Inter-specific gene flow following the naturalization of a cultivated mint promotes the formation of a coalescent complex

Jill Olofsson[1], Torbjörn Tyler[2], Luke Dunning[3], Mats Hjertson[4], Åke Rühling[2], and Anders Hansen[1]

[1]University of Copenhagen
[2]Lund University
[3]The University of Sheffield
[4]Uppsala Universitet

February 15, 2022

Abstract

Cultivation and naturalization of plants beyond their natural range can bring previously geographically isolated taxa together, thereby increasing the opportunity for hybridization and inter-specific gene flow, the outcomes of which are not predictable. These anthropogenic events therefore allow us to study how hybridization and inter-specific gene flow affect genetic and phenotypic diversity. Here, we explore the phenotypic and genomic effects of increased inter-specific gene flow following the re-introduction of the cultivated Mentha spicata (spearmint) into the ranges of the native mints M. longifolia and M. suaveolens. Using morphological analyses, we show that the cultivated M. spicata has altered trichome characters, which is likely a product of human imposed selection for a more palatable plant or a byproduct of selection on essential oil production. Using whole genome sequencing, we then show that there is extensive genetic admixture between the morphologically defined mint taxa that to some extent is mediated by the cultivated M. spicata. This has, at least partially, resulted in a breakdown of the species barriers. However, despite this breakdown, we find that genetic variants associated with the cultivated trichome morphology continue to segregate in cultivated, naturalized, and wild populations and we identify three genes that may function in the production of the characteristic aromatic oils of mints. Although hybridization can increase species richness by forming new hybrid taxa, we here show that unless reproductive barriers are strong it can also merge species into population/coalescent complexes over evolutionary time.

INTRODUCTION

Cultivation of plants beyond their native ranges has a long history (Hobhouse, 1992; Huxley, 1978; Reichard & White, 2001). Although we often associate this anthropogenic practice with our staple crops, hundreds of plants are currently cultivated. While cultivation provides valuable resources to humans, it can have detrimental effects on the native flora, especially when alien taxa escape their cultivated lifestyle and re-establish themselves in the wild within or outside their native ranges (Todesco et al., 2016; Quilodrán et al., 2020b). Naturalized taxa can for example be highly invasive and outcompete native species (Anderson, Galatowitsch, & Gomez, 2006; Downey & Richardson, 2016). In addition, naturalizations can also increase the opportunities for hybridization and inter-specific gene flow when previously isolated taxa are brought together (Owen et al., 2020).

Hybridization and gene flow are major drivers of speciation and trait evolution in plants (Anderson & Stebbins, 1954; Ellstrand & Schierenbeck, 2000; Morjan & Rieseberg, 2004; Otto & Whitton, 2000; Rieseberg, 1995; Rieseberg & Burke, 2001; Schierenbeck & Ellstrand, 2009; Soltis et al., 2009; Soltis, Marchant, Van de Peer, & Solitis, 2015; Stebbins, 1959; Yakimowski & Rieseberg, 2014). Gene flow between divergent

lineages can increase genetic diversity and establish novel stable hybrids that are reproductively isolated from their parental species, by for example allopolyploidization or clonal propagation, and thus can evolve into new species (Ainouche et al., 2009; Ellstrand & Schierenbeck, 2000; Grant, 1981; Mallet, 2007; Morjan & Rieseberg, 2004; Rieseberg & Burke, 2001; Schierenbeck & Ellstrand, 2009; Todesco et al., 2016; Quilodrán et al., 2020a). However, if the isolating barriers between hybrids and parental taxa are weak, continued gene flow between them can result in homogenization and ultimately genetic swamping with the formation of coalescent complexes (hybrid swarms; Beninde, Feldmeier, Veith, & Hochkirch, 2018; Ellstrand & Schierenbeck, 2000; Pinto et al., 2005; Todesco et al., 2016; Quilodrán et al., 2020a). Hybridization can therefore promote the loss of species by merging previously separated taxa (Owens & Samuk, 2020; Todesco et al., 2016; Quilodrán et al., 2020a). Here we explore the genomic and phenotypic consequences of frequent inter-specific gene flow following cultivation and naturalization using mints (*Mentha*) as our study system.

Mints are aromatic herbs with cosmopolitan distributions. Several mints are widely cultivated in many parts of the world for culinary use and for extraction of essential oils used as flavoring agents (Gobert, Moja, Colson, & Taberlet, 2002; Singh & Pandey, 2018; Tucker, 2012; Vining et al., 2020). Although *Mentha* is a relatively small genus, it is highly diverse due to inter-fertility between many taxa and numerous hybrid species are recognized (Gobert et al., 2002; Tucker et al., 1980; Tucker & Naczi, 2007). Most mint hybrids are sterile but can successfully propagate clonally using rhizomes (Harley & Brighton, 1977; Gobert et al., 2002). Exceptions to this are hybrids between the two diploid species *M. longifolia* (2n=24) and *M. suaveolens* (2n=24; Chambers & Hummer, 1994; Harley & Brighton, 1977; Sobti, 1965). In Europe, where these two taxa are sympatric, they can spontaneously hybridize forming the partly fertile hybrid complex*M. × rotundifolia* (Harley & Brighton, 1977). In addition, the widely cultivated *M. spicata* (spearmint; Singh & Pandey, 2018; Vining et al., 2020) is often described as an allopolyploid (2n=48) formed from hybridization of *M. longifolia* and *M. suaveolens* (Chambers & Hummer, 1994; Gobert et al., 2002; Harley & Brighton, 1977; Sobti, 1965; Tucker et al., 1980; Tucker & Naczi, 2007). However, the allopolyploid status of *M. spicata* has been questioned (Heylen, Debortoli, Marescaux, & Olofsson, 2021) and chromosome counts range between 36 and 72 (Ahmad, Tyagi, Raghuvanshi, & Bahl, 1992; Chambers & Hummer, 1994; Sobti, 1965). Similarly, there are reports of polyploid chromosome counts (2n=36/48) for *M. longifolia* (Chambers & Hummer, 1994; Harley & Brighton, 1977; Sobti, 1965).

Mentha spicata and its hybrid with *M. aquatica* , i.e.*M. × piperita* (peppermint), are the most widely cultivated mints (Salehi et al., 2018). The exact origin of *M. spicata* is unknown, but some studies suggest that it was formed in cultivation (Harley & Brighton, 1977; Tucker, 2012). Although *M. spicata* is considered native to most of continental Europe, the Middle East, and southern Asia, it has also spread widely from cultivated plants obscuring its natural range and population size (Gobert et al., 2002; Harley & Brighton, 1977; Vining et al., 2020). Outside of the presumably native range, *M. spicata* is currently naturalized in most geographic regions with a warm-temperate climate overlapping, and exceeding, that of the two hypothesized parental taxa (Harley & Brighton, 1977; Vining et al., 2020). Despite the reported differences in ploidy, *M. spicata* frequently forms hybrids with both *M. longifolia* and *M. suaveolens* but the offspring (*M. ×villosa-nervata* and *M. × villosa* , respectively) are believed to be sterile due to their theoretical triploid status (Chambers & Hummer, 1994; Harley & Brighton, 1977; Gobert et al., 2002). The re-establishment of cultivated *M. spicata* in its native range creates an excellent opportunity to study the phenotypic and genotypic consequences of increased opportunities for hybridization.

Here we take advantage of the biodiversity stored in herbarium collections that allows for evaluations of specimens collected over large geographic regions (Bieker & Martin, 2018). We first evaluate the morphological spaces of the cultivated *M. spicata* , the native species (*M. longifolia* and *M. suaveolens*), and their hybrids (*M. × rotundifolia*). We then use whole genome sequencing to (1) establish patterns of genetic admixture and gene flow between taxa, (2) infer the genomic histories of *M.* x*rotundifolia* and *M. spicata* , and (3) test the hypothesis that increased opportunities for inter-specific gene flow following the naturalization of the cultivated *M. spicata* has laid the foundation for the formation of a coalescent complex.

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); Table S1. Specimens were selected based on the plant identification by the collector(s) and/or subsequent re-identifications by botanists, with the aim to include the taxa, *M. longifolia* , *M. suaveolens* ,*M. spicata* , and the hybrid *M. × rotundifolia*(*M. longifolia × suaveolens*). However, many specimens were re-identified based on the result of our morphometric analyses (see below). From here on, the taxon 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 (Stace, 1993; Tutin, 2010). However, it should be stressed that these morphologically defined groups may not exactly correspond to the biological species that previous authors have referred to by these names, nor to their nomenclatural types. All specimens were morphologically evaluated and a subset of 93 samples were whole genome sequenced.

Morphological characters and statistical analyses

A total of 34 morphological characters were selected for the analyses and scored for all specimens (Table S2). Characters were selected to cover all parts of the plant generally available on herbarium specimens (stem, leaves, calyces, flowers, and sexual organs) and care was taken to avoid intrinsically dependent characters as well as characters directly related to the size/vigour of the plant. Indumentum characters are widely used, and generally considered important, in *Mentha*taxonomy (Šarić-Kundalić et al., 2009; Yu et al., 2018) and ten characters referring to various types of trichomes were therefore included (Table S2). All characters were treated as continuous traits, although some were scored as multiple categories (Table S2). Cases of missing data were treated by mean value imputations. The morphometric data was analyzed using Principal Component Analyses (PCA) conducted in R v.3.6.3 (R Core Team, 2021) and sub-clusters of specimens were visually identified. To illustrate and assess the importance of indumentum characters, separate PCAs were conducted first excluding all characters referring to the density or structure of trichomes and secondly only including the characters of indumentum (Table S2). In addition, detailed analyses of the distributions of indumentum characters between different taxa were conducted.

DNA extraction, library construction, and Illumina sequencing

A total of 93 specimens were whole genome sequenced at the Geogenetics Sequencing Core (GLOBE Institute, University of Copenhagen, Denmark). Total genomic DNA was extracted from approximately a 1–3 cm^2 piece of dried (herbarium mounted) material from a single leaf. In cases of small leaves, a single 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 manufacturer's protocol. The amount of extracted DNA was estimated using the Qubit double-stranded DNA High-Sensitivity Assay kit (ThermoFisher Scientific, USA).

A total of 4.6–110 ng DNA was fragmented using the 96 plate set-up in a Covaris LE220-plus (Covaris, Woburn, USA) to a target insert size of 300 bp and 32 µl of 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 quantitative PCR (Mx3005; Agilent Technologies) and libraries were amplified (25 µl reaction volume) with the desired number of PCR cycles (Table 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 (100bp or 150 bp; Table S1) on one lane of Illumina NovaSeq6000. For 24 samples additional sequencing was performed (Table S1). Libraries were re-built as above and pooled in equimolar concentrations and paired-end sequenced (150 bp) on one lane of Illumina NovaSeq6000.

For 24 samples (Table S1) single indexed libraries were built using the NEBNext Ultra II FS DNA Library Prep Kit (Illumina) adjusting the fragmentation time to 25 min for a target insert size of 160 bp and PCR amplified with 12 cycles. The resulting libraries were pooled in equimolar concentrations and single-end

sequenced (80 bp) on one lane of Illumina HiSeq4000 (Table S1).

Raw reads were quality filtered in AdaptorRemoval v.2.3.1 (Schubert, Lindgreen, & Orlando, 2016) removing adaptor sequences, ambiguous bases (N), and consecutive bases with low quality (Q<20) from both 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. Finally overlapping pair-end reads were collapsed and all paired-end reads without a mate were discarded.

Chloroplast alignment and phylogenetics

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 & Salzberg, 2012; Langmead, Wilks Antonescu, & Charles, 2019) with default settings. Mapped reads were quality filtered using SAMtools v.1.9 (Danecek et al., 2021) removing all reads mapping with a mapping quality score below 20 and 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 & 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* an outgroup (Li et al., 2016; NCBI Genebank accessions: NC_050898, MN520020, MN520021, NC_038165, MG772529, MN520019, NC_050897, MN520023, and NC_050900).

Genomic clustering and admixture analyses

The quality filtered and trimmed reads were mapped to the *M. longifolia* reference genome (GCA_001642375.1_Mlong1.0; Vining et al., 2017) using bowtie2 with default settings. Mapped reads were quality filtered as outlined above and reads previously identified to map to the plastome were removed. The quality filtered mapping files were used to call single nucleotide polymorphisms (SNPs) in angsd v.0.931 (Korneliussen, Albrechtsen, & Nielsen, 2014). Only SNPs with less than 50% missing data across all individuals were retained and maximum depth to call a SNP within an individual was set to 50. SNPs were called using the SAMtools option and only genotypes with a posterior probability above 0.95 were used in downstream analyses. Genomic clusters were evaluated from the nuclear SNPs in PCangsd v.0.95 (Meisner & Albrechtsen, 2018) with default settings. The resulting covariance matrices were converted to eigen-values and visualized in R. Genomic admixture was evaluated using the nuclear SNPs in NGSadmix v3.2 (Skotte, Korneliussen, & Albrechtsen, 2013) with K increasing from 1 to 10. The best fit K was evaluated using the method of Evanno, Regnaut, and Goudet (2005) implemented in CLUMPAK (Kopelman et al., 2015) using ten independent runs of NGSasmix for each K .

Estimating ploidy

Since the study was based on herbarium vouchers, chromosome counting or genome size estimates by flow cytometry were not possible. Instead, we bioinformatically estimated the ploidy of 32 specimens for which we obtained an estimated nuclear genome coverage of at least three following mapping. Data for each specimen were merged and used to estimate ploidy in HMMploidy, a program that combines information of sequencing depth and genotype likelihoods to estimate ploidy levels (Soraggi et al., 2021). A multi-sample mpileup file was generated in SAMtools v.1.10 for all genome scaffolds longer than 10kb using only reads with a minimum mapping quality of 30 and only calling sites with a minimum quality of 30, counting anomalous reads, and setting the maximum per-file depth to 50. Genotype likelihoods were then calculated in HMMPloidy using default settings and ploidy levels were inferred in 10kb windows (total of 5,224 windows). The percentage of 10kb windows supporting each ploidy level (1n-4n; no windows supported a ploidy level larger than 4n) were calculated. Specimens for which at least 60% of the windows supported a single ploidy were assigned to one of four categories: 1n="likely diploid", 2n="diploid", 3n="likely polyploid", or 4n="polyploid". The estimates for specimens not assigned to any of these categories were considered uncertain.

F_{ST}-outlier detection, transcriptome analyses, and evaluations of site frequencies spectra

Genetic distances between groups of specimens were evaluated by means of pairwise F_{ST} (Weir & Cockerham, 1984). The pairwise F_{ST} was calculated in VCFtools v.0.1.16 (Danecek et al., 2011) from all SNPs. Negative values ofF_{ST} were converted to zero and the meanF_{ST} value for each contig/scaffold of the reference genome was calculated as well as the mean and medianF_{ST} across all SNPs. 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.

Transcriptomes for *M. × piperita* (Figueroa-Pérez, Reymoso-Camacho, Garcia-Ortega, & Guevara-Conzález, 2018) and *M. spicata* (Jin et al., 2014) were downloaded and re-assembled. 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 base pairs 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) with default parameters. The sequences of F_{ST}-outlier contigs/scaffolds were extracted and used as databases for blast-searches of all re-assembled mint transcripts with default settings in BLAST v.2.9.0 (Altschul, Gish, Miller, Myers, & Lipman, 1990; Ye,, McGinnis, & Madden, 2006). Transcripts with top blast-hits longer than 300 bp and with an e-value below 1e-5 were extracted and annotations by extracting the top tblastx-hit (>100 amino acids and e-value < 1e-5) to the UniProt database (The UniProt Consortium, 2021).

Folded site frequency spectra (SFS) were finally calculated in angsd for each morphologically defined group of specimens (see Results) and all SNPs used in the genomic cluster and admixture analyses.

RESULTS

Trichome characters distinguish morphological groups

A PCA of all assessed morphological characters (Tables S2 and S3) separates the specimens rather well. The first component (PC1) separated the cultivated *M. spicata* from the other mints, while PC2 separates the two native mints (*M. longifolia* and *M. suaveolens*) with the hybrid *M. × rotundifolia* (*M. longifolia × suaveolens*) falling intermediate (Figure 1a). However, the separation between *M. longifolia* and *M. ×rotundifolia* is not very clear (Figure 1a). Five specimens were identified as *M. × villosa* (*M. suaveolens ×spicata*), one as *M. × villosa-nervata* (*M. longifolia × spicata*), and one as *M. capensis* , a South-African close relative of the Eurasian*M.longifolia* /*M.spicata* (Figure 1). Although the clustering of specimens in the morphological space is overall consistent with current taxonomic assignment, the distinction of *M. spicata* is largely driven by trichome characters and when these are removed there is a much more cloud-like projection (Figure 1b). There is however, a tendency for the taxonomic units to roughly separate on the first PC with *M. × rotundifolia* intermediate to *M. longifolia* and *M. suaveolens* (Figure 1b). The importance of trichome morphology is particularly evident when only characters of indumentum are considered which results in a projection of specimens resembling that of the overall analysis and where the first PC explains as much as 53.3% of the variation among specimens (Figure 1c).

Distribution analyses of the indumentum characters show that the two wild taxa have large variations in characters of indumentum, with*M. longifolia* generally being more 'hairy' than *M. suaveolens* , and branched trichomes are almost exclusively found in the latter (Figure S1 and S2). However, while the hybrid *M. × rotundifolia* show large variation in indumentum with distributions mostly resembling that of *M. longifolia* , the widely cultivated*M. spicata* shows less variation and is less hairy, mostly with distributions outside of the ranges of the native taxa (Figures S1 and S2).

Genomic clusters represent different admixture histories

In accordance with previous results, *Mentha* is resolved as monophyletic in relation to *Salvia* in the plastome phylogeny (Figure S3a; Li et al., 2016). However, the taxa within *Mentha*are polyphyletic although *M. longifolia* and *M. suaveolens*are mostly separated into different lineages (Figure S3b). Specimens identified as the hybrid *M. × rotundifolia* are found intermingled with both *M. longifolia* and *M. suaveolens*(Figure S3b).

In contrast, with the exception of a single specimen, all specimens identified as *M. spicata* are monophyletic and found together with a subset of the *M. longifolia* specimens (Figure S3b).

Genomic clustering was performed by means of PCA, and similar to the morphological analyses there is a separation of *M. longifolia*(long) and *M. suaveolens* (sua) on the first PC (11%; Figure 2a). Intermediate to these taxa there are three relatively distinct groups of specimens that are consistent with the hybrid *M. ×rotundifolia* although a few samples of *M. longifolia* are also found in these clusters (rot1-3; Figure 2a). The genetic distinction between *M. longifolia* and *M. spicata* is more elusive with specimens only somewhat separating along the second PC (4%; Figure 2a). Some specimens mostly morphologically identified as*M. spicata* (spi1) do however, form a separate cluster on the positive extremes of the second PC (spi1; Figure 2a).

Overall, the admixture analysis reflects the genomic clustering with the best-fit number of population groups (K) being two with other smaller optima at four and six (Figures 2 and S4). The specimens separating on the first genomic PC form five groups with distinct admixture profiles (Figure 2b). Based on the morphology of these specimens the three admixed groups genomically resemble F1 crosses between *M. longifolia* and *M. suaveolens* (*M. ×rotundifolia*) as well as backcrosses to either parent (Figure 2b). Most specimens morphologically recognized as *M. spicata* are also admixed (Figures 2b). However, when the number of populations (K) is increased it is clear that the admixture histories of*M. × rotundifolia* and *M. spicata* are different (Figure 2b and S4b). When three respective four populations are considered, the two *M. spicata* clusters (spi1 and spi2; Figure 2a) forms separate populations but a few specimens morphologically identified as *M. spicata* still appear admixed (Figure 2b). The specimens with more central positions in the genomic PCA show admixture proportions consistent with mixing of all gene pools (Figures 2 and S4). However, our admixture analyses cannot decisively exclude that some patterns are caused by admixture with other species not included in our study.

Although, there is an overall agreement between the morphological and genomic analyses there are also a few cases of miss-matches. In particular a few specimens morphologically identified as *M. ×rotundifolia* are found intermingled with other genomic groups, as well as intermediate to the two groups of *M. spicata* (rot4; Figure 2a) and these specimens show admixture proportions consistent with their genomic cluster assignments (Figure 2b). In addition, a few specimens morphologically identified as *M. spicata* are genomically indistinguishable from *M. longifolia* and a few specimens morphologically identified as *M. longifolia* were found to be admixed (Figure 2b).

Estimations of ploidy

The differences in genomic histories between *M. spicata* and*M. × rotundifolia* can potentially be explained by differences in ploidy (Chambers & Hummer, 1994; Harley & Brighton, 1977). The ploidy estimates using HMMploidy were variable and somewhat consistent with previously reported chromosome counts. Although our estimates are uncertain, they suggest that specimens morphologically assigned as *M. spicata* are polyploids and those assigned as*M. suaveolens* are diploids (Table 1). The estimates for*M.longifolia* and *M. × rotundifolia* are variable (Table 1). Generally, specimens morphologically assigned as *M. longifolia*and not inferred as admixed were estimated to be diploids whereas specimens morphologically assigned as *M. × rotundifolia* were estimated to be "likely polyploids" or "polyploids" (Table 1).

Low levels of genetic differentiation and shared polymorphisms

To understand the level of genetic differentiation between morphologically defined taxa and genomic clusters, analyses of pairwiseF_{ST} were performed between groups of specimens. Overall, there is low genetic differentiation between both morphological and genomic clusters with mean scaffold F_{ST}values well below 0.2 (Table 2). A large proportion of scaffolds show no differentiation (F_{ST}=0), but a few scaffolds show high F_{ST}-values (>0.5; Table 2 and Figures S5 and S6). Although *M. × rotundifolia* and*M. spicata* are overall more distant to *M. suaveolens* than to *M. longifolia* (Table 2), there is variation inF_{ST} between the different genomic groups and the respective parental taxa (Table 2). In particular, the genomic cluster spi2 is very similar to *M. longifolia* (Table 2). Despite the detected difference in admixture history between *M.*

spicata and*M. × rotundifolia* (Figure 2b) the genetic distance between them is small (Table 2).

Given the small genetic differences between groups of specimens, analyses of shared polymorphisms were conducted. Overall, there is a large overlap in the polymorphic sites between groups of specimens and*M. spicata* or *M. × rotundifolia* show no SNPs not found within either *M. longifolia* or *M. suaveolens* (Figure 3a). Despite the overall sharing of polymorphic sites, the site frequency spectra (SFS) for each morphological group do show differences in allele frequencies (Figure 3b). In particular, the SFS of *M. spicata* and *M. × rotundifolia* are shifted to higher proportions of sites with a higher frequency of the alternative allele (Figures 3b). However, a very low number of sites show complete fixation of an alternative allele (Figures 3b). This is not surprising for*M. longifolia* as such sites would be private SNPs in the reference genome. However, the low proportion of fixed differences in relation on the reference genome for the other analyzed taxa further supports their close relationships.

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

Although there overall are low levels of genetic differentiation between the morphological and genomic groups of specimens, there are 157 scaffolds with high F_{ST} between groups of specimens (top 1% outliers) potentially indicating signs of divergent selection. A total of 1,846 *Mentha* transcripts (Figueroa-Pérez et al., 2018; Jin et al., 2014) have a top BLAST match longer than 300 bp with an e-value less than 1e-5 to one of the 1%F_{ST}-outlier scaffolds. Of these, 534 transcripts have a top BLAST match longer than 100 and an e-value less than 1e-5 to a total of 116 different entries in UniProt (The UniProt Consortium, 2021), and could thus be annotated. The vast majority (82; 71%) were matches to a plant record and most of these have a described function (Table S4). Although the functions are diverse, nine transcripts were annotated to have functions relating to reproduction/fertility, and in particular to pollen recognition (Table S4). In addition, there were eight transcripts annotated to functions relating to pathogen response, eight had functions related to abiotic stress response, and there were five annotations involving DNA recombination/transposition (Table S4). Of particular interest are the three transcripts with functions in the biosynthesis of volatile compounds (VOCs; Table S4) which potentially are involved in the production of the characteristic aromatic compounds of mints.

DISCUSSION

Morphological characters associated with cultivation continues to segregate despite genomic admixture

Despite low levels of genetic differentiation, there are clear morphological separations between specimens identifiable as *Mentha longifolia* , *M. suaveolens* , and *M. spicata* (Figures 1 and 2 and Table 2). In particular, *M. spicata* shows a distinct morphology driven by a shift in indumentum toward fewer and shorter trichomes (Figures 1, S1, and S2). *Mentha spicata* is widely cultivated for culinary use and it is therefore possible that the reduction in characters of indumentum is associated with human selection for a more palatable plant (less hairy; Harley & Brighton, 1977; Munguía-Rosas, Jácome-Flores, Bello-Bedoy, Solís-Montero, & Ochoa-Estrada, 2019), but not all cultivated mints have smooth stems and leaves. The glands associated with trichomes are however, the main organs for production of the desired essential mint oils (Fahn, 1979; Jia, Liu, Gao, & Xin, 2013; McCaskill & Croteau, 1995; Mishra, Lal, Chanotiya, & Dhawan, 2017;Yu et al., 2018). It is therefore possible that the shift in trichome characters is a byproduct of selection for essential oil production in cultivation (Maffei, Gallino, & Sacco, 1986; Mishra et al., 2017; Jia et al., 2013, Szabó, Sárosi, Cserháti, & Ferenczy, 2010). The lack of previously identified genes associated with trichome function and morphology among F_{ST}outlier scaffolds (Table S4) is likely a result of the polygenic nature of these characters (Chalvin, Drevensek, Dron, Bendahmane, & Boualem, 2020; Chopra et al., 2019, Figueroa-Pérez et al., 2019; Mishra et al., 2017) coupled with the large sharing of polymorphic sites between*M. longifolia* and *M. spicata* (Figure 3). Our sample size is therefore too small to detect significant genome wide associations (Hong & Park, 2012). However, three genes; Phospho-2-dehydro-3-deoxyheptonate aldolase 2 (DAHP2; Langer et al., 2014), Benzyl alcohol O-benzoyltransferase (BEBT1; Boatright et al., 2004; Orlova et al., 2006),

and Tetrahydrocannabinolic acid synthase (THCAS; Sirikantaramas et al., 2004), involved in the production of volatile organic compounds, were found among theF_{ST}-outlier scaffold (Table S4), indicating that there indeed is high sequence divergence between different mints for some genes potentially involved in the production of their characteristic aromatic oils.

Despite the overall genetic similarities between specimens of *M. spicata* and *M. longifolia* only a few individuals morphologically identified as one taxon genetically cluster with the opposite species (Figure 2a). However, these few cases do suggest that genetic variants associated with human desirable traits can be lost despite most of the genome otherwise being of a *M. spicata* origin. More importantly, mismatches between morphology and genomic clustering suggest that the 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). In addition, some specimens of *M. × rotundifolia* show admixture profiles and genomic clustering that resemble *M. spicata* (rot4 and spi1, Figure 2). The morphological clustering of these specimens suggest that they likely represent cases where gene flow between *M. spicata* and *M. longifolia* has resulted in the loss of genetic variants associated with the cultivated morphology and that their morphology resemble *M. × rotundifolia* due to a retention of polymorphisms from *M. suaveolens* . Introgression between *M. spicata* and *M. longifolia* has previously been suggested (Harley & Brighton, 1977), but never been evaluated using multiple specimens per species collected across large geographic ranges. Here we show the genetic admixture between these two species indeed is frequent. However, despite this it is clear that genetic variants associated with the cultivated morphology continue to segregate in both cultivated and wild populations.

Independent genomic histories despite admixture from the same gene pools

Although both *M. spicata* and *M. × rotundifolia* are generally believed to stem from the same parental gene pools, our admixture analyses suggest they have independent histories (Figure 2b). While chromosome counts for *M. spicata* suggest that it is a polyploid, reports of chromosome counts for *M. × rotundifolia* are rare but do point to a diploid status (Chambers & Hummer, 1994; Harley & Brighton, 1977). The usage of herbarium material prevented cytological evaluations of our samples and we therefore bioinformatically estimated ploidy using our sequencing data. Our estimates are not conclusive but in combination with our admixture analyses they do support that *M. spicata* is an allopolyploid and that *M. longifolia* is likely cytomictic with both diploid and polyploid populations (Figure 2 and Table 1). However, the ploidy level of *M.* x *rotundifolia* is more uncertain as most were assigned as "likely polyploid" (triploids; Table 1). The existence of triploid individuals among the analyzed specimens is not unlikely. However, it is unclear to what extent HMMploidy is able to distinguish between triploids and tetraploids. For example, tetraploids formed from the merging of the very closely related genomes of *M. longifolia*and *M. suaveolens* could potentially appear as triploids due to extensive allele sharing between them (Figure 3). But it is unclear why this would more often be the case for specimens morphologically assigned as *M.* x *rotundifolia* (Table 1). It is possible that the distinct genomic histories of *M. spicata* and *M.* x*rotundifolia* are linked to differences in ploidy but no firm conclusions can be made based on our data.

Differences in frequencies of sexual reproduction can also drive the genomic distinction between *M. × rotundifolia* and *M. spicata* . While *M. × rotundifolia* exhibits a reduction in fertility, *M. spicata* is usually completely fertile (Table S3; Harley & Brighton, 1977). It is therefore possible that *M. ×rotundifolia* to a higher degree propagates clonally via rhizomes. However, such a scenario is not completely supported by our plastome analysis that suggests that most of the analyzed *M. ×rotundifolia* stem from independent hybridizations while the plastomes of *M. spicata* are largely homogeneous suggesting a single origin (Figure S3). It is however likely that human imposed selection on *M. spicata* in cultivation coupled with clonal propagation of plants with desired traits (Harley & Brighton, 1977) have caused shifts in allele frequencies in *M. spicata* and that the distinct genomic histories of these taxa, in addition to potentially be associated with ploidy differences, also are coupled to the (partly) cultivation status of *M. spicata* .

Increased opportunities for hybridization can merge species into a coalescent complex

Inter-fertility of the taxa in *Mentha* subgen. *Mentha* is well known and species barriers are possibly retained by the formation of sterile hybrids and/or usage of distinct ecological niches (Harley & Brighton, 1977; Gobert et al., 2002). However, the emergence of the fertile allopolyploid *M. spicata* followed by its re-establishment following cultivation in the native range of *M. longifolia* and*M. suaveolens* has resulted in frequent gene flow between the classically recognized taxa (Figure 2; Harley & Brighton, 1977). Indeed, all morphologically and genomically defined groups of specimens show very low levels of genetic distances not consistent with divergent species (Table 2; Meirmans & Hedrick, 2011; Roux et al., 2016). In particular, the genomic differentiation between *M. longifolia* and*M. spicata* is elusive (Figure 2 and Table 2). The extensive genetic admixture points to a breakdown of reproductive barriers between taxonomic units and morphological groups. However, given the frequent hybridization of *M. longifolia* and *M. suaveolens* in areas of sympatry (Harley & Brighton, 1977) it is unlikely that these taxa have ever been completely reproductively isolated. The emergence of the allopolyploid *M. spicata* has further promoted the introgression between *M. longifolia* and *M. suaveolens* by acting as a genomic bridge (McDonald, Parchman, Bower, Hubert, & Rahel, 2008; Sigel, 2016). The preferential crossing of *M. spicata* to *M. longifolia* has resulted in low levels of genetic differentiation between them (Table 2) and has transferred *M. suaveolens* genes into a mostly *M. longifolia* genomic background (Figure 2). Overall, our genomic analyses point to genetic swamping and an on-going merging of *M. longifolia* , *M. suaveolens* , *M. spicata* , and their hybrids into a coalescent complex (Beninde et al., 2018; Ellstrand & Schierenbeck, 2000; Pinto et al., 2005; Todesco et al., 2016; Quilodrán et al., 2020a). However, we do find that someF_{ST}-outlier scaffolds contain genes with a function in reproduction, and in particular, pollen recognition (Table S4), which do suggest that there might be some genetic barriers to cross-fertilization between the analyzed mint species, especially between the native mints.

Human assisted movements of species, including cultivation and rearing of species outside their native ranges have undoubtedly changed many ecosystems. On the one hand the addition of species to new ranges increases biodiversity by adding to the total number of species recorded in any one location (Schlaepfer, Sax, & Olden, 2011; Hallman, Olsson, & Tyler, 2022). However, although increased biodiversity is often regarded as positive, shifts in ecosystem compositions can have detrimental effects in the long term. For example, biodiversity might drastically decrease due to alien taxa out-competing native species (Pyšek et al., 2012). In addition, hybridization between native and introduced taxa can also change the local biodiversity (Wolf, Takebayashi, & Rieseberg, 2001). The exact effect of hybridization on local biodiversity is largely controlled by the strength of reproductive barriers between hybrids and parents and here we show that when the reproductive barriers are weak genetic swamping is a likely outcome (Lowry, 2008). Climate changes associated with anthropogenic activities will cause shifts in geographic ranges and ecological niches (Kelly & Goulden, 2008; Chen, Hill, Ohlemüller, & Thomas, 2011) and similar to naturalizations of cultivated taxa this will bring previously isolated taxa into contact with each other and increase the opportunities for hybridization and inter-specific gene flow. As we show here this can indeed promote gene flow between species and speed up the formation of new coalescent complexes.

Taxonomic consequences of inter-specific gene flow

Taxonomy in the genus *Mentha* is solely based on morphology (Harley & Brighton, 1977; Tucker et al., 1980; Tucker & Naczi, 2007; Vining et al., 2020). However, here we show that the results of genomic analyses are not fully congruent with a strictly morphology-based taxonomy (Figures 1 and 2). However, although our genomic analyses suggest that *M. suaveolens* , *M. longifolia* , *M. ×rotundifolia* , and *M. spicata* should be considered a coalescent complex where no taxa are completely isolated genetically, we also find that for most parts the classically recognized taxa are morphologically supported (Figures 1 and 2). Hence, the classical taxonomic units are still mostly useful although we suggest that an in-depth taxonomic revision including more taxa is made. The observed dissociations between the morphological and genomic clustering appear to be partially due to the continued segregation of genetic variants associated with characters of indumentum. Continued and on-going cultivation and escape into the wild of various mint species and varieties, including *M. spicata* , will likely continue to feed the native gene pools with alleles from non-native taxa. It is therefore likely that there will continue to be dissociations between morphological and genomic clustering of specimens.

Here, we have focused on mostly old herbarium specimens and 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 and show a genomic admixture profile that is different from what we present here. There are numerous studies on the evolutionary origin of various mint hybrids and the phylogenetic relationships within*Mentha* (eg. Bunsawatt, Elliott, Hertweck, Sproles, & Alice, 2004; Gobert et al., 2002; Heylen et al., 2021; Panjeshahin, Sharifi-Sirchi, & Samsampour, 2018). Contradictions in conclusions between previous studies can partially be explained by biased inclusions of particular specimens and/or inconsistent delimitation of taxa, but as we, show here not including all sub-populations in such analyses can also bias the results. Inclusion of multiple samples per taxonomic unit is therefore of importance, especially for taxa, such as mints, where frequent inter-fertility is reported (Heylen et al., 2021).

CONCLUSION

Here we show that the introduction of the cultivated spearmint into the range of closely related taxa has increased inter-specific gene flow between traditionally recognized mint taxa. This has caused the breakdown of species barriers and created a cradle for the birth of a coalescent complex. We thus show that when reproductive barriers are low, the likely outcome of increased hybridization is genetic swamping and merging of previously identified species causing a decrease in the overall species richness. Hence, similarly to the cases of naturalization of cultivated taxa, range shifts associated with climate changes can promote the loss of biodiversity even when habitats are protected and species are able to adapt to the new environment. Despite the large-scale genomic overlap among the studied mints, we find that genetic variants associated with a cultivated morphology continue to segregate within and among populations, and hence classical taxonomic units are mostly morphologically supported. The dissociation between genomic and morphological merging will therefore likely continue to complicate taxonomic assignments among mints.

ACKNOWLEDGEMENTS

This study was supported by the European research council (grant MSCA-IF-2017-795018). LTD is supported by a Natural Environment Research Council Independent Research Fellowship (NE/T011025/1).We thank Ida Broman Nielsen for help with the preparation of 24 sequencing librariers.

REFERENCES

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

Ainouche, M. L., Fortune, P .M., Salmon, A., Parisod, C., Grandbastein, M.-A., Fukunga, K., ... Misset M.-T. (2009). Hybridization, polyploidy and invasion: lessons from Spartina (Poaceae).*Biological Invasions, 11* , 1159. doi: https://doi.org/10.1007/s10530-008-9383-2

Altschul S. F., Gish W., Miller W., Myers E. W., & Lipman D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215* , 403-410.

Anderson E. & Stebbins, Jr. G. L. (1954). Hybridization as an Evolutionary Stimulus.*Evolution, 8* , 378-388.

Anderson, N. O., Galatowitsch, S. M., & Gomez, N. (2006). Selection strategies to reduce invasive potential in introduced plants.*Euphytica, 148* , 203–216. doi: https://doi.org/10.1007/s10681-006-5951-7

Beninde, J., Feldmeier, S., Veith, M., & Hochkirch, A. (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, 285* , 20180143. doi:http://doi.org/10.1098/rspb.2018.0143

Bieker, V. C. & Martin, M. D. (2018) Implications and future prospects for evolutionary analyses of DNA in historical herbarium collections.*Botany letters, 165* , 409-418.

Boatright, J., Negre, F., Chen, X., Kish, C. M., Wood, B., Peel, G., Orlova, I., Gang, D., Rhodes, D., & Dudareva, N. (2004) Understanding*in vivo* benzenoid metabolism in petunia petal tissue. *Plant Physiology* , 135, 1993-2011.

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

Carøe, C., Gopalakrishna, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., ... Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution,9* , 410-419.

Chalvin, C., Drevensek, S., Dron, M., Bendahmane, A., & Boualem, A. (2020). Genetic control of glandular trichome development. *Trends in Plant Science, 25* , 477-487.

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

Chen, I-C., Hill, J.K., Ohlemüller, D.B.T., & Thomas, C.D. (2011) Rapid range shifts of species associated with high levels of climate warming,*Science, 333* , 1024-1026.

Chopra, D., Mapar, M., Stephan, L., Albani, M. C., Deneer, A., Coupland, G., ... Hülskamp, M. (2019). Genetic and molecular analysis of trichome development in *Arabis alpina* . *Proceedings of the National Academy of Sciences, 116,* 12078-12083. doi: 10.1073/pnas.1819440116

Danecek, P., Auton, A., Abecasis, G., Albers, C. A. Banks, E., DePristo, M. A., ... 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics, 27* , 2156-2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., ... Li, H. (2021). Twelve years of SAMtools and BCFtools, *GigaScience, 10* , giab008.

Ellstrand, N.C. & Schierenbeck, K.A. (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences, 97* , 7043-7050. doi: 10.1073/pnas.97.13.7043

Evanno, G., Regnaut, S., & Goudet, J. (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 (pp.162-164). New York, NY: Academic Press.

Figueroa-Pérez, M. G., Reymoso-Camacho, R., Garcia-Ortega, L.F., & Guevara-Conzález, R. C. (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.

Fuchs, E. J., Martínez, A. M., Calvo, A., Muñoz, M., & Arrieta-Espinoza, G. (2016).

Genetic diversity in *Oryza glumaepatula* wild rice populations in Costa Rica and possible gene flow from *O. sativa* . *PeerJ, 4* , e1875.

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

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology, 29* , 644–652. doi: https://doi.org/10.1038/nbt.1883

Grant, V. (1981). *Plant speciation* . New York, NY: Columbia University Press.

Hallman, C., Olsson, O., & Tyler, T. (2022) Changes in south-Swedish vegetation composition over the last 200 years as described by species-specific indicator and trait values and documented by museum and literature records, *Ecological Indicators, 134* , 108486.

Harley, R. M. & Brighton, C. A. (1977). Chromosome numbers in the genus*Mentha* L. *Botanical Journal of the Linnean Society, 74* , 71-96.

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

Hobhouse, P. (1992). *Gardening through the Ages* . New York, NY: Simon and Schuster.

Hong, E. P. & Park, J. W. (2012). Sample size and statistical power calculations in genetic association studies. *Genomics & Informatics, 10* , 117-122.

Huxley, A. (1978). An Illustrated History of Gardening. New York, NY: Paddington Press.

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

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

Kelly, A.E. & Goulden, M.L. (2008) Rapid shifts in plant distribution with recent climate change, *Proceedings of the National Academy of Sciences, 105* , 11823-11826.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014) ANGSD: Analysis of next generation sequencing data. *BMC Bioinform, 15* , 356.

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

Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie2. *Nature Methods, 9* , 357-359.

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

Langer, K. M., Jones, C. R., Jaworski, E. A., Rusging, G. V., Kim, J. Y., Clark, D. G., & Colquhoun, T. A. (2014) PhDAHP1 is required for floral volatile benzenoid/phenylpropanoid biosynthesis in Petunia × hybrida cv 'Mitchell Diploid'. *Phytochemistry* , 103, 22-31.

Li, B., Cantino, P., Olmstead, R., Bramley, G. L. C., Xiang, C.-L., Ma, ... Zhang, D.-X. (2016). A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification.*Scientific Reports, 6* , 34343. doi: https://doi.org/10.1038/srep34343

Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A., & Willis, J. H. (2008). The strength and genetic basis of reproductive isolating barriers in flowering plants, *Philosophical Transactions of the Royal Society B Biological Sciences, 363* , 3009–3021. doi: http://doi.org/10.1098/rstb.2008.0064

Maffei, M., Gallino, M., & Sacco, T. (1986). Glandular trichomes and essential oils of developing leaves in *mentha viridis lavanduliodora* . *Planta Medica, 1986* (3), 163-246.

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

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

McDonald, D. B., Parchman, T. L., Bower, M. R., Hubert, W. A., & Rahel F. R. (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, 105* , 10837-10842. doi: 10.1073/pnas.0712002105

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

Mishra, A., Lal, R. K., Chanotiya, C. S., & Dhawan, S. S. (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. doi: https://doi.org/10.1007/s00709-016-1011-x

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

Morjan, C.L. & Rieseberg, L.H. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* , *13* , 1341-1356. doi: https://doi.org/10.1111/j.1365-294X.2004.02164.x

Munguía-Rosas, M.A., Jácome-Flores, M.E., Bello-Bedoy, R., Solís-Montero, V., & Ochoa-Estrada, E. (2019). Morphological divergence between wild and cultivated chaya (*Cnidoscolus aconitifolius*) (Mill.) I.M. Johnst. *Genetic Resources and Crop Evolution, 66* , 1389–1398. doi: https://doi.org/10.1007/s10722-019-00790-w

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

Otto, S.P. & Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review in Genetics* , *34* , 401-437.

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

Panjeshahin, Z. Sharifi-Sirchi, G.-R. & Samsampour, D. (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.

Pinto, M. A., Rubink, W. L., Patton, J. C., Coulson, R. N., & Johnston, J. S. (2005). Africanization in the United States: replacement of feral European honeybees (*Apis mellifera* L.) by an African hybrid swarm. *Genetics, 170* , 1653-65. doi: 10.1534/genetics.104.035030.

Pyšek, P., Jarošík, V., Hulme, P.E., Pergl, J., Hejda, M., Schaffner, U., & Vilà, M. (2012), A global assessment of invasive plant impacts on resident species, communities and ecosystems: the interaction of impact measures, invading species' traits and environment. *Global Change Biology, 18* , 1725-1737.

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

Quilodrán, C. S., Nussberger, B., Macdonald, D. W., Montoya-Burgos, J. I. & Currat, M. (2020). Projecting introgression from domestic cats into European wildcats in the Swiss Jura. *Evolutionary Applications, 13,* 2101-2112.

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

Reichard, S. H. & White, P. (2001). Horticulture as a pathway of invasive plant introductions in the United States: Most invasive plants have been introduced for horticultural use by nurseries, botanical gardens, and individuals. *BioScience, 51* , 103-113.

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. & Burke, J.M. (2001). The biological reality of species: gene flow, selection, and collective evolution. *Taxon, 50* , 47-67. doi: https://doi.org/10.2307/1224511

Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016) Shedding light on the grey zone of speciation along a continuum of genomic divergence, *PLOS Biology, 14* , e2000234.

Salehi, B., Stojanović-Radić, Z., Matejić, J., Sharopov, F., Antolak, H., Kregiel, D., Sen, S., Sharifi-Rad, M., Acharya, K., Sharifi-Rad, R., Martorell, M., Sureda, A., Martins, N., & Sharifi-Rad, J. (2018). Plants of Genus *Mentha* : From Farm to Food Factory. *Plants, 7* , 70.

Šarić-Kundalić, B., Fialoá, S., Dobeš, C.,Ölzant, S., Tekeľová, D., Grančai, D., Reznicek, G., & Saukel, J (2009). Multivariate Numerical Taxonomy of Mentha Species, Hybrids, Varieties and Cultivars.*Scientia Pharmaceutica, 77* , 851-876.

Schlaepfer, M.A., Sax, D.F, & Olden, J.D. (2011). The potential conservation value of non-native species. *Conservation Biology, 25* , 428-37.

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

Schierenbeck, K.A., & Ellstrand, N.C. (2009). Hybridization and the evolution of invasiveness in plants and other organisms.*Biological Invasions 11,* 1093. doi: https://doi.org/10.1007/s10530-008-9388-x

Sigel, E.M. (2016). Genetic and genomic aspects of hybridization in ferns. *Journal of Systematics and Evolution, 54* , 638-655. doi: https://doi.org/10.1111/jse.12226

Singh, P. & Pandey, A. K. (2018). Prospective of essential oils of the genus *mentha* as biopesticides: A review. *Frontiers in Plant Sciences, 9* , 1295.

Sirikantaramas, S., Morimoto, S., Shoyama, Y., Ishikawa, Y., Wada, Y., Shoyama, Y., & Taura F. (2004). The gene controlling Marujuana psychoactivity: Molecular cloning and heterologous expression of ?^{1}-tetrahydrocannabinolic acid synthase from*Cannabis sativa* L. *Journal of Biological Chemistry* , 279, 39767-39774.

Skotte, L., Korneliussen, T. S, & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data.*Genetics, 195* , 693-702.

Soltis, D. E., Albert, V. A, Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., ... Soltis, P. S. (2009). Polyploidy and angiosperm diversification. *American Journal of Botany* , *96* , 336-348.

Soraggi, S., Rhodes, J., Altinkaya, I., Tarrant, O., Balloux, F., Fisher, M. C., & Fumagalli, M. (2021). HMMploidy: inference of ploidy levels from short-read sequencing data. *bioRxiv*.https://doi.org/10.1101/2021.06.29.450340

Sobti, S. N. (1965). Chromosome numbers in species of *Mentha* .

Stace, C (2010) New Flora of the British Isles. 3rd edn. Cambridge University Press, Cambridge, UK.

Stamatakis, A. (2014). RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30* , 1312-1313.

Stebbins, G. L. (1959). The role of hybridization in evolution.*Proceedings of the American Philosophical Society, 103* , 231-251.

Szabó, K., Sárosi, S., Cerháti, B., & Ferenczy, A. (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., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., ... Rieseberg, L. H. (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. & Naczi, R. F. C. (2007). *Mentha* : an overview of its classification and relationships. In B. Lawrence (Rd.), *Mint: the genus Mentha* (pp.1-40). Boco Raton, FL: CRC Press.

Tucker, A. O., Harley, R. M., & Fairbrothers, D. E. (1980). The Linnean types of *Mentha* (Lamiaceae). *Taxon, 29* , 233–255. doi: 10.2307/1220285

Tutin, T. G. (1993). *Flora Europaea* . (OKS Print.) Cambridge: Cambridge University Press

The UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research, 49* , D480-D489.

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

Vining, K. J., Johnson, S. R., Ahkami, A., Lange, I., Parrish, A. N., Trapp, ... Lange B. M. (2017). Draft genome sequence of *Mentha longifolia* and development of resources for mint cultivar improvement.*Molecular Plant, 10* , 323-339.

Weir, B. S. & Cockerham C. (1984). Estimating F -statistics for the analysis of population structure. *Evolution, 38* , 1358-1370.

Wolf, D. E., Takebayashi, N., & Rieseberg, L. H. (2001). Predicting the Risk of Extinction through Hybridization. *Conservation Biology, 15* , 1039–1053.

Ye, J., McGinnis, S., & Madden, T. L. (2006). BLAST: improvements for better sequence analysis. *Nucleic Acids Res, 34* , W6-W9.

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

Author contributions : JKO, TT, and AJH designed the study. JKO and TT selected samples. TT produced and analyzed the morphological data. JKO produced and analyzed the genomic data. LTD analyzed transcriptomes. TT, MH, and AR provided samples. JKO wrote the paper with input from all co-authors. All co-authors have read and approved the final version of the paper.

Data Accessibility: Raw sequencing reads are available on Short Read Archive under BioProject accession number XXXX.

FIGURES

Figure 1. Principal component analysis (PCA) of morphological characters.

a) PCA of all 34 morphological characters and 155 specimens, b) PCA of morphological characters excluding ten traits related to indumentum, and c) PCA of ten characters of indumentum. Taxonomic assignments are indicated with colors. Circles refer to pure taxa whereas triangles refer to hybrids.

Figure 2. Genomic analyses.

Principal component analysis (PCA) (a) and admixture analysis (b) on 31,889 single nucleotide polymorphisms (SNPs) called for all 93 sequenced specimens. In (a) the taxonomic assignments based on morphology are indicated with colors and circles refer to pure taxa whereas triangles refer to hybrid species. In (b) specimens have been ordered based on the genomic cluster analysis in (a) and color under each bar indicate taxonomic assignments as presented in (a).

Figure 3. Shared number of single nucleotide polymorphisms (SNPs) and site frequency spectra (SFS).

The number of shared polymorphic sites (a) and the SFS (b) of SNPs with at least 95% posterior probability and called in at least 50% of the specimens. No – number; alt – alternative, var – variability.

SUPPORTING INFORMATION

Figure S1. Distributions of leaf and stem indumentum traits.

Distributions of indumentum on a) upper side of the leaf, b) lower side of the leaf, c), stem, and d) proportion of branched indumentum on stem and leaf for *Mentha longifolia* (orange), *M.* x*rotundifolia* (red), *M. spicata* (green), and *M. suaveolens* (cyan).

Figure S2. Distributions of flower indumentum traits.

Distributions of indumentum on a) pedicel, b) calyx, c), corolla, d) outside of the corolla, and e) the throat of the corolla for*Mentha longifolia* (orange), *M.* x *rotundifolia*(red), *M. spicata* (green), and *M. suaveolens* (cyan).

Figure S3. Chloroplast phylogeny.

Phylogeny of alignments of whole chloroplast genomes. a) Full phylogeny including the outgroup *Salvia* and b) zoom in on the *Mentha*genus. Bootstrap values above 80 are shown near nodes (* - 100 bootstrap support) and branch lengths are given as the expected number of substitutions per site.

Figure S4. Genomic admixture analysis.

Bayesian admixture analysis using NGSadmix. In a) the likelihood (+/- SD; left panel) is shown for ten runs for K 1 through 10 and the delta(K) of the second derivative of the likelihood is shown for the same runs (middle panel) and zoomed in on the y-axis (right panel) and in b) the admixture profiles for K =5 and K =6 (see Figure 2b for K =2-4).

Figure S5. Distributions of scaffoldF_{ST}-values between morphologically defined taxonomic groups.

For each scaffold the mean F_{ST}-value was calculated. The mean and median F_{ST}-values across all scaffolds are displayed with vertical red lines (full line - mean and dashed line - median).

Figure S6. Distributions of scaffoldF_{ST}-values between genomic groups.

For each scaffold the mean F_{ST}-value was calculated. The mean and median F_{ST}-values across all scaffolds are displayed with vertical red lines (full line - mean and dashed line - median).

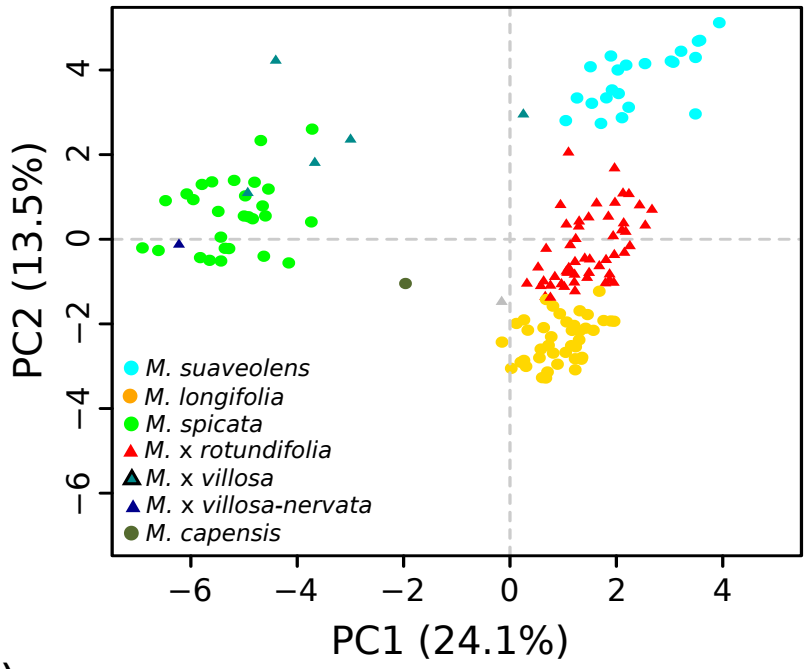
Table S1. Sample information.

Table S2. Scoring of morphological traits.

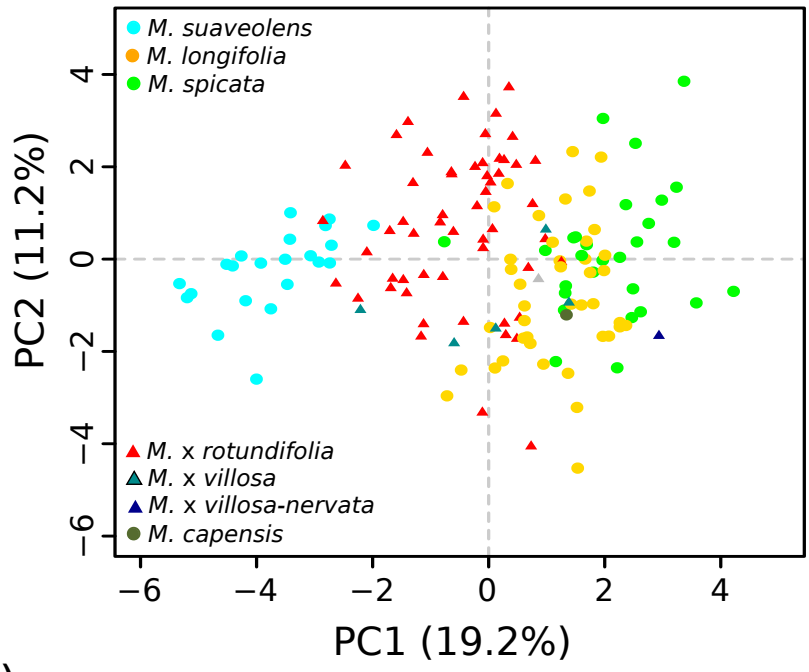
Table S3. Values for all morphological traits.

Table S4. BLAST results. F_{ST}-outlier scaffolds with a BLAST match to a *Mentha* transcript and to a plant species in UniProt.

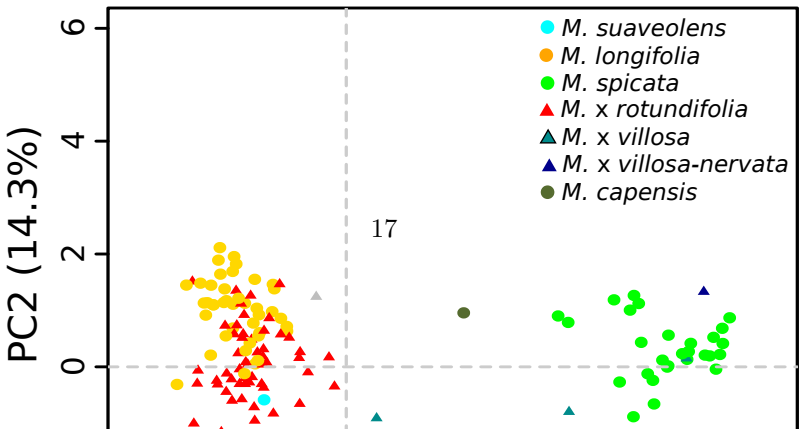
a)

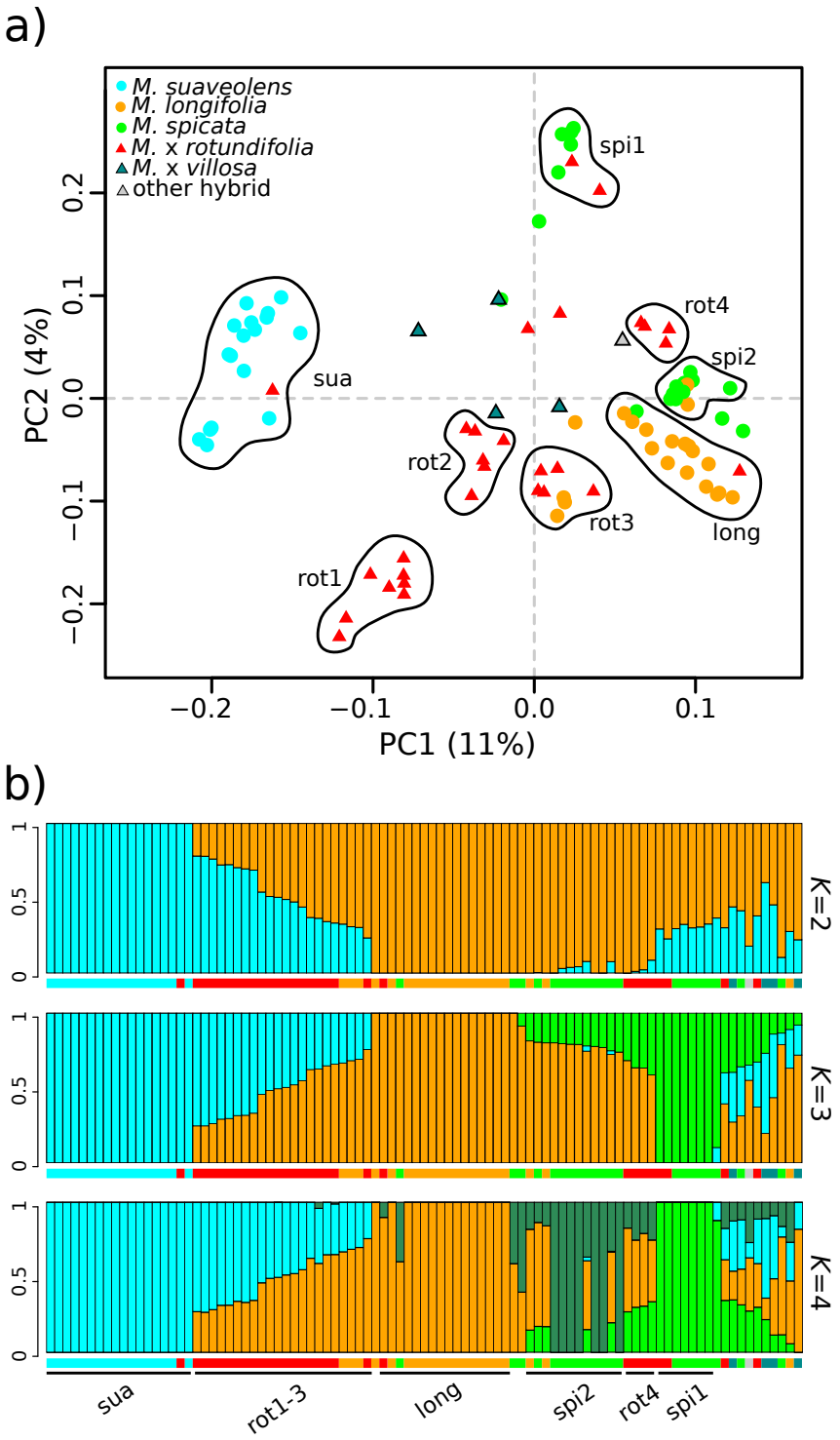


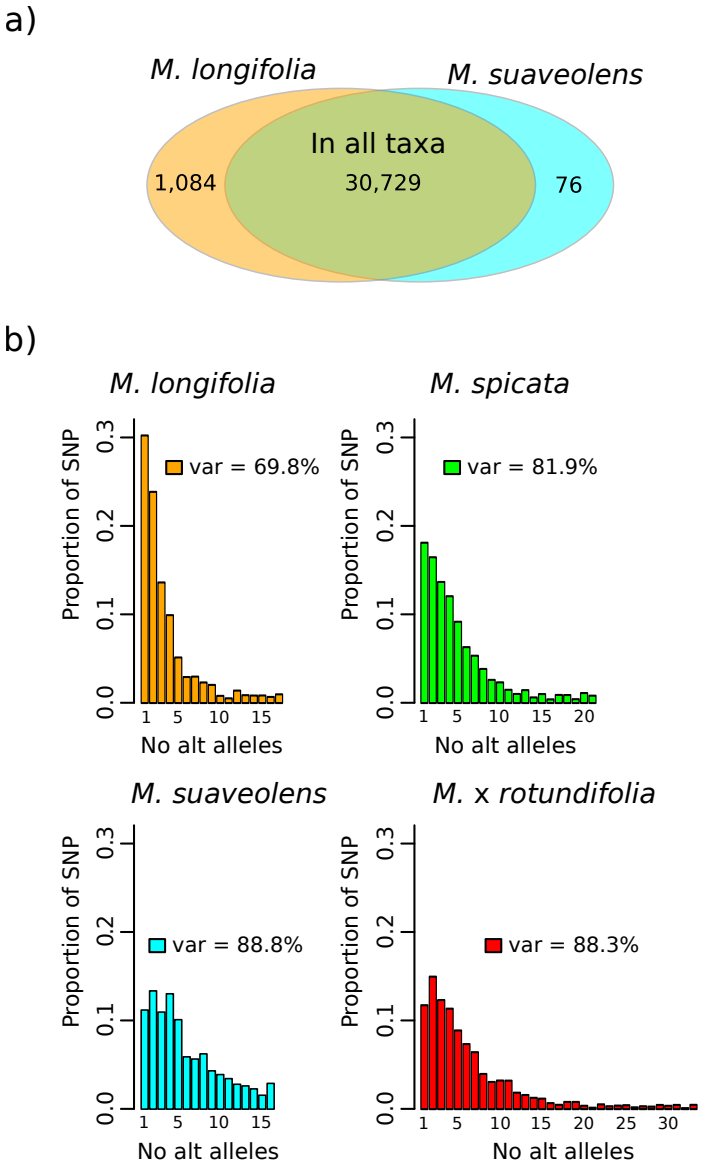
b)



c)







Hosted file

Table1_ploidy.docx available at https://authorea.com/users/460542/articles/556513-inter-specific-gene-flow-following-the-naturalization-of-a-cultivated-mint-promotes-the-formation-of-a-coalescent-complex

Hosted file

Table2_FST.docx available at https://authorea.com/users/460542/articles/556513-inter-specific-gene-flow-following-the-naturalization-of-a-cultivated-mint-promotes-the-formation-of-a-coalescent-complex