
Article:
Olofsson, J.K. orcid.org/0000-0002-9527-6573, Tyler, T. orcid.org/0000-0002-7886-7603, Dunning, L.T. orcid.org/0000-0002-4776-9568 et al. (3 more authors) (2024) Morphological and genetic evidence suggest gene flow among native and naturalized mint species. American Journal of Botany, 111 (2). e16280. ISSN 0002-9122

https://doi.org/10.1002/ajb2.16280

Received: 9 August 2023 | Accepted: 13 December 2023

DOI: 10.1002/ajb2.16280

RESEARCH ARTICLE

Botany

Morphological and genetic evidence suggest gene flow among native and naturalized mint species

Jill K. Olofsson[1] | Torbjörn Tyler[2] | Luke T. Dunning[3] | Mats Hjertson[4] | Åke Rühling[3,5] | Anders J. Hansen[1,6]

[1]Section for GeoGenetics, Globe Institute, University of Copenhagen, Øster Voldgade 5-7, Copenhagen DK-1350, Denmark

[2]Department of Biology, The Biological Museum, Lund University, Box 117, SE-221 00, Lund, Sweden

[3]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield S10 2TN, Western Bank, UK

[4]Museum of Evolution, Botany, Uppsala University, Norbyvägen 16, SE-752 36, Uppsala, Sweden

[5]Biological Museum, Gyllings väg 9, SE-572 36 Oskarshamn, Sverige

[6]Center for Evolutionary Hologenomics, Globe Institute, University of Copenhagen, Øster Farimagsgade 5, Copenhagen K 1353, Denmark

Correspondence

Jill K. Olofsson, Section for Forest and Landscape Ecology, Department of Geosciences and Natural Resource Management, Rolighedsvej 23, DK-1958 Frederiksberg C, Denmark.
Email: jko@ign.ku.dk

Torbjörn Tyler, Department of Biology, The Biological Museum, Lund University, Box 117, SE-221 00 Lund, Sweden.
Email: torbjorn.tyler@biol.lu.se

Present address

Jill K. Olofsson, Section for Forest and Landscape Ecology, Department of Geosciences and Natural Resource Management, Rolighedsvej 23, Frederiksberg CDK-1958, Denmark.

Abstract

Premise: Cultivation and naturalization of plants beyond their natural range can bring previously geographically isolated taxa together, increasing the opportunity for hybridization, the outcomes of which are not predictable. Here, we explored the phenotypic and genomic effects of interspecific gene flow following the widespread cultivation of *Mentha spicata* (spearmint), *M. longifolia*, and *M. suaveolens*.

Methods: We morphologically evaluated 155 herbarium specimens of three *Mentha* species and sequenced the genomes of a subset of 93 specimens. We analyzed the whole genomes in a population and the phylogenetic framework and associated genomic classifications in conjunction with the morphological assessments.

Results: The allopolyploid *M. spicata*, which likely evolved in cultivation, had altered trichome characters, that is possibly a product of human selection for a more palatable plant or a byproduct of selection for essential oils. There were signs of genetic admixture between mints, including allopolyploids, indicating that the reproductive barriers between *Mentha* species with differences in ploidy are likely incomplete. Still, despite gene flow between species, we found that genetic variants associated with the cultivated trichome morphology continue to segregate.

Conclusions: Although hybridization, allopolyploidization, and human selection during cultivation can increase species richness (e.g., by forming hybrid taxa), we showed that unless reproductive barriers are strong, these processes can also result in mixing of genes between species and the potential loss of natural biodiversity.

KEYWORDS

admixture, allopolyploid, hybridization, indumentum, introgression, *Mentha*, trichome

Hybridization and gene flow are major drivers of speciation and trait evolution in plants (Anderson and Stebbins, 1954; Rieseberg, 1995; Otto and Whitton, 2000; Morjan and Rieseberg, 2004; Soltis et al., 2009, 2015). Hybridization between divergent lineages can result in novel stable hybrid taxa that are reproductively isolated from their parental species, by for

example, allopolyploidization or clonal propagation (Grant, 1981; Ellstrand and Schierenbeck, 2000; Rieseberg and Burke, 2001; Schierenbeck and Ellstrand, 2009; Abbott et al., 2013). Allopolyploidization (genome duplication after hybridization) results in a change in chromosome number in the allopolyploid in relation to the parental species and often creates an instant reproductive barrier between them (reviewed by for example, Mallet, 2007; Abbott et al., 2013). Thus, allopolyplodization is an efficient mechanism of speciation and trait adaptation (Mallet, 2007; Abbott et al., 2013). However, there is evidence of instances where the isolating barriers between hybrids, including allopolyploids, and parental taxa are weak (Quilodrán et al., 2020; Schmickl and Yant, 2021), with the possibilities of continued gene flow between them (Ellstrand and Schierenbeck, 2000; Pinto et al., 2005; Todesco et al., 2016; Beninde et al., 2018; Quilodrán et al., 2020; Schmickl and Yant, 2021). Hybridization can therefore not only promote speciation but also the loss of species due to homogenization of the gene pools (Todesco et al., 2016; Owens and Samuk, 2020; Quilodrán et al., 2020).

Human cultivation has undoubtedly changed the evolutionary history of many plant species. Because polyploids often possess mega-features, e.g., large seed size or increased biomass, they are prominent among our cultivated plants (Renny-Byfield and Wendel; 2014). Although crops and cultivars often are associated with human activity, many can escape their cultivated lifestyle and become naturalized (Richardson and Pyšek; 2012; Dawson et al., 2017; van Kleunen et al., 2020). Naturalization of cultivated taxa often bring previously isolated, but closely related, taxa in contact and thus increase the opportunities for hybridization and interspecific gene flow, the consequences of which is not always clear (Abbott, 1992; Rhymer, 1996; Richardson and Pyšek, 2012; Dawson et al., 2017). Outcompetition of native taxa by alien hybrids/allopolyploids is a well-described example in some systems (e.g., Abbott, 1992; Mooney and Cleland, 2001). Here we explored the genomic and phenotypic dynamics of interspecific gene flow after recurring cultivation and naturalization in mints (genus *Mentha*).

Mints (genus *Mentha*; Lamiaceae) are mostly aromatic herbs with large distributions in Eurasia and North America and several species are widely cultivated (Gobert et al., 2002; Tucker, 2012; Salehi et al., 2018; Singh and Pandey, 2018; Vining et al., 2020; Figure 1A). Although *Mentha* is a small genus (18 species; Tucker and Naczi, 2007), it is highly diverse due to the interfertility of many taxa, their capacity for vegetative propagation, and therefore numerous hybrid species are recognized (Figure 1; Tucker et al., 1980; Gobert et al., 2002; Tucker and Naczi, 2007). Most mint hybrids have strongly reduced fertility but can successfully propagate clonally using rhizomes (Harley and Brighton, 1977; Gobert et al., 2002). However, fertile hybrids are known, e.g., between the two diploid species *M. longifolia* L. (L.) ($2n = 24$) and *M. suaveolens* Ehrh. ($2n = 24$; Sobti, 1965; Harley and Brighton, 1977; Chambers and Hummer, 1994). In southwestern Europe, where these two taxa are sympatric, and in cultivation, they hybridize spontaneously to form the partly fertile hybrid

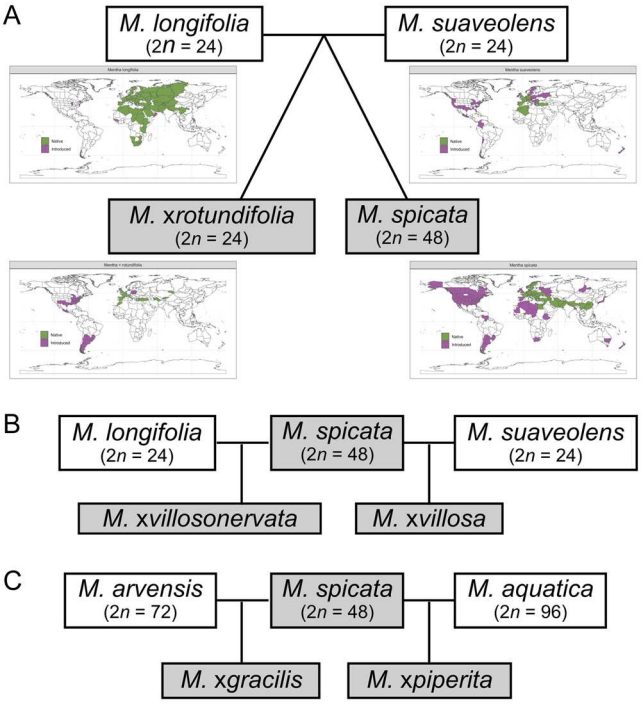

FIGURE 1 Relationships among the studied mints. Relationships between selected species and hybrids (grey boxes) in the section *Mentha* of the genus *Mentha* (mints). Previously reported chromosome numbers are indicated. Adapted from Gobert et al. (2002). Distribution maps adapted from POWO (2023). *Mentha longifolia* and *M. suaveolens* are hypothesized to hybridize to form the two taxa *M.* ×*rotundifolia* and *M. spicata* (A). *Mentha spicata* can both backcross to its presumed parents (B) and with other mints (C) to form complex hybrids.

complex *M.* ×*rotundifolia* L. (Huds.) (Figure 1A; Harley and Brighton, 1977). In addition, the widely cultivated *M. spicata* L. (spearmint; Singh and Pandey, 2018; Vining et al., 2020) is often interpreted as an allopolyploid ($2n = 48$) formed from hybridization between *M. longifolia* and *M. suaveolens* (Figure 1A; Sobti, 1965; Harley and Brighton, 1977; Tucker et al., 1980; Chambers and Hummer, 1994; Gobert et al., 2002; Tucker and Naczi, 2007). *Mentha spicata*, in turn, is known to hybridize with its presumed parental species and with several other species of the genus (Figure 1B, C).

The origin and native range of *M. spicata* is unknown, and some have argued that it arose in cultivation (Harley and Brighton, 1977; Tucker, 2012). *Mentha spicata* is found in natural habitats across most of continental Europe, the Middle East, and southern Asia (Figure 1A). It may be native in some regions (Harley and Brighton, 1977; Vining et al., 2020), but due to its wide cultivation, it is more common as a garden escape within its presumed native range and elsewhere (Harley and Brighton, 1977; Gobert et al., 2002; Vining et al., 2020). Today *M. spicata* is therefore considered naturalized in most geographic regions with a warm-temperate climate (Stace et al., 2016), and its current distribution thus widely exceeds the native ranges of its hypothesized parental taxa (Figure 1A; Harley and Brighton, 1977; Vining et al., 2020). Regardless of whether *M. spicata* arose in cultivation or in the wild, its

emergence and complex history of cultivation and garden escapes (Harley and Brighton, 1977; Gobert et al., 2002; Vining et al., 2020) creates an excellent opportunity to study the phenotypic and genotypic consequences of hybridization between hybrid species, diploids and allopolyploids, and their parental taxa, in particularly following their naturalization.

Here we took advantage of the biodiversity stored in herbarium collections that allows for evaluations of specimens collected over large geographic regions (Bieker and Martin, 2018). We first evaluated the morphological variation and delimitation of *M. spicata*, its putative parental species (*M. longifolia* and *M. suaveolens*), and their hybrids (*M. ×rotundifolia*). We then used whole-genome sequencing to (1) establish patterns of genetic admixture and gene flow between taxa, (2) infer the genomic histories of *M. ×rotundifolia* and *M. spicata*, and (3) test the hypothesis that there have been increased opportunities for interspecific gene flow after the introduction of cultivated *M. spicata*.

MATERIALS AND METHODS

Sample selection and herbarium label interpretations

A total of 155 herbarium specimens of *Mentha* subgen. *Mentha* were obtained from the herbaria at Lund University, Sweden (LD), Uppsala University, Sweden (UPS), and Oskarshamn, Sweden (OHN; Appendix S1). Specimens were selected based on the plant identification by the collector(s) and/or subsequent re-identifications by botanists during this or previous projects, with the aim to include the taxa *M. longifolia* L. (L.), *M. suaveolens* Ehrh., *M. spicata* L., and the hybrid *Mentha ×rotundifolia* L. (Huds.) (*M. longifolia ×suaveolens*). However, after detailed morphological and genomic analyses (see Results), we found two additional hybrid taxa were included in our sampling: *Mentha ×villosonervata* (*M. spicata ×longifolia*) and *Mentha ×villosa* (*M. spicata ×suaveolens*; Figure 1C). In addition, a single specimen representing *M. capensis*, a close relative of *M. longifolia*, was also sampled. Specimens across Europe and Asia and a few from North America were included, although about one third of the specimens were collected in Sweden (Appendix S1). In the herbaria, specimens were identified by historic experts, and more than 100 different names for various mints were used, most of which are now interpreted as referring to different vegetatively propagated clones of mostly hybrid origin. We therefore revised the taxonomy of all specimens, and the names of 65 specimens (41%) were changed to be concordant with the current taxonomy. In addition, in a few cases specimens were re-identified based on our new morphometric analyses (see below; Appendix S1, Table S1). From here on, the names *M. longifolia* (L.) L., *M. suaveolens* Ehrh., *M. spicata* L., and *M. capensis* Thunb. (and their hybrids) are used to denote those groups of specimens identified by the morphometric analyses that most closely correspond to the morphological descriptions of these species in standard floras (Tutin, 1993; Stace, 2010; Stace et al., 2016). However, we stress that these morphologically defined groups may not exactly correspond to the biological species that previous authors have referred to by these names (which may, in addition to morphology, partly have been based on biogeographical, reproductive, cytological, or chemical evidence), nor to their nomenclatural types (which we did not check). All specimens were morphologically evaluated, and the whole genome of 93 samples was sequenced. The genome of specimen of *M. pulegium* L. was also sequenced and used as an outgroup in the genomic analyses.

Morphological characters and statistical analyses

We scored 34 morphological characters (Appendix S2) that were selected to cover all parts of the plant generally available on herbarium specimens (stem, leaves, calyces, flowers, and sexual organs) and took care to avoid intrinsically dependent characters and characters directly related to the size/vigor of the plant. Indumentum characters are widely used and generally considered important in *Mentha* taxonomy (Šarić-Kundalić et al., 2009; Yu et al., 2018) because they change relatively little in response to environmental conditions. Ten characters referring to various types of trichomes were therefore included (Appendix S2). All characters were treated as continuous traits, although some were scored as multiple categories (Appendix S2). Cases of missing data were treated by mean value imputations. The morphometric data were analyzed using principal component analyses (PCAs) in R v.3.6.3 (R Core Team, 2021), and subclusters of specimens were visually identified. To illustrate and assess the importance of indumentum characters, we used separate PCAs, excluding all characters referring to the density or structure of trichomes in the first and in the second, only including characters of the indumentum (Appendix S2). The distributions of indumentum characters between different taxa were analyzed. We further analyzed our taxonomic assignments with a generalized canonical discriminant analysis using the function candisc in the R package candisc (Friendly and Fox, 2021).

DNA extraction, library construction, and Illumina sequencing

For 94 specimens, the whole genome was sequenced at the Geogenetics Sequencing Core (GLOBE Institute, University of Copenhagen, Denmark; Appendix S1). Total genomic DNA was extracted from approximately a 1–3-cm^2 piece of dried material from a single leaf of an herbarium specimen. In cases of small leaves, one whole leaf was used. The leaf material was ground to a powder using TissueLyser II (Qiagen, Hilden, Germany), and DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The extracted DNA was quantified using the Qubit double-stranded

DNA High-Sensitivity Assay kit (ThermoFisher Scientific, Waltham, MA, USA).

For 72 samples, 4.6–110 ng of DNA (Appendix S1) was fragmented using the 96-microTUBE Plate set-up in a Covaris LE220-plus Focused-ultrasonicator (Covaris, Woburn, MA, USA) to a target insert size of 300 bp; 32 μL of the fragmented DNA (4.2–100 ng) was used to build shot-gun sequencing libraries using the BEST protocol (Carøe et al., 2018). The final library concentrations were evaluated with qPCR (Mx3005; Agilent Technologies, Santa Clara, CA, USA), and libraries were amplified (25 μL reaction volume) with the desired number of PCR cycles (Appendix S1) adding indices to the P5 and P7 end of the libraries according to Carøe et al. (2018). A total of 41–47 compatible libraries from the same or different projects were pooled in equimolar concentrations, and each pool was paired-end sequenced (100 bp or 150 bp; Appendix S1) on one lane of Illumina NovaSeq. 6000 for a target coverage of approximately 5×. For an additional 22 samples (Appendix S1), single indexed libraries were built using the NEBNext Ultra II FS DNA Library Prep Kit (Illumina, San Diego, CA, USA), adjusting the fragmentation time to 25 min for a target insert size of 160 bp, then PCR-amplified with 12 cycles. The resulting libraries were pooled in equimolar concentrations and single-end sequenced (80 bp) on one lane of Illumina HiSeq. 4000 (Appendix S1). Twenty-one samples were further sequenced to obtain deeper sequencing (Appendix S1). Libraries were rebuilt using the BEST protocol as above, pooled in equimolar concentrations, and paired-end sequenced (150 bp) on one lane of Illumina NovaSeq. 6000.

Raw reads were quality-filtered in AdaptorRemoval v.2.3.1 (Schubert et al., 2016) removing adaptor sequences, ambiguous bases (N), and consecutive bases with low quality ($Q < 20$) from the 5′ and 3′ termini. Trimmed reads shorter than 25 bp and with more than 20 ambiguous bases after the quality filtering were discarded. All other settings were default. Overlapping pair-end reads were collapsed, and all paired-end reads without a mate were discarded.

Chloroplast alignment and phylogeny

Quality-filtered and trimmed reads were mapped to the *Mentha longifolia* plastome (GenBank NC_032054, masking one of the inverted repeats) using bowtie2 v.2.3.4.3 (Langmead and Salzberg, 2012; Langmead et al., 2019) with default settings. Mapped reads were quality-filtered using SAMtools v.1.9 (Danecek et al., 2021), removing all reads that mapped with a mapping quality score below 20. PCR-duplicated reads were removed using PicardTools v.2.25.2 (https://github.com/broadinstitute/picard). For each sample, a consensus sequence was extracted using BCFtools v.1.9 (Danecek et al., 2021), and all ambiguous bases were recorded as N. Plastomes were aligned in mafft v.7.392 (Katoh and Standley, 2013) with default settings, and a bootstrapped maximum-likelihood phylogenetic tree was inferred in RaxML v.8.2.11 (Stamatakis, 2014) with a GTR + G + I mutation model using *Salvia* as the outgroup (Li et al., 2016; NCBI GenBank accessions NC_050898, MN520020, MN520021, NC_038165, MG772529, MN520019, NC_050897, MN520023, and NC_050900).

Nuclear SNP calling and genomic analyses

The quality-filtered and trimmed reads were mapped to the *M. longifolia* reference genome (GCA_001642375.1_M-long1.0; Vining et al., 2017) using bowtie2 with default settings. The reference genome used is highly fragmented with a total scaffolds of 190,876 and N50 of 3.6 kb (Vining et al., 2017). Mapped reads were quality-filtered as outlined above, and reads that were previously mapped to the plastome were removed. The quality-filtered mapping files were used to call single nucleotide polymorphisms (SNPs) in angsd v.0.931, outputting genotype likelihoods (Korneliussen et al., 2014). For comparative purposes, all specimens were treated as diploids. Mapping polyploid samples to a diploid reference genome can be problematic, especially for allopolyploids. Differences in divergence time between subgenomes of an allopolyploid and the reference genome can cause differences in mapping success between alleles of the subgenomes, resulting in allele dropouts and overestimation of genetic similarity between some taxa. We did indeed observe differences in mapping success between different species but not to the extent that we expect systematic biases (see Results and Discussion). Similarly, we might miss low frequency alleles in polyploids, which could lead to an underestimation of genetic variation between groups of samples. However, the high level of genetic admixture detected in our analyses (see Results) indicates that such low-frequency alleles only contribute to a small fraction of the overall genetic variation among the analyzed taxa. To compensate for mapping biases, *k*-mer analyses can be useful; however, they require relatively high sequencing depth, preventing us from making efficient *k*-mer analyses for most of our samples (Appendix S1). In addition, samples from the herbarium record usually show relatively high amount of exogenous DNA (Bieker and Martin, 2018); thus, *k*-mer analyses can be inefficient as the "contamination" peak overshadows the *k*-mer spectra, an issue also observed with our data (data not shown). Thus, we chose to base our analyses on reads that mapped to the reference genome, recognizing the limitations of this method (see Discussion). Only SNPs with less than 50% missing data across all individuals were retained, and minimum and maximum depth to call a SNP within an individual was set to three and 50, respectively. The SAMtools option was used to call SNPs, and only genotypes with a posterior probability above 0.95 were used in downstream analyses. For comparative purposes, a second SNP set was called as above but allowing for any level of missing data.

A maximum likelihood phylogeny was evaluated in IQ-TREE v.1.6.12 (Nguyen et al., 2015) from the concatenated nuclear SNPs (stringent filter). To compare how clustered

SNPs affect the phylogenetic inference, we also performed the phylogenetic analysis only including the first SNP on each scaffold. Gene-tree to species-tree discordance was evaluated using 1123 benchmark universal single-copy ortholog (BUSCO) genes. The BUSCO genes were annotated to the reference genome using busco v.5.1.3 (Manni et al., 2021a, 2021b). Genotypes were extracted from mapped reads for the 1312 genes that were found to be single copy, coding ambiguous genotypes based on the IUPAC codes with missing data coded as "N". Only two alleles were allowed because all samples were treated as diploids. For rooting purposes, a single accession of *M. pulegium* was included that excluded 187 genes for which no genotypes were called in the accession. The remaining 1125 were trimmed in Trimal v1.3.1, removing sites with more than 50% missing data across the accession, which completely removed two genes. Alignments that did not include all accessions and/or were shorter than 500 bp were discarded. Individual gene trees were generated for the remaining 402 genes using SMS and PhyML v.1.8.1, and a coalescence species tree was generated using ASTRAL v.5.7.5 (Zhang et al., 2018) as previously described (Dunning et al., 2022). Individual gene tree support for the coalescence species tree was evaluated in Phyparts v.0.0.1 (Smith et al., 2015), and the results were visualized as described by Dunning et al. (2022).

Genomic clusters were evaluated from the nuclear SNPs in PCangsd v.0.95 (Meisner and Albrechtsen, 2018) with default settings. The resulting covariance matrices were converted to eigenvalues and visualized in R. Genomic admixture was evaluated using the nuclear SNPs in NGSadmix v3.2 (Skotte et al., 2013) with K increasing from 1 to 10. The best fit K was evaluated using the method of Evanno et al. (2005) implemented in CLUMPAK (Kopelman et al., 2015) using 10 independent runs of NGSasmix for each K.

To confirm our admixture results, analyses of ABBA-BABA sites (Green et al., 2010; Durand et al., 2011) were conducted in angsd v.0.935 using a (((H1,H2),H3),O) phylogenetic constellation. *Mentha pulegium* was used as the outgroup (O), and accessions morphologically identified as *M. longifolia* and not showing signs of admixture at $K = 3$ (see Results) were used as H1 and H2. Accessions of *M. spicata* were then used as H3. Jack-knifed Z-values were obtained using the jackknife.R script supplied with angsd, and a Bonferroni-adjusted significance level of 4.17e-5 (0.05/1200) was used.

F_{ST}-outlier detection and transcriptome analyses

Genetic distances between groups of specimens were evaluated by means of pairwise F_{ST} (Weir and Cockerham, 1984). The pairwise F_{ST} between morphologically and genomically defined groups of specimens (see Results) was calculated in VCFtools v.0.1.16 (Danecek et al., 2011). Analyses of F_{ST} distributions were conducted on a scaffold basis, and thus, to reduce the effect of differences in contig/scaffold length, only contigs/scaffolds longer than 5000 bp were evaluated (max length = 65,819 bp). Negative values of F_{ST} were converted to zero, and the mean F_{ST} value for each included contig/scaffold of the reference genome was calculated as well as the mean and median F_{ST} across all SNPs (including all contigs/scaffold). The distribution of contig/scaffold F_{ST} values were visualized using R, and the contigs/scaffolds with the top 1% F_{ST} values were defined as outliers.

To identify genes on the F_{ST}-outlier contigs/scaffolds, we used available transcriptome data. Transcriptomes for *M. ×piperita* (Figueroa-Pérez et al., 2019) and *M. spicata* (Jin et al., 2014) were downloaded and re-assembled to ensure that they were assembled and filtered using the same method. In short, raw reads were cleaned and trimmed in TrimGalore v.0.6.7 (https://zenodo.org/badge/latestdoi/62039322), removing bases with $Q < 20$ and reads shorter than 50 bp or containing any ambiguous base (N). The cleaned reads were then used to assemble transcripts with Trinity v.2.11.00 (Grabherr et al., 2011) using default parameters. The sequences of F_{ST}-outlier contigs/scaffolds were used as databases for blast searches of the reassembled mint transcriptomes with default settings in BLAST v.2.9.0 (Altschul et al., 1990; Ye et al., 2006). Transcripts with top blast-hits longer than 300 bp (minimum 50% matches) and with an e-value below 1e-5 were used in a blast search (tblastx) against the UniProt database (UniProt Consortium, 2021). Only, the top tblastx-hit (>100 amino acids, minimum 50% matches, and e-value < 1e-5) were considered, and the likely function of the detected genes were evaluated using the UniProt database.

RESULTS

Trichome characters distinguish morphological groups

In the PCA for all specimens based on all assessed morphological characters (Appendices S2, S3), the first component (PC1; 24.1%) separated a group of specimens of *M. spicata* from all other specimens, while PC2 (13.5%) separated *M. suaveolens* from the other taxa (Figure 2A). The hybrid *M. ×rotundifolia* mostly falls intermediate to the two parental taxa, *M. longifolia* and *M. suaveolens* (Figure 2A). However, there is no clear separation between *M. longifolia* and *M. ×rotundifolia* (Figure 2A). Five specimens were identified as *M. ×villosa* (*M. suaveolens ×spicata*), one as *M. ×villosonervata* (*M. longifolia ×spicata*), and one as *M. capensis*, a close South African relative of the Eurasian *M. longifolia*/*M. spicata* (Figures 1B and 2). Separate analyses excluding (Figure 2B) and isolating (Figure 2C) trichome traits showed that the morphological distinction of *M. spicata* is largely driven by characters of the indumentum. Distribution analyses of these indumentum characters showed a large variation of the traits in its putative parental taxa, with

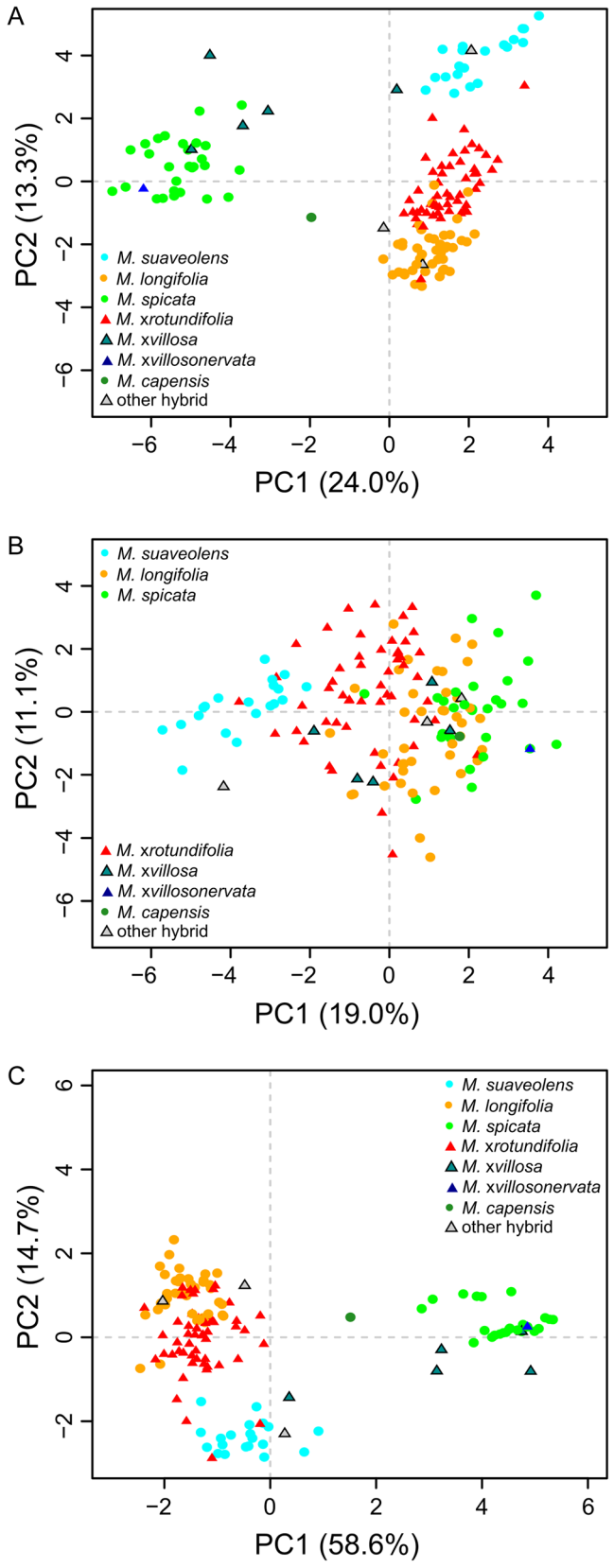

FIGURE 2 Principal component analysis (PCA) of morphological characters. (A) PCA of (A) all 34 morphological characters and 155 specimens, (B) morphological characters excluding 10 indumentum traits, and (C) 10 indumentum characters. Taxonomic assignments are indicated with colors. Circles indicate pure taxa; triangles indicate hybrids.

M. longifolia generally having longer and more densely set hairs than *M. suaveolens*, with branched trichomes almost exclusively found in the latter (Appendix S4, Figures S1, S2). However, while the hybrid *M.* ×*rotundifolia* showed large variation in indumentum with distributions mostly resembling that of *M. longifolia*, *M. spicata* showed less variation and is much less hairy (sometimes almost completely glabrous), thus mostly with distributions outside the ranges of its putative parental taxa (Appendix S4, Figures S1, S2). Canonical discriminant analyses confirmed that, apart from *M. longifolia* and *M. rotundifolia*, the taxonomic units used were well separated in the morphological space (Appendix S4, Figure S3).

Sequencing

Sequencing resulted in an average of 20.8 million raw read pairs per specimen (2.6–190 million; median = 16.6 million), and after quality filtering, an average of 12.7 million read pairs were retained (1.2–108 million; median = 10.1 million), equivalent to a theoretical coverage of the *M. longifolia* genome of ~3.6× (0.3–30×; Appendix S1). On average, 5% (1–17%) of the reads mapped to the chloroplast genome of *M. longifolia*. The variation in percentage of chloroplast reads likely reflects the quality of the sample DNA, whereas the variation in mapping success is influence by both sample quality and evolutionary distance to the reference genome. However, plastomes were obtained for all specimens, and the complete alignment was 153,822 bp, of which 1666 positions were variable among the analyzed mints.

Despite the historic nature of many of our accessions, we obtained a significant number of reads that mapped to the nuclear reference genome with an average of 68% (21–79%; Appendix S1; Appendix S4, Figure S4). This finding is in line with previous work on herbarium samples (Bieker and Martin, 2018; Olofsson et al., 2016) showing that a rather large proportion of the nuclear genome can be assessed from these vast biodiversity resources despite many specimens being old and often not so well (for DNA analyses) preserved. We observed lower mapping to the reference genome of all taxa compared to *M. longifolia*, with a significant difference (Turkey's test) between *M. longifolia* and *M.* ×*rotundifolia*, *M. longifolia* and *M. suaveolens*, and all taxa, except *M. suaveolens*, and *M.* ×*villosa* (Appendix S1; Appendix S4, Figure S4), Differences in mapping success between the different taxa could lead to a mapping bias for alleles with different evolutionary distance (subgenomes) in hybrids and allopolyploids. This can for example lead to an overestimation of gene flow between hybrids/allopolyploids and one of the parent taxa. However, the small difference in mapping success between the two suspected parent taxa (*M. longifolia* and *M. suaveolens*; Appendix S4, Figure S4) combined with the fact that significant gene flow is only detected in some trios (see below) suggests that the overall effect of any mapping bias on the SNP calling is very limited. Furthermore, the low sequencing coverage combined with our treatment of all samples as diploids likely result in missing alleles

in allopolyploids and potentially in the diploids. However, there should not be any bias in missing alleles toward one subgenome; thus, estimation of, for example, gene flow using SNPs across the whole genomes should not be biased by missing alleles. After stringent filtering, 31,889 nuclear SNPs were obtained spread across 5237 scaffolds (mean: 6 SNPs/scaffold; median: 3 SNPs/scaffold; range 1–227). For the less-stringent filtering, 1,688,670 nuclear SNPs were obtained.

Phylogenetic analyses

In accordance with previous results (Li et al., 2016), *Mentha* is resolved as monophyletic in relation to *Salvia* in the plastome phylogeny (Appendix S4, Figure S5A), but the taxa within *Mentha* are polyphyletic (Figure 3A; Appendix S4, Figure S5B). Specimens identified as the hybrid *M.* ×*rotundifolia* were found across the phylogeny together with both *M. longifolia* and *M. suaveolens* (Figure 3A; Appendix S4, Figure S5B). In contrast, except for a single specimen, all specimens identified as *M. spicata* were monophyletic, and most closely related to a subset of the *M. longifolia* specimens (Figure 3A; Appendix S4, Figure S5B).

The phylogenetic analyses using the concatenated nuclear SNPs was overall weakly supported (vast majority of bootstrap values < 90; Figure 3B; Appendix S4, Figure S6). However, *M. suaveolens* was moderately supported (>70) as monophyletic, and most accessions of *M. longifolia* belonged to a single monophyletic group (bootstrap < 70; Figure 3B; Appendix S4, Figure S6). Most accessions of *M.* ×*rotundifolia* and *M. spicata* were situated between *M. suaveolens* and *M. longifolia* (Figure 3B; Appendix S4, Figure S6). A similar phylogeny was obtained when one SNP on each scaffold was used (Appendix S4, Figure S7). Consistent with extensive gene flow among taxa as well as the well-described nonbifurcate history of mints, the gene-tree analyses (BUSCO genes) showed large incongruences with the species tree (Figure 3C; Appendix S4, Figures S8, S9). The high amount of missing data for sample JOL3108 means that the placement of this specimen and the branch lengths are highly uncertain in the nuclear phylogenies (Figure 3C; Appendix S4, Figures S8, S9).

There was no apparent geographical clustering in either the plastome or nuclear phylogenies (Appendix S4, Figures S5, S6). However, although we have included specimens across the species geographic distribution ranges, most of our samples were collected in Nordic countries, with most originating in Sweden. Thus, firm conclusions on geographic clustering are not possible.

Genomic clustering and admixture

Genomic clustering was performed by means of PCA, and there was a separation of *M. longifolia* and *M. suaveolens* on

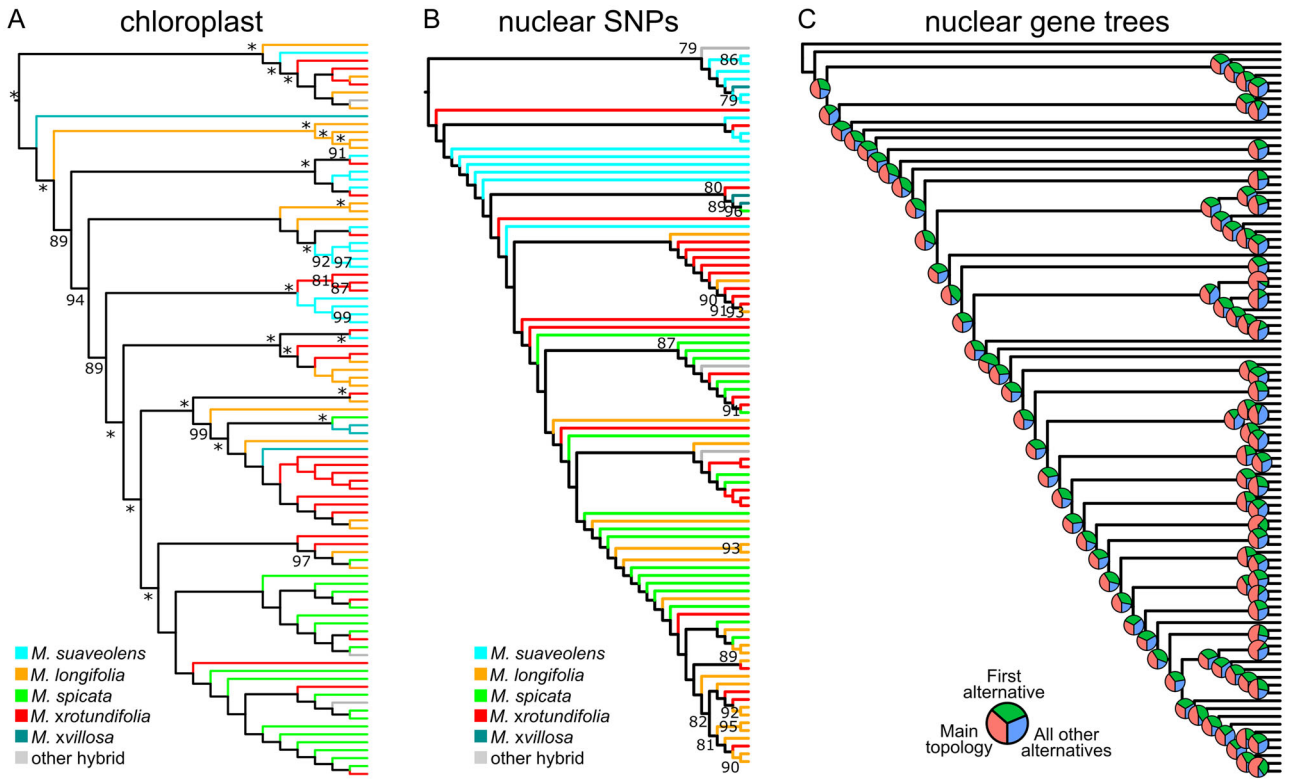

FIGURE 3 Phylogenetic relationships among *Mentha*. Phylogenetic relationships among *Mentha* cladograms of (A) chloroplast phylogeny (full phylogeny: Figure S5) and (B) 31,889 concatenated nuclear single nucleotide polymorphisms (SNPs; full phylogeny: Figure S6), and (C) the maximum likelihood topology with pie charts representing individual gene tree support.

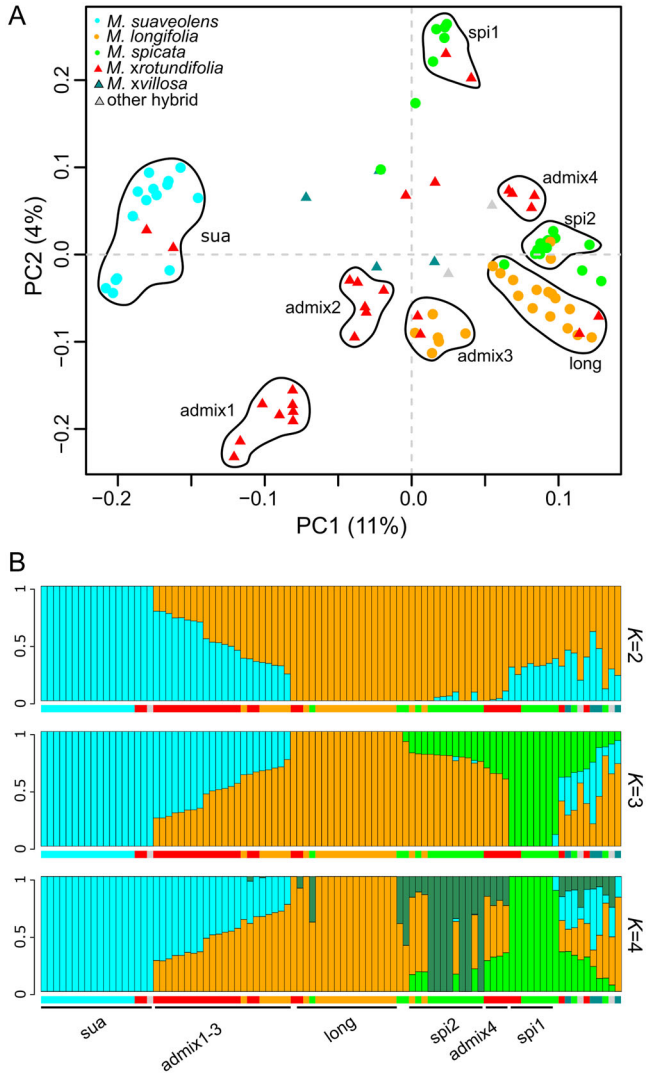

FIGURE 4 Genomic analyses. (A) Principal component analysis and (B) admixture analysis for 31,889 SNPs called for all 93 sequenced specimens. In (A), the taxonomic assignments based on morphology are indicated with colors. Circles indicate pure taxa; triangles indicate hybrid species. In (B), specimens have been ordered based on the genomic cluster analysis (PC1 and PC2) in (A); color under each bar indicates taxonomic assignments as presented in (A).

the first PC (11%; Figure 4A). Intermediate to these taxa were a number of specimens that were consistent with the hybrid *M.* ×*rotundifolia* (admix1-3; Figure 4A). The genetic distinction between *M. longifolia* and *M. spicata* was more elusive, with specimens only somewhat separating along the second PC (4%; Figure 4A). A subset of the specimens morphologically identified as *M. spicata* do however, together with some specimens identified as hybrids, form a separate cluster (spi1) on the positive extreme of the second PC (Figure 4A). The same clustering was observed with the less-stringent SNP filtering (1,688,670 SNPs; Appendix S4, Figure S10A).

Overall, the admixture analysis reflects the genomic clustering with the best-fit number of population groups (*K*) being two, with additional optima at four and six (Figure 4B; Appendix S4, Figure S11A). The specimens separating on the first genomic PC formed three groups with distinct admixture profiles that, based on morphology, represent *M. longifolia*, *M. suaveolens*, and hybrids between them (Figure 4B). Most specimens morphologically recognized as *M. spicata* are admixed (Figure 4B). However, when the number of populations (*K*) was increased, two *M. spicata* clusters appeared (spi1 and spi2; Figure 4; Appendix S4, Figure S11B). For the specimens with more central positions in the genomic PCA, admixture proportions were consistent with mixing of all gene pools (Figure 4). However, our admixture analyses cannot decisively exclude that some patterns are caused by admixture with species not included in our study because hybrids between our study species and, e.g., *M. aquatica* and *M. arvensis* are also well documented (Harley and Brighton, 1977). Similar admixture proportions were obtained with the less-stringent SNP filter (Appendix S4, Figures S10B, C). Analyses of ABBA-BABA sites confirm gene flow between *M. longifolia* and some of the *M. spicata* (Appendix S4, Figure S12). Of the 1200 trios tested, 697 (58%) showed significant gene flow (Appendix S5). A slightly higher percentage of trios including *M. spicata* specimens clustering close to *M. longifolia* (spi2; Figure 4A) showed significant gene flow, 63% of trios compared to 53% for spi1 (Figure 4A). Samples among the nonsignificant trios (503; 42%) can represent geographic subpopulations that are more distant and hence have not exchanged genes and/or are sterile clones and thus propagate vegetatively. Nonsignificant trios can also represent cases where there have been similar levels of gene flow between the populations represented by H3 and H1 and H3 and H2, respectively, or that there has been introgression with other closely related taxa in both H1 and H2.

Although there is overall agreement between the morphological and genomic analyses, there are also a few cases of mismatches. In particular, four specimens morphologically identified as *M.* ×*rotundifolia* were found intermingled with other genomic groups, and an additional four specimens were found intermediate to the two groups of *M. spicata* (admix4; Figure 4A). Finally, eight specimens morphologically identified as *M. longifolia* were admixed and genomically more similar to either *M.* ×*rotundifolia* (admix3) or *M. spicata* (spi2; Figure 4).

Low levels of genetic differentiation

To understand the level of genetic differentiation between morphologically defined taxa and genomic clusters, we analyzed pairwise F_{ST} between groups of specimens. Overall, there was low genetic differentiation between both morphological and genomic clusters with mean scaffold F_{ST} values well below 0.2 (Table 1). A large proportion of scaffolds showed no differentiation ($F_{ST} = 0$), but a few scaffolds show high F_{ST}-values (>0.5; Table 1; Appendix S4, Figures S13, S14). Although peaks of high F_{ST} should be interpreted with

TABLE 1 Mean and median scaffold F_{ST} between morphologically defined taxonomy (Figure 2) and genomic clusters (Figure 4) for *Mentha* taxa.

Morphological taxa							
Taxon	***M. suaveolens***	***M. spicata***	***M. ×rotundifolia***				
M. longifolia	0.141/0.116	0.043/0.029	0.039/0.024				
M. suaveolens	—	0.104/0.082	0.074/0.049				
M. spicata		—	0.037/0.023				
M. ×rotundifolia			—				
Genomic clusters							
Cluster	**Sua**	**spi1**	**spi2**	**admix1**	**admix2**	**admix3**	**admix4**
long	0.098/0.070	0.092/0.059	0.039/0.021	0.118/0.076	0.076/0.025	0.035/0.017	0.080/0.014
sua	—	0.117/0.086	0.100/0.075	0.103/0.056	0.087/0.042	0.089/0.062	0.142/0.096
spi1		—	0.086/0.057	0.146/0.111	0.102/0.065	0.084/0.056	0.109/0.065
spi2			—	0.119/0.079	0.076/0.038	0.046/0.028	0.071/0.019
rot1				—	0.100/0.056	0.094/0.057	0.153/0.115
rot2					—	0.056/0.021	0.095/0.036
rot3						—	0.082/0.033
rot4							—

caution, they are often associated to regions with reduced gene flow and thus important for local adaptation 2014). Although *M. ×rotundifolia* and *M. spicata* were overall more distant to *M. suaveolens* than to *M. longifolia* (Table 1), there was variation in F_{ST} between the different genomic groups and the presumed parental taxa (Table 1). In particular, the genomic cluster spi2 was very similar to *M. longifolia* (Table 1). Despite the detected difference in admixture history between *M. spicata* and *M. ×rotundifolia* (Figure 4B), the genetic distance between them was small (Table 1).

Outlier scaffolds encode genes with functions in reproduction, stress responses, and production of volatile compounds

Although overall genetic differentiation was low between groups of specimens, 63 scaffolds had high pairwise F_{ST}-values (top 1% outliers; $F_{ST} > 0.15$). A total of 782 *Mentha* transcripts (Jin et al., 2014; Figueroa-Pérez et al., 2019) had a BLAST match longer than 300 bp (minimum 50% matches) with an e-value less than 1e-5 to one of the 1% F_{ST}-outlier scaffolds. Of these, 617 transcripts had a BLAST match longer than 100 amino acids (minimum 50% matches) and an e-value less than 1e-5 to a total of 255 different entries in UniProt (UniProt Consortium, 2021), and their potential functions could thus be evaluated. A total of 112 (44%) were matches to an annotation based on plant data with diverse functions (Appendix S6). Of particular interest are the two transcripts with functions in the biosynthesis of volatile compounds (VOCs; Appendix S6), which potentially are involved in the production of the characteristic aromatic compounds of mints.

DISCUSSION

Morphological characters associated with cultivation continue to segregate despite genomic admixture

Despite low levels of genetic differentiation, there is morphological separation between specimens identifiable as *Mentha longifolia*, *M. suaveolens*, and *M. spicata* (Figures 2, 4; Table 1). In particular, *M. spicata* has fewer and shorter trichomes (Figure 2; Appendix S4: Figures S1, S2). This reduction in "hairiness" could be associated with human selection for a more palatable plant (Harley and Brighton, 1977; Munguía-Rosas et al., 2019). Our sample number prevented an extensive genomic association study, but we did not find any genes previously associated with trichome function and morphology among F_{ST}-outlier scaffolds (Appendix S6). Trichome function and morphology are, however, likely controlled by multiple genes (Holeski et al., 2010; Mishra et al., 2017; Askary et al., 2016; Chopra et al., 2019, Figueroa-Pérez et al., 2019; Chalvin et al., 2020); thus, the lack of such genes in our analyses is not surprising.

The glands associated with trichomes are the main organs for production of the desired essential mint oils (Fahn, 1979; McCaskill and Croteau, 1995; Jia et al., 2013; Mishra et al., 2017; Yu et al., 2018). It is therefore possible that selection for essential oil production and/or gland properties (not evaluated here) could have shifted trichome characters as a byproduct (Maffei et al., 1986; McCaskill and Croteau, 1995; Jia et al., 2013 Szabó et al., 2010; Mishra et al., 2017). For example, among the F_{ST}-outlier scaffolds, we found two genes, encoding benzyl alcohol *O*-benzoyltransferase (BEBT1;

Boatright et al., 2004; Orlova et al., 2006) and tetrahydrocannabinolic acid synthase (THCAS; Sirikantaramas et al., 2004), which are potentially involved in the production of volatile organic compounds (Appendix S6). However, given the large number of genes involved in the production of volatile organic compounds (Picazo-Aragonés et al., 2020), it is not surprising to find a few genes among F_{ST}-outliers, and thus, we cannot conclude that there has been any selection on these particular genes.

Despite the overall genetic similarities between specimens of *M. spicata* and *M. longifolia*, and to some extent *M.* × *rotundifolia*, only a few individuals morphologically identified as one taxon clustered genetically with another species (Figure 4A). However, these cases show that variants associated with human desirable traits can enter the genome of other mints through hybridization, similar to what has been observed in other systems (e.g., Fuchs et al., 2016; Karlsson et al., 2016). Overall, our results show that genetic variants associated with the cultivated morphology continue to segregate in both cultivated and wild populations.

Origins and genomic histories of hybrids and allopolyploids

Interfertility of the taxa in *Mentha* subgen. *Mentha* is well known, and in nature, species barriers are possibly retained by the formation of sterile hybrids and/or confinement to distinct ecological niches (Harley and Brighton, 1977; Gobert et al., 2002). Consistent with these observations, we found that all morphologically and genomically defined groups of mint specimens had low levels of genetic distances not consistent with divergent species (Table 1; Meirmans and Hedrick, 2011; Roux et al., 2016). Given the frequent hybridization of *M. longifolia* and *M. suaveolens* in areas of sympatry and in cultivation (Harley and Brighton, 1977), it is unlikely that these two taxa have ever been completely reproductively isolated. Here, we observed extensive admixture between the two species with a large proportion of specimens inferred as the hybrid taxon *M.* ×*rotundifolia* (Figure 4). However, since such hybrids are overrepresented among cultivated clones, cultivation is likely to have further increased possibilities for interspecific gene flow. There is no consistent placement of the specimens of *M.* ×*rotundifolia* with either *M. longifolia* or *M. suaveolens* in the chloroplast phylogeny (Figure 3; Appendix S4, Figure S4) suggesting that hybridization is bidirectional. However, since neither *M. longifolia*, nor *M. suaveolens* is monophyletic (Figure 3; Appendix S4, Figure S4), this conclusion is rather weak.

In accordance with previous studies, we also showed that at least some specimens taxonomically recognized as *M. spicata* are potentially hybrids between *M. longifolia* and *M. suaveolens* (Figure 4). Our sequencing is not deep enough to offer any statistical conclusion to whether they are allopolyploids, and our herbarium material prevented cytological evaluations. Thus, we cannot make any firm conclusions about ploidy, but much evidence points to them being an allopolyploid. For example, although there is some variation in published chromosome counts, most studies indicate that *M. spicata* usually has twice as many chromosomes compared to *M. longifolia* and *M. suaveolens* (Sobti, 1965; Harley and Brighton, 1977; Ahmad et al., 1992; Chambers and Hummer, 1994). In contrast to M. ×*rotundifolia*, *M. spicata* has plastomes with low genetic diversity, and all but one sample clustered in a monophyletic clade with specimens of *M. longifolia* (Figure 3; Appendix S4, Figure S5). This finding is consistent with previous reports (Bunsawatt et al., 2004) and suggests that the allopolyploid *M. spicata* has a single origin and that the maternal parent is likely *M. longifolia*. Consistent with the results of the plastid analyses, the admixture results suggest that although many specimens of *M. spicata* and *M.* ×*rotundifolia* stem from the same parental gene pools, they appear to have different evolutionary histories (Figure 4B; Appendix S4, Figure S9).

Differences in the frequency of sexual reproduction can drive genomic divergence. While the fertility of *M.* ×*rotundifolia* is reduced, *M. spicata* is usually completely fertile (Harley and Brighton, 1977). It is therefore possible that *M.* ×*rotundifolia* to a higher degree propagates clonally via rhizomes. Ploidy differences can also drive genomic divergence, and we suggest that ploidy differences and human selection on *M. spicata* in cultivation (Harley and Brighton, 1977) have caused shifts in allele frequencies between *M. spicata* and *M.* ×*rotundifolia*, resulting in the observed differences in genomic history.

Gene flow between ploidy levels

Our genomic analyses suggest that there are two genomic, and possibly morphological, groups of *M. spicata*, but the genetic distinction between them is small (4% of the genetic variation is explained by PC2; Figures 2, 4). With our samples (Appendix S1), it is difficult to deduce geographic patterns that could explain the two clusters. However, it appears that the *M. spicata* group closest to *M. longifolia* (spi2) has a limited distribution restricted to the most southern parts of Sweden, whereas the more distant spi1 group has a larger geographic spread within Sweden (Figure 4A; Appendix S1). However, preliminary results from a larger morphology-based revision of herbarium material (T. Tyler, personal observations) show that both groups are widely distributed in Europe and America, although in regionally varying frequencies. It is also possible that the two groups of *M. spicata* represent groups of cultivated versus naturalized or native samples.

Our genomic analyses suggest that, despite the hypothesized difference in ploidy, there might be gene flow between *M. longifolia* and at least part of the *M. spicata* complex (spi2; Figure 4, Table 1). Gene flow between ploidal levels has been described in other plant systems (e.g., Wang et al, 2020; Schmickl and Yant, 2021; Cao et al., 2023) and

thus might be possible for mints as well. However, chromosome counts range between 36 and 72 for *M. spicata* (Sobti, 1965; Ahmad et al., 1992; Chambers and Hummer, 1994), and although diploid ($2n = 24$) counts dominate, there are also reports of polyploid ($2n = 36/48$) *M. longifolia* (Sobti, 1965; Harley and Brighton, 1977; Chambers and Hummer, 1994). Hence, it is possible that the observed genetic admixture between *M. longifolia* and part of the *M. spicata* complex is driven by interfertility between subpopulations of the two taxa with the same ploidy levels. It is also possible that there are additional genetic contributions from other species in complex three-way hybridization patterns, e.g., involving the widely cultivated *M.* ×*piperita* (*M. aquatica* ×*spicata*; Figure 1C), or that subgenomic bias, including preferential loss of one subgenome (Feldman et al., 2012; Bird et al., 2018) in some individuals of the polyploid *M. spicata* contributes to some of the observed patterns.

The emergence of the fertile hybrid *M. spicata* possibly created an additional genomic bridge between the parental diploids *M. longifolia* and *M. suaveolens* with the potential for increased gene flow (McDonald et al., 2008; Sigel, 2016). Cultivation and recurring naturalizations of *M. spicata* in the ranges of its parental taxa have likely increased the frequency of secondary contacts between these closely related taxa and thus potentially enhanced their genomic interconnectivity. However, future large-scale population studies in mints are needed to firmly show whether there is gene flow between other taxa of this system and whether this gene flow can lead to, for example, genomic swamping (Ellstrand and Schierenbeck, 2000; Pinto et al., 2005; Todesco et al., 2016; Beninde et al., 2018; Quilodrán et al., 2020). The implications of our study are 2-fold. First, fertile cultivars (artificially induced or not) that have escaped from cultivation into more natural habitats or are cultivated near natural populations, enabling cross-pollination, can greatly change the make-up of natural populations and is especially apparent when there have been recurring naturalizations. Second, as climates change, we are likely to observe large-scale range-shifts of many species bringing previously isolated species into contact, similar to what is observed following naturalizations. Although ploidy differences can be an efficient reproductive barrier, we here showed that might not be the case for all plant systems.

Taxonomic consequences

The current taxonomy in the genus *Mentha* is based on morphology with some input from cytology (Harley and Brighton, 1977; Tucker et al., 1980; Tucker and Naczi, 2007; Vining et al., 2020). Although we here showed that *M. suaveolens*, *M. longifolia*, *M.* ×*rotundifolia*, and *M. spicata* are not completely genetically separated, we found that for the most part, the classically recognized taxa, as re-circumscribed by our morphometric analyses, are supported (Figures 1, 2, 4). Hence, the classical taxonomic units are still useful, although we suggest that an in-depth taxonomic revision be undertaken using more taxa. Ongoing cultivation and garden escapes of various mint species, hybrids, and varieties, including *M. spicata*, will continue to feed the native gene pools. It is therefore likely that there will continue to be dissociations between morphological and genomic clustering of specimens, and future taxonomists may thus have to deal with even more complex patterns than the historic ones, but merging all taxa that are close genetic relatives and exchange genes would result in extremely broadly defined species with very low informative value.

Here, we have primarily focused on old herbarium specimens; hence, we cannot be certain that we have included any of the currently widely cultivated varieties of *M. spicata*. These varieties could possibly be genomically distinct. Contradictory conclusions between studies (e.g. Gobert et al., 2002; Bunsawatt et al., 2004; Panjeshahin et al., 2018; Heylen et al., 2021) can similarly be explained with biased inclusions of particular specimens and/or inconsistent delimitation of taxa. Here we found further support for the importance of including multiple samples per taxonomic unit, especially for taxa with frequent interfertility (Heylen et al., 2021).

CONCLUSIONS

We showed that the introduction of cultivated mint taxa into the range of close relatives increased contacts between species including wild species with the potential to increase gene flow among taxa. Although we found that classically recognized taxa of mints can mostly be recognized morphologically, our results indicate that there are incomplete reproductive barriers between the studied mint species and that there is genetic admixture between species of different ploidy. Furthermore, genetic variants associated with the cultivated morphology continue to segregate within and among populations. The observed gene flow between species with different ploidy levels suggest that, despite the often-observed speciation effect of alloploidization, it can also create additional genomic bridges the outcomes of which are uncertain but, in some cases, might promote genetic homogenization and loss of biodiversity. We hypothesize that secondary contact between previously isolated, but closely related, allopolyploids and diploids will increase as the climate changes and can promote naturalizations and range shifts.

AUTHOR CONTRIBUTIONS

J.K.O., T.T., and A.J.H. designed the study. J.K.O. and T.T. selected samples. T.T. produced and analyzed the morphological data. J.K.O. produced and analyzed the genomic data. L.T.D. analyzed transcriptomes. T.T., M.H., and A.R. provided samples. J.K.O. wrote the paper with input from all co-authors. All co-authors have read and approved the final version of the paper.

ACKNOWLEDGMENTS

This study was supported by the European Research Council (grant MSCA-IF-2017-795018). L.T.D. is supported by a Natural Environment Research Council Independent Research Fellowship (NE/T011025/1). The authors thank Ida Broman Nielsen (I.B.N.) for help with the preparation of sequencing libraries and Fabio Pedercini for help with preparing the maps in Figure 1. I.B.N. is supported by the Danish National Research Foundation DNRF143. We thank the anonymous reviewers for their help improving our manuscript.

DATA AVAILABILITY STATEMENT

Raw sequencing reads are available in the Short Read Archive under BioProject accession number PRJNA936798 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA936798). All scripts describing the analyses of this work and additional data files are available on GitHub (https://github.com/jill-olofsson/GENOMINT).

ORCID

Jill K. Olofsson http://orcid.org/0000-0002-9527-6573
Torbjörn Tyler http://orcid.org/0000-0002-7886-7603
Luke T. Dunning http://orcid.org/0000-0002-4776-9568
Anders J. Hansen http://orcid.org/0000-0002-1890-2702

REFERENCES

Abbott, R. J. 1992. Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends in Ecology and Evolution* 7: 401–405.

Abbott, R., D. Albach, A. Ansell, J. W. Arntzen, S. J. E Baird, J. Boughman, C. A. Brelsford, et al. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229–246.

Ahmad, T., B. R. Tyagi, S. S. Raghuvanshi, and J. R. Bahl. 1992. Variation of nuclear DNA content in the genus *Mentha* L. (Lamiaceae). *Cytologia* 57: 359–367.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Anderson, E., and G. L. Stebbins 1954. Hybridization as an evolutionary stimulus. *Evolution* 8: 378–388.

Askary, M., S. M. Talebi, F. Amini, and A. D. B Angan. 2016. Effects of stress on foliar trichomes plasticity in *Mentha piperita*. *Nusantara Bioscience* 8: 2–38.

Beninde, J., S. Feldmeier, M. Veith, and A. Hochkirch 2018. Admixture of hybrid swarms of native and introduced lizards in cities is determined by the cityscape structure and invasion history. *Proceedings of the Royal Society, B, Biological Sciences* 285: 20180143.

Bieker, V. C., and M. D. Martin. 2018. Implications and future prospects for evolutionary analyses of DNA in historical herbarium collections. *Botany Letters* 165: 409–418.

Bird, K. A., R. VanBuren, J. R. Puzey, and P. P. Edger. 2018. The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist* 220: 87–93.

Boatright, J., F. Negre, X. Chen, C. M. Kish, B. Wood, G. Peel, I. Orlova, et al. 2004. Understanding in vivo benzenoid metabolism in petunia petal tissue. *Plant Physiology* 135: 1993–2011.

Bunsawatt, J., N. E. Elliott, K. L. Hertweck, E. Sproles, and L. A. Alice. 2004. Phylogenetics of *Mentha* (Lamiaceae): evidence from chloroplast DNA sequences. *Systematic Botany* 29: 959–964.

Cao, Y., K. Zhao, J. Xu, and Z. Xiong. 2023. Genome balance and dosage effect drive allopolyploid formation in *Brassica*. *Proceedings of the National Academy of Sciences, USA* 120: e2217672120.

Carøe, C., S. Gopalakrishna, L. Vinner, S. S. T. Mak, M. H. S. Sinding, J. A. Samaniego, N. Wales, et al. 2018. Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution* 9: 410–419.

Chalvin, C., S. Drevensek, M. Dron, A. Bendahmane, and A. Boualem. 2020. Genetic control of glandular trichome development. *Trends in Plant Science* 25: 477–487.

Chambers, H. L., and K. E. Hummer. 1994. Chromosome counts in the *Mentha* collection at the USDA-ARS National Clonal Germplasm Repository. *Taxon* 43: 423-432.

Chopra, D., M. Mapar, L. Stephan, M. C. Albani, A. Deneer, G. Coupland, E.-M. Willing, et al. 2019. Genetic and molecular analysis of trichome development in *Arabis alpina*. *Proceedings of the National Academy of Sciences, USA* 116: 12078–12083.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* 23: 3133–3157.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, et al. 2021. Twelve years of SAMtools and BCFtools, *GigaScience* 10: giab008.

Dawson, W., D. Moser, M. van Kleunen, H. Kreft, J. Pergl, P. Pysek, P. Weigelt, et al. 2017. Global hotspots and correlates of alien species richness across taxonomic groups. *Nature Ecology and Evolution* 1: 0186.

Dunning, L. T., J. K. Olofsson, A. S. T. Papadopulos, S. G. S. Hibdige, O. Hidalgo, I. J. Leitch, P. C. Baleeiro, et al. 2022. Hybridization and chloroplast capture between distinct *Themeda triandra* lineages in Australia. *Molecular Ecology* 33: 5846–5860.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* 28: 2239–2252.

Ellstrand, N. C., and K. A. Schierenbeck. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences* 97: 7043–7050.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14: 2611–2620.

Fahn, A. 1979. Secretory tissues in plants. Academic Press, NY, NY, USA.

Feldman, M., A. A. Levy, T. Fahim, and A. Korol. 2012. Genomic asymmetry in allopolyploid plants: wheat as a model. *Journal of Experimental Botany* 63: 5045–5059.

Figueroa-Pérez, M. G., R. Reymoso-Camacho, L. F. Garcia-Ortega, and R. C. Guevara-Conzález. 2019. Transcriptome profiling of peppermint (*Mentha piperita*) with improved antioxidant properties in response to salicylic acid elicitation. *Journal of Plant Biochemistry and Biotechnology* 28: 301–311.

Friendly, M., and J. Fox. 2021. candisc: Visualizing generalized canonical discriminant and canonical correlation analysis. R package version 0.8-6. Website: https://CRAN.R-project.org/package=heplots

Fuchs, E. J., A. M. Martínez, A. Calvo, M. Muñoz, and G. Arrieta-Espinoza. 2016. Genetic diversity in *Oryza glumaepatula* wild rice populations in Costa Rica and possible gene flow from *O. sativa*. *PeerJ* 4: e1875.

Gobert, V., S. Moja, M. Colson, and P. Taberlet. 2002. Hybridization in the section *Mentha* (Lamiaceae) inferred from AFLP makers. *American Journal of Botany* 89: 2017–2023.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adoconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.

Grant, V. 1981. Plant speciation. Columbia University Press, NY, NY USA.

Green, R. G., J. Krause, A. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, et al. 2010. A draft sequence of the Neandertal genome. *Science* 238: 710–722.

Harley, R. M., and C. A. Brighton. 1977. Chromosome numbers in the genus *Mentha* L. *Botanical Journal of the Linnean Society* 74: 71–96.

Heylen, O. C. G., N. Debortoli, J. Marescaux, and J. K. Olofsson. 2021. A revised phylogeny of the *Mentha spicata* clade reveals cryptic species. *Plants* 10: 819.

Holeski, L. M., R. Chase-Aline, and J. K. Kelly. 2010. The genetics of phenotypic plasticity in plant defense: tricome production in Mimulus guttatus. *The American Naturalist* 175: 391-400.

Jia, P., H. Liu, T. Gao, and H. Xin. 2013. Glandular trichomes and essential oil of *Thymus quinquecostatus*. *Scientific World Journal* 2013: 387952.

Jin, J., D. Panicker, Q. Wang, M. J. Kim, J. Liu, J.-L. Yin, L. Wong, et al. 2014. Next generation sequencing unravels the biosynthetic ability of spearmint (*Mentha spicata*) peltate glandular trichomes through comparative transcriptomics. *BMC Plant Biology* 14: 293.

Karlsson, S., O. H. Diserud, P. Fiske, K. Hindar, and H. E. W. S. Grant. 2016. Widespread genetic introgression of escaped farmed Atlantic salmon in wild salmon populations. *ICES Journal of Marine Science* 73: 2488–2498.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15: 1179-1191.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15: 356.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie2. *Nature Methods* 9: 357–359.

Langmead, B., C. Wilks, V. Antonescu, and R. Charles. 2019. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 35: 421–432.

Li, B., P. Cantino, R. Olmstead, G. L. C. Bramley, C.-L. Xiang, Z.-H. Ma, Y.-H. Tan, and D.-X. Zhang. 2016. A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. *Scientific Reports* 6: 34343.

Maffei, M., M. Gallino, and T. Sacco. 1986. Glandular trichomes and essential oils of developing leaves in *Mentha viridis lavanduliodora*. *Planta Medica* 1986: 163–246.

Mallet, J. 2007. Hybrid speciation. *Nature* 446: 279–283.

Manni, M., M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov. 2021a. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* 38: 4647–4654.

Manni, M., M. R. Berkeley, M. Seppey, and E. M. Zdobnov. 2021b. BUSCO: Assessing genomic data quality and beyond. *Current Protocols* 1: e323.

McCaskill, D., and R. Croteau. 1995. Monoterpene and sesquiterpene biosynthesis in glandular trichomes of peppermint (*Mentha* × *piperita*) rely exclusively on plastid-derived isopentenyl diphosphate. *Planta* 197: 49–56.

McDonald, D. B., T. L. Parchman, M. R. Bower, W. A. Hubert, and F. R. Rahel. 2008. An introduced and a native vertebrate hybridize to form a genetic bridge to a second native species. *Proceedings of the National Academy of Sciences, USA* 105: 10837–10842.

Meirmans, P. G., and P. W. Hedrick. 2011. Assessing population structure: *F*ST and related measures, *Molecular Ecology Resources* 11: 5–18.

Mishra, A., R. K. Lal, C. S. Chanotiya, and S. S. Dhawan. 2017. Genetic elaborations of glandular and non-glandular trichomes in Mentha arvensis genotypes: assessing genotypic and phenotypic correlations along with gene expressions. *Protoplasma* 254: 1045–1061.

Meisner, J., and A. Albrechtsen. 2018. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* 210: 719–731.

Mooney, H. A., and E. E. Cleland. 2001. The evolutionary impact of invasive species, *Proceedings of the National Academy of Sciences, USA* 98: 5446–5451.

Morjan, C. L. and L. H. Rieseberg. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* 13: 1341–1356.

Munguía-Rosas, M. A., M. E. Jácome-Flores, R. Bello-Bedoy, V. Solís-Montero, and E. Ochoa-Estrada (2019). Morphological divergence between wild and cultivated chaya (*Cnidoscolus aconitifolius*) (Mill.) I.M. Johnst. *Genetic Resources and Crop Evolution* 66: 1389–1398.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Olofsson, J. K., M. Bianconi, G. Besnard, L. T. Dunning, M. R. Lundgren, H. Holota, M. S. Vorontsova, et al. 2016. Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. *Molecular Ecology* 25: 6107-6123.

Orlova, I., A. Marshall-Colón, J. Schnepp, B. Wood, M. Varbanova, E. Fridman, J. J. Blakeslee, et al. 2006. Reduction of benzenoid synthesis in petunia flowers reveals multiple pathways to benzoic acid and enhancement in auxin transport. *Plant Cell* 18: 3458–3475.

Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. *Annual Review in Genetics* 34: 401–437.

Owens, G. L., and K. Samuk. 2020. Adaptive introgression during environmental change can weaken reproductive isolation. *Nature Climate Change* 10: 58–62.

Panjeshahin, Z., G.-R. Sharifi-Sirchi, and D. Samsampour. 2018. Genetic and morphological diversity of wild mint "*Mentha longifolia* (L.) Hudson subsp. *noeana* (Briq.) Briq." in south and southeastern Iran. *Journal of Medicinal Plants Products* 1: 105–115.

Picazo-Aragonés, J., A. Terrab, and F. Balao. 2020. Plant volatile organic compounds evolution: transcriptional regulation, epigenetics, and polyploidy. *International Journal of Molecular Sciences* 21: 8956.

Pinto, M. A., W. L. Rubink, J. C. Patton, R. N. Coulson, and J. S. Johnston. 2005. Africanization in the United States: replacement of feral European honeybees (*Apis mellifera* L.) by an African hybrid swarm. *Genetics* 170: 1653–1665.

POWO. 2023. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Website: http://www.plantsoftheworldonline.org/ [retrieved 23 October 2023].

Quilodrán, C. S., J. A. Montoya-Burgos, and M. Currat. 2020. Harmonizing hybridization dissonance in conservation. *Communications Biology* 3: 391.

R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: https://www.R-project.org/

Renny-Byfield, S., and J. F. Wendel. 2014. Doubling down on genomes: polyploid and crop plants. *American Journal of Botany* 101: 1711–1725.

Rhymer, J. M. 1996. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* 27: 83–109.

Richardson, D. M., and P. Pyšek. 2012. Naturalization of introduced plants: ecological drivers of biogeographic patterns, *New Phytologist* 196: 383–396.

Rieseberg, L. H. 1995. The role of hybridization in evolution: old wine in new skins. *American Journal of Botany* 82: 944–953.

Rieseberg, L. H., and J. M. Burke. 2001. The biological reality of species: gene flow, selection, and collective evolution. *Taxon* 50: 47–67.

Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence, *PLoS Biology* 14: e2000234.

Salehi, B., Z. Stojanović-Radić, J. Matejić, F. Sharopov, H. Antolak, D. Kręgiel, S. Sen, et al. 2018. Plants of genus *Mentha*: from farm to food factory. *Plants* 7: 70.

Šarić-Kundalić, B., S. Fialoá, C. Dobeš, S. Ölzant, D. Tekeľová, D. Grančai, G. Reznicek, and J. Saukel. 2009. Multivariate numerical taxonomy of *Mentha* species, hybrids, varieties and ultivars. *Scientia Pharmaceutica* 77: 851–876.

Schierenbeck, K. A., and N. C. Ellstrand. 2009. Hybridization and the evolution of invasiveness in plants and other organisms. *Biological Invasions* 11: 1093.

Schmickl, R., and L. Yant. 2021. Adaptive introgression: how polyploidy reshapes gene flow landscapes. *New Phytology* 230: 457–461.

Schubert, M., S. Lindgreen, and L. Orlando. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes* 9: 88.

Sigel, E. M. 2016. Genetic and genomic aspects of hybridization in ferns. *Journal of Systematics and Evolution* 54: 638–655.

Singh, P., and A. K. Pandey. 2018. Prospective of essential oils of the genus *Mentha* as biopesticides: a review. *Frontiers in Plant Sciences* 9: 1295.

Sirikantaramas, S., S. Morimoto, Y. Shoyama, Y. Ishikawa, Y. Wada, Y. Shoyama, and F. Taura. 2004. The gene controlling marijuana psychoactivity: molecular cloning and heterologous expression of Δ^1-tetrahydrocannabinolic acid synthase from *Cannabis sativa* L. *Journal of Biological Chemistry* 279: 39767–39774.

Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195: 693–702.

Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15: 150.

Sobti, S. N. 1965. Chromosome numbers in species of *Mentha*. *Proceedings of the Indian Academy of Sciences, B* 62: 145–148.

Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. Zheng, D. Sankoff, et al. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

Soltis, P. S., D. B. Marchant, Y. Van de Peer, and D. E. Soltis. 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics and Development* 35: 119–125.

Stace, C. 2010. New flora of the British Isles, 3rd ed. Cambridge University Press, Cambridge, UK.

Stace, C. A., C. D. Preston, and D. A. Pearman. 2016. Hybrid flora of the British Isles. BSBI, Bristol, UK.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 130: 1312–1313.

Szabó, K., S. Sárosi, B. Cerháti, and A. Ferenczy. 2010. Can glandular hair density be a breeding marker for *origanum vulgare* subsp. *hirtum* with high essential oil content? *Natural Product Communications* 5: 1437–1440.

Todesco, M., M. A. Pascual, G. L. Owens, K. L. Ostevik, B. T. Moyers, S. Hübner, S. M. Heredia, et al. 2016. Hybridization and extinction. *Evolutionary Applications* 9: 892–908.

Tucker, A. O. 2012. Genetics and breeding of the genus *Mentha*: a model for other polyploid species with secondary constituents. *Journal of Medicinally Active Plants* 1: 19–29.

Tucker, A. O., and R. F. C. Naczi. 2007. *Mentha*: an overview of its classification and relationships. In B. Lawrence [ed.], Mint: the genus Mentha, 1–40. CRC Press, Boca Raton, FL, USA.

Tucker, A. O., R. M. Harley, and D. E. Fairbrothers. 1980. The Linnean types of *Mentha* (Lamiaceae). *Taxon* 29: 233–255.

Tutin, T. G. 1993. Flora europaea. Cambridge University Press, Cambridge, UK.

UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* 49: D480–D489.

van Kleunen, M., X. Xu, Q. Yang, N. Maurel, Z. Zhang, W. Dawson, F. Essl, et al. 2020. Economic use of plants is key to their naturalization success. *Nature Communications* 11: 3201.

Vining, K. J., K. E. Hummer, N. V. Bassil, B. M. Lange, C. K. Khoury, and D. Carver. 2020. Crop wild relatives as germplasm resources for cultivar improvement in mint (*Mentha* L.). *Frontiers in Plant Science* 19: 1217.

Vining, K. J., S. R. Johnson, A. Ahkami, I. Lange, A. N. Parrish, S. C. Trapp, R. B. Croteau, et al. 2017. Draft genome sequence of *Mentha longifolia* and development of resources for mint cultivar improvement. *Molecular Plant* 10: 323–339.

Wang, J., S. Dong, L. Yang, A. Harris, H. Schneider, and M. Kang. 2020. Allopolyploid speciation accompanied by gene flow in a tree fern. *Molecular Biology and Evolution* 37: 2487–2502.

Weir, B. S., and C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.

Ye, J., S. McGinnis, and T. L. Madden. 2006. BLAST: improvements for better sequence analysis. *Nucleic Acids Research* 34: W6–W9.

Yu, X., C. Liang, H. Fang, X. Qi, W. Li, and Q. Shang. 2018. Variation of trichome morphology and essential oil composition of seven *Mentha* species. *Biochemical Systematics and Ecology* 79: 30–36.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19 (Suppl 6): 153.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Sample information.

Appendix S2. Scoring of morphological traits.

Appendix S3. Values for all morphological traits.

Appendix S4. Supplementary Figures S1–S14.

Appendix S5. Results of ABBA-BABA analyses (D-statistics).

Appendix S6. BLAST results.

How to cite this article: Olofsson, J. K., T. Tyler, L. T. Dunning, M. Hjertson, Å. Rühling, and A. J. Hansen. 2024. Morphological and genetic evidence suggest gene flow among native and naturalized mint species. *American Journal of Botany* 111(1): e16280. https://doi.org/10.1002/ajb2.16280