The Genome of *Microthlaspi erraticum* (Brassicaceae) Provides Insights Into the Adaptation to Highly Calcareous Soils

Bagdevi Mishra1,2, Sebastian Ploch1, Fabian Runge1, Angelika Schmuker3, Xiaojuan Xia1,2, Deepak K. Gupta1,2, Rahul Sharma1 and Marco Thines1,2*

1 Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft für Naturforschung, Frankfurt am Main, Germany, 2 Goethe University, Department for Biological Sciences, Institute of Ecology, Evolution and Diversity, Frankfurt am Main, Germany, 3 Institute of Botany, University of Hohenheim, Stuttgart, Germany

*Correspondence: Marco Thines m.thines@thines-lab.eu

**Specialty section:**
This article was submitted to Computational Genomics, a section of the journal Frontiers in Plant Science

**Received:** 13 March 2020  
**Accepted:** 10 June 2020  
**Published:** 03 July 2020

Micro*thlaspi erraticum* is widely distributed in temperate Eurasia, but restricted to Ca2+-rich habitats, predominantly on white Jurassic limestone, which is made up by calcium carbonate, with little other minerals. Thus, naturally occurring *Microthlaspi erraticum* individuals are confronted with a high concentration of Ca2+ ions while Mg2+ ion concentration is relatively low. As there is a competitive uptake between these two ions, adaptation to the soil condition can be expected. In this study, it was the aim to explore the genomic consequences of this adaptation by sequencing and analysing the genome of *Microthlaspi erraticum*. Its genome size is comparable with other diploid Brassicaceae, while more genes were predicted. Two Mg2+ transporters known to be expressed in roots were duplicated and one showed a significant degree of positive selection. It is speculated that this evolved due to the pressure to take up Mg2+ ions efficiently in the presence of an overwhelming amount of Ca2+ ions. Future studies on plants specialized on similar soils and affinity tests of the transporters are needed to provide unequivocal evidence for this hypothesis. If verified, the transporters found in this study might be useful for breeding Brassicaceae crops for higher yield on Ca2+-rich and Mg2+-poor soils.

**Keywords:** Brassicaceae, evolution, genomics, magnesium transporters, *Microthlaspi erraticum*

**INTRODUCTION**

The plant family Brassicaceae includes many economically important ornamental and crop species. Members of the family are mostly herbaceous, and many can be easily grown in the laboratory, such as *Arabidopsis thaliana*, the first plant to have its genome sequenced, as it is widely used as a model organism for flowering plants. In addition, several other Brassicaceae genomes have been sequenced, facilitating comparative studies (Slotte et al., 2011; Yang et al., 2016; Mandáková et al., 2017). In this study, *Microthlaspi erraticum* of the tribe Coluteocarpeae was targeted for genome sequencing.

Many members of the Coluteocarpeae are able to grow on highly Ca2+-rich carbonate soils, and several are heavy metal accumulators, such as *Noccaea caerulescens* (Mandáková et al., 2015). Here,
the genome assembly of *M. erraticum* is reported. *M. erraticum* is an interesting plant on which to study environmental adaptation, as it has a wide distribution range throughout warm temperate Europe and Central Asia (Ali et al., 2016a; Ali et al., 2016b; Ali et al., 2017). The species occurs almost exclusively in soil derived from calcium carbonate-rich bedrock und usually grows on well-drained loamy, somewhat open areas (Ali et al., 2017). Similar to *A. thaliana*, *M. erraticum* usually is a winter annual, but has longer seed dormancy, requires vernalisation, and so does not produce a second flowering generation in autumn (Baskin and Baskin, 1979). In nature, the plant hibernates in the rosette stage, but at the southern limits of the distribution, seeds may directly germinate in winter or early spring to the flowering stage without going through the rosette stage (unpublished observations). In the laboratory, the time from seed germination to seed maturation is 4–5 months.

Growing on Ca²⁺-rich soil can be challenging for plants, if the soil is at the same time Mg²⁺-deficient, due to the low specificity of channels for bivalent cations. This would lead to an imbalance of Ca²⁺ and Mg²⁺ ions, if the more specific transporters of the MRS2/MGT family cannot provide enough selectivity to counter this (Schock et al., 2000; Li et al., 2001). There is strong evidence that the MRS2/MGT family members, and in particular the root-expressed genes, are vital for the fitness of plants in conditions where there is an overwhelming amount of Ca²⁺ in comparison to Mg²⁺ ions (Gebert et al., 2009). As *M. erraticum* is almost completely restricted to such soils derived from very pure Calcium Carbonate rocks, such as the upper Jurassic limestone deposited in the Tethys Ocean (Kimmig et al., 2001), we hypothesized that this could be mirrored in its MRS2/MGT genes.

*M. erraticum* is easy to grow, as it is a rather small flowering plant without going through the rosette stage. Seeds were separated from the others and used as new mother plant. This way, six generations of selfing were done to create the inbred line *LIMBURG*.

For genome sequencing, plants were grown from seeds of the 7th generation as described above, but for two months without vernalisation. Then leaves were collected, surface-sterilized for 1 min in 3% sodium hypochlorite solution with 0.1% Tween, and rinsed in sterile water to remove the disinfectant. Subsequently, DNA and RNA were extracted from this material as described previously (Mishra et al., 2018). As the RNA sequencing was done to guide and improve gene predictions rather than quantifying expression, only a single extraction was done. After checking the integrity and purity of the extracted nucleic acids using agarose gels, DNA and RNA extracts were sent to Eurofins Genomics (Erlangen, Germany) for library preparation (Illumina shotgun libraries with 300 and 800 bp inserts, 3, 8, and 20 kbp LDJ libraries, as well as PacBio shotgun libraries for the RSII instrument) and sequencing.

### Read Trimming and Correction

Genomic paired-end Illumina reads were trimmed for adaptors and bad quality ends using Trimmomatic (v 0.32) (Bolger et al., 2014) with the following parameters: TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:60. Afterwards, read pairs containing ambiguous bases were removed from the dataset using a perl script. The remaining reads were evaluated for their quality using FastQFS (Sharma and Thines, 2015) and only paired sequences with an average quality score of 30 and a length greater than 70 were considered for further analyses. Preliminary contigs were constructed using velvet, version 1.2.10 (Zerbino and Birney, 2008) and subsequently aligned using BLAST against a local NT database (downloaded from NCBI: 3/10/2014). Possible contaminations were found to be *Lachancea thermotolerans*, Cloning vector pUC19 and Synthetic construct clone G1 from *Pseudomonas* species, all probably derived from artefacts during sequencing, as none of these contaminants are present in our laboratory. The reads that were matching to the contamination were removed from the dataset. The cleaned Illumina reads were used to correct the PacBio reads using proovread (Hackl et al., 2014).

Quality control of the transcriptomic single-end Illumina reads were also performed as for the genomic reads but using the “TruSeq3-SE.fa” file for the adapter trimming in trimmomatic.

### Assembly

A hybrid assembly was built on the basis the Illumina-corrected PacBio reads using Canu (Koren et al., 2017). Contigs made up...
by only 2 to 5 PacBio reads were discarded, but the reads were further used for scaffolding the assembly using SSpace-Long 
(Boetzer & Pirovano, 2014). SSpace standard (Boetzer et al., 
2011) was used afterwards to scaffold the assembly using SG Illumina reads and LJD reads. KGBaseblerm (Ma et al., 2012) 
was used to build a final pseudo-chromosome level assembly, 
using the karyotype of the related species, *Eutrema salsugineum*, 
as a template. The genomic data for *M. erraticum* can be found 
under the following accession numbers – Bioproject ID 
PRJEB35998, BioSample: SAMEA6449025, SRA: ERS4214584, 
GenBank assembly: GCA_902728155.2.

**Assembly Assessment, Gene Prediction, 
and Annotation**

The final assembly was subjected to both CEGMA (Parra et al., 
2009) and BUSCO (Simão et al., 2015) genome completeness 
assessments. Transcriptomic reads were mapped onto the 
assembled genome using tophat2 (Kim et al., 2013) which uses 
GenMark-ET and Augustus to predict the gene space. Homology search for the predicted genes was 
performed using Blastp against a locally stored NR (non-
redundant protein sequences) database (downloaded from 
NCBI 25/03/2017). Interpro ids of all predicted genes were 
filed by running InterPro via the webservice option in 
Blast2GO (Conesa et al., 2005). Inside the Blast2GO 
framework, blast results and InterPro annotations were merged 
and GO ids were assigned to the sequences. The most generic GO 
ids (top level) were removed from the annotation and sequences 
were further annotated according to their predicted localization. 
Repeatscout (Price et al., 2005) was used for de-novo 
identification of repeat elements and for creating a repeat element database. This database was used in repeatmasker 
(Smit et al., 1996-2010) to predict repeat elements in the 
genome. Putative repeats were further filtered on the basis of 
their copy numbers and only those repeats present with more 
than 10 copies in the genome were annotated as repeats. This 
way, repeat domain families were identified in *M. erraticum* 
and four other Brassicaceae genomes (*Table S1*) downloaded from 
the JGI genome portal (https://phytozome.jgi.doe.gov/pz/portal. 
html). InteProScan was run over the gene-set from these species 
and the number of sequences from each species matching to 
specific repeat domain families were obtained.

**Positive Selection Analysis**

The protein and the nucleotide sequences of the one-to-one 
orthologs were fetched from five Brassicaceae species (*A. 
thaliana, A. lyrata, Capsella rubella, Eutrema salsugineum*, and 
*M. erraticum*) for a list of 1:1 orthologs generated using 
OrthoMCL, v2.0.9 (Li et al., 2003). The protein sequences were 
aligned using mafft, v7 (Katoh and Standley, 2013) with default 
parameters. The protein alignment and the nucleotide sequences 
were used in the program transalign from EMBOSS (version: 
6.4.0.0) (Rice et al., 2000) to produce codon alignments. For 
phylogenetic analyses raxmlHPC-PTHREED-SSE3 of RAxML 
v8.1.17 (Stamatakis, 2014) was used with the algorithm 
parameter -f a, which runs rapid Bootstrap analysis and the 
search for the best-scoring ML tree in one program run. The 
substitution model was selected as -m GTRGAMMAI which uses 
GTR, plus an optimization of substitution rates, plus a GAMMA 
model of rate heterogeneity, plus an estimation of the proportion 
of invariant sites. The program was run with 1,000 bootstrap 
replicates (Felsenstein, 1985). Codon alignments of the coding 
sequences and newick-formatted phylogenetic trees were used to 
run positive selection analyses with the codeML module of 
PAML, v4.8 (Yang, 2007). The site model was run to identify 
positively selected genes, and the branch-site model was run to 
identify species-specific positive selection of the genes. As 
multiple hypotheses were tested in the branch-site model of 
codeML, q values were calculated for false discovery rate (FDR) 
testing using q values (Bass et al., 2015) calculated with 
Bioconductor in an R environment, v3.4.1.

The same approach as followed for the one-to-one orthologs 
in the five species, was also used for their *MRS2/MGT* genes to 
alise the selection pressure on the Mg$^{2+}$ transporters.

**Annotations of MRS2/MGT Mg$^{2+}$ 
Transporter Genes**

Functional *MRS2/MGT* Mg$^{2+}$ transport genes from *A. thaliana* 
were used for the identification of the potential Mg$^{2+}$ 
transporters in *M. erraticum* by homology search. The criteria 
for the homology search were as follows: evalue < 10e-5; 
percentage identity > 50%; length of match > 50%. A 
phylogenetic tree was constructed using *A. thaliana* Mg$^{2+}$ 
transporters and their *M. erraticum* orthologs by using 
mafft v7 (Katoh and Standley, 2013) for multiple alignment 
and raxmlHPC-PTHREED-SSE3 of RAxML v8.1.17 (Stamatakis, 
2014) with 1,000 bootstraps for the tree construction. All genes 
were inspected for *MRS2/MGT* specific domains and re-
annotated according to the results. The *MRS2/MGT* genes 
from *M. erraticum* were blasted against genes from the three 
additional Brassicaceae genomes considered in this study, i.e. 
*Arabidopsis lyrata, Eutrema salsugineum*, and *Capsella rubella*, 
and blast hits with an e-value lower than 10e-5 and an at least 
50% length match with more than 50% identity were taken 
as putative *MRS2/MGT* genes. Further, the presence of two 
transmembrane domains was checked using the TPred 
at https://embnet.vital-it.ch/software/TPRED_form. 
html. The presence of a GMN domain at the end of the first 
transmembrane domain was checked manually. The *MRS2/MGT* 
genes from the five Brassicaceae genomes were further aligned 
using mafft v7 (Katoh and Standley, 2013) and a phylogenetic 
tree was built using the method reported above.

**RESULTS**

**Assembly**

The 2C value and genome size for *M. erraticum* LIMBURG had 
been estimated using flow cytometry with *Glycine max* as size
have a substantially lower percentage of interspersed repeats with 16 and 17%, respectively. The outcrosser *E. salsugineum* has the highest percentage of interspersed repeats with 52% (Table 1). All five genomes have a similar percentage of simple repeats with around 2% (Table 1) of the genome. Thus, the proportion of the coding space to the genome size in *M. erraticum* is similar to that of selfing plants but the interspersed repeat regions are higher in proportion.

Repeat domain family associated genes known to have role in biotic and abiotic stress (see discussion) were analyzed in the five species used for comparisons. *M. erraticum* has substantially more members of Pentatricopeptide (PPR), Leucine-rich repeat (LRR and LRR-2) and Kelch repeat domain families in comparison to all other species in this study while having similar number of genes in the Armadillo, HEAT, Ankyrin, Tetratricopeptide (TPR), RCC1, WD40 repeat domain families (Figure S2).

Of the 819 LRR and LRR-2 genes of *M. erraticum*, the majority are F-box proteins (342 proteins) and receptors (314 proteins). Out of the latter, 110 are classified as probable serine threonine-kinase receptors and several as involved in plant defence, acting as disease resistance genes (61 proteins), out of which 25 are annotated as nucleotide-binding site (NBS)-leucine-rich repeat (LRR) domain containing R genes. Apart from the functional annotation of Blast2Go, a separate domain search revealed that a total of 49 genes have both NBS and LRR domains. A similar search in *A. thaliana* indicated the presence of 40 NBS and LRR domain containing genes. The detailed numbers of genes containing NBS and LRR domains in five species are presented in the Table 2.

In *M. erraticum*, out of 259 proteins containing the Kelch repeat domain, 206 are F-box proteins (FBK). The majority of the non-F-box Kelch repeat domain proteins belonged to Galactose oxidase Kelch repeat superfamily and few are receptors to different chemicals and viral substrates. FBK proteins in *M. erraticum* are around twice in number when compared to the other species in this study.

**Positive Selection Analyses of the One-to-One Orthologs**

In the test of positive selection using the site model from codeML, out of 6,725 one-to-one core orthologs, 92 were inferred as positively selected, with at least one amino acid being positively selected according to Bayes Empirical Bayes

---

**TABLE 1** | Details of genome features in respect to size, genes, coding regions, and repeat regions for *M. erraticum* and four other Brassicaceae species.

| Genome   | Genome size (Mb) | Gene numbers | Coding space (Mb) | Coding space (%) | Simple repeats (%) | Interspersed repeats (%) | Reference                  |
|----------|------------------|--------------|------------------|------------------|-------------------|------------------------|----------------------------|
| *M. erraticum* | 170.42          | 51309        | 55.19            | 32.38            | 1.3               | 33.93                  | This study.                |
| *A. lyrata*    | 206.66          | 33132        | 38.61            | 18.68            | 1.51              | 35.86                  | Lamesch et al. (2012)      |
| *A. thaliana*  | 119.66          | 35386        | 43.55            | 36.39            | 1.59              | 15.9                   | Rawat et al. (2015)        |
| *C. rubella*   | 134.83          | 28447        | 35.66            | 26.45            | 2.12              | 16.88                  | Slotte et al. (2013)       |
| *E. salsugineum* | 243.11        | 29284        | 36.12            | 14.86            | 1.22              | 51.81                  | Yang et al. (2013)         |
(BEB) analysis (Yang et al., 2005) with \( p > 95\% \). An additional 305 genes had omega values > 1, but no amino acid position in those genes had a significant BEB value. In the test of positive selection using the branch-site model, positively selected genes in individual species were identified. Figure S3 shows a bar plot showing the numbers of positively selected genes in the individual species. Though the number of positively selected genes in \( M. \) erraticum is slightly higher than the other species, the difference is not pronounced.

**MRS2/MGT Gene Family (Mg\(^{2+}\) Ion Transporters)**

In the Blast2GO pipeline, 13 genes were assigned to the \( MRS2/MGT \) gene family out of which two were discarded, one being an isoform giving rise to the same gene product and the other lacking a functional GMN domain, resulting in a total of 11 \( MRS2/MGT \) genes in \( M. \) erraticum (Table S3). All of these 11 genes had two transmembrane domains and one GMN domain; and all were homologous to the \( MRS2/MGT \) genes in \( A. \) thaliana. The genes in \( MRS2/MGT \) gene family in plants are grouped into 5 clades, named from A to E. In \( M. \) erraticum Clade-A and Clade-C have 1 gene each and Clade-B, Clade-D, and Clade-C have 4, 2 and 3 genes in each, respectively. Details on these genes are presented in Table S2. The \( MRS2/MGT \) genes were also mined from the other Brassicaceae genomes used in this study and details of these genes are given in Table S2. \( Microthlaspi erraticum \) had the highest number of \( MRS2/MGT \) genes in comparison to the other Brassicaceae species. A phylogenetic tree using all the mined \( MRS2/MGT \) genes from \( M. \) erraticum and related species is presented in the Figure S4. Interestingly, duplications in two clusters of \( MRS2/MGT \) genes were also observed for \( M. \) erraticum and one of these genes was added in the \( M. \) genome.

**DISCUSSION**

**Genome Size and Gene Space**

The assembled genome size of \( Microthlaspi erraticum \) is 170.42 Mb, which is larger than the genomes of other selfing plants included in this study, \( A. \) thaliana (119.66 Mb) and \( Capsella rubella \) (134.83 Mb), but is smaller than the genomes of outcrossers, \( Arabidopsis lyrata \) (206.66 Mb) and \( Eutrema salsugineum \) (243.11 Mb). Generally, selfing plants have less transposable elements in comparison to outcrossing plants, causing genome size differences between them (Johnston et al., 2005; Wright et al., 2008).

There is evidences that many repeat domain proteins have roles in coping with abiotic stress conditions such as the Armadillo gene family in rice (Sharma et al., 2014), the mitochondrial PPR-PGN protein (PPR repeat protein for germination on NaCl) in \( A. \) thaliana (Laluk et al., 2011), and proteins of the LRR repeat family in \( A. \) thaliana (Osakabe et al., 2005; Park et al., 2014). The presence of excess of interspersed repeats in the genome of \( M. \) erraticum might indicate possible genomic and genic rearrangements in \( M. \) erraticum that might have emerged to cope with the stress resulting from the harsh abiotic factors the plant is experiencing in its habitat. In line with this assumption, compared to other species, the proportion of genic repeats in \( M. \) erraticum was found to be substantially higher (Figure S5).

**Genes**

Positive selection analyses of one-to-one orthologous genes does not suggest any drastic difference in level of positive selection in \( M. \) erraticum in comparison to the other species in this study. A comparison of the number of members in the 10 repeat domain family genes that have a known role in biotic and abiotic stress conditions, indicates similar number of genes for the six species in all families except PPR, LRR, LRR-2, and Kelch, for which \( M. \) erraticum has a substantially higher number of genes. In \( M. \) erraticum, 672 genes are found to have PPR repeats. Proteins from the PPR repeat family have a role in growth and development of plants, but many PPR proteins are also known to be biotic and abiotic stress regulators. They have roles in high salinity, drought, and cold stress tolerance (Laluk et al., 2011; Yuan and Liu, 2012; Zhu et al., 2014; Jiang et al., 2015). As \( M. \) erraticum grows in environments that face both frost in winter and drought during seed maturation, it could be possible that this is reflected by the high PRR gene content.

| Gene ID | Annotation | peptide length | # exons | # trans-membrane domains | GMN domain | clades | location on chromosomes |
|---------|------------|----------------|---------|--------------------------|-------------|--------|------------------------|
| g25930.t1 | MRS2-11/MGT710 | 456           | 13      | 2                        | Yes         | clade-A | Chr.Ud1 |
| g20081.t1 | MRS2-10/MGT71 | 443           | 3       | 2                        | Yes         | clade-B | Chr.Ud1 |
| g32461.t1 | MRS2-10/MGT71 | 411           | 4       | 2                        | Yes         | clade-B | 5         |
| g7336.t1 | MRS2-5/MGT73 | 399           | 6       | 2                        | Yes         | clade-C | 1         |
| g24065.t1 | MRS2-1/MGT72 | 443           | 4       | 2                        | Yes         | clade-D | 6         |
| g1307.t1 | MRS2-3/MGT74 | 471           | 6       | 2                        | Yes         | clade-E | 6         |
| g8719.t1 | MRS2-4/MGT76 | 414           | 3       | 2                        | Yes         | clade-F | 6         |
| g17909.t1 | MRS2-4/MGT76 | 425           | 3       | 2                        | Yes         | clade-G | 5         |
| g351.t1 | MRS2-7/MGT77 | 616           | 14      | 2                        | Yes         | clade-H | 6         |
| g3572.t1 | MRS2-7/MGT77 | 384           | 11      | 2                        | Yes         | clade-I | 6         |
| g1982.t1 | MRS2-2/MGT79 | 396           | 10      | 2                        | Yes         | clade-J | 2         |
In *M. erraticum*, 819 genes are classified to belong to the Leucine-rich repeat family proteins (LRR and LRR-2), which is far more than in the other species analysed (Figure S5). LRR and LRR-2 family genes are signalling molecules in plants and also have a role in plant development (Hsu et al., 2000) and pathogen defence (Li and Chory, 1997; Deyoung and Clark, 2008). Expression level studies of LRR repeat domain proteins in rice (Park et al., 2014) and *Arabidopsis* (Osakabe et al., 2010) have shown that LRR repeat proteins also positively regulate genes involved in coping with various abiotic stress conditions. This is further supported by Van der Does et al. (2017), who found that MIK2/LRR-KISS is involved in sensing cell-wall integrity changes in response to both biotic and abiotic stress in line with LRR-receptors acting to recognise both pathogen associated molecular patterns and danger signals (Boller and Felix, 2009). It is tempting to speculate that the very rich LRR complement of *M. erraticum* is not only due to the frequent presence of downy mildew in its populations, but also due to the often open slopes on which *M. erraticum* occurs with frequent soil movements, which might need an enhanced and precise danger recognition that responds to root injury. However, more detailed analyses and functional tests will be needed to provide a solid ground to investigate this interesting pattern further.

Kelch repeat domains are found mostly in the C-terminus of F-box proteins. F-Box coupled Kelch (FBK) proteins are abundant in plants, with very few non-plant representatives (Schumann et al., 2011), and are associated with several vital plant molecular mechanisms. These are associated with growth and development (Zhang et al., 2013), secondary metabolism, Circadian clock and photoperiodic flowering (Nelson et al., 2000) by taking part in signal transduction in various pathways. They also play a role in coping with abiotic stress conditions (Jia et al., 2012; Chen et al., 2014). The finding that *M. erraticum* contains about twice as many FBK genes (206) as the other plants investigated in this study might again indicate an adaptation to stressful environmental conditions. This is also reflected by the fact that *M. erraticum* is often among the few or even the only plant that is present in some open slopes it colonises (unpublished observations).

### Uptake and Transport of Cations in *M. erraticum* in Calcium-Rich Soil

#### Uptake and Transport of Ca\(^{2+}\) Ions

Two-pore channel 1 (TPC1) is responsible for transport of Ca\(^{2+}\) from vacuoles to the cytoplasm and expression of TPC1 regulates the storage capacity of Ca\(^{2+}\) in the vacuoles (Pottosin et al., 2009; Gilliam et al., 2011). Each of the species that we included in this study have one gene each that codes for TPC1. The more specific Cyclic Nucleotide-Gated Ion Channel, AtCNGC2 has been reported to have crucial role in adaptation to Ca\(^{2+}\) Stress in plants (Chan et al., 2003 & Wang et al., 2016). AtCNGC2 is coded by a single gene in *A. thaliana* and has one homolog in each *M. erraticum* and other Brassicaceae species included in this study. Also for other Ca\(^{2+}\) channels, no unusual variation was found. This probably reflects the high Ca\(^{2+}\) supply that has also been described to be advantageous (Yamazaki et al., 2000; Sugimoto et al., 2010) and thus does not necessitate enhanced channel specificity, duplication or other forms of adaptation.

#### Uptake and Transport of Mg\(^{2+}\) ions

As *Microthlaspi erraticum* is found almost exclusively in soil derived from Ca\(^{2+}\)-rich but Mg\(^{2+}\)-poor bedrock (Kimmig et al., 2001; Ali et al., 2017), we speculated that an adaptation regarding the targeted uptake of Mg\(^{2+}\) might have evolved that gives the species an evolutionary advantage over other Brassicaceae species. Mg\(^{2+}\) is an essential bivalent ion with vital functions as a co-factor with ATP in various enzymatic reactions and as central ion in the porphyrine ring of chlorophyll molecules.
Different types of Mg\textsuperscript{2+} transporters interactively transport the ion across membranes in plant tissues to maintain homeostasis. In the presence of excessive Ca\textsuperscript{2+} ions in soil solution, specialized Mg\textsuperscript{2+} transporters might be playing a major adaptive role. The MRS2/MGT (Schock et al., 2000; Li et al., 2001) gene family is known to harbour various proteins that transport Mg\textsuperscript{2+} across membranes. MRS2/MGT Mg\textsuperscript{2+} transporters have two trans-membrane domains at the C-terminus with a characteristic GMN domain at the end of the first trans-membrane domain. In M. erraticum 11 potential MRS2/MGT genes were identified with two transmembrane domains and a GMN motif. In comparison A. thaliana and rice for 10 such genes, while 9 are reported from maize (Li et al., 2016). The MRS2/MGT gene AtMRS2-10, has been shown to be expressed in the root in the plasma membrane (Gebert et al., 2009). For this gene, two homologs are found in M. erraticum, meaning that this gene has been duplicated. All other species have only one homolog except C. rubella in which no homolog for MRS2-10/MGT1 was found (Figure S4). Single knock-out experiments and a double knock-out of MRS2-1/MGT2 and MRS2-5/MGT3, as well as MRS2-5/MGT3 and MRS2-10/MGT1 (Gebert et al., 2009), had no visible effect under normal growth conditions, pointing at functional redundancy of the MRS2 gene family members. In a phylogenetic analysis, it was shown that MRS2-1/MGT2 and MRS2-10/MGT1 form a sub-clade of the family and plants with double knock-out of MRS2-1/MGT2 and MRS2-10/MGT1 have a high demand of Mg\textsuperscript{2+} for normal growth (Lenz et al., 2013). Thus, the presence of a third member in this sub-clade might indicate a genomic adaptation to the high Ca\textsuperscript{2+}/low Mg\textsuperscript{2+} soil condition.

MRS2-4/MGT6 and MRS2-6/MGT5 form a sub-clade in A. thaliana. All species investigated in this study have two genes in this subclade except E. salsugineum which has only one. Considering the phylogenetic distance of this gene from the two A. thaliana members of this group, it can be assumed that E. salsugineum is missing MRS2-6/MGT5. MRS2-4/MGT6 had previously been identified to localize on either chloroplast or mitochondria in shoots (Gebert et al., 2009; Conn et al., 2011), but a later study has identified this gene to be localised in root plasma membrane under lowered Mg\textsuperscript{2+} conditions and that in Mg\textsuperscript{2+}-deficient experimental conditions the transcript levels of this gene in the root increased eight-fold (Mao et al., 2014). Thus, it seems possible that the retaining of the duplication of MRS2-4/MGT6 in M. erraticum is advantageous in Mg\textsuperscript{2+}-poor conditions.

Another subclade in A. thaliana comprise of MRS2-2/MGT9, MRS2-7/MGT7 and MRS2-8/MGT8. In some ecotypes in A. thaliana MRS2-8/MGT8 has been found to be a pseudogene (Gebert et al., 2009). MRS2-7/MGT7 from this clade, an ER-localized transporter, is known to be expressed in roots and to promote growth in plants growing in Mg\textsuperscript{2+} deficient soil (Gebert et al., 2009). Its expression was found to be essential for germination in a solution culture system and for normal growth in low Mg\textsuperscript{2+} conditions (Gebert et al., 2009; Conn et al., 2011). In our analyses, we found a duplication of MRS2-7/MGT7 in M. erraticum in this clade (Figure 1). One of these genes in M. erraticum was found to be positively selected with significant p- and q-values in the branch site model of codeml (Figure 1). As MRS2-7/MGT7 has shown to be important in Mg\textsuperscript{2+}-deficient conditions, we speculate that its duplication might again be an adaptation of M. erraticum to Ca\textsuperscript{2+}-rich but Mg\textsuperscript{2+}-poor soils.

**CONCLUSION**

In conclusion, the genome sequence of M. erraticum provided several indications of adaptation to stressful abiotic conditions, which is in line with its ephemeral growth in habitats with shallow soil and little vegetation cover, exposing it to a variety of adverse environmental conditions. Probably the most striking characteristic of the preferred habitat of M. erraticum is that its soil is derived usually from white Upper Jurassic limestone, a bedrock that is extremely rich in Ca\textsuperscript{2+} but rather poor in Mg\textsuperscript{2+} (Kimmig et al., 2001), creating an environment in which vital Mg\textsuperscript{2+} ion uptake is difficult to achieve. The duplication of two Mg\textsuperscript{2+} transporters that have been shown to be important for Mg\textsuperscript{2+} uptake in Mg\textsuperscript{2+}-deficient conditions is indicate an adaptive response to this. Further experiments are necessary to carry out transgenic and affinity assays to underpin this assumption. Should heterologous expression and affinity experiments support this hypothesis, the MRS2/MGT family of M. erraticum could be an interesting target for improving crop yield on highly calcareous soils.

**DATA AVAILABILITY STATEMENT**

The datasets generated for this study can be found under the accession number NCBI PRJEB35998 (https://www.ncbi.nlm.nih.gov/bioproject/PRJEB35998).

**AUTHOR CONTRIBUTIONS**

MT conceived the study. MT, AS, and FR created the Limburg plant genome. BM and MT interpreted the data and wrote the manuscript, with contributions from the other authors.

**FUNDING**

This study has been supported by LOEWE in the framework of the Centre of Translational Biodiversity Genomics and by the Max Planck Society through a fellowship awarded to MT.

**ACKNOWLEDGMENTS**

Tahir Ali is gratefully acknowledged for critical discussions and editing.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00943/full#supplementary-material
Mandakóvá, T., Li, Z., Barker, M. S., and Lysak, M. A. (2017). Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. Plant J. 91, 23–31. doi: 10.1111/tpc.13553

Mao, D., Chen, J., Tian, L., Liu, Z., Yang, L., Tang, R., et al. (2014). Arabidopsis transporter MGT76 mediates magnesium uptake and is required for growth under magnesium limitation. Plant Cell. 26, 2234–2248. doi: 10.1105/tpc.114.124628

Mishra, B., Gupta, D. K., Penninger, M., Hickler, T., Langer, E., Nam, B., et al. (2018). A reference genome of the European beech (Fagus sylvatica L.). GigaScience 7, ggy063. doi: 10.1093/gigascience/ggy063

Nelson, D. C., Lasswell, J., Rogg, L. E., Cohen, M. A., and Bartel, B. (2000). FKF1, a Clock-Controlled Gene that Regulates the Transition to Flowering in Arabidopsis. Cell 101, 331–340. doi: 10.1016/S0092-8674(00)08842-9

Osakabe, K., Osakabe, Y., and Toki, S. (2010). Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. PNAS U. S. A. 107, 12034–12039. doi: 10.1073/pnas.1000234107

Parra, G., Bradnam, K., Ning, Z., Keane, T., and Korf, I. (2009). Assessing the gene space in draft genomes. Nucleic Acids Res. 37, 289–297. doi: 10.1093/nar/gkn916

Pavlopoulou, A., Wherrett, T., and Shabalova, S. (2009). SV channels dominate the vacuolar Ca2+ release during intracellular signaling. FEBS Lett. 583, 921–926. doi: 10.1016/j.febslet.2009.02.009

Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. Bioinformatics 21, i351–i358. doi: 10.1093/bioinformatics/bti1018

Price, A. L., Zhang, L., Pevzner, P. A., and Pop, M. (2005). Map remotely related genomes. Genome Biol. 6, R58. doi: 10.1186/gb-2005-6-6-r58

Randall, J. E., Kalkman, T. P., Zhu, J., Lu, Y., and Vicient, C. (2012). Large-scale functional analysis of the Arabidopsis T-Box transcription factor family. Plant Cell. 24, 4401–4417. doi: 10.1105/tpc.112.107062

Sapkota, B., Babu, M., and Arora, P. (2018). De novo assembly of the resistance to Pseudomonas syringae pv. tomato strain DC3000 in wild tomato species, Lycopersicon pennellii and L. chilense. Plant Physiol. 176, 1651–1663. doi: 10.1104/pp.17.00520

Smit, A. F. A., Hubley, R., and Green, P. (2000). RepeatMasker Open-3.0. Repeatmasker.org. http://www.repeatmasker.org/ (accessed December 1, 2017).

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Sugimoto, T., Watanabe, K., Yoshida, S., Aino, M., Furuki, M., Shiono, M., et al. (2010). Field application of calcium to reduce phytophthora stem rot of soybean, and calcium distribution in plants. Plant Dis. 94, 812–819. doi: 10.1094/PDIS-94-7-0812

Van der Does, D., Boutrot, F., Engelsdorf, T., Rhodes, J., McKenna, J. F., Vernhettes, S., et al. (2017). The Arabidopsis leucine-rich repeat receptor kinase MIK2/LRR-KISS connects cell wall integrity sensing, root growth and response to abiotic and biotic stresses. PloS Genet. 13, e1006832. doi: 10.1371/journal.pgen.1006832

Wang, Y., Kang, Y., Ma, C., Miao, R., Wu, C., Long, Y., et al. (2016). CNGC2 is a Ca2+ influx channel that prevents accumulation of apoplastic Ca2+ in the leaf. Plant Physiol. 173, 1342–1354. doi: 10.1094/pp.16i0222

Wright, S., Ness, R., Foxe, J., and Barrett, S. (2008). Genomic Consequences of Outcrossing and Selfing in Plants. Int. J. Plant Sci. 169, 105–118. doi: 10.1086/523366

Yamazaki, H., Kikuchi, S., Hoshina, T., and Kimura, T. (2000). Effect of calcium concentration in nutrient solution on development of bacterial wilt and population of its pathogen Ralstonia solanacearum in grafted tomato seedlings. Soil Sci. Plant Nutr. 46, 535–539. doi: 10.1080/0300768. 2000.10408807

Yang, Z., Wong, W. S., and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107–1118. doi: 10.1093/molbev/msi097

Yang, R., Jarvis, D. E., Chen, H., Beilstein, M. A., Grimwood, J., Jenkins, J., et al. (2013). The Reference Genome of the Halophytic Plant Eutrema salinaigeum. Front. Plant Sci. 4, 46. doi: 10.3389/fpls.2013.00046

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., et al. (2016). The genome sequence of allopolyploid Brassica juncea and analysis of differential homoeolog gene expression influencing selection. Nat. Genet. 48, 1225–1232. doi: 10.1038/ng.3657

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24, 1586–1591. doi: 10.1093/molbev/msm088

Yuan, H., and Liu, D. (2012). Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in Arabidopsis. Plant J. 70, 432–444. doi: 10.1111/j.1365-313X.2011.04883.x

Zerbino, D. R., Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829. doi: 10.1101/gr.071442.107

Zhang, X., Gou, M., and Liu, C. J. (2013). Arabidopsis Kelch Repeat F-Box Proteins Regulate Phytoantheropgranoid Biosynthesis via Controlling the Turnover of Phenylalanine Ammonia-Lyase. Plant Cell. 25, 4994–5010. doi: 10.1105/tpc.113.119644

Zhao, Q., Degarudeyn, J., Zhang, C., Mühlenbock, P., Eastmond, P. J., Valcke, R., et al. (2017). The Reference Genome of the Halophytic Plant Eutrema salinaigeum. Front. Plant Sci. 4, 46. doi: 10.3389/fpls.2013.00046

Zhu, Q., Dugardeyn, J., Zhang, C., Müllerbock, P., Eastmond, P. J., Valcke, R., et al. (2014). The Arabidopsis thaliana RNA editing factor SLO2, which affects the mitochondrial electron transport chain, participates in multiple stress and hormone responses. Mol. Plant 7, 290–310. doi: 10.1093/mp/sst012

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Mishra, Ploch, Runge, Schmucker, Xia, Gupta, Sharma and Thines. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.