



Construction of the First SNP-Based Linkage Map Using Genotyping-by-Sequencing and Mapping of the Male-Sterility Gene in Leaf Chicory

Fabio Palumbo¹, Peng Qi^{2,3}, Vitor Batista Pinto⁴, Katrien M. Devos^{2,3} and Gianni Barcaccia^{1*}

¹ Laboratory of Genomics for Plant Breeding, Department of Agronomy Food Natural Resources Animals and Environment (DAFNAE), University of Padova, Legnaro, Italy, ² Institute of Plant Breeding, Genetics and Genomics, Department of Plant Biology, University of Georgia, Athens, GA, United States, ³ Institute of Plant Breeding, Genetics and Genomics, Department of Crop and Soil Sciences, University of Georgia, Athens, GA, United States, ⁴ Department of General Biology, Federal University of Viçosa, Viçosa, Brazil

OPEN ACCESS

Edited by:

Roberto Papa,
Marche Polytechnic University, Italy

Reviewed by:

Therese Bengtsson,
Swedish University of Agricultural
Sciences, Sweden

Ezio Portis,
University of Turin, Italy

*Correspondence:

Gianni Barcaccia
gianni.barcaccia@unipd.it

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 14 December 2018

Accepted: 20 February 2019

Published: 11 March 2019

Citation:

Palumbo F, Qi P, Pinto VB,
Devos KM and Barcaccia G (2019)
Construction of the First SNP-Based
Linkage Map Using
Genotyping-by-Sequencing
and Mapping of the Male-Sterility
Gene in Leaf Chicory.
Front. Plant Sci. 10:276.
doi: 10.3389/fpls.2019.00276

We report the first high-density linkage map construction through genotyping-by-sequencing (GBS) in leaf chicory (*Cichorium intybus* subsp. *intybus* var. *foliosum*, $2n = 2x = 18$) and the SNP-based fine mapping of the linkage group region carrying a recessive gene responsible for male-sterility (*ms1*). An experimental BC₁ population, segregating for the male sterility trait, was specifically generated and 198 progeny plants were preliminary screened through a multiplexed SSR genotyping analysis for the identification of microsatellite markers linked to the *ms1* locus. Two backbone SSR markers belonging to linkage group 4 of the available *Cichorium* consensus map were found genetically associated to the *ms1* gene at 5.8 and 12.1 cM apart. A GBS strategy was then used to produce a high-density SNP-based linkage map, containing 727 genomic loci organized into 9 linkage groups and spanning a total length of 1,413 cM. 13 SNPs proved to be tightly linked to the *ms1* locus based on a subset of 44 progeny plants analyzed. The map position of these markers was further validated by sequence-specific PCR experiments using an additional set of 64 progeny plants, enabling to verify that four of them fully co-segregated with male-sterility. A mesosyteny analysis revealed that 10 genomic DNA sequences encompassing the 13 selected SNPs of chicory mapped in a peripheral region of chromosome 5 of lettuce (*Lactuca sativa* L.) spanning about 18 Mbp. Since a MYB103-like gene, encoding for a transcription factor involved in callose dissolution of tetrads and exine development of microspores, was found located in the same chromosomal region, this orthologous was chosen as candidate for male-sterility. The amplification and sequencing of its CDS using accessions with contrasting phenotypes/genotypes (i.e., 4 male sterile mutants, *ms1ms1*, and 4 male fertile inbreds, *Ms1Ms1*) enabled to detect an INDEL of 4 nucleotides in its second exon,

responsible for an anticipated stop codon in the male sterile mutants. This polymorphism was subsequently validated through allele-specific PCR assays and found to fully co-segregate with male-sterility, using 64 progeny plants of the same mapping BC₁ population. Overall, our molecular data could be practically exploited for genotyping plant materials and for marker-assisted breeding schemes in leaf chicory.

Keywords: *Cichorium intybus*, genetic linkage map, male sterility, *ms1* locus, single nucleotide polymorphism (SNP) markers, genotyping-by-sequencing (GBS), transcription factor MYB103

INTRODUCTION

Linkage maps based on molecular markers play a key role in the study of the genetics and genomics of crop plants. Among the possible applications, the development of high-density linkage maps has simplified the discovery of Mendelian genes (Kaur et al., 2014; Zhao et al., 2015; Huang et al., 2016) and quantitative trait loci (QTL) (Colasuonno et al., 2014; Marcotuli et al., 2017; Schumann et al., 2017). The first genetic linkage map of chicory (*Cichorium intybus* subsp. *intybus* L.), a leafy vegetable crop belonging to the family Asteraceae and widely cultivated in many European countries, consisted of 431 SSR and 41 EST markers, and covered 878 cM (Cadalen et al., 2010).

Chicory is a diploid plant species ($2n = 18$) that is naturally allogamous, due to an efficient sporophytic self-incompatibility system (Barcaccia et al., 2003; Lucchin et al., 2008). In addition, outcrossing is promoted by a number of traits, including: (i) a floral morpho-phenology (i.e., proterandry, with the anthers maturing before the pistils) unfavorable to selfing in the absence of pollen donors (Pécaut, 1962; Desprez et al., 1994); and (ii) a competitive advantage of allo-pollen grains and tubes (i.e., pollen genetically diverse from that produced by the seed parents) (Desprez and Bannerot, 1980; Eenink, 1982). Two main botanical varieties can be recognized within *C. intybus* subsp. *intybus* to which all the cultivated types of chicory belong. The first is var. *foliosum*, which traditionally includes Witloof chicory, Pain de sucre, Catalogne and Radicchio and all the cultivar groups whose commercial products are the leaves (i.e., leaf chicory). The second is var. *sativum* and comprises all the types whose commercial product, either destined to industrial transformation or direct human consumption, is the root (i.e., root chicory). In root chicory, Gonthier et al. (2013) identified molecular markers associated with the Nuclear Male-Sterility 1 (NMS1) locus and the Sporophytic Self-Incompatibility (SSI) locus. These two loci were both mapped to narrow genomic regions belonging, respectively, to linkage groups 5 and 2 of the genetic map developed by Cadalen et al. (2010). Similarly, in leaf chicory, Barcaccia and Tiozzo Caenazzo (2012, 2014) mapped molecular markers linked to the male-sterility gene (*ms1*) within linkage group 4, according to the map by Cadalen et al. (2010).

Recently, a chicory genetic linkage map spanning 1,208 cM was developed by Muys et al. (2014) using an F₂ population composed of 247 plants. This map comprised 237 markers (i.e., 170 AFLP, 28 SSR, 27 EST-SNP, and 12 EST-SSR markers) and covered about 84% of the chicory genome. The markers were then used to find potential orthologous based on sequence homology in mapped lettuce EST clones from the Compositae Genome

Project Database (Muys et al., 2014). A total of 27 putative orthologous pairs were retained, pinpointing seven potential blocks of synteny that covered 11% of the chicory genome and 13% of the lettuce genome, opening new avenues for the comparative analysis of these two species.

Mapping of the self-incompatibility and male-sterility mechanisms in chicory is important, not only to understand the genetic basis of the main reproductive barriers that act in flowering plants, but also because of the potential applications of these loci for breeding F₁ hybrid varieties. In fact, although in the past chicory varieties were mainly synthetics produced by intercrossing a number of phenotypically superior plants, selected on the basis of morpho-phenological and commercial traits, recently private breeders and seed firms have developed methods for the development of F₁ hybrids.

In the last century, male sterile mutants have allowed the exploitation of heterosis (i.e., hybrid vigor) through the development of F₁ hybrid varieties in many agricultural and horticultural crops. In general, male-sterility is defined as the failure of plants to develop anthers or to form functional pollen grains and it is more prevalent than female-sterility. In nature, male sterile plants have reproduction potentials because they can still set seeds, as female-fertility is unaffected by most of the mutations responsible for male-sterility. This behavior is known to occur spontaneously via mutations in nuclear and/or cytoplasmic genes involved in the development of anthers and pollen grains. Barcaccia and Tiozzo Caenazzo (2012, 2014) have recently identified and characterized a spontaneous male sterile mutation in cultivated populations of leaf chicory, namely Radicchio (*C. intybus* subsp. *intybus* var. *foliosum* L.). Cytological analyses revealed that microsporogenesis proceeds regularly up to the development of tetrads when the microspores arrest their developmental program showing a collapse of the exine. At the beginning of microgametogenesis, non-viable shrunken microspores were clearly visible within anthers. Moreover, genetic segregation data derived from replicated F₂ and BC₁ populations clearly supported a nuclear origin, monogenic control and recessive nature of the male-sterility trait in the leaf chicory mutants (Barcaccia and Tiozzo Caenazzo, 2012, 2014).

In this work, taking advantage of the method of genotyping based on 27 mapped microsatellite marker loci scattered throughout the linkage groups of leaf chicory (Ghedina et al., 2015) and the first genome sequence draft of leaf chicory with the functional annotation of more than 18,000 unigenes (Galla et al., 2016), we successfully constructed a high-density linkage map and finely mapped the *ms1* locus in leaf chicory. After a preliminary genetic mapping of the *ms1* locus using SSR and

EST markers, a Genotyping-By-Sequencing (GBS) methodology was used to narrow down the chromosomal window around the *ms1* gene, first of all developing well-saturated linkage groups for this species and then selecting molecular markers and candidate genes for male-sterility exploitable for marker-assisted breeding and gene cloning programs.

MATERIALS AND METHODS

Plant Materials and Genomic DNA Extraction

Several male sterile mutants sharing the same genotype at the *ms1* locus were discovered by T&T Srl Agricola (Blumen Group SpA) within local varieties of radicchio “Red of Chioggia” stemmed from recurrent phenotypic selection programs (Barcaccia and Tiozzo Caenazzo, 2012, 2014). A backcross (BC₁) population segregating 1:1 for the male-sterility trait and comprising 198 individual plants was generated as follows. A male-sterile mutant plant (genotype *msms*), belonging to a cultivated population of radicchio was crossed with a male-fertile plant (genotype *MsMs*) selected from a local accession of wild chicory, in order to maximize genetic diversity and polymorphism levels. An individual F₁ male-fertile plant heterozygous at the male-sterility locus (genotype *Msms*) was selfed and one F₂ male-sterile progeny (genotype *msms*) was then backcrossed as seed parent to a sister F₁ male-fertile plant (genotype *Msms*) used as pollen donor (Figure 1). The agronomic field trials for plant phenotyping (discrimination of male-fertile and male-sterile phenotypes) were conducted at the experimental farm “Lucio Toniolo” of the University of Padova, located in Legnaro (Padova, Italy) (45°21′5,68″N 11°57′2,71″E – 8 m above the sea level). Seeds were sown in February and grown in heated greenhouse, under light/dark cycle conditions of 12/12 h and temperature of 20 °C. Uniformly sized, 4 week-old seedlings were transplanted in the field under polyethylene tunnels and under controlled pollination conditions. The soil texture was the following: 46 sand, 24% clay and 30% loam; pH = 7.9; electric conductivity 112 μS; organic carbon 1.1%. Flowering started around 150 days after sowing and all individuals of the BC₁ population were phenotyped using three flowers per plant by visual observations of the anthers (*in vivo* screening for the presence/absence of pollen) and cytological investigations (*in vitro* staining of pollen). More in details, cytological investigations were accomplished using both aceto-carmin and DAPI solutions. The aceto-carmin staining technique was used to measure discriminant phenotypic descriptors, such as shape, size and coloration of pollen grains, as reported by Janssen and Hermsen (1976) and implemented by Barcaccia et al. (1998), while DAPI staining was accomplished as described by Barcaccia and Tiozzo Caenazzo (2012). Microscopy characterized a mutant phenotype by shapeless, small and shrunken microspores as compared to the wild-type ones (Figure 1). It is worth mentioning that in mutant plants at the stage of dehiscent anthers, microspores were arrested in their development at the uninucleate stage, and collapsed before their release from the tetrads. Viable pollen grains were never detected in mature

anthers, demonstrating full expression of the male-sterility trait (Barcaccia and Tiozzo Caenazzo, 2012, 2014).

Total genomic DNA of the parents and progeny was isolated from 100 mg of fresh leaf tissue using the DNeasy® Plant mini-kit (Qiagen, Hilden, Germany) following the recommendations of the manufacturer. Quality and concentration of DNA samples were estimated by spectrophotometric analysis (NanoDrop 2000c UV-Vis, Thermo Fisher Scientific, San Jose, CA, United States) and quality was also assayed by agarose gel electrophoresis (1.0% w/v agarose TAE 1× gel containing 1× SYBR® Safe, Thermo Fisher Scientific).

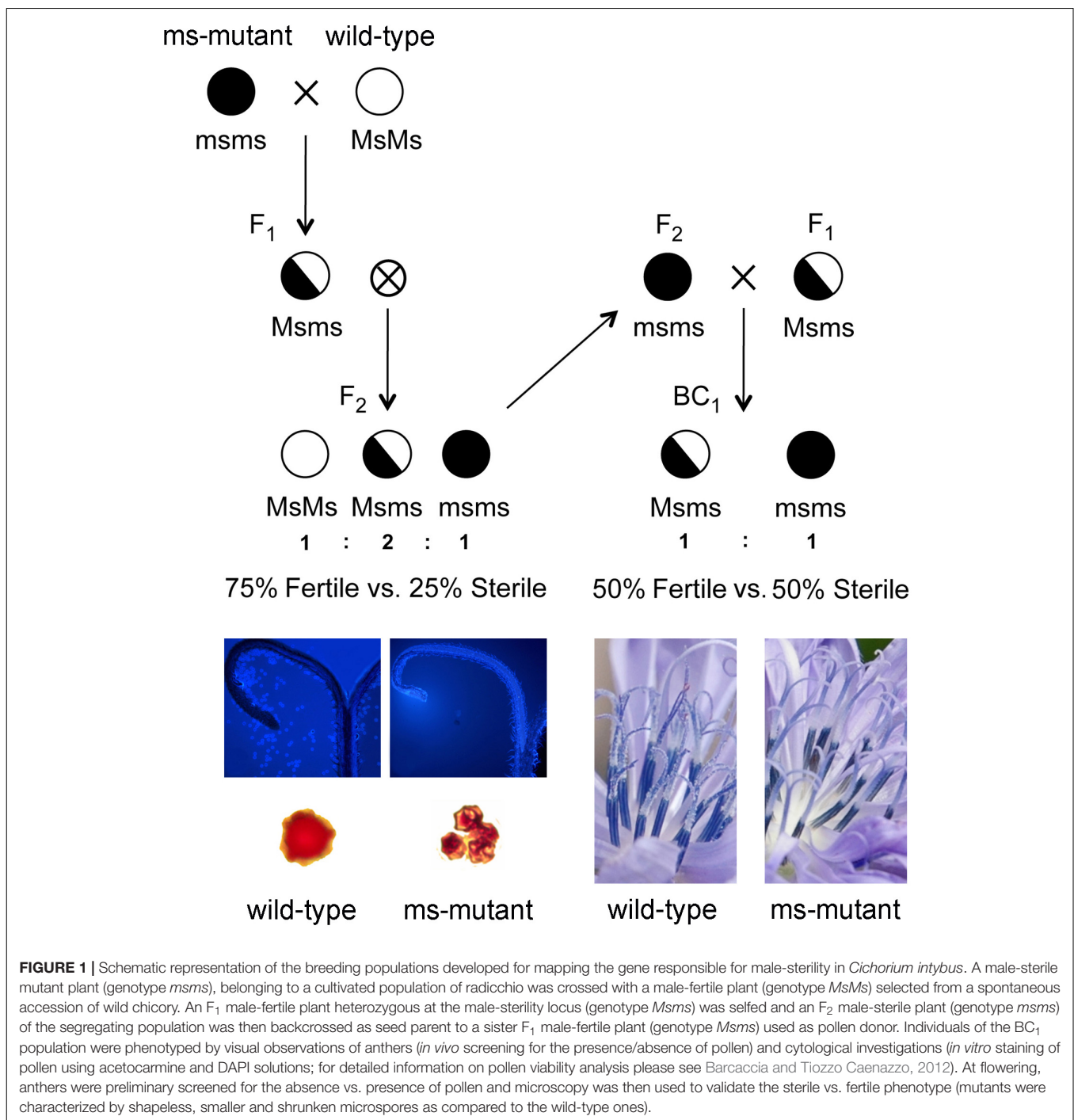
Preliminary Genetic Mapping of the *ms1* Locus With Simple Sequence Repeat (SSR) and Cleaved Amplified Polymorphic Sequence (CAPS) Markers

The entire BC₁ population of leaf chicory was used for genetic mapping of the male-sterility gene using three selected SSR markers (M4.12, M4.11b, and M4.10b) and one EST-derived CAPS marker previously mapped on linkage group 4 (Cadalen et al., 2010; Barcaccia and Tiozzo Caenazzo, 2012, 2014). This step was preliminarily applied in order to determine and validate upstream and downstream backbone DNA markers encompassing the *ms1* locus.

An AFLP-derived amplicon corresponding to marker E02M09 (Barcaccia and Tiozzo Caenazzo, 2012), whose sequence was found to encompass a microsatellite region (GenBank accession JF748831), was converted into a SSR marker and renamed as M4.12 by Ghedina et al. (2015). This sequence-tagged site marker was mapped on linkage group 4 and was therefore included in the genetic analysis of the BC₁ population. Among the microsatellite marker loci publicly available for the leaf chicory genome (Cadalen et al., 2010) and associated to the linkage group 4, the M4.11b (Ghedina et al., 2015) (synonym EU03H01) contained an imperfect [(TG)₅CG(TG)₇] microsatellite motif (GenBank accession KF880802) and M4.10b (Ghedina et al., 2015) [synonym EU07G10 (Cadalen et al., 2010)] carried a (CT)₈TT(CT)₅CC(CT)₃TT(CT)₇ microsatellite motif (GenBank accession KX534081).

For SSR amplification, the three-primer strategy reported by Schuelke (2000) was adopted, with some modifications. Briefly, forward primers were tagged at their 5′ end with universal sequences (M13 or PAN2, unpublished) and used in PCR reactions in combination with sequence-specific reverse primers, and M13 and PAN2 oligonucleotides labeled with the fluorophores 6-FAM and NED, respectively.

The PCR reactions were performed in a total volume of 10 μl containing approximately 20 ng of gDNA template, 1× Platinum® Multiplex PCR Master Mix (Applied Biosystems, Carlsbad, CA, United States), GC enhancer 10% (Applied Biosystems), 0.05 μM tailed forward primer (Invitrogen Corporation, Carlsbad, CA, United States), 0.1 μM reverse primer (Invitrogen Corporation) and 0.23 μM universal primer (Invitrogen Corporation). The following thermal conditions were adopted for all reactions: 2 min at 95°C for the initial denaturation step,



45 cycles at 95°C for 30 s, 55°C for 30 s and 72°C for 45 s. A final extension step at 72°C for 30 min terminated the reaction, to fill-in any protruding ends of the newly synthesized strands.

Amplicons were initially separated and visualized on 2% agarose gels in 1× TAE gel containing 1× Sybr Safe DNA stain (Life Technologies, Carlsbad, CA, United States). The remainder of the fluorescent labeled PCR products (8 μl) was subjected to capillary electrophoresis on an ABI PRISM 3130xl Genetic

Analyzer (Thermo Fisher). LIZ500 (Applied Biosystems) was used as molecular weight standard.

The CAPS marker was developed from a MADS-box gene (GenBank accession AF101420) which was initially considered (Cadalen et al., 2010) but later disproven (Barcaccia et al., 2016) to be a candidate gene for male-sterility. Several primer pairs were designed for nested PCR assays using PerlPrimer v1.1.21 and used to amplify the full-length sequence and sub-regions of the MADS-box gene from plants of the BC₁ population phenotyped

for male-sterility. Amplification reactions were performed in a 9700 Thermal Cycler (Applied Biosystems) with the following conditions: initial denaturation at 94°C for 5 min followed by 30 cycles at 94°C for 30 s, 57°C for 30 s, 72°C for 60 s and a final extension of 10 min at 72°C, and then held at 4°C. The quality of PCR products was assessed on a 2% (w/v) agarose gel stained with 1× SYBR® Safe™ DNA Gel Stain (Life Technologies). Amplicons were Sanger-sequenced to detect SNPs potentially associated with male-sterility and to locate restriction sites. PCR products were then cleaved using *NcoI* (Promega, Madison, WI, United States) as endonuclease specific for single nucleotide variants, following the protocol suggested by the manufacturer. Amplicons were digested at 37°C for 2 h. CAPS variants were visualized on 2.5% (w/v) agarose gels (Life Technologies) stained with 1× SYBR® Safe™ DNA Gel Stain (Life Technologies).

Segregation data from the three SSR markers and the MADS-box gene-specific CAPS marker were analyzed with JoinMap® v. 2.0 (Stam, 1993) using the BC₁ population type option. Genetic association between each of the markers and the male-sterility trait was assessed by recording the target *ms1* locus as a qualitative trait. The grouping module was applied with a LOD threshold of 3 and a maximum recombination frequency *r* of 40%. The genetic distance between each pair-wise comparison of marker locus and target locus, expressed in centiMorgans (cM), was calculated from the recombination frequency corrected with the Kosambi's mapping function (Kosambi, 1943). MapChart v.2.3 (Voorrips, 2002) was used to display the map.

Construction of a SNP-Based Linkage Map Using Genotyping-by-Sequencing (GBS)

Genomic DNA from 22 male sterile progeny and 22 male fertile progeny from the BC₁ population, along with 2 DNA samples from the parental plants were quantified with the dsDNA BR assay on a Qubit® 1.0 fluorometer (Invitrogen, Carlsbad, CA, United States) and DNA concentrations were normalized to 20 ng/μl. DNA samples were shipped to LGC Genomics (Berlin, Germany) for GBS library preparation, sequencing and subsequent bioinformatic analysis. Briefly, DNA samples were digested with the restriction enzyme *MspI* (NEB, Beverly, MA, United States) and the indexed Illumina libraries were prepared using the Ovation Rapid DR Multiplex System (Nugen, Leek, Netherlands) according to the protocol provided by the manufacturer. The 46 libraries were pooled removing PCR primers and small amplicons by Agencourt XP bead purification (Beckman Coulter, High Wycombe, United Kingdom) and then normalized using the Trimmer Kit (Evrogen, Moscow, Russia). The normalized library pool was amplified using MyTaq polymerase (Bioline, Taunton, MA, United States) and standard Illumina TrueSeq amplification primers (Illumina Inc., San Diego, CA, United States). The nGBS library was finally size selected on a LMP-Agarose gel, removing fragments smaller than 300 bp and larger than 500 bp, and sequenced on a single lane of an Illumina NextSeq 500 v2 (2 × 150 bp, Illumina Inc., San Diego, CA, United States).

Raw reads were de-multiplexed and split according to their barcodes using Illumina bcl2fastq 2.17.1.14 software (Illumina, San Diego, CA, United States) and sequencing adapter remnants were clipped. After this step, using proprietary LGC Genomics software reads were processed as follows: (i) trimming of the 3'-end (to get a minimum average Phred quality score of 20 over a window of ten bases); (ii) discarding reads with 5'-ends not matching the restriction enzyme site; (iii) removing all reads containing undetermined (N) bases; (iv) discarding reads with final length < 64 bases. CD-HIT EST (Fu et al., 2012) clustered all the processed sequences in a reference catalog of consensus loci, allowing up to 5% difference. BWA v.0.7.12¹ was used to align the reads from each sample against the newly constituted reference catalog and FreeBayes v1.0.2-16² was then employed at default settings for variant discovery. The raw SNP variants were filtered by applying customized Perl scripts according to the following rules: (i) minimum allele count exceeding eight reads; (ii) allele frequency across all samples between 5 and 95%; (iii) genotypes observed in at least 32 samples; (iv) discard adjacent SNPs and SNPs with more than two alleles (i.e., only biallelic SNPs were taken into account).

Segregation data, analyzed using a modified version of MapMaker v3 software³ consisted of GBS-SNP data, genotypic scores for SSR markers M4.10b, M4.12, M2.6, and M2.4 (Ghedina et al., 2015) and qualitative scores for the male-sterility locus *ms1*. SSRs M2.6 and M2.4 were included because they have been shown by Gonthier et al. (2013) to be associated with the self-incompatibility locus. Linkage groups were formed at a LOD threshold of 5. Marker orders within single linkage groups were determined using the MapMaker functions 'order', 'try', and 'ripple', and were checked manually to ensure optimal placement of the marker loci. Genetic distances were calculated using the Kosambi mapping function and graphically represented using MapChart v2.3 (Voorrips, 2002).

The newly developed genetic map was enriched by locating markers on the first genome draft of leaf chicory (Galla et al., 2016) using a BLASTN approach (similarity >90%, *E*-value < 1E-50). Putative functional annotation of genes present in the mapped contigs was performed using BLASTX against the TAIR10 (Berardini et al., 2015) database (*E*-value < 1E-5).

Validation of SNP Variants Linked to the Male-Sterility Locus Through Allele Specific (AS)-PCR Assays

All SNPs potentially associated with the male-sterility *ms1* locus and exhibiting a maximum of three recombinant events with *ms1* were validated in a larger number of progeny through allele-specific PCR (AS-PCR) assays.

A total of 64 genomic DNA samples (i.e., 32 male sterile plants and 32 male fertile plants) from the same BC₁ population but not included in the GBS analysis were used for amplification using two sets of primers for each male-sterility associated SNP. Each primer set consisted of a different allele-specific

¹<http://bio-bwa.sourceforge.net/>

²<https://github.com/ekg/freebayes#readme>

³<http://rna-informatics.uga.edu/malmberg/rlmlab>

forward primer and a common locus-specific reverse primer. The two allele-specific PCR primers were designed so that the 3' nucleotide was complementary to one allele of the putative polymorphism. Amplicons were produced in a 9700 Thermal Cycler (Applied Biosystems) with the following conditions: initial denaturation at 94°C for 5 min followed by 30 cycles at 94°C for 30 s, 55°C for 30 s, 72°C for 60 s and a final extension of 10 min at 72°C, and then held at 4°C. PCR products were separated on 2.0% w/v agarose TAE 1× gels containing 1× SYBR® Safe stain (Thermo Fisher Scientific). Segregation data for *ms1* and *ms1*-associated markers obtained in the entire set of 108 (44 used in GBS + 64 used in AS-PCR) BC₁ progeny were used to build a new genetic linkage map for the chromosomal block carrying the target locus with JoinMap® v. 2.0 (Stam, 1993). The BC₁ population type option was adopted and the Kosambi mapping function was used to calculate genetic distances. The resulting map was drawn with MapChart v.2.3 (Voorrips, 2002).

Micro-Syntenic Comparison Between *Lactuca sativa* and *Cichorium intybus* Homologous Chromosomal Segments

Considering that the recent release of the high quality *Lactuca sativa* genome assembly (Reyes-Chin-Wo et al., 2017) represents a reference assembly for the whole Asteraceae family, a BLASTN approach (*E*-value < 1E-20) was performed aligning the chicory contigs carrying the 13 SNPs associated to the *ms1* locus against the genome of lettuce. Robust evidences of micro-syntenic between these two species, made MYB103 worth to be investigated due to its possible functional involvement in the male sterility (Zhang et al., 2007). Since contig_119275 from the chicory genome draft, showed the best match with *Lsat_1_v5_gn_5_561* locus, annotated in *L. sativa* as MYB103, it was used to design three couple of primers, covering the entire coding DNA sequence (CDS, **Table 1**). Amplicons of four commercial male sterile accessions (namely D49, A89, B86, and D17) and four male fertile inbred lines (namely 2111, 202, 231, and 334) were Sanger-sequenced to detect polymorphisms between these two groups with contrasting microgametogenesis. According to a four nucleotides insertion observed in all male sterile accessions within the putative MYB103 gene, two additional couples of primer were used in an AS-PCR assay to test the 64 genomic DNA samples (i.e., 32 male sterile plants and 32 male fertile plants) from the BC₁ population previously used to validate the *ms1*-associated SNPs (**Table 1**). In details two allele-specific forward primers were designed so that one amplified the wild type allele (without the insertion) while the other amplified only the mutated allele (with the insertion). Two reverse primers were designed within a conserved region of the second exon in order to produce amplicons of different size, distinguishable by standard agarose gel electrophoresis (i.e., MYB103_MS and MYB103_MF, **Table 1**). The two PCR reactions specifically designed for the wild type and the mutated alleles were performed separately for each DNA sample at the same conditions used for the previous AS-PCR analysis. The two PCR outputs of each individual were mixed and run together on a

2.0% w/v agarose TAE 1× gels containing 1× SYBR® Safe stain (Thermo Fisher Scientific).

RESULTS

Molecular Mapping of the SSR and CAPS Markers in the Linkage Group Carrying the Male-Sterility Locus

The fine genetic mapping of three selected SSR markers (**Table 1**) was successfully pursued and the genetic recombination estimates were validated using 198 BC₁ individual plants of the segregating population on the basis of chi-square values against independent assortment patterns (**Supplementary Table S1**). The genetic distances between the male-sterility gene and the microsatellite markers M4.12 (EU02M09) and M4.11b (EU03H01) mapped apart from the *ms1* locus were equal to 5.8 and 12.1 cM, respectively. An additional mapped microsatellite marker, M4.10b (EU07G10), belonging to the same linkage group, was located downstream the *ms1* locus at a genetic distance of 31.2 cM (**Figures 2A,B**). The DNA sequences of the genomic regions containing these SSR markers were deposited in GenBank as accessions KX534081, KF880802, and JF748831. According to the nomenclature of Cadalen et al. (2010), the gene responsible for male-sterility was associated to linkage group 4, as predicted by Barcaccia and Tiozzo Caenazzo (2012, 2014).

Sequence data published by Galla et al. (2016) on the genome draft of leaf chicory were used to discover, predict and annotate all genes of the genomic contigs encompassing the three molecular markers used for the SSR analysis. Regarding M4.12 (Ghedina et al., 2015), the mapped marker closest to the *ms1* locus, its sequence was found to match with contig_84164, long 36,250 nucleotides and putatively linked with gene models AT3G11330, AT4G03600, AT5G50170, and AT5G63130 (**Table 2**). It is worth noting that the former one encodes for PIRL9, a member of the Plant Intracellular Ras-group-related LRRs (Leucine rich repeat proteins) and is required for differentiation of microspores into pollen grains.

Concerning the MADS-box L2/R21 gene (AF101420) as candidate, the alignment of sequences of part of its 5'-UTR region, exon 1 and the early region of intron 1 recovered from both male-sterile mutants and wild-type plants enabled to map the MADS-box locus on the linkage group 4 (Cadalen et al., 2010) by means a CAPS markers. In fact, three SNPs determining a restriction site were discovered in the amplified sequence of the first exon of the MADS-box gene. The cleavage site of the six-base cutter *NcoI* endonuclease was found to include a polymorphism at position 61 of the nucleotide sequence of the male fertile genotype when compared with the male sterile genotype (GenBank accessions KX455840 and KX455841). The amplification-restriction protocol for the detection of the CAPS marker alleles was applied to the total 198 BC₁ individual plants of the mapping population. In particular, 14 individuals scored recombinant genotypes when compared with the male-sterile and male-fertile phenotypes, so that the MADS-box gene was mapped at a genetic distance of 7.3 cM apart from the

TABLE 1 | List of primer pairs used to amplify the mapped markers and the candidate gene MYB103 of linkage group 9 [corresponding to linkage group 4 of Cadalen et al. (2010)].

Locus name	GenBank ID	Polymorphism	Forward primer	Reverse primer
EU02M09 ³ M4.12 ²	JF748831 ³	(TC) _n	GGCATCGGGATAGAAAAACA	TCAATGCCTCAACAGAAATCC
EU03H01 ^{1,3} M4.11b ²	KF880802 ^{1,2,3}	(TG) _n CG(TG) _n	GCCATTCTTTCAAGAGCAG	AACCCAAAACCGCAACAATA
EU07G10 ¹ M4.10b ²	KX534081 ⁴	(CT) _n CATA/(CA) _n CT(CA) _n	CATCCATTATTGGGCAG	CACCAACGAACTCCTTACAAA
MADS box L2/R2 ¹	AF101420 ¹	<i>Nco</i> I	TTTTTGTGGGGTTTTGATTTAGA	TGAGATTGCATGAATGAGAACA
CAPS marker ⁴	KX45584/0/1 ⁴			
T11292 ⁴	KX789069 ⁴	g.205A > C	GAATGAAAATTGACATAATC GATGCATTAACATGGGTCT	TTTGTGATTCTGTTCCTG GACTGATGGATGTCCAAT
T4403 ⁴	KX789070 ⁴	g.253A > T	GCGAACRAATGAGGATATATGAG	GTGTTGATTGAGTGAAATCT GTGTTGATTGAGTGAAATCA
T4390 ⁴	KX789071 ⁴	g.102C > T*; g.139T > C	TGTAACGCCCGTAAACCCAA	GACATGTTTACTAAGGTGATGATAATATAA GACATGTTTACTAAGGTGATGATAATATAG
T4391 ⁴	KX789072 ⁴	g.150C > T	TAAATGTGCAATACCATGAAGC TAAATGTGCAATACCATGAAGT	AAGTGAAGTAAAGTGGTTGATTC
T4393 ⁴	KX789073 ⁴	g.137A > G	AACACATGAAGGMACTCTAG AGGTCCCTCATATTAAGG	GATGGGTATTGAACTTATG GATGTTTGTGAATGATGTTT
T4392 ⁴	KX789074 ⁴	g.68C > T	TTGTTGGAAGTGATGAGGTGT	TGTTATTAAGTGTTCCTGATATAG TGTTATTAAGTGTTCCTGATATAA
T4399 ⁴	KX789075 ⁴	g.24T > A	AACATGATTGTCTGCCA GACATTTTTGGAACACTTTTTTA	CATCACAATATTCTATCCAAA ACTGTTACATAATGGCTAG
T4394 ⁴	KX789076 ⁴	g.132T > C; g.235G > A*	GTTGGTCTGTTGATTGTTGG	CATGTATATTTGGAAGTTATCAACA CATGTATATTTGGAAGTTATCAACG
T4401 ⁴	KX789077 ⁴	g.89T > A	GTAAGAGTGGGAGTCATATA GTAAGAGTGGGAGTCATATT	CTTGTTCCTTTGTTCTAGA GTTTTCAACATCMTTCACT
T4402 ⁴	KX789078 ⁴	g.166T > G	CAGTGGCAATGCACAT GGAAGTAAACTATACTATCGC	ATAGTCTCATGGTATGCGAGT CATGACCACAAGGTAACC
T4400 ⁴	KX789079 ⁴	g.231G > A	CTTAATTAGGACAGAACTTAATATAGACAG CTTAATTAGGACAGAACTTAATATAGACAA	GGAGTGGGGTTAAAGGGAA
T4395 ⁴	KX789080 ⁴	g.179C > T*; g.214T > C	GATTCTGCAATCGGATATCGCCTTT GATTCTGCAATCGGATATCGCCTTC	ACACCGCTGATAACCACCACC
T4398 ⁴	KX789082 ⁴	g.175C > T	TGCTGCCCGGAGAAATGTTAG	TTTAGGCAACCGAAGTAAGGCAGTG TTTAGGCAACCGAAGTAAGGCAGTA
MYB103_exon1 ⁴	MK285053– MK285054	–	GAGCGAGAGTGGTGGCTAT	TTACCTCACCCAACCAGAAA
MYB103_intron1 ⁴		–	CAGAGTCATCGCGTTTCTA	GTGTGAAGAAGGGTCAGTGG
MYB103_exon2 ⁴		–	GATCACAGGCGTCAAGGTTA	CGGAGATAGCCACCCTCT
MYB103_MS ⁴	–	–	CGTTTGCTTCATGCTTAATT	AAGAACCATTGGAACACGAA
MYB103_MF ⁴	–	–	CGTTTGCTTCATGCTTAGA	GGAGATAGCCACCCTCT

¹Cadalen et al., 2010; ²Ghedina et al., 2015; ³Barcaccia and Tiozzo Caenazzo, 2012, 2014; ⁴Present paper.

The name of the locus, the GenBank accession of the trait amplified, the polymorphism observed in *ms* mutants and the primer pairs are reported. SNPs marked with an asterisk were not tested with allele-specific primer combinations.

ms1 locus (Figures 2A,B). A genomic sequence, corresponding to contig_95308, long 8,023 nucleotides, was found to match with the sequence encompassing the CAPS marker that, from a BLASTX approach with the TAIR database, proved to be annotated as AT4G24540, encoding for a protein involved in flowering (Table 2).

The First SNP-Based Linkage Map of Chicory Through GBS

A GBS approach was applied to a subset of the BC₁ population, consisting of 22 male sterile and 22 male fertile plants, in order

to build the first SNP-based linkage map in *Cichorium* spp. and to identify markers associated with the *ms1* locus. NextSeq 500 v2 Illumina platform produced 419,884,246 raw reads. After quality and adapter trimming, we obtained 339,081,743 reads that were used to create a reference catalog of 1,192,451 consensus loci. A raw pool of 16,353 SNPs was identified using Freebayes v1.0.2-16. After removal of (1) SNPs with more than 30% of missing data, (2) SNPs with a sequence depth $\leq 8\times$, (3) tri- and tetra-allelic SNPs, and (4) SNPs with allele frequencies across all samples $\leq 5\%$ and $\geq 95\%$, 1,995 SNPs were retained for the construction of a genetic linkage map. A total of 727 SNPs clustered and mapped into 9 linkage groups spanning a total

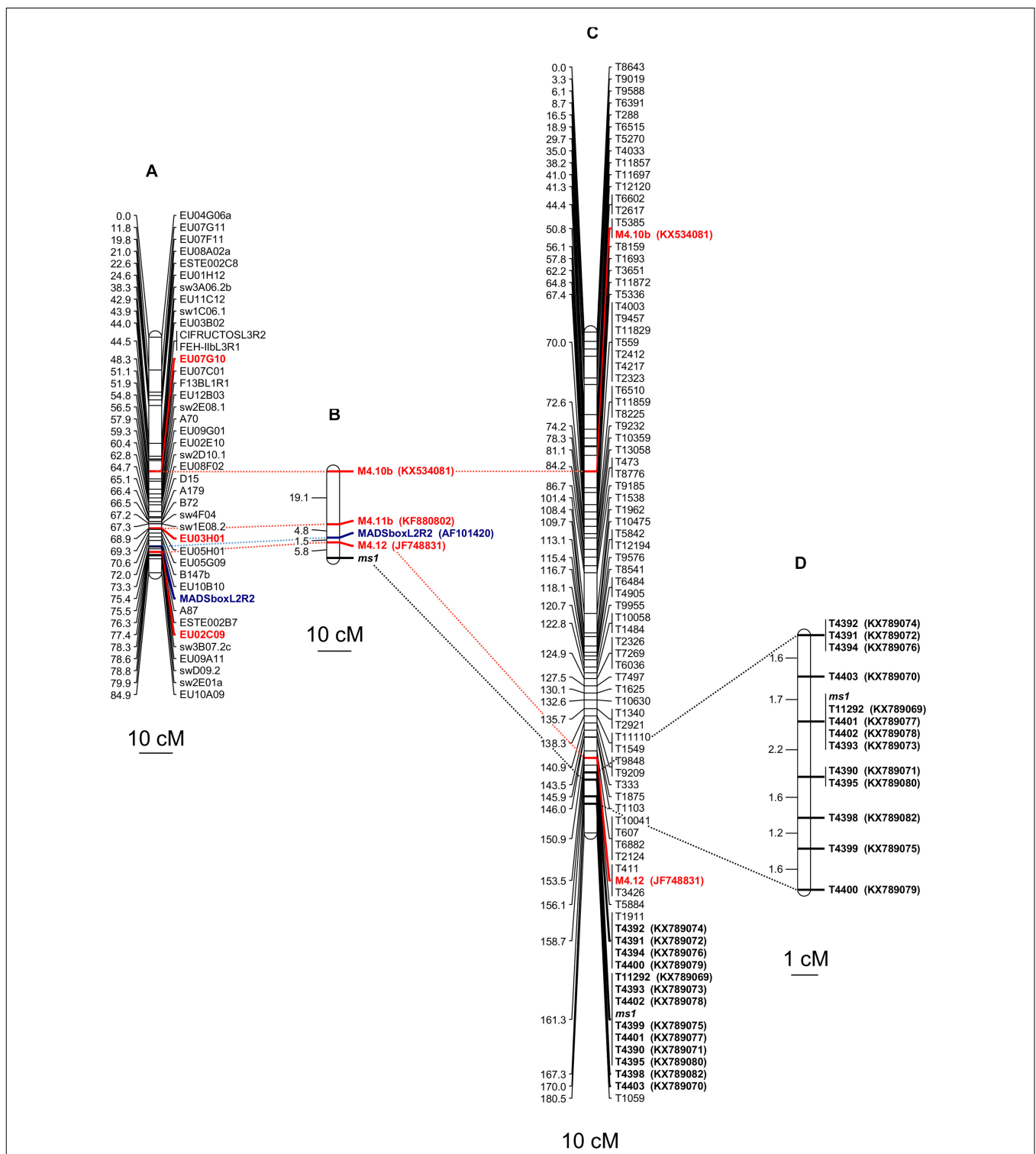


FIGURE 2 | Molecular markers mapped on linkage group 4 and found associated to the male sterility locus (*ms1*) of *C. intybus*. **(A)** Linkage group 4 constructed from Cadalen et al. (2010). **(B)** Genetic distances, expressed in cM, among three SSR markers (in red), one CAPS marker derived from the MADS-box region (in blue) and the male-sterility locus, whose estimates were calculated using a BC₁ population (198 samples) segregating 1:1 for the male-sterility trait. The SSR markers were retrieved from Ghedina et al. (2015) while the CAPS marker was developed in the present work. **(C)** Linkage group 9 [corresponding to linkage group 4 of Cadalen et al. (2010)] constructed using the data from the Genotyping-By-Sequencing (GBS) approach and based on 44 samples of the BC₁ population mentioned above. In red are reported the two SSR already used for the first SSR-approach, in bold are listed the 13 SNP showing ≤ 3 recombinants with the *ms1* locus. **(D)** Chromosomal region around the *ms1*, considering the recombinant data of the aforesaid 13 SNP and a total number of 108 BC₁ samples.

TABLE 2 | List of markers that, aligning against contigs of the first genome draft (Galla et al., 2016), were functionally annotated on the bases of matches with *Arabidopsis* protein database (TAIR10).

Marker map reference	GenBank ID	Marker GenBank ID	Marker localization	Gene model name	Predicted function
E02M09 ³ M4.12 ²	KX639717	JF748831 ³	Upstream	AT5G50170	C2 calcium/lipid-binding and GRAM domain containing protein
	KX639715		Upstream	AT3G11330	Plant Intracellular Ras group-related LLR (PIRL)
	KX639716		Upstream	AT5G63130	Octicosapeptide/Phox/Bem1p family like protein
	KX639719		Upstream	AT5G62230	ERECTA like protein
MADs box L2/R2 ¹	KX639714	AF101420 ¹	Included	AT4G24540	Agamous-like 24 (AGL24)
CAPS marker ⁴	KX639718	KX5455840/1 ⁴	Upstream (5' UTR)	AT2G22540	SVP like protein
EU03H01 ^{1,3} M4.11b ²	KX639711	KF880802 ^{1,2,3}	Downstream	AT1G01620	Plasma membrane intrinsic like protein
EU07G10 ¹ M4.10b ²	KX639712	KX534081 ⁴	Upstream	AT5G39000	Malectin/receptor-like protein kinase family protein
T11292 ⁴	na	KX789069 ⁴	Downstream	AT3G61700	Helicase with zinc finger domain
T4403 ⁴	na	KX789070 ⁴	Downstream	AT4G27280	Calcium-binding EF-hand family protein
T4390 ⁴	na	KX789071 ⁴	na	nd	na
T4391 ⁴	na	KX789072 ⁴	Included (CDS-synonymous)	AT5G47910	NADPH/respiratory burst oxidase protein D (RBOHD)
T4393 ⁴	na	KX789073 ⁴	na	nd	na
T4392 ⁴	na	KX789074 ⁴	Downstream	AT5G55350	MBOAT (membrane bound O-acyltransferase) family protein
T4399 ⁴	na	KX789075 ⁴	Upstream	AT5G61890	ERF subfamily B-4, from ERF/AP2 transcription factor family.
T4394 ⁴	na	KX789076 ⁴	na	nd	na
T4401 ⁴	na	KX789077 ⁴	Downstream	AT3G61420	BSD domain (BTF2-like transcription factors, Synapse-associated proteins and DOS2-like proteins)
T4402 ⁴	na	KX789078 ⁴	Included (CDS-non-synonymous)	XP_022024718*	Transposase of the MuDR family
T4400 ⁴	na	KX789079 ⁴	Included (intron)	AT1G15740	Leucine-rich repeat family protein
T4395 ⁴	na	KX789080 ⁴	Upstream	AT5G56120	RNA polymerase II elongation factor
T4398 ⁴	na	KX789082 ⁴	Included (CDS- synonymous)	AT3G25620	ABC-2 type transporter family protein (ABCG9)

¹Cadalen et al., 2010; ²Ghedina et al., 2015; ³Barcaccia and Tiozzo Caenazzo, 2012, 2014; ⁴Present paper; *For T4402 is reported the best match with the NR database; na, not available; nd, no significant match detected. Localization of the marker within the genic region is also reported.

length of 1,413 cM (**Figure 3** and **Supplementary Table S2**). Each mapped SNP-carrying read was used to anchor the first genome draft of leaf chicory by conducting a default BLASTN analysis. A total of 688 out of 727 genetically mapped reads (95%) strongly matched (similarity >90%, *E*-value < 1E-50) sequences in at least one genomic contig. Considering top hits only, 3.7 Mb (0.3% of the whole genome) of the chicory genome sequence was anchored. Among the mapped contigs, 18.6% aligned with expressed regions from the TAIR Database (BLASTX, *E*-value < 1E-5). This allowed organizing 128 coding regions over the 9 linkage groups (**Supplementary Table S3**). Among them, it was possible to identify 46 different enzymatic proteins, 8 membrane proteins and 4 transcription factors.

The *ms1* locus, along with the M4.10b and M4.12 SSR markers, mapped to linkage group 9 of our genetic map (**Figures 2C, 3**), allowing us to associate it with the linkage group 4 from Cadalen et al. (2010) (see also **Figure 2A**). In particular,

marker M4.12 co-segregated with the target gene and it was mapped at 7.8 cM from the *ms1* locus (**Figure 2C**). Thirteen SNPs exhibited ≤ 3 recombination events with the target *ms1* locus, seven of which (T11292, T4393, T4402, T4399, T4401, T4390, and T4395) cosegregated with *ms1* in the population of 44 BC₁ progeny (**Figure 2C**). The GBS reads corresponding to these 13 markers have been deposited in GenBank under accession numbers KX789069-KX789080, KX789082. Ten of the chicory contigs carrying the mapped SNPs showed a significant match (*E*-value < 10⁻⁵) with TAIR database (**Table 2** and **Supplementary Table S3**).

Since a recent study located the SSI locus (Gonthier et al., 2013) in linkage group 2 from Cadalen et al. (2010), two SSR markers from this group (namely M2.6 and M2.4) were used to genotype the BC₁ samples employed for the GBS strategy. This analysis allowed us to associate the S-locus of leaf chicory to our linkage group 5 and the genetic distance between the two SSR markers resulted equal to 7.9 cM (**Figure 3**).

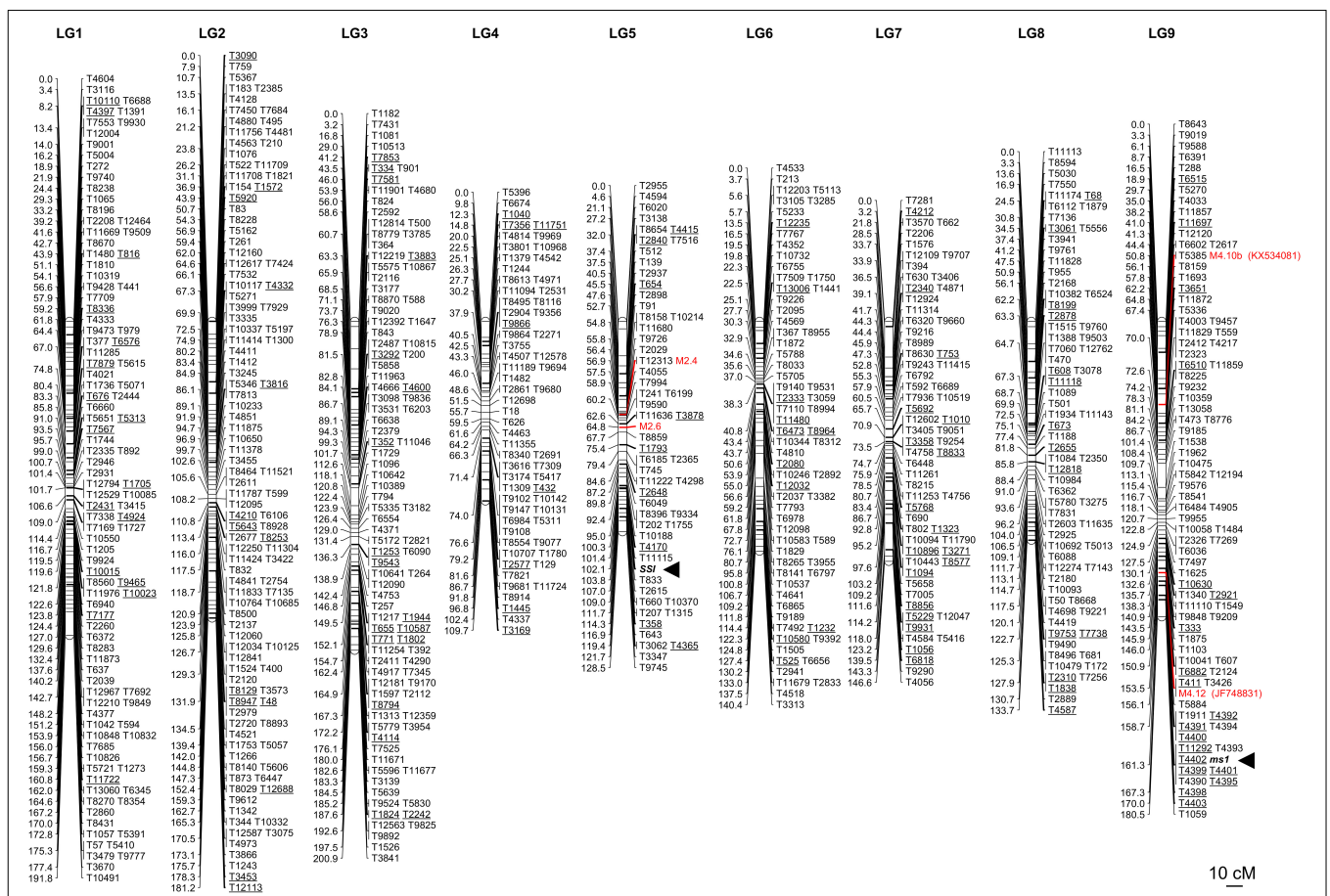


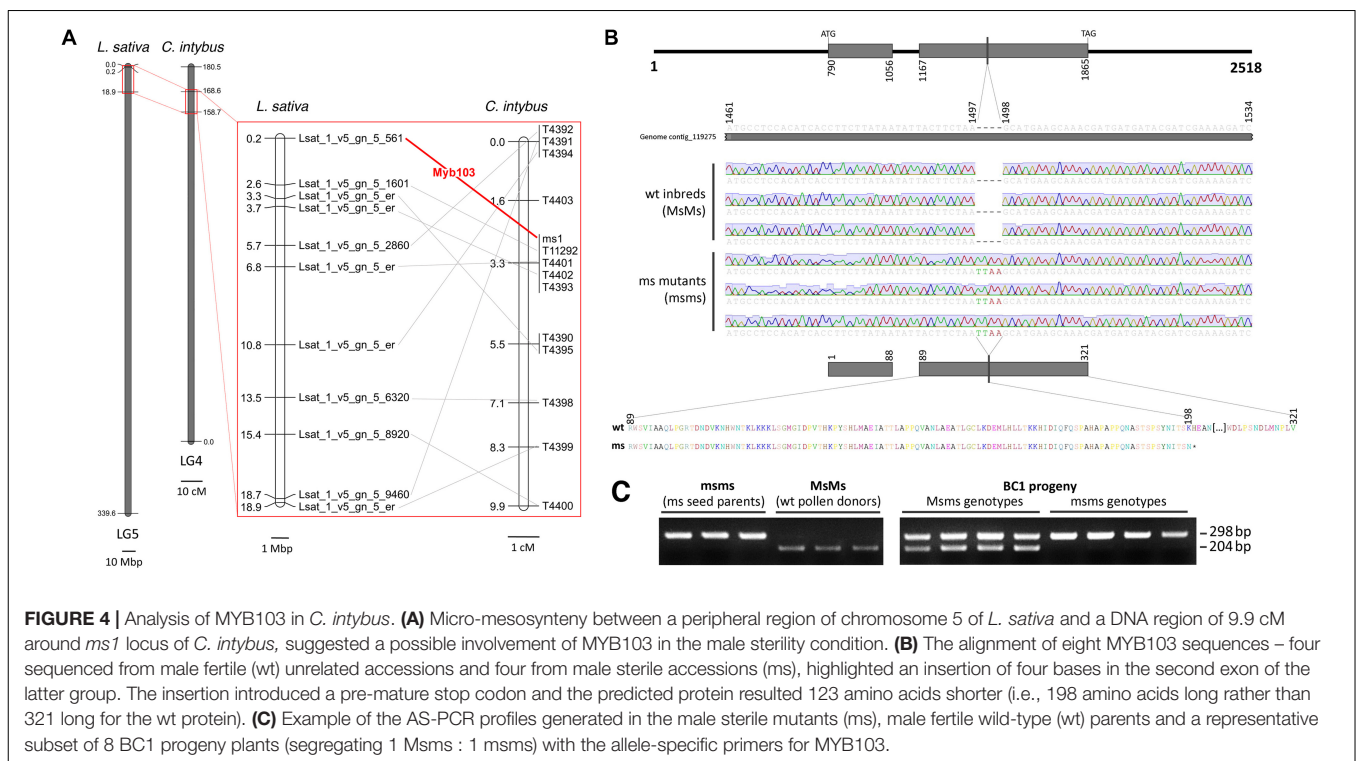
FIGURE 3 | The first SNP-based linkage map in *C. intybus*. Tags underlined represent those mapped sequences showing a significant match (BLASTN, similarity >90%, *E*-value < 1E-50) with contigs from the first genome draft of chicory (Galla et al., 2016) that in turn aligned against protein from the TAIR database (BLASTX, *E*-value < 1E-5). The correspondence between each underlined tag and the best TAIR match is reported in **Supplementary Table S3**. The male-sterility locus (*ms1*) was assessed by recording the target locus as a putative gene fully co-segregating with the trait. Four SSR markers, M4.10b, M4.12 [from linkage group 4 of Cadalen et al. (2010)], M2.6 and M2.4 [from linkage group 2 of Cadalen et al. (2010)] were used to genotype the same samples employed for the SNP-based map and integrated in the linkage map. M4.10b and M4.12 were chosen because co-segregating with the *ms1* (Barcaccia and Tiozzo Caenazzo, 2012, 2014), M2.6 and M2.4 were selected because were found to be associated with the sporophytic self-incompatibility (SS) locus by Gonthier et al. (2013).

Validation of the SNPs Linked to the *ms1* Locus

The map positions of the 13 male sterility-associated SNPs were validated by analyzing an additional 64 BC₁ progeny (32 male sterile and 32 male fertile). This brought the number of BC₁ progeny analyzed to 108, including the initial pool of 44 BC₁ samples analyzed by GBS. For each of the SNP markers, two pairs of primers targeting the two alleles were used in separate reactions. Because BC₁ progeny are either heterozygous or homozygous for the recurrent parent allele, the primer set that amplified the recurrent parent allele generated amplification products in all BC₁ progeny and hence acted as positive control. The other primer set amplified the alternate allele which was present only in heterozygous progeny (**Supplementary Figure S1**). Linkage analysis across the 108 progeny showed that all 13 SNPs mapped to a 9.9 cM region on linkage group 9 [corresponding to linkage group 4 according to Cadalen et al. (2010), **Figure 2D**]. Four SNPs (T11292, T4393, T4401,

and T4402) co-segregated with the *ms1* locus (**Figure 2D** and **Supplementary Table S4**).

The GBS reads carrying the SNP markers T11292 (KX789069) and T4401 (KX789077) aligned against contig_55191 and contig_71514 in the chicory draft genome assembly (Galla et al., 2016). These contigs had significant matches with gene models AT3G61700 (contig_55191) and AT3G61420 (contig_71514) (**Table 2** and **Supplementary Table S3**) that encoded, respectively, for a helicase with zinc-finger protein and a BSD domain (BTF2-like transcription factors, synapse-associated proteins and DOS2-like proteins). The SNPs were located downstream from the coding region. The read carrying SNP T4402 (KX789078) aligned with contig_2496, but this sequence did not show any significant BLAST hits with proteins in the TAIR database. Extending the BLASTX analyses to NCBI's NR database, contig_2496 was found to match significantly (*E*-value = 1E-43, Similarity 44%) with a locus from *Helianthus annuus* (Asteraceae family) encoding for a transposase of the



MuDR family (XP_022024718). Moreover, the discriminant SNP resulted to be non-synonymous and it was located in the zinc-finger domain (ZNF_PMZ) of the protein. The read carrying SNP T4393 (KX789073) did not show any significant match with either TAIR or the NR database.

Micro-Syteny Relationships Between *L. sativa* and *C. intybus* Narrow Chromosomal Segments Enabled the Isolation of MYB103 as Candidate Gene for Male-Sterility

A BLASTN approach, performed using the lettuce genome as database, highlighted that 10 out of the 13 contigs carrying the *ms1*-associated SNPs, aligned (E -value < $1E-20$) with as many loci of *L. sativa*, all located in a peripheral region of ~18 Mbp of chromosome 5 (Figure 4A). In the same chromosomal region of *L. sativa* mapped a transcription factor from the MYB family, known as MYB103 and functionally required for anther development in *Arabidopsis thaliana* (Zhang et al., 2007). From a new BLASTN alignment, carried out using the lettuce genome as database, contig_119275 showed the 92% of similarity (E -value = 0.0) with the MYB103 locus of *L. sativa* (namely Lsat_1_v5_gn_5_561). Since a reciprocal BLASTN search accomplished using the chicory genome draft as database and Lsat_1_v5_gn_5_561 as query confirmed the match (91% similarity, E -value = 0.0), the CDS region entirely included within contig_119275 was considered the putative orthologous of MYB103 in chicory. The DNA segment carrying the two exons of the gene as well as the intron between them was amplified

by target-sequence PCR and Sanger-sequenced in eight chicory accessions (4 male sterile - namely D49, A89, B86, D17 – and four male fertile inbred lines, namely 2111, 202, 231, and 334) using a strategy of overlapping primer walking. The multiple sequence alignment enabled to detect an insertion of four nucleotides (TTAA) in position 1497 of the contig, within the second exon of the male sterile individuals (Figure 4B). Moreover, based on the AS-PCR assay (Figure 4C) performed using 64 BC₁ accessions (32 male sterile and 32 male fertile), MYB103 was found to fully co-segregate with *ms1* and no recombinants were detected. Sequence data of the entire MYB103 gene were deposited in GenBank with accession no. MK285053–MK285054.

DISCUSSION

To the best of our knowledge, this is the first time that male-sterile mutants have been genetically characterized in leaf chicory (*C. intybus* subsp. *intybus* var. *foliosum*). The main goals of our study were to build the first SNP-based linkage map in this species and to accomplish the molecular mapping of the male-sterility trait.

A deep comprehension of nuclear male-sterility systems is extremely important for their exploitation in plant breeding, as male-sterility is one of the most effective methods to produce F₁ hybrid varieties in crop plants (Rajeshwari et al., 1994; Havey, 2004; Barclay, 2010; Acquaah, 2012). F₁ hybrids are usually developed by crossing two highly homozygous parental lines, selected to obtain highly heterozygous progeny, which are characterized not only by high uniformity of phenotypic traits, but also by strong heterosis in terms of productivity.

A recent cytological study of a naturally occurring male-sterile mutant in leaf chicory has shown that micro-sporogenesis proceeds regularly up to the development of tetrads (Barcaccia and Tiozzo Caenazzo, 2012, 2014). After that, all microspores arrest their developmental program. At the beginning of micro-gametogenesis, non-viable shrunken microspores were clearly visible within anthers. Moreover, detailed investigations indicated the occurrence of meiotic abnormalities in the male-sterile mutants, especially at prophase I. Abnormal pairings and chromosomal loops were observed during pachytene. This new mutant, whose male-sterility is caused by a recessive nuclear gene, has been applied in the production of F₁ hybrids of Radicchio, and has been recently subjected to patenting (Barcaccia and Tiozzo Caenazzo, 2012, 2014). However, beyond the fact that the NMS was discovered and mapped in root chicory (Desprez et al., 1994; Gonthier et al., 2013) and leaf chicory (Barcaccia and Tiozzo Caenazzo, 2012, 2014) within linkage groups 5 and 4, respectively, no genetic information is available about this locus.

Exploiting an SSR-based approach, the gene responsible for male-sterility in leaf chicory was found genetically linked to the genomic locus M4.12 (JF748831), an AFLP-derived amplicon encompassing an SSR region, about 5.8 cM apart from the *ms1* locus. Two additional SSR markers, corresponding to genomic loci M4.10b (KX534081) and M4.11b (KF880802), were mapped in the same linkage group at a genetic distance equal to, respectively, 31.2 and 12.1 cM from the *ms1* locus. Although mapped on a chromosomal region surrounding the gene of interest, the three discovered SSR markers were found loosely linked to the *ms1* gene. Considering the functional annotation of the three SSR-containing genomic sequences, E02M09, the closest marker to the *ms1* locus, was found to be putatively linked with a TAIR gene (AT3G11330, PIRL 9 protein) required for differentiation of microspores into pollen grains (Forsthoefel et al., 2013), which definitely is the phenomenon described as disrupted in our *ms1* mutants based on cytological observations (Barcaccia and Tiozzo Caenazzo, 2012, 2014). However, the relatively high number of recombinants detected in the BC₁ population, supporting a genetic distance of 5.8 cM from the *ms1* locus, proved that it could not be responsible for male-sterility in leaf chicory.

The diagnostic CAPS marker derived from the MADS-box L2/R2 gene (KX455840–KX455841) was found genetically linked at 7.3 cM apart from the *ms1* locus. This marker encompasses a genomic block that includes a MADS-box protein (AT4G24540) and a protein that acts as a floral repressor (AT4G22540). Functional analyses by molecular genetic studies in model eudicots, such as *A. thaliana* L., have shown that the proteins encoded by these two genes are both essential for the regulation of various aspects of flower development (Yamaguchi and Hirano, 2006) but no information about their involvement in the male-sterility mechanism is available. Moreover, the relatively high number of recombinants detected in the BC₁ population raises some doubts regarding its role in the pollen development in leaf chicory.

A genotyping-by-sequencing approach allowed us to construct the first SNP-based linkage map and to narrow down the genomic window around the *ms1* locus in leaf chicory. A total of 727 reads-carrying SNPs were clustered into 9 linkage

groups. The first genome draft of leaf chicory was then used to anchor 688 contigs and 128 coding regions to the genetic map. Among the enzymes mapped, it was possible to identify proteins involved in the biosynthesis of *N*-glycan (AT5G19690), cutin, suberin and wax (AT5G55340), diterpenoid (AT5G25900) and amino acids, including valine, leucine and isoleucine (AT1G31180). Other enzymes resulted specifically involved in metabolism processes like glycine, serine and threonine metabolism (AT4G29840), glyoxylate and dicarboxylate metabolism (AT2G05710), inositol phosphate (AT5G42810) and ascorbate and aldarate metabolism (AT5G56490). Two different proteins VAMP713 (AT5G11150) and VAMP714 (AT5G22360), with a key role in the vacuolar trafficking during salt stress (Leshem et al., 2006) were mapped in the linkage groups 7 and 8, respectively. Finally, a noteworthy SNP marker mapped in the linkage group 9 was associated to a genomic contig that, in turn, matched with an ERF BUD ENHANCER in *Arabidopsis* (EBE, AT5G61890). This transcription factor, member of the APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF) transcription factor superfamily, was found to promote cell proliferation, leading to enhanced callus growth, to stimulate axillary bud formation and outgrowth, and to affect shoot branching, acting in cell cycle regulation and dormancy breaking (Mehrnia et al., 2013).

Curiously, 22 mapped contigs matched with as many mitochondrial genes of *Arabidopsis* and, in particular, ATMG00300 (14 matches with as many contigs) and ATMG00750 (3 matches with as many contigs). From TAIR database they resulted annotated as ‘Gag-Pol-related retrotransposon family protein’ and ‘Gag-Pol-Env polyprotein,’ respectively. We found that these two classes of mitochondrial retrotransposons were detected in multiple copies throughout the nuclear genome of several species. At this regards, according to what reported in GenBank database, at least three ATMG00300-like copies were located within chromosomes 3, 7, and 15 of *Malus domestica* as well as in the linkage groups 2, 5, and 6 of *Glycine max*. In *A. thaliana*, the same locus was found also within chromosome 2. This is in accordance with large and unexpected organellar-to-nuclear gene-transfer events highlighted in species like rice (Ueda et al., 2005) and *Arabidopsis* (Lin et al., 1999). Conversely, we cannot exclude an opposite situation where transposable elements were originally transferred from the nuclear genome to the mitochondrial one as highlighted in *Malus x domestica* (Goremykin et al., 2012). However, the abundance of retroelements sequences identified within the chicory map (e.g., 24 out of 128 coding regions were retrotransposon family proteins) is coherent with what already reported in other species like rice or maize where the retroelements represented, respectively, 14% (Sasaki et al., 2002) and 49% (Meyers et al., 2001) of the whole genome.

Two SSR markers (M2.6 and M2.4), mapped on the linkage group 2 by Cadalen et al. (2010), known for carrying the SSI locus (Gonthier et al., 2013) in root chicory, were used to genotype the 44 samples employed for the GBS. This analysis enabled to associate the self-incompatibility locus to the linkage group 5 of leaf chicory. The genetic distance between these two mapped SSR markers was 7.9 cM whereas they were 14.9 cM apart in the genetic map developed by Cadalen et al. (2010).

Recording the target *ms1* locus as a putative gene fully co-segregating with the trait mapped and using two SSR markers [i.e., M4.10b and M4.12, according to Ghedina et al. (2015)] co-segregating with the *ms1*, enabled to overlap our linkage group 9 with linkage group 4 from Cadalen et al. (2010) (see **Figures 2A,B**). At least 13 SNPs were found tightly linked to the target locus, exhibiting three or less recombinants (**Figure 2C**). In particular, seven out of 13 SNPs (i.e., T11292, T4393, T4402, T4399, T4401, T4390, and T4395) did not show recombinants. To increase the robustness of this finding, an AS-PCR assay was developed focusing on these 13 loci and increasing the number of BC₁ samples assayed up to 108. This new round of analysis proved to be fast, cheap and highly efficient and, most importantly, allowed us to finely map all the SNPs in a chromosomal DNA region spanning 9.9 cM (**Figure 2D**) in length, both upstream and downstream the *ms1* locus. The allelic variants of four SNPs proved still to fully co-segregate with the target trait and the corresponding reads were mapped at 0 cM from the *ms1*. Interestingly, two of these genomic sequences, namely T11292 (KX789069) and T4401 (KX789077), retrieved a significant match with two TAIR gene models: a helicase with zinc-finger protein (AT3G61700) and a BSD domain (BTF2-like transcription factors, synapse-associated proteins and DOS2-like proteins, AT3G61420) characterizing the RNA polymerase II transcription factor B. The fact that these two genes resulted to be strictly associated also in the chromosome 3 of *A. thaliana* (~100 kb one from the other), strengthens the possibility that they may interact synergistically. Moreover, according to Honys and Twell (2004), both genes resulted differentially expressed during micro-gametogenesis. Nevertheless, no information about their involvement in the male-sterility mechanism is available and the position of the two SNPs downstream of the coding regions raises some doubts regarding their role in the pollen development.

Patterns of conserved synteny from the genomes of different organisms play a central undertaking in the field of molecular biology (Veltri et al., 2016). At this regard, the 13 contigs carrying the SNPs associated to the *ms1* locus were aligned against the lettuce genome, currently considered the reference assembly for the Asteraceae family (Reyes-Chin-Wo et al., 2017). Ten of these contigs significantly matched with as many loci mapping in a small peripheral region, representing the 5.5% of chromosome 5 of *L. sativa*. It is worth noting that the 10 contigs of chicory spanned a chromosome region of 9.9 cM, constituting again the 5.5% of linkage group 4 of *C. intybus*. Looking at the micro mesosynteny observed between these two species – defined as the conservation of gene content but not gene order or orientation (Ohm et al., 2012) – transcription factor MYB103 stood out as the most interesting locus to investigate. In fact, according to Zhang et al. (2007), a mutation in the first exon of MYB103 in *A. thaliana* is responsible for a male sterile phenotype. In particular, the tapetum development and callose dissolution resulted altered in MYB103 defective plants. Moreover, most of the microspores in mature anthers were degraded and surviving microspores lacked exine. This is totally in agreement with the cytological findings reported in chicory by Barcaccia and Tiozzo Caenazzo (2012, 2014) as well as in

other crops like wheat, rice, canola, and cotton (Phan et al., 2012). A BLASTN approach enabled to identify the orthologous of MYB103 in chicory and the nucleotide similarity with the same gene in lettuce resulted very high (91%, *E*-value = 0.0). From a BLASTX alignment, the amino acid similarity resulted very high (90%, *E*-value = 1E-84) also between the putative MYB103 of chicory and the orthologous gene in *A. thaliana*. The sequence conservation observed among these three species is coherent with the conservation of the gene function already reported by Phan et al. (2012). Preliminary analyses carried out on eight unrelated accessions (four male sterile and four male fertile inbred samples) highlighted within the male sterile group, the occurrence of a four nucleotides insertion in the second exon of the putative CDS (see **Figure 4**). The insertion introduced a pre-mature stop codon in the MYB103 coding sequence and the protein predicted *in silico* resulted 123 amino acids shorter (i.e., 198 amino acids long for the mutants rather than 321 as for the wild types). Zhang et al. (2007) highlighted that mutations of MYB103 negatively affect the expression of both A6 – a putative β -1,3-glucanase involved in callose dissolution – and MS2 – a fatty acid reductase putatively involved in sporopollenin synthesis and therefore in exine formation. Our hypothesis is that a truncated variant of MYB103, lacking more than a third of the amino acid sequence, may affect the process of pollen development in leaf chicory. As a matter of fact, we demonstrated that MYB103 fully co-segregates with *ms1*, further corroborating the evidence that it could be the responsible for the male sterility in *C. intybus*.

CONCLUSION

The male-sterility gene (*ms1*) of leaf chicory was firstly confined to a chromosomal region spanning 5.8 cM through an SSR-based approach. The construction of a SNP-based linkage map and the application of an AS-PCR assay enabled to narrow down the genomic window around the target locus and to select SNPs mapped at 0 cM from *ms1* locus. Moreover, the newly developed genetic linkage map combined with the micro-mesosynteny analysis enabled to identify MYB103, a gene encoding for a transcription factor of the MYB family, fully co-segregating with the *ms1* locus, which could be considered a primary candidate gene for male-sterility.

DATA AVAILABILITY

The datasets generated for this study can be found in GenBank (Accession Numbers: KX789069, KX789070, KX789071, KX789072, KX789073, KX789074, KX789075, KX789076, KX789077, KX789078, KX789079, KX789080, KX789082, MK285053, and MK285054).

AUTHOR CONTRIBUTIONS

FP and GB designed the research. FP and VP conducted and controlled the experiments. PQ carried out the bioinformatics

analyses. FP and KD analyzed data. FP, GB, and KD wrote the manuscript. All authors contributed to editing the manuscript.

FUNDING

This project was financially supported by T&T Srl Agricola, Rosolina (Rovigo, Italy), within a Research Contract stipulated with DAFNAE, University of Padova (Italy), and titled “Development of new radicchio F1 hybrid varieties using molecular marker-assisted breeding methods”.

ACKNOWLEDGMENTS

We wish to thank Mr. Roberto Tencani, CEO of Blumen Group SpA (Bologna, Italy), to have kindly allowed the utilization of data related to the European Patent EP2713705-B1 “*Cichorium* spp. male-sterile mutants”, Intl. application no. PCT/EP2011/058765, for scientific purposes. This research was carried out in partial

fulfillment of the Ph.D. Program of FP by taking advantage of the Doctoral Research Fellowship funded by the University of Padova (Italy). VP, who worked for 6 months at the Laboratory of Genomics for Plant Breeding, University of Padova (Italy), received financial support from CAPES/PSDE, Brazil (Process No. 88881.132586/2016-01). We also wish to thank Drs. Alice Patella and Mariano Giannone, Blumen Group SpA (Italy), for their assistance with breeding materials of leaf chicory, and Dr. Ibrahim Hmham, Faculty of Agriculture, University of Cairo (Egypt), for his invaluable technical help with some of the preliminary SNP-specific PCR experiments performed during December 2016 at the Laboratory of Genomics for Plant Breeding, University of Padova (Italy).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00276/full#supplementary-material>

REFERENCES

- Acquaah, G. (ed.). (2012). “Breeding hybrid cultivars,” in *Principles of Plant Genetics and Breeding* (Chichester: Wiley), 355–373. doi: 10.1002/9781118313718.ch18
- Barcaccia, G., Ghedina, A., and Lucchin, M. (2016). Current advances in genomics and breeding of leaf chicory (*Cichorium intybus* L.). *Agriculture* 6:50. doi: 10.3390/agriculture6040050
- Barcaccia, G., Rosellini, D., Falcinelli, M., and Veronesi, F. (1998). Reproductive behaviour of tetraploid alfalfa plants obtained by unilateral and bilateral sexual polyploidization. *Euphytica* 99, 199–203. doi: 10.1023/A:1018359310706
- Barcaccia, G., and Tiozzo Caenazzo, S. (2012). New male sterile *Chicorium* spp. mutant, parts or derivatives, where male sterility is due to a nuclear recessive mutation linked to a polymorphic genetic marker, useful for producing mutant F1 hybrids of *Chicorium* spp. EU Patent No. WO2012163389-A1. doi: 10.13140/2.1.4749.5044
- Barcaccia, G., and Tiozzo Caenazzo, S. (2014). New male sterile mutant of leaf chicory, including Radicchio, used to produce chicory plant and seeds with traits such as male sterility exhibiting cytological phenotype with shapless, small and shrunken microspores in dehiscent anthers. Patent No. US2014157448-A1. doi: 10.13140/2.1.3176.6400
- Barcaccia, G., Varotto, S., Soattin, M., Lucchin, M., and Parrini, P. (2003). “Genetic and molecular studies of sporophytic self-incompatibility in *Cichorium intybus* L.,” in *Proceedings of the Eucarpia meeting on Leafy Vegetables Genetics and Breeding*, Noordwijkerhout.
- Barclay, A. (2010). Hybridizing the world. *Rice Today* 9, 32–35.
- Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., et al. (2015). The *Arabidopsis* information resource: making and mining the “gold standard” annotated reference plant genome. *Genesis* 53, 474–485. doi: 10.1002/dvg.22877
- Cadalen, T., Mörchen, M., Blassiau, C., Clabaut, A., Scheer, I., Hilbert, J. L., et al. (2010). Development of SSR markers and construction of a consensus genetic map for chicory (*Cichorium intybus* L.). *Mol. Breed.* 25, 699–722. doi: 10.1007/s11032-009-9369-5
- Colasuonno, P., Gadaleta, A., Giancaspro, A., Nigro, D., Giove, S., Incerti, O., et al. (2014). Development of a high-density SNP-based linkage map and detection of yellow pigment content QTLs in durum wheat. *Mol. Breed.* 34, 1563–1578. doi: 10.1007/s11032-014-0183-3
- Desprez, B., Delesalle, L., Dhellemmes, C., and Desprez, M. F. (1994). Genetics and breeding of industrial chicory. *Comptes Rendus l'Academie d'Agriculture Fr.* 80, 47–62.
- Desprez, F., and Bannerot, H. (1980). “A study of pollen tube growth in witloof chicory,” in *Proceedings of the Eucarpia meeting on leafy vegetables*, Littlehampton, 47–52.
- Eenink, A. H. (1982). Compatibility and incompatibility in witloof-chicory (*Cichorium intybus* L.). 3. gametic competition after mixed pollinations and double pollinations. *Euphytica* 31, 773–786. doi: 10.1007/BF00039217
- Forsthoefel, N. R., Klag, K. A., Simeles, B. P., Reiter, R., Brougham, L., and Vernon, D. M. (2013). The *Arabidopsis* plant intracellular ras-group LRR (PIRL) family and the value of reverse genetic analysis for identifying genes that function in gametophyte development. *Plants* 2, 507–520. doi: 10.3390/plants2030507
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- Galla, G., Ghedina, A., Tiozzo, S. C., and Barcaccia, G. (2016). “Toward a First High-quality Genome Draft for Marker-assisted Breeding in Leaf Chicory, Radicchio (*Cichorium intybus* L.),” in *Plant Genomics*, ed. I. Y. Abdurakhmonov (Rijeka: InTech). doi: 10.5772/61747
- Ghedina, A., Galla, G., Cadalen, T., Hilbert, J.-L., Caenazzo, S. T., and Barcaccia, G. (2015). A method for genotyping elite breeding stocks of leaf chicory (*Cichorium intybus* L.) by assaying mapped microsatellite marker loci. *BMC Res. Notes* 8:831. doi: 10.1186/s13104-015-1819-z
- Gonthier, L., Blassiau, C., Mörchen, M., Cadalen, T., Poirat, M., Hendriks, T., et al. (2013). High-density genetic maps for loci involved in nuclear male sterility (NMS1) and sporophytic self-incompatibility (S-locus) in chicory (*Cichorium intybus* L., *Asteraceae*). *Theor. Appl. Genet.* 126, 2103–2121. doi: 10.1007/s00122-013-2122-9
- Goremykin, V. V., Lockhart, P. J., Viola, R., and Velasco, R. (2012). The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. *Plant J.* 71, 615–626. doi: 10.1111/j.1365-313X.2012.05014.x
- Havey, M. (2004). “The use of cytoplasmic male sterility for hybrid seed production,” in *Molecular Biology and Biotechnology of Plant Organelles*, eds H. Daniell and C. Chase (New York, NY: Springer), 623–634. doi: 10.1007/978-1-4020-3166-3_23
- Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol.* 5:R85. doi: 10.1186/gb-2004-5-11-r85
- Huang, S., Liu, Z., Yao, R., Li, D., Zhang, T., Li, X., et al. (2016). Candidate gene prediction for a petal degeneration mutant, pdm, of the chinese cabbage (*Brassica campestris* ssp. *pekinensis*) by using fine mapping and transcriptome analysis. *Mol. Breed.* 36, 1–10. doi: 10.1007/s11032-016-0452-4

- Janssen, A. W. B., and Hermsen, J. G. T. (1976). Estimating pollen fertility in *Solanum* species and haploids. *Euphytica* 25, 577–586. doi: 10.1007/BF00041595
- Kaur, S., Cogan, N. O. I., Stephens, A., Noy, D., Butsch, M., Forster, J. W., et al. (2014). EST-SNP discovery and dense genetic mapping in lentil (*Lens culinaris* Medik.) enable candidate gene selection for boron tolerance. *Theor. Appl. Genet.* 127, 703–713. doi: 10.1007/s00122-013-2252-0
- Kosambi, D. (1943). The estimation of map distances from recombination values. *Ann. Eugen.* 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x
- Leshem, Y., Melamed-Book, N., Cagnac, O., Ronen, G., Nishri, Y., Solomon, M., et al. (2006). Suppression of *Arabidopsis* vesicle-SNARE expression inhibited fusion of H₂O₂-containing vesicles with tonoplast and increased salt tolerance. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18008–18013. doi: 10.1073/pnas.0604421103
- Lin, X., Kaul, S., Rounsley, S., Shea, T. P., Benito, M. I., Town, C. D., et al. (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402, 761–768. doi: 10.1038/45471
- Lucchin, M., Varotto, S., Barcaccia, G., and Parrini, P. (2008). “Chicory and endive,” in *Handbook of Plant Breeding, Vegetables I: Asteraceae, Brassicaceae, Chenopodiaceae*, eds J. Prohens-Tomás and F. Nuez (New York, NY: Springer), 1–46.
- Marcotuli, I., Gadaleta, A., Mangini, G., Signorile, A. M., Zacheo, S. A., Blanco, A., et al. (2017). Development of a high-density SNP-based linkage map and detection of QTL for β -glucans, protein content, grain yield per spike and heading time in durum wheat. *Int. J. Mol. Sci.* 18:E1329. doi: 10.3390/ijms18061329
- Mehrnia, M., Balazadeh, S., Zanor, M.-I., and Mueller-Roeber, B. (2013). EBE, an AP2/ERF transcription factor highly expressed in proliferating cells, affects shoot architecture in *Arabidopsis*. *Plant Physiol.* 162, 842–857. doi: 10.1104/pp.113.214049
- Meyers, B. C., Tingey, S. V., Morgante, M., Meyers, B. C., Tingey, S. V., and Morgante, M. (2001). Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res.* 11, 1660–1676. doi: 10.1101/gr.188201
- Muys, C., Thienpont, C. N., Dauchot, N., Maudoux, O., Draye, X., Cutsem, P. V., et al. (2014). Integration of AFLPs, SSRs and SNPs markers into a new genetic map of industrial chicory (*Cichorium intybus* L. var. *sativum*). *Plant Breed.* 133, 130–137. doi: 10.1111/pbr.12113
- Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., et al. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen dothideomycetes fungi. *PLoS Pathog* 8:e1003037. doi: 10.1371/journal.ppat.1003037
- Pécaut, P. (1962). Etude sur le système de reproduction de l’endive (*Cichorium intybus* L.). *Ann Amélior Plantes* 12, 265–296.
- Phan, H. A., Li, S. F., and Parish, R. W. (2012). MYB80, a regulator of tapetal and pollen development, is functionally conserved in crops. *Plant Mol. Biol.* 78, 171–183. doi: 10.1007/s11103-011-9855-0
- Rajeshwari, R., Sivaramakrishnan, S., Smith, R. L., and Subrahmanyam, N. C. (1994). RFLP analysis of mitochondrial DNA from cytoplasmic male sterile lines of pearl millet. *Theor. Appl. Genet.* 88, 441–448. doi: 10.1007/BF00223658
- Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikiti, S., Song, C., et al. (2017). Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* 8:14953. doi: 10.1038/ncomms14953
- Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., et al. (2002). The genome sequence and structure of rice chromosome 1. *Nature* 420, 312–316. doi: 10.1038/nature01196. Published
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments a poor man’s approach to genotyping for research and high-throughput diagnostics. *Nat. Biotechnol.* 18, 233–234. doi: 10.1038/72708
- Schumann, M. J., Zeng, Z.-B., Clough, M. E., and Yencho, G. C. (2017). Linkage map construction and QTL analysis for internal heat necrosis in autotetraploid potato. *Theor. Appl. Genet.* 130, 2045–2056. doi: 10.1007/s00122-017-2941-1
- Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J.* 3, 739–744. doi: 10.1111/j.1365-313X.1993.00739.x
- Ueda, M., Tsutsumi, N., and Kadowaki, K. I. (2005). Translocation of a 190-kb mitochondrial fragment into rice chromosome 12 followed by the integration of four retrotransposons. *Int. J. Biol. Sci.* 1, 110–113. doi: 10.7150/ijbs.1.110
- Veltri, D., Wight, M. M., and Crouch, J. A. (2016). SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species. *Nucleic Acids Res.* 44, W41–W45. doi: 10.1093/nar/gkw330
- Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77
- Yamaguchi, T., and Hirano, H.-Y. (2006). Function and diversification of MADS-Box genes in rice. *Sci. World J.* 6, 1923–1932. doi: 10.1100/tsw.2006.320
- Zhang, Z. B., Zhu, J., Gao, J. F., Wang, C., Li, H., Li, H., et al. (2007). Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. *Plant J.* 52, 528–538. doi: 10.1111/j.1365-313X.2007.03254.x
- Zhao, X., Han, Y., Li, Y., Liu, D., Sun, M., Zhao, Y., et al. (2015). Loci and candidate gene identification for resistance to *Sclerotinia sclerotiorum* in soybean (*Glycine max* L. Merr.) via association and linkage maps. *Plant J.* 82, 245–255. doi: 10.1111/tpj.12810

Conflict of Interest Statement: 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 © 2019 Palumbo, Qi, Pinto, Devos and Barcaccia. 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.