



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

Bagdevi Mishra^{1,2}, Sebastian Ploch¹, Fabian Runge¹, Angelika Schmuker³, Xiaojuan Xia^{1,2}, Deepak K. Gupta^{1,2}, Rahul Sharma¹ and Marco Thines^{1,2*}

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

OPEN ACCESS

Edited by:

Nunzio D'Agostino,
University of Naples Federico II, Italy

Reviewed by:

Neil Graham,
University of Nottingham,
United Kingdom
Thomas D. Alcock,
University of Nottingham,
United Kingdom

*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

Citation:

Mishra B, Ploch S, Runge F, Schmuker A, Xia X, Gupta DK, Sharma R and Thines M (2020) The Genome of *Microthlaspi erraticum* (Brassicaceae) Provides Insights Into the Adaptation to Highly Calcareous Soils. *Front. Plant Sci.* 11:943. doi: 10.3389/fpls.2020.00943

Microthlaspi 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: Brassicales, 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 and 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 produce a flowering plant 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 Ca^{2+} -rich soil can be challenging for plants, if the soil is at the same time Mg^{2+} -deficient, due to the low specificity of channels for bivalent cations. This would lead to an imbalance of Ca^{2+} and Mg^{2+} 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 Ca^{2+} in comparison to Mg^{2+} 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.

Microthlaspie erraticum is easy to grow, as it is a rather small plant with a short life cycle, but neither a genome nor molecular tools are available for the species, yet. As the edaphic niche of *M. erraticum* is very well defined, the species provides a good model to investigate environmental adaptation apart from soil effects. Therefore, it was the general aim of this study to provide a well-assembled genome of the species and a specific aim to unravel potential genomic and genetic adaptations to Ca^{2+} -rich but Mg^{2+} -poor soil, with specific reference to the limiting Mg^{2+} ion uptake.

MATERIAL AND METHODS

Plant Material, DNA and RNA Extraction, and Sequencing

For genome sequencing, a six times inbred line of the *Microthlaspie erraticum* was used. This line was named LIMBURG, after the original collection site beneath the ruins of the stronghold Limburg of Berthold mit dem Barte. He was the founder of the house of Zähringen, an influential medieval house in what are now South-Western Germany, Eastern France, Switzerland, Austria, and Northern Italy. The original mother plant was collected in spring 2007 from beneath the Limburg ruins (near Kirchheim unter Teck, Swabian Alb, Germany) in the flowering stage and kept in a climate chamber at 16°C, 60%

humidity and a 14 h light 10 h darkness cycle, taking care that the soil remained always moist. After seed maturation, seeds were collected and air-dried for three months at room temperature. Subsequently, seeds were sown into standard gardening soil and placed into a climate chamber with the same conditions as reported before. Six weeks after germination, when the seedlings had produced a small rosette, plants were transferred to a refrigerator for two weeks. Subsequently, they were brought back to the climate chamber, where they flowered to produce seeds within about two months. From the plants one was 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 genepredictions 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. KGBassebmler (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). The mapping results were used in the BRAKER1 pipeline (Hoff et al., 2016) 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 fetched 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>). Interproscan 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-PTHREADS-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 invariable 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 analyse the selection pressure on the Mg²⁺ transporters.

Annotations of *MRS2/MGT* Mg²⁺ Transporter Genes

Functional *MRS2/MGT* Mg²⁺ transport genes from *A. thaliana* were used for the identification of the potential Mg²⁺ 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* Mg²⁺ transporters and their *M. erraticum* orthologs by using mafft v7 (Katoh and Standley, 2013) for multiple alignment and raxmlHPC-PTHREADS-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 TMpred at https://embnet.vital-it.ch/software/TMPRED_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

standard (Ali et al., 2016b). The 2C value was 0.44 pg, corresponding to a diploid genome size of 422 Mbp, implying a haploid genome size of 211 Mbp. The hybrid assembly generated using the Illumina and PacBio data was of 170 Mbp in length, out of which 156 Mbp were assembled into 7 pseudo-chromosomes. The rest of the contigs were built into two pseudo-molecules, one with predominantly genic regions and the other with predominantly non-genic regions. The placement of 1,940 scaffolds within the 9 units and the sizes of the pseudo-chromosomes are listed in **Table S2**. Only one gene was not identified when using CEGMA, implying more than 99% completeness of the genic space. The assembled genome had 0.61% of N and the GC content was 37.87%. In the BUSCO analysis out of a total of 1,375 genes that were considered to check the completeness of the gene space, and 98.5% of the genes were present in the genome, out of which 8.2% were duplicated. A total of 0.4% of the genes were present as fragments and a total of 1.1% of the genes were missing. The shotgun Illumina reads were mapped onto the final genome using Bowtie2 (Langmead and Salzberg, 2012) and the reads mapped to 169 Mb of the assembled genome. The PacBio reads were also mapped back to the genome using Blasr (Chaisson and Tesler, 2012) and the reads mapped back to 160 Mb of the assembled genome. Out of 170 Mb, 159 Mb of the total genome were mapped by both Illumina short reads and PacBio reads.

Genomic Features

The number of genes predicted in *M. erraticum* was 51,309 (with coding space of 55.19Mb), which is higher than in other Brassicaceae, but the proportion of coding space to non-coding space in the genome is similar to that of *A. thaliana* (**Table 1**). The 51,309 genes included 1,372 splice variants, resulting in 49,937 unique genic locations out of which 49,060 were complete genes with both start and stop codons and without in-frame stop codons. A total of 34,370 genes were found to have transcript support when transcriptomic reads were mapped to the genes. In the blast2go annotation process, 1,521 genes were not annotated. The number of genes annotated with GO terms of molecular functions, cellular components and biological processes is given in **Figure S1**, with top functions under each category. A total of 35,042 genes was provided with a GO term.

The genome of *M. erraticum* contains 34% of Interspersed repeats, a percentage comparable to *A. lyrata* which has 36% of interspersed repeats in its genome. *A. thaliana* and *C. rubella*

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
<i>M. erraticum</i>	170.42	51309	55.19	32.38	1.3	33.93	This study.
<i>A. lyrata</i>	206.66	33132	38.61	18.68	1.51	35.86	Lamesch et al. (2012)
<i>A. thaliana</i>	119.66	35386	43.55	36.39	1.59	15.9	Rawat et al. (2015)
<i>C. rubella</i>	134.83	28447	35.66	26.45	2.12	16.88	Slotte et al. (2013)
<i>E. salsugineum</i>	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 (Mg²⁺ 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-E have 4, 2 and 3 genes in each, respectively. Details on these genes are presented in **Table 3**. 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**. *Microthlaspie 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 observed for *M. erraticum* and one of these genes was positively selected (**Figure 1**).

TABLE 2 | Numbers of genes in five Brassicaceae genomes that contain nuclear binding site (NBS) and/or leucine-rich repeat (LRR) domains.

Genome	NBS domain	LRR domains	NBS+LRR domain
<i>M. erraticum</i>	279	819	49
<i>A. lyrata</i>	126	417	19
<i>A. thaliana</i>	208	558	40
<i>C. rubella</i>	139	590	30
<i>E. salsugineum</i>	129	522	30

TABLE 3 | Details of the *MRS2/MGT* genes identified in *Microthlaspie erraticum* genome.

Gene ID	Annotation	peptide length	# exons	# trans-membrane domains	GMN domain	clades	location on chromosomes
g25930.t1	MRS2-11/MGT10	456	13	2	Yes	clade-A	ChrUd1
g20081.t1	MRS2-10/MGT1	443	3	2	Yes	clade-B	ChrUd1
g32461.t1	MRS2-10/MGT1	411	4	2	Yes	clade-B	5
g7336.t1	MRS2-5/MGT3	399	6	2	Yes	clade-B	ChrUd1
g24065.t1	MRS2-1/MGT2	443	4	2	Yes	clade-B	1
g1307.t1	MRS2-3/MGT4	471	6	2	Yes	clade-C	ChrUd1
g8719.t1	MRS2-4/MGT6	414	3	2	Yes	clade-D	ChrUd1
g17909.t1	MRS2-4/MGT6	425	3	2	Yes	clade-D	5
g571.t1	MRS2-7/MGT7	616	14	2	Yes	clade-E	6
g572.t1	MRS2-7/MGT7	384	11	2	Yes	clade-E	6
g1982.t1	MRS2-2/MGT9	396	10	2	Yes	clade-E	2

DISCUSSION

Genome Size and Gene Space

The assembled genome size of *Microthlaspie 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 evidence 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 PPR gene content.

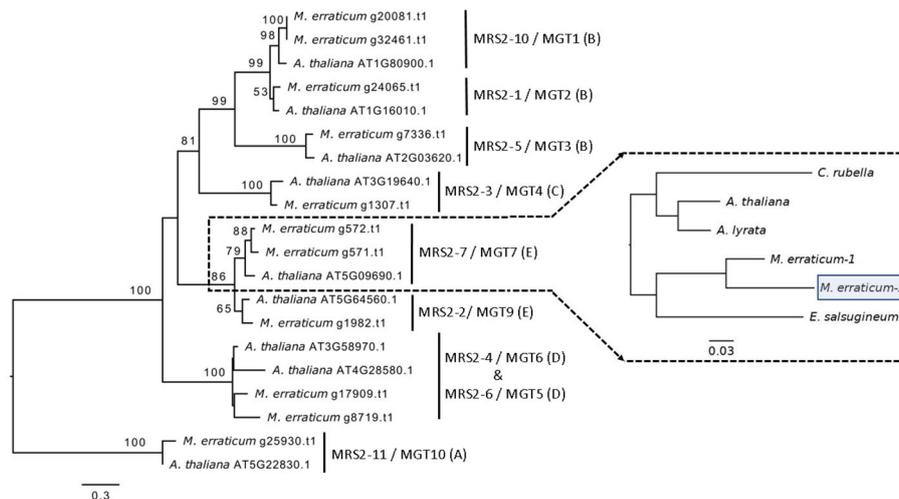


FIGURE 1 | Phylogenetic tree of 20 *MRS2/MGT* genes (11 from *M. erraticum* and 9 from *A. thaliana*) with the main monophyletic groups highlighted as A–E, showing duplication of *MRS2-7/MGT7* and *MRS2-10/MGT1* genes in *M. erraticum*. An inset of the *MRS2-7/MGT7* clade with five Brassicaceae species is shown, which was used for a positive selection analysis. One of the two *MRS2-7/MGT7* genes in *M. erraticum* was positively selected and is highlighted in the figure.

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. Many 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; Gilliham 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^{2+} transporters interactively transport the ion across membranes in plant tissues to maintain homeostasis. In the presence of excessive Ca^{2+} ions in soil solution, specialized Mg^{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^{2+} across membranes. *MRS2/MGT* Mg^{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 code 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^{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^{2+} /low Mg^{2+} soil condition.

MRS2-4/MGT6 and *MRS2-6/MGT5* form a subclade 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^{2+} conditions and that in Mg^{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^{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^{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^{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^{2+} -deficient conditions, we speculate that its

duplication might again be an adaptation of *M. erraticum* to Ca^{2+} -rich but Mg^{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^{2+} but rather poor in Mg^{2+} (Kimmig et al., 2001), creating an environment in which vital Mg^{2+} ion uptake is difficult to achieve. The duplication of two Mg^{2+} transporters that have been shown to be important for Mg^{2+} uptake in Mg^{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 inbred line. SP and XX performed laboratory experiments. BM, RS, and DG analyzed the 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>

REFERENCES

- Ali, T., Runge, F., Dutbayev, A., Schmuker, A., Solovyeva, I., Nigrelli, L., et al. (2016a). *Microthlaspie erraticum* (Jord.) T. Ali et Thines has a wide distribution, ranging from the Alps to the Tien Shan. *Flora* 225, 76–81. doi: 10.1016/j.flora.2016.09.008
- Ali, T., Schmuker, A., Runge, F., Solovyeva, I., Nigrelli, L., Paule, J., et al. (2016b). Morphology, phylogeny, and taxonomy of *Microthlaspie* (Brassicaceae: Coluteocarpeae) and related genera. *Taxon* 65, 79–98. doi: 10.12705/651.6
- Ali, T., Muñoz-Fuentes, V., Buch, A. K., Çelik, A., Dutbayev, A., Gabrielyan, I., et al. (2017). Genetic patterns reflecting Pleistocene range dynamics in the annual calcicole plant *Microthlaspie erraticum* across its Eurasian range. *Flora* 236–237, 132–142. doi: 10.1016/j.flora.2017.09.014
- Baskin, J. M., and Baskin, C. C. (1979). The ecological life cycle of *Thlaspie perfoliatum* and a comparison with published studies on *Thlaspie arvensis*. *Weed Res.* 19 (5), 285–292. doi: 10.1111/j.1365-3180.1979.tb01540.x
- Bass, J. D., Swcf, A. J., Dabney, A., and Robinson, D. (2015). qvalue: Q-value estimation for false discovery rate control. R package version 2.10.0. github.com. <http://github.com/jdstorey/qvalue> (accessed September 1, 2016).
- Boetzer, M., and Pirovano, W. (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinf.* 15, 211. doi: 10.1186/1471-2105-15-211
- Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27, 578–589. doi: 10.1093/bioinformatics/btq683
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Ann. Rev. Pl. Biol.* 60, 379–406. doi: 10.1146/annurev.arplant.57.032905.105346
- Chaisson, M. J., and Tesler, G. (2012). Mapping single molecule sequencing reads using Basic Local Alignment with Successive Refinement (BLASR): Theory and Application. *BMC Bioinf.* 13, 238. doi: 10.1186/1471-2105-13-238
- Chan, C., Schorrak, L., Smith, R., Bent, A., and Sussman, M. (2003). A Cyclic Nucleotide-Gated Ion Channel, CNGC2, Is Crucial for Plant Development and Adaptation to Calcium Stress. *Plant Physiol.* 132, 728–731. doi: 10.1104/pp.102.019216
- Chen, R., Guo, W., Yin, Y., and Gong, Z.-H. (2014). A Novel F-Box Protein CaF-Box Is Involved in Responses to Plant Hormones and Abiotic Stress in Pepper (*Capsicum annuum* L.). *Int. J. Mol. Sci.* 15, 2413–2430. doi: 10.3390/ijms15022413
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Conn, S. J., Conn, V., Tyerman, S. D., Kaiser, B. N., Leigh, R. A., and Gilliam, M. (2011). Magnesium transporters, *MGT2/MRS2-1* and *MGT3/MRS2-5*, are important for magnesium partitioning within *Arabidopsis thaliana* mesophyll vacuoles. *New Phytol.* 190, 583–594. doi: 10.1111/j.1469-8137.2010.03619.x
- Deyoung, B. J., and Clark, S. E. (2008). BAM receptors regulate stem cell specification and organ development through complex interactions with *CLAVATA* signalling. *Genetics* 180, 895–904. doi: 10.1534/genetics.108.091108
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Gebert, M., Meschenmoser, K., Svidová, S., Weghuber, J., Schweyen, R., Eifler, K., et al. (2009). A root-expressed magnesium transporter of the *MRS2* gene family in *Arabidopsis thaliana* allows for growth in low-Mg²⁺ environments. *Plant Cell* 21, 4018–4030. doi: 10.1105/tpc.109.070557
- Gilliam, M., Athman, A., Tyerman, S. D., and Conn, S. J. (2011). Cell-specific compartmentation of mineral nutrients is an essential mechanism for optimal plant productivity—another role for TPC1? *Plant Signaling Behav.* 6, 1656–1661. doi: 10.4161/psb.6.11.17797
- Hackl, T., Hedrich, R., Schultz, J., and Förster, F. (2014). proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* 30, 3004–3011. doi: 10.1093/bioinformatics/btu392
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32, 767–769. doi: 10.1093/bioinformatics/btv661
- Hsu, S. Y., Kudo, M., Chen, T., Nakabayashi, K., Bhalla, A., van der Spek, P. J., et al. (2000). The three subfamilies of leucine-rich repeat-containing G protein-coupled receptors (LGR): identification of LGR6 and LGR7 and the signalling mechanism for LGR7. *Mol. Endocri.* 14, 1257–1271. doi: 10.1210/mend.14.8.0510
- Jia, Y., Gu, H., Wang, X., Chen, Q., Shi, S., Zhang, J., et al. (2012). Molecular cloning and characterization of an F-box family gene CarF-box1 from chickpea (*Cicer arietinum* L.). *Mol. Biol. Rep.* 39, 2337–2345. doi: 10.1007/s11033-011-0984-y
- Jiang, S. C., Mei, C., Liang, S., Yu, Y. T., Lu, K., Wu, Z., et al. (2015). Crucial roles of the pentatricopeptide repeat protein SOAR1 in Arabidopsis response to drought, salt and cold stresses. *Plant Mol. Biol.* 88, 369–385. doi: 10.1007/s11033-015-0327-9
- Johnston, J. S., Pepper, A. E., Hall, A. E., Chen, Z. J., Hodnett, G., Drabek, J., et al. (2005). Evolution of genome size in Brassicaceae. *Ann. Bot.* 95, 229–235. doi: 10.1093/aob/mci016
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36. doi: 10.1186/gb-2013-14-4-r36
- Kimmig, B., Werner, W., and Aigner, T. (2001). Highly pure limestone in the Upper Jurassic of the Swabian Jura: composition, distribution, applications. *Z. Angew. Geol.* 47 (2), 101–108.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116
- Laluk, K., Abuqamar, S., and Mengiste, T. (2011). The Arabidopsis mitochondria-localized pentatricopeptide repeat protein PGN functions in defence against necrotrophic fungi and abiotic stress tolerance. *Plant Physiol.* 156, 2053–2068. doi: 10.1104/pp.111.177501
- Lamesch, P., Berardini, T. Z., Li, D., Swarbrick, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090
- Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lenz, H., Dombinov, V., Dreistein, J., Reinhard, M. R., Gebert, M., and Knoop, V. (2013). Magnesium deficiency phenotypes upon multiple knockout of *Arabidopsis thaliana* *MRS2* clade B genes can be ameliorated by concomitantly reduced calcium supply. *Plant Cell Physiol.* 54, 1118–1131. doi: 10.1093/pcp/pct062
- Li, J., and Chory, J. (1997). A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* 90, 929–938. doi: 10.1016/S0092-8674(00)80357-8
- Li, L., Tutone, A. F., Drummond, R. S., Gardner, R. C., and Luan, S. (2001). A novel family of magnesium transport genes in Arabidopsis. *Plant Cell* 13, 2761–2775. doi: 10.1105/tpc.010352
- Li, L., Stoeckert, C. J., Jr., and Roos, D. S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Li, H., Du, H., Huang, K., Chen, X., Liu, T., and Gao, S. (2016). Identification and functional and expression analyses of the *CORA/MRS2/MGT*-type magnesium transporter family in maize. *Plant Cell Physiol.* 57, 1153–1168. doi: 10.1093/pcp/pcw064
- Ma, C., Chen, H., Xin, M., Yang, R., and Wang, X. (2012). KGBassembler: a karyotype-based genome assembler for Brassicaceae species. *Bioinformatics* 28, 3141–3143. doi: 10.1093/bioinformatics/bts586
- Mandáková, T., Singh, V., Krämer, U., and Lysak, M. A. (2015). Genome Structure of the Heavy Metal Hyperaccumulator *Noccaea caerulea* and Its Stability on Metalliferous and Nonmetalliferous Soils. *Plant Physiol.* 169, 674–689. doi: 10.1104/pp.15.00619

- Mandáková, 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, 3–21. doi: 10.1111/tpj.13553
- Mao, D., Chen, J., Tian, L., Liu, Z., Yang, L., Tang, R., et al. (2014). Arabidopsis transporter *MGT6* 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., Pfenninger, M., Hickler, T., Langer, E., Nam, B., et al. (2018). A reference genome of the European beech (*Fagus sylvatica* L.). *GigaScience* 7, giv063. doi: 10.1093/gigascience/giv063
- 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)80842-9
- Osakabe, Y., Maruyama, K., Seki, M., Satou, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2005). Leucine-rich repeat receptor-like kinase1 is a key membrane-bound regulator of abscisic acid early signalling in Arabidopsis. *Plant Cell.* 17, 1105–1119. doi: 10.1105/tpc.104.027474
- 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
- Park, S., Moon, J. C., Park, Y. C., Kim, J. H., Kim, D. S., and Jang, C. S. (2014). Molecular dissection of the response of a rice leucine-rich repeat receptor-like kinase (LRR-RLK) gene to abiotic stresses. *J. Plant Physiol.* 171, 1645–1653. doi: 10.1016/j.jplph.2014.08.002
- 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
- Pottosin, I., Wherrett, T., and Shabala, 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
- Rawat, V., Abdelsamad, A., Pietzenku, B., Seymour, D. K., Koenig, D., Weigel, D., et al. (2015). Improving the Annotation of *Arabidopsis lyrata* Using RNA-Seq Data. *PLoS One* 10, e0137391. doi: 10.1371/journal.pone.0137391
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- Schock, I., Gregan, J., Steinhäuser, S., Schweyen, R., Brennicke, A., and Knoop, V. (2000). A member of a novel *Arabidopsis thaliana* gene family of candidate Mg²⁺ ion transporters complement a yeast mitochondrial group II intron-splicing mutant. *Plant J.* 24, 489–501. doi: 10.1046/j.1365-313x.2000.00895.x
- Schumann, N., Navarro-quezada, A., Ullrich, K., Kuhl, C., and Quint, M. (2011). Molecular evolution and selection patterns of plant F-box proteins with C-terminal kelch repeats. *Plant Physiol.* 155, 835–850. doi: 10.1104/pp.110.166579
- Sharma, R., and Thines, M. (2015). FastQFS - A Tool for evaluating and filtering paired-end sequencing data generated from high throughput sequencing. *Mycol. Prog.* 14, 60. doi: 10.1007/s11557-015-1077-4
- Sharma, M., Singh, A., Shankar, A., Pandey, A., Baranwal, V., Kapoor, S., et al. (2014). Comprehensive expression analysis of rice Armadillo gene family during abiotic stress and development. *DNA Res.* 21, 267–283. doi: 10.1093/dnares/dst056
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Slotte, T., Bataillon, T., Hansen, T. T., St Onge, K., Wright, S. I., and Schierup, M. H. (2011). Genomic determinants of protein evolution and polymorphism in Arabidopsis. *Genome Biol. Evol.* 3, 1210–1219. doi: 10.1093/gbe/evr094
- Slotte, T., Hazzouri, K. M., Ågren, J. A., Koenig, D., Maumus, F., Guo, Y. L., et al. (2013). The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* 45, 831–835. doi: 10.1038/ng.2669
- Smit, A. F. A., Hubley, R., and Green, P. “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., Furiki, 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 MK2/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: 10.1104/pp.16.01222
- 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: 10.1080/00380768.2000.10408807
- 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 salsgineum*. *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., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107
- Zhang, X., Gou, M., and Liu, C. J. (2013). Arabidopsis Kelch Repeat F-Box Proteins Regulate Phenylpropanoid Biosynthesis via Controlling the Turnover of Phenylalanine Ammonia-Lyase. *Plant Cell.* 25, 4994–5010. doi: 10.1105/tpc.113.119644
- Zhu, Q., Dugardeyn, J., Zhang, C., Mühlenbock, 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/sst102

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.