Thieme

Variability of Chalcone Synthase in Chamomile Accessions *(Matricaria chamomilla)*#

Authors
Melanie Novak, Dijana Jovanovic, Johannes Novak

Affiliations
Institute of Animal Nutrition and Functional Plant Compounds, University of Veterinary Medicine Vienna, Vienna, Austria

Key words
Chamomile, *Matricaria chamomilla*, Asteraceae, chalcone synthase, flavonoids

received July 31, 2023
accepted after revision November 1, 2023

Bibliography
Planta Med 2024; 90: 534–545
DOI 10.1055/a-2214-0940
ISSN 0032-0943

Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Correspondence
Prof. Johannes Novak
Institute of Animal Nutrition and Functional Plant Compounds, University of Veterinary Medicine Vienna
Veterinärplatz 1, 1210 Vienna, Austria
Phone: + 43 12 50 77 31 00, Fax: + 43 12 50 77 31 90
Johannes.Novak@vetmeduni.ac.at

ABSTRACT

Chamomile (*Matricaria chamomilla*) is an important medicinal plant whose beneficial activities partly rely on certain flavonoids. The first dedicated step in flavonoid biosynthesis is chalcone synthase (CHS, EC 2.3.1.74). The genomic DNA of CHS was studied in six chamomile specimens from different genotypes to describe interspecimen, as well as interspecific, variability. One specimen of *M. discoidea* was included as an outgroup. The two exons of CHS of *M. chamomilla* (McCHS) and *M. discoidea* (MdCHS) were 188 bp and 1,011 bp long, separated by an intron of variable length between 192 and 199 bp in McCHS and 201 bp in MdCHS, respectively. The two exons with 5.3 and 6.2 mutations per 100 bp, respectively, were more conserved than the intron with 11.5 mutations per 100 bp. In total, 96 SNPs were detected in both species, of which 12 SNPs were only present in MdCHS and 80 SNPs only in McCHS. Overall, 70 haplotypes (multilocus genotypes, MLGs) were detected. The samples could be classified into two groups, a 'compact' group of a low number and diversity of haplotypes and a 'variable' group of a high number and diversity of haplotypes. Of the 74 SNPs in McCHS, only six SNPs were non-synonymous. However, the amino acid changes did not affect critical areas of the enzyme. The combination of the six SNPs resulted in nine translated amino acid MLGs. The CHS network located MdCHS, due to the crossing barrier, quite distant from chamomile. MdCHS docked to McCHS at a position from where McCHS divergently evolved into two directions.

Introduction

Matricaria chamomilla L. (syn. *Matricaria recutita* L., *Chamomilla recutita* (L.) Rauschert, Asteraceae; common names: true chamomile, German chamomile, Hungarian chamomile, blue chamomile, further on referenced here as 'chamomile') is one of the most important medicinal plants and has its origin in the near-east to southwestern Europe and is now widespread over the world. *M. chamomilla* is an annual plant with long and narrow bipinnate to tripinnate leaves, white ray, and yellow tubular flowers. *Matricaria discoidea* DC. (syn. *M. matricarioides* (Less.) Porter; common names: pineappleweed, disc or rayless mayweed) is another member of the small *Matricaria* genus and looks like *M. chamomilla* with differences in smaller flower heads, the absence of ray flowers, greenish tubular flowers and a four-toothed corolla [1].

The medicinal activity is attributed to essential oil compounds and flavonoids both concentrated in the flowers. The flavonoids of chamomile have antispasmodic and antiphlogistic activity [2]. Flavonoids are formed through the phenylpropanoid pathway starting from phenylalanine over 4-coumaroyl-CoA to an uncycled chalcone, which – after cyclisation – is modified by enzymes of different functional classes into the different flavonoids; 4-coumaroyl-CoA is the starter for a stepwise condensation of three acetate units to an uncycled chalcone, a condensation performed by chalcone synthase (CHS, EC 2.3.1.74), a type III polyketide synthase ubiquitous in plants, but also present in fungi and bacteria.

This work is dedicated to Professors Rudolf Bauer, Chlodwig Franz, Brigitte Kopp, and Hermann Stuppner for their invaluable contributions and commitment to Austrian pharmacognosy.

▶ **Table 1** Genotypes, number of clones sequenced, total lengths (bp), origin and source of genotypes. Accessions 1, 3, 5 and 6 were obtained from Pharmaplant (Germany), 2 from Vilora (Slovak Republic) and 4 and 6 from the Genebank Gatersleben (Germany). Number of MLG and eMLG: DNA level, in bracket amino acid (AA) level.

Species	Accession	Ploidy level	No. of clones	Length(s) (bp)	MLG DNA (AA)	eMLG DNA (AA)	Origin
M. chamomilla	Argenmilla	2×	12	1391	4 (3)	3.8 (2.8)	Argentina
	Bona	2×	6	1391	5 (1)	5 (1)	Slovak Republic
	Degumille	2×	9	1391 1394	9 (3)	9 (3)	Germany
	MAT 16	2×	21	1392 1394 1396 1398	21 (5)	9.6 (4.6)	Bulgaria
	PNOS	4×	13	1391 1394	13 (4)	9.4 (3.7)	Poland
	Promyk	2×	21	1391 1394	21 (4)	10 (3.5)	Poland
M. discoidea	–	2×	26	1400	6 (2)	6 (1.4)	Austria
Mean			15.43		14.1 (3.1)	9 (2.8)	
Total			108		70 (11)	50.6 (5.2)	

CHS is a homodimeric protein with a conserved structure of one intron at a conserved position [3]. The rare exceptions are CHS with two (*Antirrhinum majus* [4] and *Alpinia calcarata* [5]) and three introns (*Polygonum cuspidatum* [6]). CHS is a gene family variable in number starting from three members in pecan (*Carya illinoinensis*) [7] to 27 member in rice (*Oryza sativa*) [8, 9].

Many hundred CHS and CHS-like genes are described [3]. Studies on intraspecific variability, however, are scarce. In this study, we analysed the CHS variability in six specimens of *M. chamomilla*, each one from a different genotype to assess variability between genotypes but also within specimens. To determine the starting point of CHS evolution within *M. chamomilla*, one specimen of *M. discoidea* was included as an outgroup.

Results

In this study, genomic DNA of CHS was analysed in six *M. chamomilla* plants (McCHS), every plant coming from a different genotype (▶ **Table 1**). The sample set was complemented by one specimen of *M. discoidea* as an outgroup (MdCHS). McCHS and MdCHS were cloned and sequenced from a variable number of clones per specimen where the number of clones depended on different cloning efficiency. For cloning, an overlapping primer set was developed by assembling chamomile NGS sequences (Submission ID: SUB5046906/BioProject ID: PRJNA515664) with CHS from *Chrysanthemum* x *morifolium* (MW368977) as reference.

One intron divided the coding sequence of McCHS and MdCHS into a shorter exon 1 and a longer exon 2. The lengths of the two exons were conserved in all samples (188 bp and 1009 bp, respectively), while the intron had three polymorphic indels leading to variable lengths between 192 and 199 bp (McCHS) and 201 bp (MdCHS), respectively.

In total, 96 SNPs were detected in the 7 specimens from 2 *Matricaria* species (▶ **Table 2**). Ten mutations were found in exon 1 (5.3 mutations per 100 bp), 22 in the intron (length depending between 11 and 11.5 mutations per 100 bp) and 63 in exon 2 (6.2 mutations per 100 bp). Six SNPs had three alleles; all the others had two alleles. Twelve SNPs were only present in MdCHS (homozygous in McCHS); 80 SNPs were homozygous in MdCHS. Simpson's index of diversity showed the lowest diversity for g.225A>T (0.036). The highest diversities are those linked to the occurrence of three alleles and are the only ones over 0.5 with a maximum of 0.545 in g.135 G>C. g.224A>G was the SNP with lowest evenness (0.390). An almost perfect evenness of 0.988 was found for g.775A>T and g.802C>T.

Some of the SNPs were perfectly linked and could be classified into nine linkage groups. Linkage groups I, III, IV, VI and VII were groups of neighbouring SNPs. Linkage group II was a group only present ('private') in *M. discoidea*. Twenty private alleles were found in Mat16, 17 in *M. discoidea* and 11 in Argenmilla.

The 108 sequenced clones grouped into 70 haplotypes (MLGs) (▶ **Table 3**). Since the number of sequenced clones per specimen were unequal, the standardized eMLG – a figure of expected MLG at the smallest sample size ≥ 10 based on rarefaction – was more suitable to compare the samples. *M. discoidea* and 'Argenmilla' had the lowest number of eMLGs (3.78 and 3.82, respectively), followed by 'Bona' (5). In all other accessions, numerous MLGs were found, reflected by eMLGs between 9 and 10.

The 70 MLG resulted in a complex network where the outgroup *M. discoidea* formed a compact group, while all other specimens from *M. chamomilla* formed a complex and widespread net-

▶ **Table 2** CHS loci polymorphic in two *Matricaria* species (DNA). Position numbering follows the alignment of GenBank accession numbers OR634854–OR634923.

	No.	SNP	alleles	1-D	even-ness	linkage group	private allele (frequency in the respective accession)
Exon1	1	g.27 T>C	2	0.198	0.590		*M. discoidea* (46%)
	2	g.36 G>T	2	0.409	0.844		
	3	g.45 G>A	2	0.401	0.832		
	4	g.49C>A	2	0.393	0.820		
	5	g.62C>A	2	0.356	0.769		*M. discoidea* (96%)
	6	g.96 T>C	2	0.088	0.470	I	
	7	g.99 T>C	2	0.088	0.470	I	
	8	g.135 G>C	3	0.545	0.845		Mat16 (38%)
	9	g.150A>C	2	0.105	0.490		Argenmilla (50%)
	10	g.157 G>A	2	0.431	0.878		
Intron	11	g.193 T>C	2	0.153	0.543		Mat16 (43%)
	12	g.196C>T	2	0.456	0.919		
	13	g.202 T>A	2	0.366	0.782	II	*M. discoidea* (100%)
	14	g.223 G>A	2	0.479	0.960	III	
	15	g.224A>G	2	0.036	0.390		*M. discoidea* (8%)
	16	g.225A>T	2	0.479	0.960	III	
	17	g.228 T>A	2	0.479	0.960	III	
	18	g.236A>T	2	0.479	0.960	III	
	19	g.237A>G	2	0.461	0.928	IV	
	20	g.251 G>T	2	0.461	0.928	IV	
	21	g.255A>C	2	0.461	0.928	IV	
	22	g.288 T>A	2	0.461	0.928		
	23	g.320C>G	2	0.252	0.648		
	24	g.323A>T	2	0.479	0.960		
	25	g.336 G>T	2	0.366	0.782	II	*M. discoidea* (100%)
	26	g.346A>C	2	0.450	0.909		
	27	g.348 G>T	2	0.366	0.782	II	*M. discoidea* (100%)
	28	g.363 T>G	2	0.366	0.782	II	*M. discoidea* (100%)
	29	g.366 G>A	2	0.456	0.919	V	
	30	g.370C>A	2	0.153	0.543		Mat16 (43%)
	31	g.382 T>G	2	0.444	0.899		
	32	g.374A>G	2	0.456	0.919	V	

continued next page

work (▶ **Fig. 1**). The outgroup *M. discoidea* showed two major haplotypes represented by 46% and 35% of the *M. discoidea* clones, respectively, and four 'satellite' haplotypes with low haplotype frequencies (one with 8%, the other three with 4% of clones, respectively). All outgroup haplotypes were closely related (▶ **Fig. 1** and **Table 3**). The outgroup had a large distance to *M. chamomilla* and docked to the *M. chamomilla* network at a clone from 'PNOS'. The *M. chamomilla* network was bifurcating from the docking point to *M. discoidea* and further specialising into evolutionary newer lines. 'Argenmilla' was the most compact specimen located on one side of the network with only four haplotypes, followed by "Bona' bridging the distance between 'Argenmilla' and the other genotypes with five haplotypes. 'Degumille', 'Promyk' and 'PNOS' were widespread over the centre of the network and consisted of

▶ **Table 2** CHS loci polymorphic in two *Matricaria* species (DNA). Position numbering follows the alignment of GenBank accession numbers OR634854–OR634923. *continued*

	No.	SNP	alleles	1-D	even-ness	linkage group	private allele (frequency in the respective accession)
Exon2	33	g.403C>T	2	0.137	0.526		Mat16 (38%)
	34	g.412 T>A	2	0.438	0.889		
	35	g.421A>G	2	0.431	0.878		
	36	g.424 G>A	2	0.466	0.937		
	37	g.454 T>A	2	0.409	0.844		
	38	g.475C>T	2	0.466	0.937		
	39	g.481C>T	2	0.431	0.878		
	40	g.490A>G	2	0.438	0.889		
	41	g.505C>T	2	0.153	0.543		Mat16 (43%)
	42	g.514C>T	2	0.461	0.928		
	43	g.520C>T	2	0.417	0.855		
	44	g.553A>G	2	0.183	0.575		Mat16 (52%)
	45	g.565C>T	2	0.168	0.559		Mat16 (48%)
	46	g.580A>G	2	0.450	0.909		
	47	g.598C>T	2	0.153	0.543	VI	Mat16 (43%)
	48	g.602C>T	2	0.153	0.543	VI	Mat16 (43%)
	49	g.625 T>C	2	0.226	0.619		
	50	g.637 T>C	2	0.438	0.889		
	51	g.652A>G	2	0.168	0.559	VII	Mat16 (48%)
	52	g.655C>T	2	0.168	0.559	VII	Mat16 (48%)
	53	g.658C>A	2	0.456	0.919		
	54	g.670A>C	2	0.366	0.782	II	*M. discoidea* (100%)
	55	g.676 T>C	2	0.366	0.782		*M. discoidea* (100%)
	56	g.682A>G	3	0.585	0.889		*M. discoidea* (100%)
	57	g.688A>G	2	0.489	0.978		
	58	g.740C>T	2	0.489	0.978		
	59	g.775A>T	2	0.494	0.988		
	60	g.781C>T	3	0.665	0.995		
	61	g.782 T>C	2	0.366	0.782	II	
	62	g.802C>T	2	0.494	0.988		
	63	g.832 T>C	2	0.497	0.995		
	64	g.868 T>C	2	0.366	0.782	II	*M. discoidea* (100%)
	65	g.886 T>C	2	0.393	0.820		
	66	g.901A>T	2	0.183	0.575	VIII	Mat16 (52%)
	67	g.911 T>G	2	0.183	0.575	VIII	Mat16 (52%)
	68	g.925C>T	3	0.540	0.805		Mat16 (52%)
	69	g.929 T>C	2	0.183	0.575		Mat16 (52%)
	70	g.961 T>C	2	0.366	0.782	II	*M. discoidea* (100%)

continued next page

▶ **Table 2** CHS loci polymorphic in two *Matricaria* species (DNA). Position numbering follows the alignment of GenBank accession numbers OR634854–OR634923. *continued*

	No.	SNP	alleles	1-D	even-ness	linkage group	private allele (frequency in the respective accession)
Exon2	71	g.982A>G	2	0.471	0.945		
	72	g.1006 T>G	2	0.461	0.928		
	73	g.1024C>A	2	0.393	0.820		
	74	g.1039 T>A	3	0.651	0.968		*M. discoidea* (96%)
	75	g.1051C>T	2	0.183	0.575		Mat16 (52%)
	76	g.1060C>A	2	0.366	0.782	II	*M. discoidea* (100%)
	77	g.1090A>T	2	0.168	0.559		Mat16 (48%)
	78	g.1093 G>A	2	0.475	0.953		
	79	g.1114 G>C	2	0.183	0.575		Mat16 (52%)
	80	g.1126 T>C	2	0.366	0.782	II	*M. discoidea* (100%)
	81	g.1138 T>G	2	0.313	0.717		
	82	g.1150A>T	2	0.401	0.832	IX	
	83	g.1153 G>T	3	0.582	0.873		Mat16 (57%)
	84	g.1162 T>G	2	0.198	0.590		Mat16 (57%)
	85	g.1168 T>G	2	0.401	0.832	IX	
	86	g.1210C>T	2	0.409	0.844		
	87	g.1246 T>C	2	0.409	0.844		
	88	g.1249 T>C	2	0.483	0.967		
	89	g.1282A>G	2	0.356	0.769		*M. discoidea* (96%)
	90	g.1306C>T	2	0.456	0.919		
	91	g.1357 G>A	2	0.456	0.919		
	92	g.1366C>T	2	0.409	0.844		
	93	g.1369 G>A	2	0.497	0.995		
	94	g.1375C>T	2	0.366	0.782	II	*M. discoidea* (100%)
	95	g.1390C>T	2	0.431	0.878		
	96	g.1403 T>A	2	0.313	0.717		
	mean		2.062	0.371	0.796		

numerous haplotypes, while 'Mat16', opposite to 'Argenmilla', evolved two major side lines with numerous haplotypes.

Variation in *M. chamomilla* was 60% higher than between the two species (40%) (▶ **Table 4**). The genotypes were quite different (26% variation) but higher, and so was the variation within the specimens (34%).

Of the 73 SNPs in both chamomile exons, only six changed the corresponding amino acid, none of them affecting positions that may indicate a change in function [10] (▶ **Table 5**). The combination of the six polymorphic amino acids in this dataset resulted in 11 MLGs (▶ **Fig. 2**). *M. discoidea* consisted of two closely related AA MLGs, none of them shared with chamomile. One AA only present in *M. discoidea* at a frequency of 96% was Thr21 (instead of Asn). Further 'private' amino acid changes were Glu50Asp in 'Argenmilla' and Leu223Val in 'Mat16', both at frequencies of 50% and 52% within the two specimens, respectively. Two neighbouring amino acid changes (Asp12Glu and Lys17Gln) were completely linked. The AMOVA results were like the DNA results, allocating 38% between the two *Matricaria* species, 19% between specimens and 43% within genotypes (▶ **Table 4**).

Discussion

CHS is a key enzyme in flavonoid biosynthesis. Duplication events could therefore counteract mutations in creating functional redundancy. CHS is known already as a multicopy gene family with copy number varying between three members in pecan (*Carya illinoinensis*) [7] to 27 members in rice (*Oryza sativa*) [8, 9]. In this

▶ **Table 3** Affiliation of DNA haplotypes to the respective specimens (chs_mdxx … *M. discoidea* haplotypes, chs_mcxx … M. chamomilla haplotypes).

haplotype	Argenmilla	Bona	Degumille	Mat16	PNOS	Promyk	*M. discoidea*
chs_md01	0	0	0	0	0	0	2
chs_md02	0	0	0	0	0	0	9
chs_md03	0	0	0	0	0	0	1
chs_md04	0	0	0	0	0	0	1
chs_md05	0	0	0	0	0	0	1
chs_md06	0	0	0	0	0	0	12
chs_mc01	6	2	1	0	1	1	0
chs_mc02	3	0	0	0	0	0	0
chs_mc03	2	0	0	0	0	0	0
chs_mc04	1	0	0	0	0	0	0
chs_mc05	0	1	0	0	0	1	0
chs_mc06	0	1	0	0	0	0	0
chs_mc07	0	1	0	0	0	0	0
chs_mc08	0	1	0	0	0	0	0
chs_mc09	0	0	1	0	0	0	0
chs_mc10	0	0	1	0	0	0	0
chs_mc11	0	0	1	0	0	0	0
chs_mc12	0	0	1	0	0	0	0
chs_mc13	0	0	1	0	0	0	0
chs_mc14	0	0	1	0	0	0	0
chs_mc15	0	0	1	0	0	0	0
chs_mc16	0	0	1	0	0	0	0
chs_mc17	0	0	0	2	0	0	0
chs_mc18	0	0	0	2	0	0	0
chs_mc19	0	0	0	1	0	0	0
chs_mc20	0	0	0	1	0	0	0
chs_mc21	0	0	0	1	0	0	0
chs_mc22	0	0	0	1	0	0	0
chs_mc23	0	0	0	1	0	0	0
chs_mc24	0	0	0	1	0	0	0
chs_mc25	0	0	0	1	0	0	0
chs_mc26	0	0	0	1	0	0	0
chs_mc27	0	0	0	1	0	0	0
chs_mc28	0	0	0	1	0	0	0
chs_mc29	0	0	0	1	0	0	0
chs_mc30	0	0	0	1	0	0	0
chs_mc31	0	0	0	1	0	0	0
chs_mc32	0	0	0	1	0	0	0
chs_mc33	0	0	0	1	0	0	0

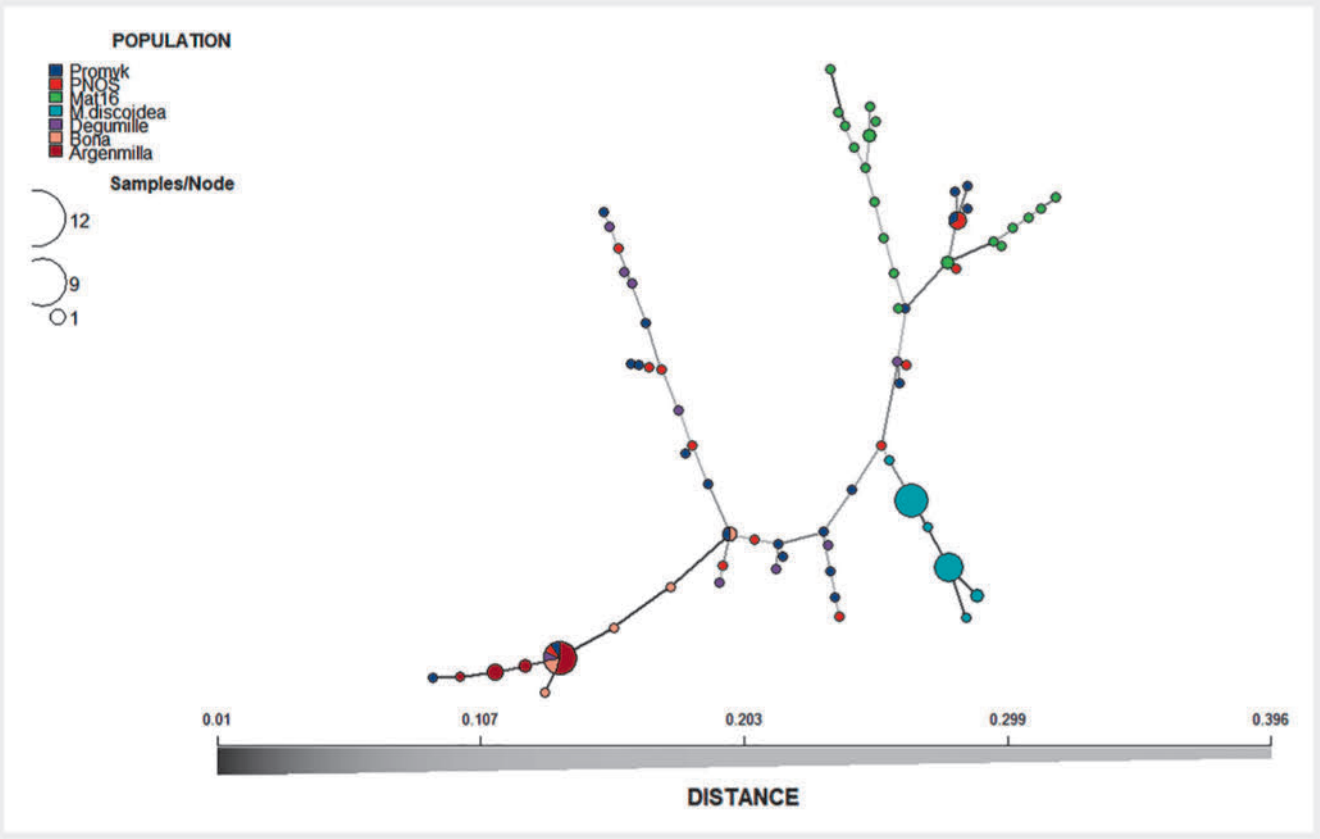
continued next page

▶ **Table 3** Affiliation of DNA haplotypes to the respective specimens (chs_mdxx … *M. discoidea* haplotypes, chs_mcxx … M. chamomilla haplotypes). *continued*

haplotype	Argenmilla	Bona	Degumille	Mat16	PNOS	Promyk	*M. discoidea*
chs_mc34	0	0	0	1	0	0	0
chs_mc35	0	0	0	1	0	0	0
chs_mc36	0	0	0	0	2	1	0
chs_mc37	0	0	0	0	1	0	0
chs_mc38	0	0	0	0	1	0	0
chs_mc39	0	0	0	0	1	0	0
chs_mc40	0	0	0	0	1	0	0
chs_mc41	0	0	0	0	1	0	0
chs_mc42	0	0	0	0	1	0	0
chs_mc43	0	0	0	0	1	0	0
chs_mc44	0	0	0	0	1	0	0
chs_mc45	0	0	0	0	1	0	0
chs_mc46	0	0	0	0	1	0	0
chs_mc47	0	0	0	0	0	1	0
chs_mc48	0	0	0	0	0	1	0
chs_mc49	0	0	0	0	0	1	0
chs_mc50	0	0	0	0	0	1	0
chs_mc51	0	0	0	0	0	1	0
chs_mc52	0	0	0	0	0	1	0
chs_mc53	0	0	0	0	0	1	0
chs_mc54	0	0	0	0	0	1	0
chs_mc55	0	0	0	0	0	1	0
chs_mc56	0	0	0	0	0	1	0
chs_mc57	0	0	0	0	0	1	0
chs_mc58	0	0	0	0	0	1	0
chs_mc59	0	0	0	0	0	1	0
chs_mc60	0	0	0	0	0	1	0
chs_mc61	0	0	0	0	0	1	0
chs_mc62	0	0	0	0	0	1	0
chs_mc63	0	0	0	0	0	1	0
chs_mc64	0	0	0	0	0	1	0

work, we cloned and sequenced McCHS from six specimens of *M. chamomilla* from different accessions and included one specimen of *M. discoidea* (MdCHS) as 'outgroup' to determine the most ancient McCHS haplotype. CHS in both chamomile species had with one intron the same conserved structure as found in almost all of the known CHSs [3]. Also, the lengths of exons, intron and amino acid are not exceptional. In the two exons, no indels were detected. Therefore, no obvious pseudogenes by frameshifts were present in chamomile CHS. Phylogenetic trees of chamomile CHS were not well supported by bootstrap values. The network approach presented here reflected the complex relationship of different CHS variants within and between the specimens much better. Interestingly, the specimens divided into compact specimens ('Argenmilla', 'Bona' and *M. discoidea*) with a limited number of closely related haplotypes where eMLGs were below 5 and diverse specimens ('PNOS′, 'Promyk', 'Mat16′) with many haplotypes differing to a greater extent with eMLGs around 10.

Identifying all haplotypes in an specimen by cloning should have been possible in our approach if gene copies would have been in the range known so far with up to 27 members (rice, *Ory-*

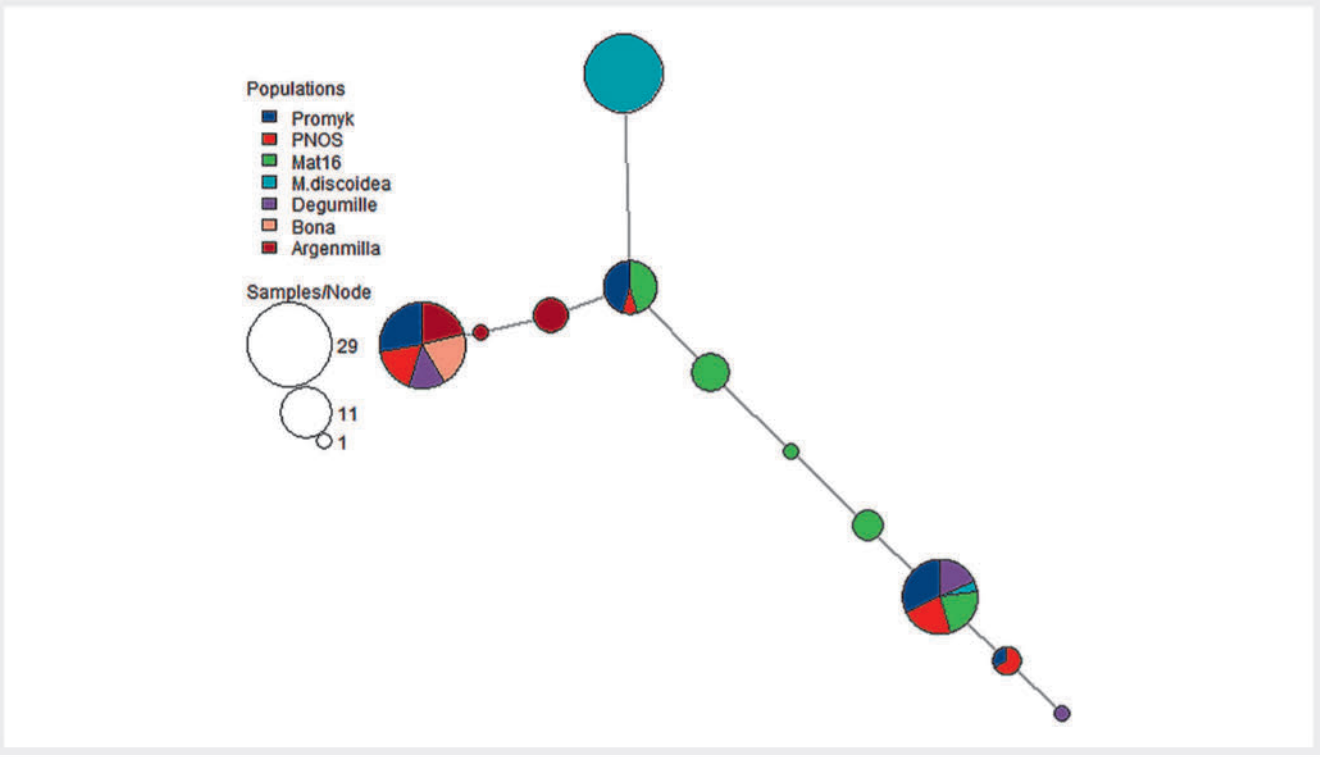


▶ **Fig. 1** CHS haplotype network of six specimens of *M. chamomilla* and one specimen of *M. discoidea* based on DNA sequences. The size of the bubbles indicates the number of samples per haplotype, the grey scale of edges the distance between haplotypes, while distances between nodes are not representing genetic distances.

▶ **Table 4** AMOVA of DNA and AA polymorphisms of CHS between the two *Matricaria* species and genotypes (specimens).

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
DNA				
Between species	1	1,159	20.822	40%
Between specimens	5	981	13.591	26%
Within specimen	101	1,704	16.878	34%
Total	107	3,845	51.291	
AA				
Between species	1	52.07	0.988	38%
Between specimens	5	39.02	0.507	19%
Within specimen	101	111.65	1.105	43%
Total	107	202.74	2.601	

za sativa) [8, 9]. From a pragmatic point of view, the completeness of haplotypes can be defined as observing all haplotypes at least twice, which – in a strict sense – still does not guarantee completeness but may be a good compromise between exactness and costs. All singletons were excluded from the dataset to not introduce sequencing errors as artificial variations. Based on this criterion for completeness, all haplotypes were identified for the specimens from 'Argenmilla', 'Bona' and *M. discoidea*. All other specimens, however, were so variable that some haplotypes could not be confirmed. The higher variability in this second group is



▶ **Fig. 2** Amino acid (AA) haplotype network of six specimens of *M. chamomilla* and one specimen of *M. discoidea* based on the deduced amino acids. The size of the bubbles indicates the number of samples per haplotype, the grey scale of edges the distance between haplotypes (almost all distances are equal).

▶ **Table 5** Deduced polymorphic amino acids (AA) in two *Matricaria* species.

No.	polymorphic AA	1-D	evenness	linkage group	private allele (frequency in the respective accession)
1	Asp12Glu	0.41	0.84	I	
2	Lys17Gln	0.39	0.82	I	
3	Asn21Thr	0.36	0.77		*M. discoidea* (96%)
4	Glu50Asp	0.10	0.49		Argenmilla (50%)
5	Val53Ile	0.43	0.88		
6	Leu223Val	0.18	0.57		Mat16 (52%)
mean		0.31	0.73		

possibly due to a (much) higher gene copy number. Another advantage of a higher CHS copy number, in addition to a higher redundancy, would be a possible gene dose effect resulting in higher flavonoid contents. However, the second enzyme in flavonoid biosynthesis, chalcone isomerase (CHI, EC 5.5.1.6), is also known to be rate-limited [11]. So, a higher copy number of CHS alone might not increase the flavonoid content. Following this lead could possibly increase the content of bioactive flavonoids in chamomile by selecting for high CHS and/or CHI gene copy numbers.

Often CHS variants are designated with consecutive numbering (e.g., 'Schs_mcHS1′, 'Schs_mcHS2′ and 'Schs_mcHS3' for *Silybum marianum* [12]). Such a simple classification was not possible for chamomile because of the continuous variation found even on the amino acid level.

Basically, the genetic CHS relationships between the specimens are very similar to those of the mitochondrial relationships [13] with the couple 'Argenmilla'/'Bona' being closely related and both very distant from the Bulgarian 'Mat16'. 'Argenmilla' is so close to 'Bona' that it may be an Argentinian reselection from 'Bona'. 'PNOS' goes back to Polish landraces [14], and 'Promyk' is a younger Polish cultivar with origins in 'PNOS', registered in 1992 [15]. The two specimens from the two cultivars are very diverse

▶ **Table 6** List of primers used for this study. Δ12 part of the sequence underlined.

name	5′-3′ sequence	length	tm (°C)
mr_CHS_-283F	ACACCCGCTCGTAGACAAAC	20	64.1
mr_CHS_1535R	GCTCGTAAACCTAGAGYCCA	20	58.6
Δ12_mr_CHS_-283F	AAAAAGCAGGCTACACCCGCTCGTAGACAAAC	32	76.5
Δ12_mr_CHS_1535R	AGAAAGCTGGGTGCTCGTAAACCTAGAGYCCA	32	75.3
attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCT	29	72.8
attB2	GGGGACCACTTTGTACAAGAAAGCTGGGT	29	74.6

and bridge the extremes, 'Argenmilla'/'Bona' and 'Mat16'. 'Bona' is a cultivar bred by crossing the cultivar 'Bohemia' (originating from wild chamomile from Bohemia and Moravia) with wild populations from Spain [16]. From such a breeding background with geographically diverse origins, higher CHS variability should have been expected. So, geographical diversification is possibly not the only driving force that explains the CHS distance of the Bulgarian 'Mat16' from most of the other chamomile variants. All our genotypes were outcomes or a part of breeding approaches of chamomile that usually started with the collection of diverse accessions, cultivating them under the same conditions and selecting the best genotypes, often after uncontrolled cross-fertilisation blurred our view of the true relationships in the wild.

The wide variation of CHS within chamomile cannot be compared to *M. discoidea* since only one specimen from that species was included in our study. Interestingly, *M. discoidea* docks to *M. chamomilla* in the centre of its CHS network, indicating a diverging bifurcating evolution in chamomile immediately after speciation. Assuming a centre of chamomile origin in the West Balkan area [17], a northern ('Bona') and eastern ('Mat16′) route could explain the bifurcation. The two Polish varieties coming from farther north than 'Bona', however, contradict this argument or could indicate the immigration from the east rather than the south. Such guesses, however, are inappropriate if no wild accessions are analysed. However, the complexity of McCHS could be a good tool for identifying migration routes.

If geographical distance alone does not explain the heterogeneity of McCHS, perhaps interspecific hybridisation was the (additional) driving force for the hypervariability of McCHS. *M. discoidea* was chosen as an outgroup because of an overlapping distribution area with *M. chamomilla* and morphological resemblance with only some differences (smaller flower heads, the absence of ray flowers, greenish tubular flowers and a four-toothed corolla in *M. discoidea* [1]). *M. discoidea* (syn. *M. matricarioides*) is a self-fertile plant, not well adapted to cross-fertilization, while *M. chamomilla* (syn. *M. recutita*) is self-incompatible [18, 19], which may explain the compact MdCHS and variable McCHS. Repeated artificial crosses between *M. chamomilla* and *M. discoidea* were not successful [19]. Crossability of *M. chamomilla* with other species from the small *Matricaria* genus is unknown, but hardly any of the other *Matricaria* species except for *M. discoidea* have a wider natural overlap, making interspecific hybridisations explaining the high McCHS variability rather unlikely.

Material and Methods

Plant material and DNA extraction

Seeds of chamomile (▶ **Table 1**) were cultivated in the greenhouse. Leaves of selected specimens were harvested when the plants were about 10 cm tall and dried at room temperature. *Matricaria discoidea* (syn. *M. matricarioides*) was included as outgroup to determine the most 'ancient' haplotype in chamomile [17]. The botanical identities of all specimens were checked by sequencing ITS (GenBank accession nos. OR710969–OR710975) and aligning them with sequences from [20].

Development of primer sets

Chamomile NGS data from a previous study (Submission ID: SUB5046906/BioProject ID: PRJNA515664) were assembled to CHS from *Chrysanthemum* x *morifolium* (GenBank accession no. MW368977) as reference sequence. An overlapping primer set was developed in Geneious Prime 2023.0.4 (https://www.geneious.com) and was synthesised by Merck (▶ **Table 6**). The reversed primer was finally designed after preliminary tests (data not shown) revealed a polymorphic site at a primer binding position in one of the primers (mr_CHS_1535R). This primer was developed as the 'wobble' primer.

DNA Extraction

DNA was extracted with a modified CTAB extraction protocol [21]. Quantity and quality of the DNA were determined on a spectrophotometer (DS-11 Series, DeNovix Inc.) and on a 1.4% agarose gel stained with peqGREEN (VWR International). The DNA was dissolved in TE buffer and stored at − 20 °C until further usage.

PCR

For a 20 µL PCR reaction, 10 ng of genomic DNA was added to a master mix containing 5% DMSO in the 1× PCR buffer HF, 1.6 U Phusion Hot Start II DNA Polymerase (both Thermo Fisher Scientific), 200 µM dNTPs (Solis BioDyne) and 200 nM of the respective forward and reverse primers (Merck). The PCR conditions included one cycle with a denaturation step at 98 °C for 50 seconds, followed by 35 cycles at 98/58/72 °C for 10/30/90 seconds and a final elongation step at 72 °C for 7 minutes. PCR products were checked on 1% agarose gels stained with peqGREEN (2 µL/100 mL agarose solution).

Gel extraction

The PCR products were cleaned up with the InnuPREP DOUBLEpure Kit (IST Innuscreen GmbH) where necessary as recommended in the manufacturer's instructions and used for PCR or the BP cloning reaction dependent of the next step.

Cloning and Sequencing

PCR products were cloned into the Gateway pDONR221 Vector using the Gateway cloning protocol. The DNA fragment containing the chalcone synthase gene was initially amplified by PCR, followed by a PCR using the same primers elongated with Δ12 sequence (▶ **Table 6**). The last PCR step was undertaken by using primers with the appropriate attB sites. Gateway cloning with the Gateway pDONR221 vector and the Gateway BP Clonase II Enzyme Mix (Thermo Fisher Scientific) was undertaken according to the manufacturer's protocol, except for the duration of the BP reaction (18 hours) to enable a higher recombination rate. The plasmid was introduced into chemically competent *E. coli* (One Shot TOP10 Chemically Competent *E. coli*; Thermo Fisher Scientific) using the heat shock method. *E. coli* was plated on LB agar plates containing 5 µg/mL kanamycin sulphate. *E. coli* colonies were picked from the agar plate and cultivated in liquid LB broth with 5 µg/mL kanamycin sulphate overnight. The next day, the plasmids were extracted with the innuPREP Plasmid Mini Kit 2.0 (IST Innuscreen GmbH) according to the manufacturer's instructions. Quantity and quality of the DNA were determined on a spectrophotometer and diluted to 80 ng/µL and sent to Microsynth for sequencing.

Bioinformatics

Sequence chromatograms were controlled and assembled with Geneious Prime 2023.0.4 (https://www.geneious.com). The sequences were submitted to GenBank (accession numbers OR634854–OR634923). All singletons (mutations not confirmed by the same mutation in another sequence) were regarded as sequencing errors and eliminated. The sequences were aligned and first analysis performed in MEGA X [22]. Further analyses were conducted in R version 4.3.0 under RStudio 2023.06.1, packages pegas [23], poppr [24, 25] and igraph [26]. Clusters of the haplotype network were identified with the Leiden cluster (resolution on DNA: 0.0001, resolution on AA: 0.03) [27].

Contributors' Statement

MN: analysis, statistical evaluation, preparation of manuscript. DJ: adaptation of methods, laboratory analysis, preparation of manuscript. JN: planning, statistical evaluation, preparation of manuscript.

Acknowledgements

We are thankful to Tamara Teufl for her assistance in the laboratory.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Fischer M, Oswald K, Adler W. Exkursionsflora für Österreich, Liechtenstein und Südtirol. 3rd ed. Linz: Land Oberösterreich, Biologiezentrum der oberösterreichischen Landesmuseen; 2008

[2] Schilcher H, Imming P, Goeters S. Pharmacology and Toxicology. In: Franke R, Schilcher H, eds. Chamomile. Vol. 42. Medicinal and Aromatic Plants – Industrial Profiles. Boca Raton, FL, USA: CRC Press; 2005: 245–264

[3] Pandith SA, Ramazan S, Khan MI, Reshi ZA, Shah MA. Chalcone synthases (CHSs): The symbolic type III polyketide synthases. Planta 2019; 251: 15. doi:10.1007/s00425-019–03307-y

[4] Sommer H, Saedler H. Structure of the chalcone synthase gene of *Antirrhinum majus*. Mol Gen Genet 1986; 202: 429–434. doi:10.1007/BF00333273

[5] Radhakrishnan EK, Varghese RT, Vasudevan SE. Unusual intron in the second exon of a Type III polyketide synthase gene of *Alpinia calcarata* Rosc. Genet Mol Biol 2010; 33: 141–145. doi:10.1590/S1415–47572010000100024

[6] Ma LQ, Pang XB, Shen HY, Pu GB, Wang HH, Lei CY, Wang H, Li GF, Liu BY, Ye HC. A novel type III polyketide synthase encoded by a three-intron gene from *Polygonum cuspidatum*. Planta 2009; 229: 457–469. doi:10.1007/s00425-008–0845-7

[7] Zhang C, Yao X, Ren H, Wang K, Chang J. Isolation and characterization of three chalcone synthase genes in pecan (*Carya illinoinensis*). Biomolecules 2019; 9: 236. doi:10.3390/biom9060236

[8] Han Y, Cao Y, Jiang H, Ding T. Genome-wide dissection of the chalcone synthase gene family in *Oryza sativa*. Mol Breeding 2017; 37: 1–12. doi:10.1007/s11032-017–0721-x

[9] Hu L, He H, Zhu C, Peng X, Fu J, He X, Chen X, Ouyang L, Bian J, Liu S. Genome-wide identification and phylogenetic analysis of the chalcone synthase gene family in rice. J Plant Res 2017; 130: 95–105. doi:10.1007/s10265-016–0871-7

[10] Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nat Struct Biol 1999; 6: 775–784. doi:10.1038/11553

[11] Yin YC, Zhang XD, Gao ZQ, Hu T, Liu Y. The research progress of Chalcone Isomerase (CHI) in plants. Mol Biotechnol 2019; 61: 32–52. doi:10.1007/s12033-018–0130-3

[12] Sanjari S, Shobbar ZS, Ebrahimi M, Hasanloo T, Sadat-Noori SA, Tirnaz S. Chalcone synthase genes from milk thistle (*Silybum marianum*): isolation and expression analysis. J Genet 2015; 94: 611–617. doi:10.1007/s12041-015–0560-7

[13] Ruzicka J, Hacek M, Novak J. Mitochondrial relationships between various chamomile accessions. J Appl Genet 2021; 62: 73–84. doi:10.1007/s13353-020–00602-3

[14] Das M. Chamomile. Boca Raton, FL, USA: CRC Press; 2014

[15] Albrecht S, Otto LG. *Matricaria recutita* L.: True Chamomile. In: Novak J, Blüthner WD, eds. Medicinal, Aromatic and Stimulant Plants. Vol. 12. Springer eBook Collection. Cham: Springer International Publishing; Imprint Springer; 2020: 313–331

[16] Salamon I. Growing condition and breeding of chamomile (*Chamomilla recutita* (L.) Rauschert) regarding the essential oil qualitative-quantitative characteristics in Slovakia. Herba Polonica 1994; 40: 68–74

[17] Ruzicka J, Baumschlager G, Jovanovic D, Novak J. Three major chlorotype lineages in chamomile (*Matricaria chamomilla* L., Asteraceae). Genet Resour Crop Evol 2023. doi:10.1007/s10722-023–01625-5

[18] Kay Q. Experimental and Comparative Ecological Studies of Selected Weeds [Dissertation]. Oxford: University of Oxford; 1965

[19] Mitsuoka S, Ehrendorfer F. Cytogenetics and evolution of *Matricaria* and related genera (Asteraceae-anthemideae). Oesterr Bot Z 1972; 120: 155–200. doi:10.1007/BF01373265

[20] Inceer H, Garnatje T, Hayırlıoğlu-Ayaz S, Pascual-Díaz JP, Vallès J, Garcia S. A genome size and phylogenetic survey of Mediterranean Tripleurospermum and Matricaria (Anthemideae, Asteraceae). PLoS One 2018; 13: e0203762. doi:10.1371/journal.pone.0203762

[21] Schmiderer C, Lukas B, Novak J. Der Einfluss verschiedener DNS-Extraktionsmethoden und Verdünnungen auf den Amplifikationserfolg in der PCR von verschiedenen Arznei- und Gewürzpflanzen. ZAG 2013; 18: 65–72

[22] Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 2021; 38: 3022–3027. doi:10.1093/molbev/msab120

[23] Paradis E. pegas: An R package for population genetics with an integrated-modular approach. Bioinformatics 2010; 26: 419–420. doi:10.1093/bioinformatics/btp696

[24] Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2014; 2: e281. doi:10.7717/peerj.281

[25] Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 2015; 6: 208. doi:10.3389/fgene.2015.00208

[26] Csárdi G, Nepusz T. The igraph software package for complex network research. Complex Systems 2006: 1695

[27] Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: Guaranteeing well-connected communities. Sci Rep 2019; 9: 5233. doi:10.1038/s41598-019–41695-z