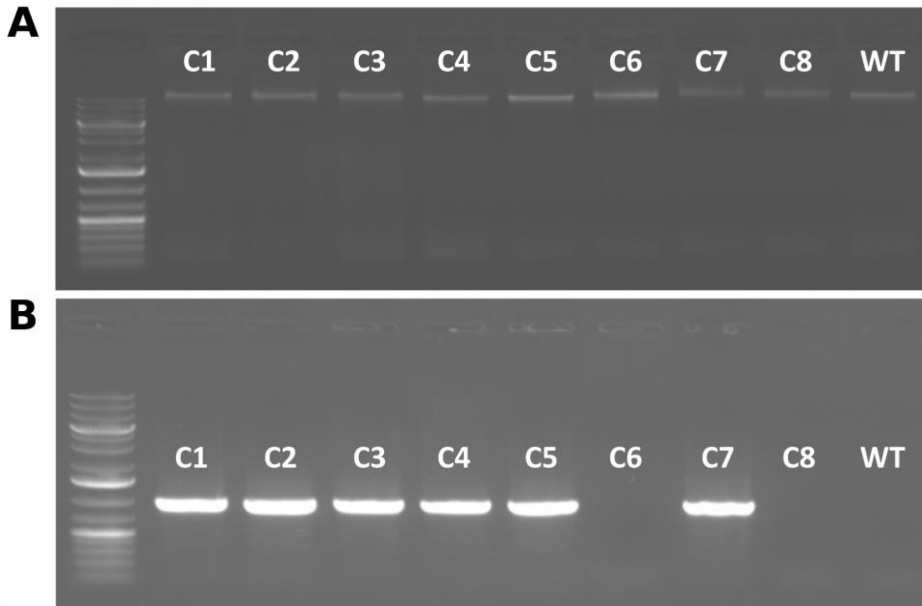

Fig. S5. Tissue specificity of the promoter of MpACR3. Complementation line C1 was used to assess the expression pattern of the MpACR3::3xCitrine protein fusion in the gemmae of the Mp*acr3* mutant. From left to right panels, MpACR3::3xCitrine, green channel; propidium iodide at 0,005%, magenta channel; Merged of Citrine and PI; Merged of bright field (BF, gray) and chlorophyll autofluorescence (AF, cyan) images. Scale bars = 25µm. A: On the dorsal side of the gemmae, the MpACR3 promoter is expressed at high levels in the area surrounding the apical notch meristem (white arrow). B: MpACR3 promoter is at best weakly expressed in rhizoids on the ventral side (white arrowhead).

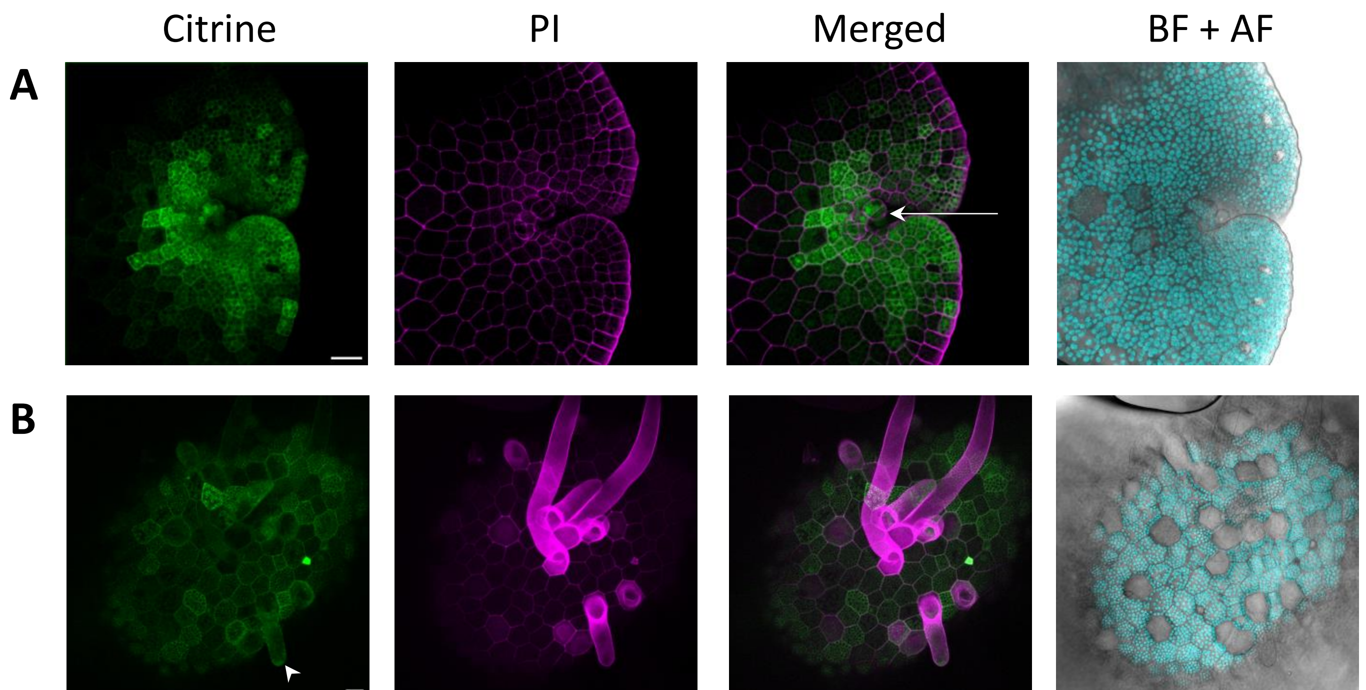

Table S1. Species, accession numbers and protein codes used for phylogenetic analysis.

SPECIES	ACESSION No.	CODE
Adiantum capillus-veneris	KAI5080765.1	AcACR3
Adiantum nelumboides	MCO5566233.1	AnACR3_1
Adiantum nelumboides	MCO5590287.1	AnACR3_2
Adiantum nelumboides	MCO5594610.1	AnACR3_3
Adiantum nelumboides	MCO5603891.1	AnACR3_4
Ceratodon purpureus	KAG0559408.1	CpACR3
Ceratopteris richardii	KAH7279060.1	CrACR3
Marchantia paleacea	KAG6545781.1	MpaACR3
Marchantia polymorpha subsp. *ruderalis*	OAE35577.1	MpACR3
Pteris vittata	XP_024362327.1	PpACR3
Pteris vittata	FJ751631.1	PvACR3
Pteris vittata	FJ751632.1	PvACR3_1
Pteris vittata	MW447114.1	PvACR3_2
Pteris vittata	MW447115.1	PvACR3_3
Pteris vittata	transcript_28534 [1]	PvACR3_4
Sphagnum fallax	KAH8961229.1	SfACR3
Sphagnum magellanicum	KAH9561924.1	SmACR3

[1] from Sun et al. (2023) Journal of Hazardous Materials 458:132034. https://doi.org/10.1016/j.jhazmat.2023.132034.

Table S2. List of primers used in this study.

Primer name	Sequence (5' - 3')	Purpose
MpACT_RT_For	AGGCATCTGGTATCCACGAG	RT-PCR
MpACT_RT_Rev	ACATGGTCGTTCCTCCAGAC	RT-PCR
MpAPT_RT_For	CGAAAGCCCAAGAAGCTACC	RT-PCR
MpAPT_RT_Rev	GTACCCCCGGTTGCAATAAG	RT-PCR
MpACR3_RT_For	GGAGGTAAGGGAATTGATGTGG	RT-PCR
MpACR3_RT_Rev	GATGAACGGGAGGAATTTGG	RT-PCR
MpACR3_gR1_For	CTCGAGAGACCAGTGGATAATAC	CRISPR gRNA1
MpACR3_gR1_Rev	AAACGTATTATCCACTGGTCTCT	CRISPR gRNA1
MpACR3_gR2_For	CTCGACCTTGTCCGAGTCCGGTA	CRISPR gRNA2
MpACR3_gR2_Rev	AAACTACCGGACTCGGACAAGGT	CRISPR gRNA2
MpACR3_For	CACCGACGTGAAGTACAATGAGGG	Cloning (CDS)
MpACR3_Rev	GGACTCCAATAGAAAGTATACGAG	Cloning (CDS)
MpACR3_Comp_For	CACCATTGTGCTCCAATTTCCGTC	Cloning (genomic locus)
MpACR3_Comp_Rev	AGCTTGTTCTTTTGAGAGCCAT	Cloning (genomic locus)
MpACR3_gR1_SeqF	GACTTTGAAGGGATGGTGGTAG	Genotyping
MpACR3_gR1_SeqR	CAATTACGGTCAGGGAACAGA	Genotyping
MpACR3_gR2_SeqF	GTAGTGCCGAGGTAAGGCGA	Genotyping
MpACR3_gR2_SeqR	CAGATCATCCAAATAATCGTCAG	Genotyping
M13_For	CCCAGTCACGACGTTGTAAAACG	Genotyping
M13_Rev	AGCGGATAACAATTTCACACAGG	Genotyping