JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12743

Research Article

Combining population genomics and ecological niche modeling to assess taxon limits between *Carex jemtlandica* and *C. lepidocarpa*

Malene Nygaard[1]*, Petri Kemppainen[1,2], James D. M. Speed[1], Reidar Elven[3], Kjell Ivar Flatberg[1], Leif P. Galten[1], Narjes Yousefi[1,4], Heidi Solstad[1,3], and Mika Bendiksby[1]

[1]NTNU University Museum, Norwegian University of Science and Technology, Trondheim NO-7491, Norway
[2]Ecological Genetics Research Unit, Organismal and Evolutionary Biology Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Viikinkaari 1, Helsinki 00790, Finland
[3]Natural History Museum, University of Oslo, Oslo NO-0318, Norway
[4]Department of Systematic and Evolutionary Botany, University of Zurich, Zurich 8008, Switzerland
*Author for correspondence. E-mail: malene.o.nygard@ntnu.no
Received 1 October 2020; Accepted 18 March 2021; Article first published online 23 March 2021

Abstract *Carex* section *Ceratocystis* (Cyperaceae) is a group of recently evolved plant species, in which hybridization is frequent, introgression is documented, taxonomy is complex, and morphological boundaries are vague. Within this section, a unified taxonomic treatment of the *Carex jemtlandica–Carex lepidocarpa* species complex does not exist, and Norway may currently be the sole country accepting species rank for both. *Carex jemtlandica* is mainly confined to Fennoscandia and is thus a Fennoscandian conservation responsibility. This motivated us to test the principal hypothesis that both *C. jemtlandica* and *C. lepidocarpa* represent evolutionary significant units, and that both deserve their current recognition at species level. We investigated their evolutionary distinctiveness in Norway, using restriction site-associated DNA sequencing and ecological niche modeling. Our genomic results reveal two genetic clusters, largely corresponding to *C. jemtlandica* and *C. lepidocarpa* that also remain distinct in sympatry, despite clear indications of ongoing hybridization and introgression. The ecological niche modeling suggests that they occupy different environmental niches. Jointly, our results clearly show that *C. jemtlandica* and *C. lepidocarpa* represent separately evolving entities that should qualify recognition as evolutionary significant units. Given the high level of introgression compared to other hybridizing species pairs in *Carex* we recommend treating *C. jemtlandica* as a subspecies of *C. lepidocarpa*.

Key words: Cyperaceae, ecological niche modelling, natural hybridization, population genetics, RAD sequencing, species delineation, taxonomy.

1 Introduction

Species is the most common unit used in biological studies and ecological researches, as well as in nature management and biodiversity assessments (Coates et al., 2018). Species are, however, rarely uniform in their attributes. Accordingly, taxonomy does not always reflect the underlying genetic diversity (Avise, 1989). Genetic diversity is believed to be critical for supporting population fitness and adaptive potential, including the ability to evolve resistance to emerging infectious diseases and pathogens (Lande & Shannon, 1996; Frankham, 2003). Hence, to strengthen species' viability in a changing environment, it is essential to preserve both ecological and genetic diversity. Units of conservation must, however, be delimited and named to be identified and communicated. Evolutionary significant units (ESUs) are evolutionary lineages (often intraspecific) that are considered distinct for purposes of conservation (see Fraser & Bernatchez, 2001 and references therein). Herein we follow the unified framework of defining such conservation units (i.e., ESUs), as defined by Fraser & Bernatchez (2001), implying restricted interlineage gene flow. Such lineages represent independent evolutionary trajectories that, due to the highly restricted gene flow, will have limited or no impact on the evolution, genetic variance, and demography of other such lineages.

Interspecific gene flow often causes intermediate phenotypes. Various crossing experiments and genetic analyses have revealed extensive amounts of ESUs within morphologically defined species, a phenomenon very common in fungi (Lücking et al., 2014), but also occurring in plants (e.g., Grundt et al., 2006). Species boundaries are further blurred by incomplete lineage sorting, a common phenomenon in young species complexes (Maddison & Knowles, 2006).

Carex section *Ceratocystis* Dumort. (1827; Cyperaceae) represents a group of young plant species, in which hybridization is common, introgression is documented, taxonomy is complex,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Systematics and Evolution* published by John Wiley & Sons Australia, Ltd on behalf of Institute of Botany, Chinese Academy of Sciences.

and morphological boundaries between entities are vague. In this section a general lack of qualitative discriminating phenotypic traits makes it nearly impossible to identify closely related species based on morphology. There is also evidence in section *Ceratocystsis* that repeated backcrossing with parental species has led to hybrids with morphologies and fertility similar to the parental species, a phenomenon known as stabilized cryptic backcrossing (Schmid, 1982).

Section *Ceratocystis* has recently been subjected to two molecular phylogenetic investigations based on DNA sequence data (Jiménez-Mejías et al., 2012, 2017; Derieg et al., 2013). The nuclear ribosomal (ETS, ITS) and plastid (*rps16* and 5′ *trnK*) DNA sequence data used by them did not resolve all phylogenetic relationships within section *Ceratocystis*. Allozyme variation, however, seems to be more consistent with the taxonomic treatments (Hedrén & Prentice, 1996; Hedrén, 2002; Blackstock, 2007), including moderate allele frequency differences between the two closely related *Carex jemtlandica* (Palmgr.) Palmgr. (1937) and *Carex lepidocarpa* Tausch (1834). This indicates that methods that more broadly fingerprint the genome have higher probability of capturing loci that are coalesced at species level. A genome-wide fingerprinting method that is applicable to non-model species, for which reference genomes are rarely available, is restriction site-associated DNA sequencing (RADseq; Miller et al., 2007; Baird et al., 2008). In RADseq, the DNA is first digested with restriction enzymes and the ends of the resulting fragments are sequenced. Compared with whole genome sequencing, such reduced-representation sequencing increases the number of reads per locus dramatically at the expense of number of loci, with the potential to still produce thousands of polymorphic markers across genomes at a low cost.

To the best of our knowledge, Norway may be the sole country that currently accepts *C. jemtlandica* and *C. lepidocarpa* at species level. Both *C. jemtlandica* and *C. lepidocarpa* are red-listed in Norway due to loss of suitable habitats, mainly caused by ditching of fens. Currently, *C. lepidocarpa* is accepted at species level by most specialists. This is argued on the basis of distinctiveness and divergence from other members of the section in morphology (Jiménez-Mejías et al., 2014), ecology (Więcław, 2014), karyology (chromosome number; Rotreklová et al., 2011), allozymes (Hedrén, 2002; Blackstock, 2007), and molecular markers (Jiménez-Mejías et al., 2012). Additional detail on the history of contradictory taxonomic treatments of *C. lepidocarpa* can be found in the study of Blackstock (2007). The delimitation of *C. jemtlandica*, however, is still a matter of debate. Some botanists (Schmid, 1983; Crins & Ball, 1989) do not recognize *C. jemtlandica* at any taxonomic level and consider it as merely part of the intraspecific variation of *C. lepidocarpa*. Regardless of taxonomic level of acceptance, current taxonomic conclusions are based on studies of only one or few of the following attributes: morphology, chromosome number, hybrid fertility, ecological preference, allozymes, and DNA sequence data from a limited number of ribosomal and plastid loci.

Carex lepidocarpa occurs widely throughout northern and central Europe as well as in sub-oceanic regions of eastern Canada (Crins & Ball, 1989). As opposed to this amphi-Atlantic distribution, *C. jemtlandica* is a Fennoscandian endemic, occurring from Norway through Sweden and Finland to the Baltic States and westernmost European Russia (Palmgren, 1959; Hedren, 1990). It largely replaces *C. lepidocarpa* in this part of the boreal zone (Hedrén & Prentice, 1996). In Norway, the two taxa are largely allopatric: *C. lepidocarpa* occurring scattered along the coast from the eastern (Østfold) to the northern regions (Sør-Troms) and *C. jemtlandica* mainly confined to the more continental parts of southeast and central Norway (Elven, 2013). Their respective distributions overlap only in the southeast (Buskerud) and central parts (Trøndelag) of Norway (Elven, 2013; Nygård, 2016). Both *C. lepidocarpa* and *C. jemtlandica* are calciphile (Schmid, 1984; Pykälä & Toivonen, 1994), preferring base-rich mires. Both are mainly found growing in lawns, carpets, and partly also in the wetter parts of the mire. Ecological studies of *C. jemtlandica* and *C. lepidocarpa* that systematically assess distribution patterns and ecological preferences are lacking. In particular, the niche width and variation of *C. jemtlandica* is poorly known.

Ecological niche modeling (ENM) can provide insights into the ecology of species (Elith & Leathwick, 2009). Niche modeling typically uses presence observations of a target (e.g., species or lineage), together with environmental data, to predict the relative probability of target presence in areas where the target has not been recorded. Combined with genetics, ENM has the potential to illuminate how species level versus intraspecific ESUs may differ in their habitat preferences (e.g., Pearman et al., 2010; Balint et al., 2011; Bendiksby et al., 2014). Norway is a highly appropriate model area for ecological niche modeling, because it encompasses more than one-third of the regional bioclimatic variation in Europe (Bakkestuen et al., 2008).

In this study we use population genomic data (RADseq) combined with ENM to test the following five hypotheses regarding the taxonomy, ecology, and evolution of *C. jemtlandica* and *C. lepidocarpa*: (i) The two taxa hybridize in sympatry, as is very common among species in section *Ceratocystis*. (ii) The morphology-based taxonomic assignments are corroborated by molecular data. (iii) The differing distribution patterns of the two can be explained by adaptations to different environments. (iv) *Carex jemtlandica* and *C. lepidocarpa* qualify as two distinct evolutionary significant units (ESUs), as opposed to representing two extremes of an intraspecific clinal variation from continental to coastal areas, and (v) *C. jemtlandica* and *C. lepidocarpa* both deserve recognition at species level.

2 Materials and Methods

2.1 Sampling and morphological identification

In 2014–2015, we sampled specimens of *C. jemtlandica* and *C. lepidocarpa* from eastern and central Norway in addition to two localities in southern Sweden (see Nygård, 2016). In the current study, we have included data from Nygård (2016) as well as a broader sampling of section *Ceratocystis* from the same collection sites. The added taxa include *C. demissa* Hornem (1806), *C. flava* L. (1753), *C. hostiana* DC (1813), *C. viridula* Michx. (1803) var. *viridula*, *C. viridula* var. *pulchella* (Lönnr.) Schmid (1983), and putative interspecific hybrids. We have also included one population of *C. flava* (population 1501 in Table S1) from outside the *C. jemtlandica* and *C. lepidocarpa* collection sites in central Norway (i.e., Røros). These taxa were included to enable detection of potential interspecific gene flow that could have led to misidentifications in our *C. lepidocarpa* and *C. jemtlandica* material. In total, the data for the present study comprise 192

individuals, for which leaf material was dried and stored on silica gel. All associated vouchers are deposited in the herbaria O and TRH (Table S1).

In addition to the morphology-based field determinations, we re-examined all included specimens in four parallels, unaware of prior identifications in each parallel. In this blind test, we examined more thoroughly the diagnostic characters suggested by Palmgren (1959; Table 1). Our initial hybrid assignments, based on intermediate morphologies and reduced pollen fertility (aborted/reduced anthers), were reassessed in light of the molecular results.

2.2 DNA extraction, library building, and sequencing

We extracted genomic DNA using the E.Z.N.A. SP Plant DNA Kit (Omega Bio-tek, Norcross, GA, USA) following the manufacturer's protocol. For the DNA extraction, we used 10-mg silica-dried leaf tissue per sample, which was grinded into fine powder using two 2-mm tungsten beads (Qiagen, Hilden, Germany) on a Retsch MM301 mixer mill (GmbH & Co., Haan, Germany) for 2 × 2 min, first at 18 and then 20 oscillations/s. We eluted the samples twice in 50-μL elution buffer in separate tubes, one as a back-up aliquot. We prepared eight 24-plexed RADseq libraries using the restriction enzyme *sbfI* and the protocol of Etter et al. (2012) with minor modifications, as described by Yousefi et al. (2017). The RADseq libraries were sequenced with 100-bp paired-end sequencing over two lanes on an Illumina HiSeq 2500 (Illumina, Inc., San Diego, CA, USA).

2.3 Processing RADseq data, SNP calling, and filtering

Raw read quality, length, and base composition were assessed using FastQC v. 0.11.4 (Andrews, 2010). We then used the Stacks v. 2.0 Beta 9 (Catchen et al., 2013) *process_radtags* program for demultiplexing raw reads according to barcodes. We allowed one mismatch in the barcode sequence and enabled the option to rescue barcodes (*–r*). We removed adapter sequences with up to two mismatches in the quality filtering of the reads (*–c*, *–q*). Samples retaining few reads (<500 000; Table S2) were discarded before genotyping. We used the Stacks *denovo_map.pl* pipeline to assemble and genotype the reads. Optimal values of the Stacks key parameters (*m*, *M*, *n*) were chosen on the basis of the *r80* method described by Paris et al. (2017), maximizing the number of polymorphic loci in at least 80% of the population (*–r* = 0.8 in the Stacks *population* program). For each parameter iteration, we used shell scripts for collecting the resulting number of assembled loci, polymorphic loci, and SNPs present in each individual (Script S1). For parameter *m*, we also collected data on mean sequence depth per individual using the option *–depth* in VCFtools v. 0.1.17 (Danecek et al., 2011). We varied the minimum number of identical reads required to form stacks within individuals (putative alleles; *–m*) from 1 to 15, disabling the calling of haplotypes from secondary reads (*–H*) when $m > 6$. We also tested the maximum nucleotide distance allowed among stacks within individuals (*–M*) from 1 to 9. Only one parameter was varied at a time, while fixing the rest at default values ($m = 3$, $M = 2$, $n = 1$). After selecting values for *m* and *M*, we varied the maximum distance allowed among stacks between individuals (putative loci; *–n*) from $M - 1$ to $M + 1$.

To reduce the extent of linkage between SNPs, we extracted only one random SNP per locus (*--write-random-snp*) using the *populations* program in Stacks. We used VCFtools to calculate mean missing data per locus (*--missing-site*) and per individual (*--missing-indv*), as well as to per site averaged coverage across all individuals (*--site-depth*). We then filtered the data in the following order by removing: (i) SNPs with >50% missing data, (ii) individuals with >50% missing data, (iii) SNPs with read depth >475, and (iv) singletons. On the basis of the filtered data, we made two datasets: one containing only morphologically identified *C. jemtlandica*, *C. lepidocarpa*, and *C. jemtlandica* × *lepidocarpa* (STRICT), and one dataset containing all sampled taxa of section *Ceratocystis* (BROAD).

2.4 Genetic structure and potential misidentification

We assessed population genetic structuring on both datasets (STRICT and BROAD) using three methods: ADMIXTURE v.1.3.0 (Alexander et al., 2009), fineRADstructrue v.0.3.2 (Malinsky et al., 2018), and principal component analyses (PCA). If the genetic signal in the data is robust, all three methods (with different assumptions) should produce concordant results.

We used ADMIXTURE to estimate genetic clustering and level of admixture by maximum likelihood estimations of genetic ancestry. We tested 1 to 15 genetic clusters for STRICT ($K = 1$–15) and 1 to 20 genetic clusters for BROAD ($K = 1$–20), with 10 replicates for each *K*. A 10-fold cross-validation test, as implemented in ADMIXTURE, which estimates the prediction errors at any given *K*, was used to assess the optimal number of genetic clusters (*K*). Individual ancestry proportions of each cluster produced by ADMIXTURE were visualized using R. We used the estimated ancestry proportions at the optimal *K* to assign individuals in the STRICT dataset into genetic groups. Individuals with >0.8 ancestry proportions of the *C. jemtlandica* or *C. lepidocarpa* clusters were defined as pure, and those with <0.8 as admixed. Non-Norwegian samples and individuals with >0.2 ancestry proportions of genetic clusters belonging to non-focal taxa were excluded from further analyses. To estimate the proportion of morphological misidentification, these genetic groupings were further compared with their corresponding morphological assignments (described above).

We used FineRADstructure to estimate recently shared ancestry based on patterns of genomic similarity. First, we used the *sampleLD.R* script provided in fineRADstructre to order RAD loci according to linkage disequilibrium. We then ran the MCMC chain with a burn-in of 10 000, 100 000 iterations, and a thinning interval of 1000. The clustered fineRADstructure coancestry matrix was visualized in R with scripts provided with the software (https://github.com/millanek/fineRADstructure).

We performed PCA based on allele frequencies using the function *dudi.pca* in the *adegenet* R package v.1.3-1 (Jombart & Ahmed, 2011). Before this, *scaleGen* was used to scale allele frequencies to mean zero, and missing genotypes were replaced by the mean allele frequency among the individuals. The eigenvalues were extracted and visualized using the *fviz_eig* function in *factoextra* R package v.1.0.7 (Kassambara & Mundt, 2020).

The genetic information contained in the coancestry matrix produced by fineRADstructure can be used as a source for PCA. We then compared the genetic information from the first principal component (PC1) based on allele frequencies and coancestry by linear regression. High

correlation coefficient and a 1:1 ratio in this comparison would indicate high concordance with respect to the signal of population genetic structuring in the data. Similarly, the ancestry proportion of *C. jemtlandica* from the ADMIXTURE analysis was regressed against PC1 based on allele frequencies.

2.5 Hybrid detection, genetic diversity and species cohesiveness

The first-generation hybrids (F1) of a given cross between different species are expected to be heterozygous for all loci that are fixed for different alleles in the two parental species. To test the hypothesis that genetically intermediate individuals in the *C. jemtlandica*–*C. lepidocarpa* complex constitute hybrids between two genetically distinct (but hybridizing) species, we estimated observed heterozygosity (proportion of heterozygous loci in a given individual) and plotted this against the ancestry proportion of *C. jemtlandica*. These results were compared with 1 000 F1-hybrids simulated using HybridLab v1.1 (Nielsen et al., 2006) by sampling alleles from pure *C. jemtlandica* and *C. lepidocarpa*. For estimating population-level diversity indices, we needed to increase the per population sample sizes for pure *C. lepidocarpa*, *C. jemtlandica*, and their putative hybrids. Taking into account the Wahlund effect (apparent excess of homozygotes due to unaccounted population structure; Nielsen & Slatkin, 2013), we grouped sampling sites based on geographic affinity into six larger populations rather than the entire distribution range (Table 2). We then estimated expected and observed heterozygosity (H_{exp} and H_{obs}, respectively), the inbreeding coefficient, F_{IS} ($1 - H_{obs}/H_{exp}$), and the fixation index (F_{ST}) using *basic.stats* (*hierfstat* R package v. 0.5-7; Goudet, 2005). If the genetically intermediate individuals comprise hybrids between two reproductively isolated lineages, we expect F_{IS} to be lower in this group, compared with the pure *C. jemtlandica* and *C. lepidocarpa* (i.e., to show an excess of heterozygous loci as compared to what is expected under Hardy–Weinberg equilibrium).

If *C. jemtlandica* and *C. lepidocarpa* represent two extreme forms of a continuum of intraspecific clinal variation within their distribution range, we expect most of the genetic variation in the STRICT dataset to be explained by geography rather than taxonomy (genetic lineages). If *C. jemtlandica* and *C. lepidocarpa* represent ESUs, we expect the opposite (i.e., a larger proportion of the genetic variation is explained by taxonomy rather than geography). To test this hypothesis, we performed discriminant analysis of principal components (DAPC; Jombart et al., 2010) on the STRICT dataset. The DAPC combines a PCA with a DA for assessing the relationship between clusters, focusing on genetic differentiation between predefined groups rather than the total variance of the samples. The analysis was performed twice, retaining 10 PC and DA axes. For the first run, we predefined groups according to geography using the six above-defined populations. For the second run, we used pure *C. jemtlandica* and *C. lepidocarpa*, and their putative hybrids as three separate predefined groups.

2.6 Ecological niche modeling

Species occurrence data of preserved specimens were downloaded from the Global Biodiversity Information Facility (GBIF) for both *C. jemtlandica* and *C. lepidocarpa* (GBIF.org, 2019). Specimen-based occurrence data are preferred over observation data alone, as the former allow for quality control and taxonomic updates (Speed et al., 2018). According to GBIF, there are 208 preserved specimens of *C. jemtlandica* and 785 of *C. lepidocarpa* from Norway in different herbaria (13 in institutions outside

Table 1 List of selected morphological traits used by Palmgren (1959) for discriminating between *C. jemtlandica* and *C. lepidocarpa*

Character	*C. jemtlandica*	*C. lepidocarpa*
Vegetative leaves	Long relative to the culm (>50%), broad	Short relative to the culm (<50%), narrow
Culm leaves	More (3–6), uppermost attached over center of culm	Less (≤3), uppermost attached below center of culm
Culm	Smooth beneath the inflorescence	Rough beneath the inflorescence
Bracts of female spikes	Long and wide. *Long compared with distance between first and second female spike*^{H}	Short (especially the uppermost) and narrow, strongly bent downwards. *Short compared with distance between first and second female spike*^{H}
Female spikes	Uppermost ± close. Shape globose to ovate	±*separated*^{H}. Shape ovate to cylindrical
Utricles	±lax. Utricles ***longer***^{BH}, with ***longer***^{BH} straight beaks	±dense. Utricles ***shorter***^{B} and *wider*^{H}, with ***shorter***^{B} deflexed beaks
Peduncle of male spike	***Short***^{B} and vertical	***Long***^{BH} (*relative to male spike*^{H}) and diagonal

Characters that differ considerably between the two taxa are marked in italic and those differences reported as significant in bold. References are abbreviated: H, Hedrén (2002), B, Blackstock (2007).

Norway). Most of these records are based on field identifications (i.e., morphology alone). Due to the difficult field determination of these species, we expect many incorrectly identified *C. jemtlandica* and *C. lepidocarpa* specimens in the herbaria. Misidentifications will, if they go undetected, confound downstream analyses. We therefore strictly filtered the occurrence data and included in our analyses only records registered or controlled by experts. We also included records from known localities. The final dataset included 174 occurrence records of *C. jemtlandica* and 546 of *C. lepidocarpa* (Table S3).

Three main axes describe the majority of bioclimatic variation in Norway: (i) temperature, (ii) total precipitation and temperature seasonality, and (iii) precipitation seasonality (Speed & Austrheim, 2017). Altogether, the occurrences of *C. jemtlandica* and *C. lepidocarpa* encompassed a very high proportion of the total bioclimatic space of Norway. Hence, we selected three uncorrelated bioclimatic variables to represent each main axis: Mean temperature of the warmest quarter (i.e., mean summer temperature, MST; Fig. S1A), mean annual precipitation (MAP; Fig. S1B), and precipitation seasonality (coefficient of variance of monthly precipitation, CV; Fig. S1C), respectively. We used WorldClim (https://www.worldclim.org/bioclim) data for Norway and resampled the bioclimatic variables to a 1-km grid using a nearest neighbor approach (Fick & Hijmans, 2017). Soil pH was downloaded from SoilGrids (Fig. S1D; Hengl et al., 2014; https://www.isric.org/explore/soilgrids). Background data were created by sampling 1 000 random points across Norway, weighted by distribution of occurrence data of all vascular plants in Norway (Fig. S1E).

Due to the rarity of *C. jemtlandica* occurrences, compared with *C. lepidocarpa*, we used a multiple ensemble modeling approach (Guisan et al., 2017). We used the *sdm* package (Naimi & Araújo, 2016) in the R environment to run seven different models: generalized linear model (GLM), generalized additive model (GAM), random forest (RD), gradient boosting machines (GBM), mixture discriminant analysis (MDA), flexible discriminant analysis (FDA), and boosted regression trees (BRT). Each model was cross-validated with five replicate runs. We then averaged the results and predictions across the methods and across the five replicates of each method using a weighted average based upon the model area under the curve (AUC).

3 Results

3.1 Data processing and SNP calling

Two lanes of Illumina sequencing produced 556 059 762 reads of 100 bp for our 192 samples. The number of reads per sample varied from 8 341 to 5 263 015 with a median of 2 049 384 (±1 181 081 SD; Table S2). Only samples with the P2-barcode TGCAT retained less than 500 000 reads (ranged from 10 538 to 39 988) due to high amounts of barcode mismatches. We therefore excluded all 48 samples containing this barcode before *de novo* assembly of loci in Stacks. The Stacks parameter combination $m = 5$, $M = 1$, and $n = 1$ yielded the highest amount of broadly shared polymorphisms (*r80* value; Figs. S2–S4; Table S4) and was used for the final *de novo* assembly. With the optimized parameters, Stacks assembled in total 66 416 polymorphic loci, of which 16 359 contained more than one SNP (Table S5).

The following data were removed during filtering: 65 738 SNPs containing >50% missing data, 10 individuals missing >50% data, 3 SNPs displaying depth >475, and 303 singletons. After filtering, we retained 372 informative SNPs distributed across 134 individuals (Table 2). Of these individuals, 93 were morphologically assigned to the *C. jemtlandica–C. lepidocarpa* complex with an average read depth of 215.31 and 28.2% average missing data. This relatively low number of SNPs is likely due to the large number of different taxa included in the analyses, which may result in polymorphism in the restriction recognition sites and a large number of null alleles.

3.2 Genetic structure and potential misidentification

The ADMIXTURE cross-validation test returned $K = 3$ as the optimal number of clusters for the STRICT dataset (Figs. S5, S6). Two of the obtained clusters (Fig. 1, see Fig. S7 for individual IDs) largely corresponded to morphologically identified *C. jemtlandica* (blue) and *C. lepidocarpa* (red). Several individuals, including most phenotypic *C. jemtlandica* x *lepidocarpa* hybrids, displayed varying degrees of admixed ancestry between these two clusters. The third cluster (light blue) contained individuals of both morphologically identified *C. jemtlandica* and *C. lepidocarpa*. Nine of these, all with ancestry proportions >0.2 of the third cluster (Fig. 1), grouped together with *C. flava* in the BROAD ADMIXTURE results at optimal $K = 4$ (Figs. S8, S9).

On the basis of the ancestry proportions obtained by ADMIXTURE (Fig. 1), 17 individuals were defined as genetically pure *C. lepidocarpa* (>0.8 red cluster), 51 as pure *C. jemtlandica* (>0.8 blue cluster), 13 as admixed between the two (<0.8 red and blue cluster), and 10 as admixed with or pure *C. flava* (>0.2 light blue cluster). These genetic groups were in large congruence with our fineRADstructure results (Fig. S6). The coancestry tree contained three major branches, one of which contained all

Table 2 Number of retained samples and SNPs, read depth, and missing data after final filtering of the two datasets

			Read depth per individual			Missing data per individual		
Dataset	Individuals	SNPs	Mean	Min	Max	Mean	Min	Max
BROAD	134	372	207.86	25.60	440.47	29.67%	13.98%	51.88%
STRICT	93	372	215.31	25.60	440.47	28.20%	13.98%	51.88%

The BROAD dataset includes all sampled taxa of section *Ceratocystis*, and STRICT only phenotypic *C. jemtlandica*, *C. lepidocarpa*, and *C. jemtlandica* × *lepidocarpa*.

but one *C. flava*-related sample. The two remaining branches constituted pure *C. jemtlandica* and *C. lepidocarpa* separately, however, with admixed individuals occurring within both of them. A single *C. flava*-related sample did also occur in the *C. lepidocarpa* branch.

Compared with the genetic assignment, 22 individuals (20% of our samples) were morphologically misidentified (Table 3). The majority of the misidentifications were due to not recognizing admixed individuals between *C. lepidocarpa* and *C. jemtlandica* (27%) and mistaking young specimens of *C. flava* for *C. jemtlandica* (41%). In the following analyses, all individuals with more than 0.2 ancestry proportions of the *C. flava* cluster were excluded.

3.3 Hybrid detection, genetic diversity and species cohesiveness

After excluding the misidentified *C. flava* specimens, the PCA analysis based on allele frequencies detected two genetic clusters corresponding to *C. jemtlandica* and *C. lepidocarpa* (Fig. 2A), with PC1 explaining 13.6% of the variation (Fig. S10). Non-introgressed populations of *C. jemtlandica* and *C. lepidocarpa* were connected by individuals with intermediate genetic signals (Fig. 2A). In the PCA plot, the PC1 was strongly correlated with the ancestry proportions estimated by ADMIXTURE (see also Fig. S11A). The PC2 accounted for 5.8% of the variation in the data but did not further segregate the two taxa. Similar patterns were also observed when using the fineRADstructure coancestry matrix as input for the PCA analysis. The first PCs from PCA analyses based on allele frequencies and coancestry were highly correlated at approximately 1:1 ratio (Fig. S11B). This shows that the genetic distinctiveness of *C. jemtlandica* and *C. lepidocarpa* is not sensitive to the analysis used, nor was there any lack of resolution in our results due to the small number of SNPs. As all three methods produced highly concordant results, in the following, we only present results from the PCA analyses based on allele frequencies.

The genetically intermediate individuals harbor, on average, higher levels of heterozygosity than the pure taxa, yet lower than all simulated F1-hybrids (Fig. 2B). High frequencies of null alleles could potentially lead to a lack of heterozygote genotypes in these hybrids (i.e., when the alternative allele in a heterozygote is a null allele). High frequencies of null alleles also lead to missing data. The lack of a significant correlation between observed heterozygosity (proportion of heterozygous genotypes within individual loci) and per locus missing data indicates, thus, that null alleles were not common and thus cannot explain the discrepancy in heterozygosity between putatively real and simulated hybrids (Fig. S12).

The mapped ancestry proportions indicate that *C. jemtlandica* and *C. lepidocarpa* hybridize in all geographic areas where they are sympatric (Fig. 2C). About half (46%) of the genetically defined hybrids grew in sampling sites where both pure forms existed. A large proportion of admixed individuals (54%) were, however, also present in sites where only one of the parental taxa occurred. After pooling sampling sites into six larger populations (Fig. 2C), only one population did not contain any hybrid individuals (pop 5 with seven *C. lepidocarpa* and no *C. jemtlandica*; Table 4). The level of heterozygosity (H_{obs} and H_{exp}) calculated among the hybrid populations was higher, compared with genetically pure *C. jemtlandica* and *C. lepidocarpa* (Table 5). Lowest heterozygosity was found in *C. jemtlandica*. Nevertheless, hybrid F_{IS} was much lower than in either of the two pure groups (Table 5).

Genetic differentiation was higher between *C. jemtlandica* and *C. lepidocarpa*, in comparison to within. The mean F_{ST} between pure populations of *C. jemtlandica* and *C. lepidocarpa* was 0.28, when standardized by expected heterozygosity, $F_{ST}' = 0.43$. In comparison, we found an overall F_{ST} of 0.18 ($F_{ST}' = 0.22$) between populations of *C. jemtlandica* and 0.03 ($F_{ST}' = 0.05$) between populations of *C. lepidocarpa*.

The DAPC analysis did not succeed in separating the six populations when grouping according to geography (Fig. 3A). The first discriminant function explained the majority of the genetic variation, even so, it only accounted for 46%. In contrast, 97% of the genetic variation was explained by the first DA eigenvalue when grouping according to taxonomy (Fig. 3B). The resulting clusters of pure *C. jemtlandica* and *C. lepidocarpa* were clearly separated in genomic space. The hybrid cluster fell in between the pure groups along the first discriminant function (Fig. 3B).

3.4 Ecological niche modeling

Similar models fit across species (*C. lepidocarpa* AUC = 0.86 ± 0.06 SD, *C. jemtlandica* = 0.89 ± 0.05 SD). Of the four included environmental variables, precipitation seasonality and mean summer temperature (MST) were the most important spatial predictors of *C. jemtlandica* and *C. lepidocarpa*, respectively (Fig. S13). For *C. jemtlandica*, precipitation seasonality accounted for ~60% of the variable importance, and MST ~70% for *C. lepidocarpa*. The resulting spatial prediction of *C. jemtlandica* and *C. lepidocarpa* distributions in Norway was largely non-overlapping (Figs. 4A, 4B).

According to the response curves for *C. jemtlandica* and *C. lepidocarpa*, annual precipitation (MAP) and soil pH did not greatly contribute to the niche suitability of either species (Fig. S14). Nevertheless, there was a tendency for increased niche suitability with increasing soil pH for *C. lepidocarpa*. Higher precipitation seasonality had a positive effect on habitat suitability for *C. jemtlandica* but negative for *C. lepidocarpa* (Fig. S14). For *C. lepidocarpa*, niche suitability increased strongly with temperature (with high niche suitability at temperatures >10 °C; Fig. S14). This relation was less pronounced for *C. jemtlandica*. The niche space use plot, based on the two most influential environmental variables (precipitation seasonality and MST), clearly separates *C. jemtlandica* and *C. lepidocarpa* (Fig. 4C). Hence, the species exist in distinct niches (Figs. 4A, 4B): *C. lepidocarpa* is located in warmer regions with constant rainfall, whereas *C. jemtlandica* is located in regions with a greater temperature and higher precipitation seasonality.

4 Discussion

In Scandinavia, *C. jemtlandica* and *C. lepidocarpa* have been recognized as different taxa for close to a century. Initially,

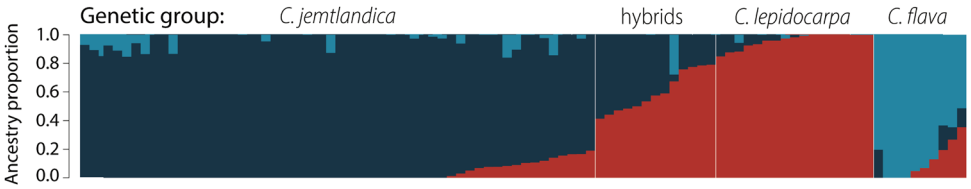


Fig. 1. ADMIXTURE results at optimal $K = 3$ for the STRICT dataset comprising 372 SNPs and 93 individuals of phenotypic *C. jemtlandica*, *C. lepidocarpa*, and putative hybrids. Individuals are represented by vertical bars and the ancestry proportion of each genetic cluster by the size of the color segment. The light blue cluster includes samples that grouped together with *C. flava* in the BROAD ADMIXTURE analysis and likely represent morphologically misidentified or admixed specimens (see Fig. S8). We assigned individuals to genetic groups according to ancestry proportions: pure *C. jemtlandica* (>0.8 blue cluster), pure *C. lepidocarpa* (>0.8 red cluster), hybrids between the two (<0.8 red and blue cluster), and *C. flava*-related (>0.2 light blue cluster), as indicated in the figure. A more detailed figure with sample IDs is provided as Fig. S6.

Palmgren (1926) treated the two as subspecies of *C. lepidocarpa*. Subsequently, he elevated both to the level of species (Palmgren, 1937). The taxonomic treatment of the two species globally, however, varies considerably, even among the Nordic countries. Herein, we demonstrate that the two entities form two genetic clusters that largely correspond to the morphologically identified *C. jemtlandica* and *C. lepidocarpa*. We further show that genetically intermediate forms are associated with intermediate morphologies as well as higher individual and population level heterozygosity, indicative of hybridization. Moreover, our results show that both taxa remain distinct in sympatric populations. Finally, our ENM suggests that *C. jemtlandica* and *C. lepidocarpa* are associated with different ecological niches (separated by precipitation, temperature, and soil pH). On the basis of the sum of evidence from our integrative approach, we argue for recognizing *C. jemtlandica* as a subspecies of *C. lepidocarpa*.

4.1 Hybrid detection and genetic diversity

Our results reveal that around half (46%) of the admixed individuals occur in sampling sites where both *C. jemtlandica* and *C. lepidocarpa* are present (Fig. 2C). In addition, admixed individuals displayed varying degrees of ancestry proportions between the *C. jemtlandica* and *C. lepidocarpa* clusters (Fig. 1). This provides support for our first hypothesis that hybridization and introgression is ongoing between *C. jemtlandica* and *C. lepidocarpa* in sympatry. Occurrences of admixed individuals in localities with only one parental taxon present can potentially be explained by sampling effort. Hybridization is further supported by the larger proportion of heterozygous loci in the admixed individuals (Fig. 2B) as well as lower F_{IS}, compared with the pure individuals (Table 5).

Our results agree with previous studies of hybridization in this species complex. Hedrén's (2002) allozyme studies showed that individuals that were morphological intermediates of *C. jemtlandica* and *C. lepidocarpa* were also genetically admixed. Schmid (1982) demonstrated that F1-hybrids, resulting from several species crosses in section *Ceratocystis*, could produce fertile pollen, enabling backcrossing with the parental species, a prerequisite for genetic introgression to occur. It is therefore widely understood that introgression is possible and relatively common between members of section *Ceratocystis* growing in geographic proximity (Schmid, 1983). Blackstock & Ashton (2010) studied natural hybrid populations resulting from section *Ceratocystis* taxa crossings in 36 sites in North America, Europe, and the British Isles. On the basis of allozyme and morphometric studies, they found that hybrids between *C. flava* and *C. lepidocarpa* with some pollen fertility were able to backcross with *C. lepidocarpa*. The same backcrossing pattern in natural hybrid populations was later documented by morphometric studies also in Poland (Więcław & Wilhelm, 2014). Neither of these two studies included *C. jemtlandica*.

As *C. lepidocarpa* is considered more closely related to *C. jemtlandica* than it is to *C. flava*, we would expect

Table 3 Misidentification within the *C. jemtlandica*–*C. lepidocarpa* complex based on differences in morphological and genetic assignment of 91 Norwegian individuals

Morphological assignment	New genetic assignment	Misidentification within taxa (%)
C. lepidocarpa (16)	*C. jemtlandica* (1)	6.25
	C. lepidocarpa x *jemtlandica* (1)	6.25
	C. flava/*C. flava* admixed (1)	6.25
C. jemtlandica (65)	*C. lepidocarpa* (1)	1.54
	C. lepidocarpa × *jemtlandica* (5)	7.69
	C. flava/*C. flava* admixed (9)	13.85
C. lepidocarpa x *jemtlandica* (10)	*C. lepidocarpa* (2)	20
	C. jemtlandica (1)	10
	C. flava/*C. flava* admixed (0)	0

The number of samples is given within brackets.

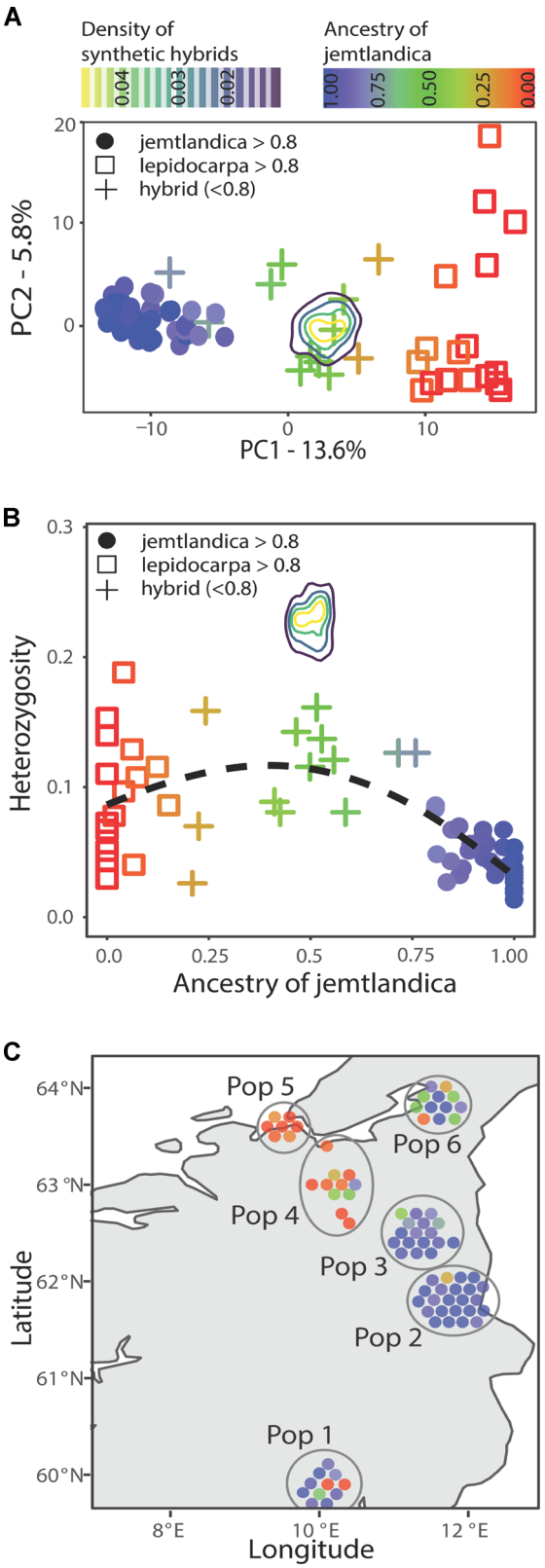


Fig. 2. Continued

Fig. 2. Hybrid detection between *C. jemtlandica* and *C. lepidocarpa* in Norway, based on 372 SNPs and 81 samples. Individuals are displayed as symbols and colored by their respective ADMIXTURE ancestry proportions of *C. jemtlandica* ranging from 1 (pure *C. jemtlandica*) to 0 (pure *C. lepidocarpa*). **A,** Principal component analysis displaying genetic variation explained by the first two PCs. **B,** Ancestry of *C. jemtlandica* plotted as a function of observed heterozygosity. Smoothed conditional mean is shown by the stippled line. Plots A and B include 1 000 synthetic F1-hybrids, simulated using HybridLab and displayed as bands with color representing the density of occurrences. **C,** Distribution of *C. jemtlandica* ancestry across geographically defined populations. Points are spread out to prevent overlap. Hence, individual sampling sites are not precisely depicted.

introgression between *C. lepidocarpa* and *C. jemtlandica* to be even more common. Furthermore, putative hybrid swarms between *C. jemtlandica* and *C. lepidocarpa*, which locally obscured their morphological distinctiveness, have been reported in parts of Sweden (Gotland, Uppland, Dalarna; Hedrén & Prentice, 1996). Backcrossing with parental taxa could potentially also explain why the admixed individuals in our study had lower levels of heterozygosity than the simulated F1-hybrids (Fig. 2B). Relatively high and positive F_{IS} in pure *C. lepidocarpa* and *C. jemtlandica* indicates some level of inbreeding in these two species (Table 5). However, while F_{IS} was lower in the admixed individuals than in both pure clusters (consistent with hybrid origin), it was nevertheless not negative ($F_{IS} = 1 - H_{obs}/H_{exp}$). It is possible that inbreeding (e.g., self-fertilization) among the admixed individuals counteracts the increased heterozygosity after hybridization. Self-compatibility is indeed widespread in *Carex* (Friedman & Barrett, 2009). This could also explain the lower heterozygosity among the admixed individuals, relative to the simulated hybrids.

On the basis of our current sampling, *C. lepidocarpa* is genetically more diverse than *C. jemtlandica* (Figs. 2A, 2B; Table 5), corroborating findings by Hedrén & Prentice (1996), who demonstrated lower allozyme diversity in *C. jemtlandica*. They interpreted this result as an indication that *C. jemtlandica* is

Table 4 Overview of the geographically defined populations and their number of assigned individuals from each genetic group.

Population	Total	*C. jemtlandica*	*C. lepidocarpa*	Hybrids
1	11	8	2	1
2	23	22	0	1
3	17	14	0	3
4	11	1	7	3
5	7	0	7	0
6	12	6	1	5

Table 5 Estimates of genetic diversity for the *C. jemtlandica*–*C. lepidocarpa* complex in Norway, presented as the mean and standard deviation (SD) of all geographically defined populations (Fig. 2C; Table 4)

Genetic group	N pops[†]	H_{obs}[‡] (SD)	H_{exp}[§] (SD)	F_{IS}[¶] (SD)
C. jemtlandica	5	0.08 (0.046)	0.098 (0.047)	0.210 (0.109)
C. lepidocarpa	4	0.135 (0.038)	0.160 (0.034)	0.165 (0.061)
Hybrids	5	0.16 (0)	0.164 (0.021)	0.010 (0.152)

[†]Number of populations, [‡] Observed heterozygosity, [§] Expected heterozygosity assuming HWE across the total population, [¶]Inbreeding coefficient.

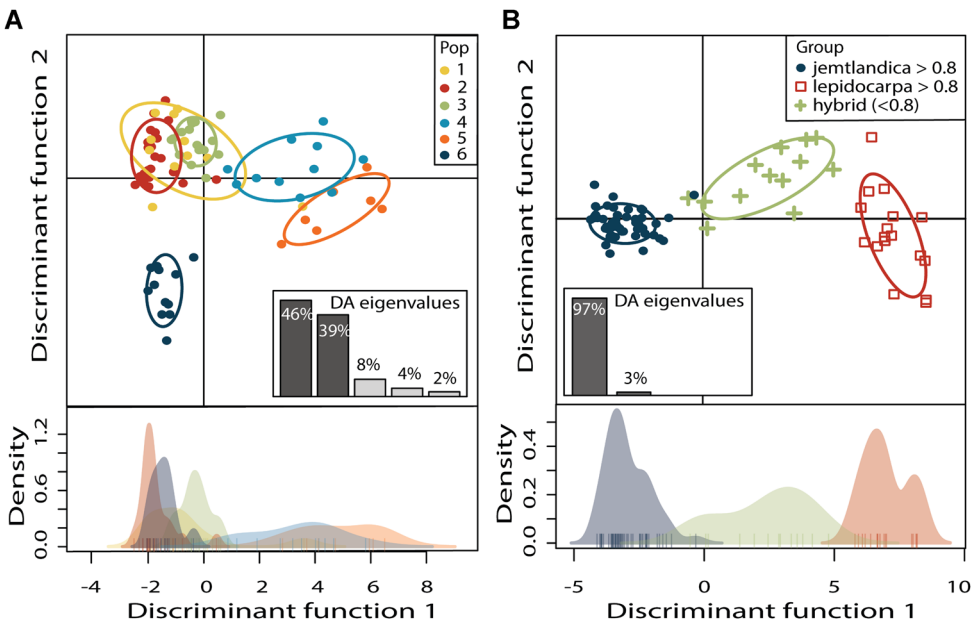


Fig. 3. Discriminant analysis of principal components (DAPC) based on 372 SNPs distributed across 81 Norwegian individuals of the *C. jemtlandica*–*C. lepidocarpa* complex. Samples were a priori separated in (**A**) six geographical populations and (**B**) three genetic groups (pure *C. jemtlandica*, pure *C. lepidocarpa*, and hybrids). In the scatter plot (upper), individuals are represented as symbols, and predefined groups are shown by different colors and inertia ellipses. The barplot with DA eigenvalues (middle) displays the proportion of genetic information that is explained by each consecutive discriminant function. The density plot (lower) presents the distribution of each predefined group on the first discriminant function, in their respective colors.

currently in the process of evolving/diverging from *C. lepidocarpa* or from a common ancestor. The higher level of genetic variability in seemingly pure *C. lepidocarpa* may, however, rather be a result of frequent hybridization between *C. lepidocarpa* and other species of section *Ceratocystis*. Stabilized cryptic backcrossing has been shown to result in retention of typical *C. lepidocarpa* morphology, despite of introgressed genetic material (Schmid, 1982). In such cases, the introgression can only be revealed by molecular data. One supporting factor in this respect may be that *C. lepidocarpa* has more overlap in its natural geographical distribution, including niche preference, with other species of section *Ceratocystis* than *C. jemtlandica*, and hence is more likely to experience introgession through hybridization. Results from our admixture analyses, when increasing the number of taxa included, do indeed suggest introgression from *C. demissa* to *C. lepidocarpa* (Fig. S8).

Our second hypothesis that the morphology-based taxonomic assignments are corroborated by molecular data also gains support from our investigations. Our blind-test assessment of taxonomic assignment shows that individuals of the blue-colored cluster are morphologically typical *C. jemtlandica* and that those of the red-colored cluster are typical *C. lepidocarpa* (Fig. 1). Among a total of 91 inspected specimens, three genetically "pure" specimens were incorrectly assigned to the other species and three were wrongly identified as hybrids (Table 2). More in-depth studies of morphological plasticity related to varying ecological conditions might provide some explanatory power to these few exceptions. Morphometric studies of *C. jemtlandica* and *C. lepidocarpa* (Pykälä & Toivonen, 1994; Hedrén, 2002; Blackstock, 2007) have identified extensive overlap in all morphological characters separating them. Moreover, Więcław (2014, 2017) found that vegetative traits in section *Ceratocystis* in Poland were highly plastic and varied considerably under different habitat conditions. The distinguishing morpho-

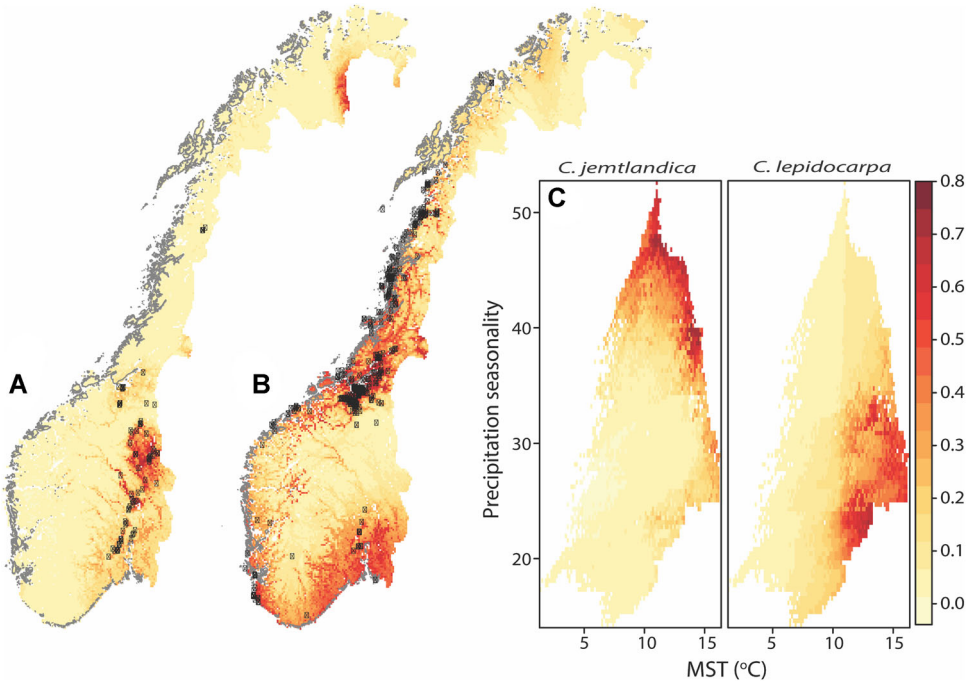


Fig. 4. Predicted habitat suitability for (**A**) *C. jemtlandica* and (**B**) *C. lepidocarpa* across Norway. Highest habitat suitability is indicated by dark red color and lowest suitability by pale yellow color. Black marks on the map represent the occurrence data of *C. jemtlandica* (**A,** 174 records) and *C. lepidocarpa* (**B,** 546 records) used in the ecological niche modeling. **C,** Visualizations of habitat suitability niches for *C. jemtlandica* (left) and *C. lepidocarpa* (right) using the two most influential bioclimatic predictors: precipitation seasonality (CV% of monthly precipitation) and mean summer temperature (°C). Darker red colors represent higher suitability.

logical characters described by Palmgren (1959; Table 1), which we used for identification, include both vegetative and generative traits, in addition to shape of organs and the ratios between them. The three latter (generative traits, shape, and ratio) have shown to be less prone to plasticity and, hence, should be the most useful for species identification (Hedrén, 1990; Więcław, 2014, 2017). Among our total of 91 inspected specimens, only three genetically admixed specimens were incorrectly assigned as "pure" species. According to our results, all individuals with intermediate morphologies (except the three incorrectly assigned as hybrids) do indeed correspond with the strongly admixed individuals (0.4–0.6 ancestry proportions). Hence, whereas the pure individuals of both species are likely to be correctly assigned based on morphology alone, there seems to be an unbalanced cline of intermediate morphologies among the admixed individuals. Overall, however, the genetic variation does indeed seem to reflect the morphological variation of *C. jemtlandica* and *C. lepidocarpa*.

4.2 Niche suitability

Species distribution patterns can be shaped by historical, biotic, and abiotic factors (Dupin & Smith, 2019). On the basis of the distribution patterns of *C. jemtlandica* and *C. lepidocarpa*, respectively, we hypothesized that adaptations to different environmental conditions may be a likely explanation. Our ENM results show high divergence in spatial distribution and habitat suitability between *C. jemtlandica* and *C. lepidocarpa* (Fig. 4). According to our findings, the *C. jemtlandica* niche is anchored in regions with high precipitation seasonality (strongest explanatory factor; Fig. S13), whereas the niche of *C. lepidocarpa* is in regions with lower precipitation seasonality and higher temperatures. High precipitation seasonality is typically found in the continental parts of Norway. The modeling also identified a potentially suitable climatic niche space for *C. jemtlandica* in Finnmark (Fig. 4B). Occurrence data of *C. jemtlandica* have not (yet) been registered from Finnmark. The species may have been overlooked, as sampling efforts and species observations in poorly populated parts of northern Norway are generally low (Speed et al., 2018). Alternatively, the realized distribution of the species has not yet reached its fundamental niche, or other dispersal or environmental factors may limit the species in this region.

Soil pH was not as strong a predictor of niche suitability as we had expected (Figs. S13, S14). This may be related to the spatial scale at which soil pH was quantified in this study (1 km cells) versus the scale at which species respond to edaphic conditions. Both *C. jemtlandica* and *C. lepidocarpa* are calciphile (Schmid, 1984; Pykälä & Toivonen, 1994) and occur in rich fens. However, the two species usually grow in separate parts of the mire: *C. jemtlandica* along the margins, which are more strongly influenced by calcareous spring water, and *C. lepidocarpa* more on open parts (Palmgren, 1959; Pykälä & Toivonen, 1994). Recent studies of soil

conditions have shown that *C. lepidocarpa* is associated with high soil Ca and $CaCO_3$, as well as high pH and C:N ratio (Więcław, 2014, 2017). *Carex jemtlandica* was not included in that study. It is possible, therefore, that their respective distributions are determined more by local variation of abiotic conditions, such as the level of pH and/or nutrient availability. Wetlands, which are typical habitats for both *C. jemtlandica* and *C. lepidocarpa*, have high levels of fine-scale environmental heterogeneity (Larkin, 2018). Available environmental data lack the necessary detail for inferring any effect of such fine-scale abiotic data in the present study. Future studies should collect detailed environmental data from sampling sites to assess the impact and explanatory power of local variation in environment factors.

4.3 Species cohesiveness, taxonomy, and implications for conservation

Our fourth hypothesis states that *C. jemtlandica* and *C. lepidocarpa* are genetically distinct to the degree that they may deserve recognition as ESUs. Alternatively, they could represent a clinal morphological and genetic variation primarily partitioned by geography (genetic isolation by distance, IBD). Compared with previous studies (Hedrén & Prentice, 1996; Hedrén, 2002; Blackstock, 2007), we herein find a clearer genetic separation of *C. jemtlandica* from *C. lepidocarpa*. Apart from the individuals that likely represent hybrids with *C. flava* or misidentified young specimens of this species (Fig. 1: light blue cluster), all visualizations of the RADseq data indicate two genetically distinct clusters corresponding to *C. jemtlandica* and *C. lepidocarpa*, respectively (Fig. 1: blue and red clusters). As *C. flava* occurs in sympatry with *C. jemtlandica* and/or *C. lepidocarpa* in the majority of the collection sites, the appearance of a third cluster consisting of *C. flava* hybrids and misidentified immature specimens was not unexpected.

According to the DAPC analysis, 97% of the genetic variation in our data was accounted for by the first discriminant function when grouping our samples as pure *C. jemtlandica*, pure *C. lepidocarpa*, and hybrids (Fig. 3B). On the contrary, grouping according to geography (using the six large populations defined by geographical affinity; Fig. 3A) explained only 46% of the genetic variation, rejecting the alternative hypothesis of clinal variation.

Finally, we wanted to evaluate whether *C. jemtlandica* and *C. lepidocarpa* both deserve recognition at species level. This study shows that *C. jemtlandica* and *C. lepidocarpa* represent two independent evolutionary trajectories that remain distinct despite ongoing hybridization. They additionally occupy different environmental niches and partially distinct geographical ranges. As such, it becomes clear that *C. jemtlandica* and *C. lepidocarpa* qualify as taxonomic units rather than ecotypes, the latter defined as geographically isolated populations occupying a particular ecological niche (Silva et al., 2017). But which taxonomic rank is appropriate? Should *C. jemtlandica* and/or *C. lepidocarpa* be recognized at the level of species, subspecies, or varieties? Species delineation, and ultimately taxonomic rank, will depend on the applied species concept. This is particularly relevant for young or early diverging lineages, as the defining properties of the different species concepts may not necessarily develop at the same time or in a fixed order (de Queiroz, 2007), which may result in conflicting alternative species concepts. This is probably one of the reasons for the numerous existing, and incongruent, taxonomic treatments within section *Ceratocystis*, including that of *C. jemtlandica* and *C. lepidocarpa* (see examples in Blackstock, 2007).

The ecological species concept requires species to occupy different ecological niches (Van Valen, 1976; Andersson, 1990). Accordingly, our ENM results lend support for the recognition of *C. jemtlandica* and *C. lepidocarpa* at species level. Following the biological-, genotypic cluster-, and morphological species concepts (Mayr, 1942; Cronquist, 1978; Mallet, 1995), the "species status" of *C. jemtlandica* is weakened by the presence of genetic and phenotypic intermediates, even though a proportion of these genetically admixed individuals displayed reduced pollen fertility. On the other side, fertile, or partly fertile, interspecific F1-hybrids have been documented in some plant genera (e.g., in *Geum* and *Draba*; Gajewski, 1959; Grundt et al., 2006), and interspecific gene flow occurs between "good" plant species (see Rieseberg et al., 2004, and references therein). Apparently, the gene flow between *C. jemtlandica* and *C. lepidocarpa* does not substantially affect their evolutionary distinctiveness. Compared with accepted species within other sections of *Carex* known for interspecific hybridization (e.g., the section *Phacocystis* and *Vesicariae*; Pedersen et al., 2016; Nowak et al., 2020), however, the level of introgression between *C. jemtlandica* and *C. lepidocarpa* is high.

Concluding on the appropriate taxonomic level for *C. jemtlandica* and *C. lepidocarpa* is clearly not straightforward. Following the general metapopulation lineage species concept (De Queiroz, 2007), any one clear evidence of lineage separation may be sufficient to infer species limits. Multiple and complementary lines of evidence, however, will give higher confidence. Considering the relatively high amount of gene flow, compared to between other hybridizing species pairs in *Carex*, and that Norway may be the sole country to currently acknowledge *C. jemtlandica* at species level, we conclude on rejecting our last hypothesis. Hence, based on our integrative approach, and to contribute toward a stable global taxonomy, we recommend lowering the taxonomic status of *C. jemtlandica* to subspecies. Further insight on the evolutionary history of *C. jemtlandica* and *C. lepidocarpa* requires expanded geographic sampling.

Both *C. jemtlandica* and *C. lepidocarpa* occur on the current national red list of species in Norway. The low genetic diversity in *C. jemtlandica* makes this taxon potentially more vulnerable to environmental change than *C. lepidocarpa*. Therefore, alongside recommendation of lowering the taxonomic status of *C. jemtlandica* to subspecies, we strongly argue for the recognition of both subspecies as ESUs worthy of conservation. Highly restricted interlineage gene flow is the criterion set by Fraser & Bernatchez (2001) for qualifying ESU. Their framework does, however, emphasizes a case-by-case flexibility and the importance of considering both ecological and genetic data. We argue that even though there is gene flow between *C. jemtlandica* and *C. lepidocarpa*, we have justified their recognition as ESUs based on multiple sources of evidence. As indicated by our DAPC results (Fig. 3) and F_{ST} calculations, the gene flow barriers between *C. jemtlandica* and *C. lepidocarpa* are sufficient to retain their genetic distinctiveness despite ongoing hybridization.

Acknowledgements

The authors acknowledge the O and TRH herbaria for kindly providing loans of material for their study. They thank Erik Boström for technical support in the molecular lab and Kristine Westergaard for constructive feedback on an early draft of the manuscript. They also acknowledge Hans Stenøien for fruitful input during initial stages of the project and NTNU Genomic Core Facility (https://www.ntnu.edu/dmf/gcf) for sequencing their RADseq libraries.

References

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664. https://doi.org/10.1101/gr.094052.109

Andersson L. 1990. The driving force: Species concepts and ecology. *Taxon* 39: 375–382.

Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data. Available from http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Avise JC. 1989. A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology & Evolution* 4: 279–281.

Bakkestuen V, Erikstad L, Halvorsen R. 2008. Step-less models for regional environmental variation in Norway. *Journal of Biogeography* 35: 1906–1922.

Balint M, Domisch S, Engelhardt CHM, Haase P, Lehrian S, Sauer J, Theissinger K, Pauls SU, Nowak C. 2011. Cryptic biodiversity loss linked to global climate change. *Nature Climate Change* 1: 313–318.

Bendiksby M, Mazzoni S, Jørgensen MH, Halvorsen R, Holien H. 2014. Combining genetic analyses of archived specimens with distribution modelling to explain the anomalous distribution of the rare lichen *Staurolemma omphalarioides*: Long-distance dispersal or vicariance? *Journal of Biogeography* 41: 2020–2031.

Blackstock N. 2007. *A reassessment of the yellow sedges*–Carex flava *L. agg. (Cyperaceae) in the British Isles*. Ph.D. Dissertation. Lancaster: University of Lancaster.

Blackstock N, Ashton PA. 2010. Genetic markers and morphometric analysis reveal past hybridization and introgression in putative *Carex flava* L. s.str. (Cyperaceae) hybrid populations. *Plant Systematics and Evolution* 287: 37–47.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cres WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3: e3376.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology* 22: 3124–3140.

Coates DJ, Byrne M, Moritz C. 2018. Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution* 6: 165.

Crins WJ, Ball PW. 1989. Taxonomy of the *Carex flava* complex (Cyperaceae) in North America and northern Eurasia. II. Taxonomic treatment. *Canadian Journal of Botany* 67: 1048–1065.

Cronquist A. 1978. Once again, what is a species? In: Knutson LV ed. *Biosystematics in agriculture*. Montclair, NJ: Allenheld Osmin. 3–20.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry TS, McVean G, Durbin R. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* 56: 879–886.

Derieg NJ, Weil SJ, Reznicek AA, Bruederle LP. 2013. *Carex viridistellata* sp. nov. (Cyperaceae), a new cryptic species from Prairie Fens of the Eastern United States. *Systematic Botany* 38: 82–91.

Dupin J, Smith SD. 2019. Integrating historical biogeography and environmental niche evolution to understand the geographic distribution of Datureae. *American Journal of Botany* 106: 667–678.

Elith J, Leathwick JR. 2009. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40: 677–697.

Elven R. 2013. *Carex jemtlandica* (Palmgr.) Palmgr. In: Elven R, Fremstad E, Pedersen O eds. *Distribution maps of Norwegian vascular plants. IV. The Eastern and Northeastern elements*. Trondheim: Akademika Publishing. 123–125.

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. 2012. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: Orgogozo V, Rockman M eds. *Molecular methods for evolutionary genetics. Methods in molecular biology (methods and protocols)*. Totowa, NJ: Humana Press. 772: 157–178.

Fick SE, Hijmans RJ. 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302–4315.

Frankham R. 2003. Genetics and conservation biology. *Comptes Rendus Biologies* 326: 22–29.

Fraser DJ, Bernatchez L. 2001. Adaptive evolutionary conservation: Towards a unified concept for defining conservation units. *Molecular Ecology* 10(12): 2741–2752. https://onlinelibrary.wiley.com/doi/epdf/10.1046/j.0962-1083.2001.01411.x

Friedman J, Barrett SCH. 2009. The consequences of monoecy and protogyny for mating in wind-pollinated *Carex*. *New Phytologist* 181: 489–497.

Gajewski W. 1959. Evolution in the Genus *Geum*. *Evolution* 13: 378–388.

GBIF.org. 19 November 2019. GBIF Occurrence Download. Available from https://doi.org/10.15468/dl.rnqnfh

Goudet J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 5: 184–186.

Grundt HH, Kjølner S, Borgen L, Rieseberg LH, Brochmann C. 2006. High biological species diversity in the arctic flora. *Proceedings of the National Academy of Sciences USA* 103: 972–975.

Guisan A, Thuiller W, Zimmermann NE. 2017. *Habitat suitability and distribution models: With applications in R*. Cambridge: Cambridge University Press.

Hedrén M. 1990. Problems in *Carex jemtlandica* and *C. bergrothii* (Cyperaceae) in Sweden. *Sommerfeltia* 11: 109–115.

Hedrén M. 2002. Patterns of allozyme and morphological differentiation in the *Carex flava* complex (Cyperaceae) in Fennoscandia. *Nordic Journal of Botany* 22: 257–301.

Hedrén M, Prentice HC. 1996. Allozyme variation and racial differentiation in Swedish *Carex lepidocarpa* s.l. (Cyperaceae). *Biological Journal of the Linnean Society* 59: 179–200.

Hengl T, de Jesus JM, MacMillan RA, Batjes NH, Heuvelink GBM, Ribeiro E, Samuel-Rosa A, Kempen B, Leenaars JGB, Walsh MG,

Gonzalez MR. 2014. SoilGrids1km—Global soil information based on automated mapping. *PLoS One* 9: e105992.

Jiménez-Mejías P, Benítez-Benítez C, Fernández-Mazuecos M, Martín-Bravo S. 2017. Cut from the same cloth: The convergent evolution of dwarf morphotypes of the *Carex flava* group (Cyperaceae) in Circum-Mediterranean mountains. *PLoS One* 12: e0189769.

Jiménez-Mejías P, Luceño M, Martín-Bravo S. 2014. Species boundaries within the Southwest old world populations of the *Carex flava* group (Cyperaceae). *Systematic Botany* 39: 117–131.

Jiménez-Mejías P, Martín-Bravo S, Luceño M. 2012. Systematics and taxonomy of *Carex* sect. *Ceratocystis* (Cyperaceae) in Europe: A molecular and cytogenetic approach. *Systematic Botany* 37: 382–398.

Jombart T, Ahmed I. 2011. Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070–3071.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics* 11: 94. https://bmcgenomdata.biomedcentral.com/articles/10.1186/1471-2156-11-94

Kassambara A, Mundt F. 2020. Factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7. https://CRAN.R-project.org/package=factoextra

Lande R, Shannon S. 1996. The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution* 50: 434–437.

Larkin DJ. 2018. Wetland heterogeneity. In: Finlayson CM, Everard M, Irvine K, McInnes RJ, Middleton BA, van Dam AA, Davidson NC eds. *The wetland book*. Dordrecht: Springer. 177-182.

Lücking R, Dal-Forno M, Sikaroodi M, Gillevet PM, Bungartz F, Moncada B, Yánez-Ayabaca A, Chaves JL, Coca LF, Lawrey JD. 2014. A single macrolichen constitutes hundreds of unrecognized species. *Proceedings of the National Academy of Sciences USA* 111: 11091–11096.

Maddison WP, Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21–30.

Malinsky M, Trucchi E, Lawson DJ, Falush D. 2018. RADpainter and fineRADstructure: Population inference from RADseq Data. *Molecular Biology and Evolution* 35: 1284–1290.

Mallet J. 1995. A species definition for the modern synthesis. *Trends in Ecology & Evolution* 10: 294–299.

Mayr E. 1942. *Systematics and the origin of species*. New York: NY: Columbia University Press.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17: 240–248.

Naimi B, Araújo MB. 2016. sdm: A reproducible and extensible R platform for species distribution modelling. *Ecography* 39: 368–375.

Nielsen EE, Bach LA, Kotlicki P. 2006. HYBRIDLAB (version 1.0): A program for generating simulated hybrids from population samples. *Molecular Ecology Notes* 6: 971–973.

Nielsen R, Slatkin M. 2013. *An introduction to population genetics*. Sunderland, MA: Sinauer Associates.

Nowak MD, Pedersen ATM, Brysting AK, Schrøder-Nielsen A, Elven R, Bjorå CS. 2020. Testing hypotheses of hybrid origins for two seashore species of *Carex* section *Phacocystis* (Cyperaceae). *Botanical Journal of the Linnean Society* 194: 100–117.

Nygård MØ. 2016. *Population Genetic Study of* Carex jemtlandica *and* C. lepidocarpa *sect.* Ceratocystis *(Cyperaceae), using RAD-seq*. Master thesis. NTNU Open. Available from http://hdl.handle.net/11250/2407589

Palmgren A. 1937. *Carex jemtlandica*, en ny *Carex*-art. *Memoranda Societatis pro Fauna et Flora Fennica* 13: 126–128.

Palmgren A. 1926. *Carex lepidocarpa* ssp. *jemtlandica*. In: Lindman CAM ed. *Svensk Fanerogam flora*. 2nd ed. Stockholm: P. A. Norstedt & Söners Förlag. 153.

Palmgren A. 1959. *Carex-gruppen Fulvellae Fr. i Fennoskandien I*. Helsingfors: Societas pro fauna et flora fennica.

Paris JR, Stevens JR, Catchen JM. 2017. Lost in parameter space: A road map for STACKS. *Methods in Ecology and Evolution* 8: 1360–1373.

Pearman PB, D′Amen M, Graham CH, Thuiller W, Zimmermann NE. 2010. Within-taxon niche structure: Niche conservatism, divergence and predicted effects of climate change. *Ecography* 33: 990–1003.

Pedersen ATM, Nowak MD, Brysting AK, Elven R, Bjorå CS. 2016. Hybrid origins of *Carex rostrata* var. *borealis* and *C. stenolepis*, two problematic taxa in *Carex* section *Vesicariae* (Cyperaceae). *PLoS One* 33: 990–1003.

Pykälä J, Toivonen H. 1994. Taxonomy of the *Carex flava* complex (Cyperaceae) in Finland. *Nordic Journal of Botany* 14: 173–191.

Rieseberg LH, Church SA, Morjan CL. 2004. Integration of populations and differentiation of species. *New Phytologist* 161: 59–69.

Rotreklová O, Bureš P, Řepka R, Grulich V, Šmarda P, Hralová I, Zedek F, Koutecký T. 2011. Chromosome numbers of *Carex*. *Preslia* 83: 25–58.

Schmid B. 1982. Karyology and hybridization in the *Carex flava* complex in Switzerland. *Feddes Repertorium* 93: 23–59.

Schmid B. 1983. Notes on the nomenclature and taxonomy of the *Carex flava* group in Europe. *Watsonia* 14: 309–319.

Schmid B. 1984. Niche width and variation within and between populations in colonizing species (*Carex flava* group). *Oecologia* 63: 1–5.

Silva TL, Vale CG, Godinho R, Fellous A, Hingrat Y, Alves PC, Abáigar T, Brito JC. 2017. Ecotypes and evolutionary significant units in endangered North African gazelles. *Biological Journal of the Linnean Society* 122: 286–300.

Speed JDM, Austrheim G. 2017. The importance of herbivore density and management as determinants of the distribution of rare plant species. *Biological Conservation* 205: 77–84.

Speed JDM, Bendiksby M, Finstad AG, Hassel K, Kolstad AL, Prestø T. 2018. Contrasting spatial, temporal and environmental patterns in observation and specimen based species occurrence data. *PLoS One* 13: e0196417.

Van Valen L. 1976. Ecological species, multispecies, and oaks. *Taxon* 25: 233–239.

Więcław H. 2014. *Carex flava* agg. (section *Ceratocystis*, Cyperaceae) in Poland: Taxonomy, morphological variation, and soil conditions. *Biodiversity Research and Conservation* 33: 3–47.

Więcław H. 2017. Within-species variation among populations of the *Carex flava* complex as a function of habitat conditions. *Plant Ecology & Diversity* 10: 443–451.

Więcław H, Wilhelm M. 2014. Natural hybridization within the *Carex flava* complex (Cyperaceae) in Poland: Morphometric studies. *Annales Botanici Fennici* 51: 129–147.

Yousefi N, Hassel K, Flatberg KI, Kemppainen P, Trucchi E, Shaw AJ, Kyrkjeeide MO, Szövényi P, Stenøien HK. 2017. Divergent

evolution and niche differentiation within the common peatmoss *Sphagnum magellanicum*. *American Journal of Botany* 104: 1060–1072.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12743/suppinfo:

Table S1. Overview of samples from *Carex* section *Ceratocystis* used for the present molecular study. Vouchers are sorted according to our collection ID. Abbreviations in sample ID: d= *C. demissa*, f = *C. flava*; h= *C. hostiana*; j = *C. jemtlandica*; l = *C. lepidocarpa*; v = *C. viridula* var. *viridula*, p = *C. viridula* var. *pulchella*, cf. = uncertain determination. Combination of abbreviations (d, f, h, j, l, v, p) in collection ID indicates putative hybrids. Collectors: AJ = Arne Jakobsen, HS = Heidi Solstad, KIF = Kjell Ivar Flatberg, LG = Leif Galten, MB = Mika Bendiksby, MØN = Malene Østreng Nygård, RE = Reidar Elven. Vouchers not yet registered are marked with x in column Herb.number.

Table S2. Overview of retained reads after demultiplexing for individual samples, and number of SNPs, mean depth, and missing data after final filtering. Samples with empty columns did not meet our filtering criteria and were excluded from the dataset.

Table S3. Filtered occurrence data for Norwegian *C. jemtlandica* and *C. lepidocarpa* used in ecological niche modeling.

Table S4. Overview of assembled loci, polymorphic loci, and SNPs present in at least 80% of the population (r80) for different values of parameter *n* (maximum allowed distance between loci from different individuals to be considered homologs) in Stacks (Catchen et al., 2013).

Table S5. Total number of assembled loci and polymorphisms, in addition to distribution of SNPs across loci under default and optimized key parameters in Stacks (m5M1n1; Catchen et al., 2013).

Table S6. Overview of the genetic groups defined for the STRICT dataset. Abbreviations: lep = *C. lepidocarpa*, jemt = *C. jemtlandica*, fla = *C. flava*, anc. = ancestry proportions, Hobs = observed heterozygosity. In sample ID: j = *C. jemtlandica*, l = *C. lepidocarpa*, lj/jl = *C. jemtlandica* x *lepidocarpa*, cf. = uncertain determination.

Fig. S1. Environmental variables and background data used for ecological niche modeling (ENM) of *C. jemtlandica* and *C. lepidocarpa* across Norway. A, Mean temperature of the warmest quarter (°C), B, mean annual precipitation (mm), C, precipitation seasonality (coefficient of variance of monthly precipitation, CV%), D, soil pH, and E, background data created by sampling 1 K random points across Norway weighted by distribution of occurrence data of all vascular plants in Norway. The bioclimatic data are scaled to 1-km grid cells.

Fig. S2. Mean sequence depth merged over all loci for individual samples under different values of parameter *m* (minimum reads required to form a stack) in Stacks (Catchen et al., 2013). All other parameters were kept at default values ($M = 2$, $n = 1$).

Fig. S3. Average number of assembled loci (first row), polymorphic loci (second row), and SNPs (in thousands; third row) obtained for individual samples at different values of parameter *m* and *M* in Stacks (Catchen et al., 2013). Only one variable was varied at a time, all others were kept at default values ($m = 3$, $M = 2$, $n = 1$). The fourth row displays the r80 values: the number of assembled loci (tags), polymorphic loci (alleles), and SNPs present in at least 80% of the population for the respective parameter values.

Fig. S4. Number of polymorphic loci added for each iteration of Stacks parameter *M* (Catchen et al., 2013) sheared among at least 80% of the samples (r80).

Fig. S5. A, Cross-validation error and B, loglikelihood for the STRICT dataset when testing *K* from 1 to 15 using ADMIXTURE (Alexander et al., 2009). We performed 10 replicates for every given value of *K*. Optimal number of clusters was found at *K*=3.

Fig. S6. Genetic structure of the STRICT dataset, including 372 SNPs and 93 individuals of phenotypic *C. jemtlandica*, *C. lepidocarpa*, and hybrids. The fineRADstructure (Malinsky et al., 2018) heat map depicts pairwise coancestry among individuals. Dark color (black) indicate high relatedness between individuals and bright color (yellow) represents low relatedness. Posterior probabilities are displayed on the branches of the fineRADstructure tree. ADMIXTURE (Alexander et al., 2009) results are provided for $K = 2$ to $K = 15$. Individuals are represented by separate vertical bars, and the ancestry proportion of each cluster by the size of the color segment. Optimal number of clusters was $K = 3$ according to the cross-validation test (Fig. S5A). Samples are ordered according to fineRADstructure. Abbreviations in sample ID represent phenotypic identification: j = *C. jemtlandica*, l = *C. lepidocarpa*, lj/jl = *C. jemtlandica* x *lepidocarpa*, and cf. = uncertain determination. Combination of abbreviations (j, l) in collection ID indicates putative hybrid.

Fig. S7. ADMIXTURE (Alexander et al., 2009) results at optimal $K = 3$ for the STRICT dataset comprising 372 SNPs and 93 individuals of phenotypic *C. jemtlandica*, *C. lepidocarpa*, and hybrids. Individuals are represented by vertical bars and the ancestry proportion of each cluster by the size of the color segment. The light blue cluster includes samples that grouped together with *C. flava* in the BROAD ADMIXTURE analysis and likely represent morphologically misidentified or admixed specimens (see Fig. S8). We assigned individuals to genetic groups according to ancestry proportions: pure *C. jemtlandica* (>0.8 of blue cluster), pure *C. lepidocarpa* (>0.8 of red cluster), hybrids (<0.8 of blue and red cluster), and *C. flava*-related (>0.2 of light blue cluster), as indicated by the figure. Abbreviations in sample ID represent phenotypic identification: j = *C. jemtlandica*, l = *C. lepidocarpa*, lj/jl = *C. jemtlandica* x *lepidocarpa*, cf. = uncertain determination.

Fig. S8. Genetic structure of the BROAD dataset, including 372 SNPs and 134 individuals of *Carex* sect. *Ceratocystis*. ADMIXTURE (Alexander et al., 2009) results are provided for $K = 2$ to $K = 20$. Individuals are represented by separate vertical bars, and the ancestry proportion of each cluster by the size of the color segment. Optimal number of clusters was found at $K = 4$ (Fig. S9). Samples are ordered according to the fineRADstructure

results for the BROAD dataset (not included). Individuals comprising the third cluster in the STRICT dataset (light blue; Figs. 1 and S7) are marked with black arrows. Abbreviations in sample ID represent phenotypic identification: d = *C. demissa*, f = *C. flava*, h = *C. hostiana*, j = *C. jemtlandica*, l = *C. lepidocarpa*, p = *C. viridula* var. *pulchella*, v = *C. viridula* var. *viridula*, and cf. = uncertain determination. Combination of abbreviations (d, f, h, j, l, v, p) in sample name indicates putative hybrids.

Fig. S9. A, Cross-validation error and B, loglikelihood for the BROAD dataset when testing *K* form to 20 using ADMIXTURE (Alexander et al., 2009). We performed 10 replicates for every given value of *K*. Optimal number of clusters was found at $K = 4$.

Fig. S10. Amount of genetic variation explained by the successive principal components of the PCA analysis (Fig. 2A).

Fig. S11. Robustness of genetic signal in the STRICT dataset. A, Correlation between the first principal component of the PCA analysis based on allele frequencies and ancestry proportions estimated by ADMIXTURE (Alexander et al., 2009). B, Correlation between the first principal component of the PCA analyses based on allele frequencies and coancestry estimates from fineRADstructure (Malinsky et al., 2018).

Fig. S12. Screening for null alleles. A, Observed heterozygosity in putative hybrids as a function of allele frequencies, B, observed heterozygosity in synthetic hybrids as a function of allele frequencies, and C, observed heterozygosity as a function of missing data per locus. The presence of null alleles does not explain the discrepancy in heterozygosity between putatively and simulated hybrids (Fig. 2B), as indicated by the lack of a significant correlation between observed heterozygosity and per locus missing data.

Fig. S13. Variable importance in spatial predictions for *C. jemtlandica* and *C. lepidocarpa* across Norway, based on, respectively, 174 and 546 species occurrence records downloaded from the Global Biodiversity Information Facility (GBIF). The first three variables represent the three main axes of bioclimatic variation within Norway: Mean temperature of the warmest quarter (MST), mean annual precipitation (MAP), and precipitation seasonality (Speed & Austrheim, 2017).

Fig. S14. Species response curves for the selected environmental variables. Increasing response yields higher habitat suitability and decreasing response yields lower habitat suitability. Red lines denote *C. jemtlandica* and black lines denote *C. lepidocarpa*. In general, *C. lepidocarpa* displays higher habitat suitability as compared with *C. jemtlandica* due to the unbalanced number of included occurrence records, respectively, 546 and 174 occurrences. MST, mean temperature of the warmest quarter; MAP, mean annual precipitation.

Script S1. Shell scripts used for collecting information on assembled alleles, polymorphic loci, and SNPs under different Stacks key parameters.