*Botanical Journal of the Linnean Society*, 2020, **194**, 100–117. With 5 figures.



# Testing hypotheses of hybrid origins for two seashore species of *Carex* section *Phacocystis* (Cyperaceae)

MICHAEL D. NOWAK[1,#,*], A. TIRIL M. PEDERSEN[1,#], ANNE K. BRYSTING[2], AUDUN SCHRØDER-NIELSEN[1], REIDAR ELVEN[1] and CHARLOTTE S. BJORÅ[1]

[1]*Natural History Museum, University of Oslo, PO Box 1172, Blindern, N-0318 Oslo, Norway*
[2]*Centre for Ecological and Evolutionary Synthesis, PO Box 1066, Blindern, N-0316 Oslo, Norway*

*Received 17 September 2019; revised 11 December 2019; accepted for publication 13 April 2020*

Taxonomists have proposed numerous hybrid species in plants, but to gain a better understanding of the role that hybridization may play in plant diversification, such taxonomic hypotheses must be tested using genomic data. In this study, we employ ddRAD sequence data to test taxonomic hypotheses of hybrid origins in *Carex salina* and *C. ramenskii* (*Carex* section *Phacocystis*). Sequence data from multiple Norwegian and Icelandic populations of the putative hybrid and parental species were generated for hundreds of ddRAD loci. These data were used to estimate geographical structuring of genetic diversity and admixture and to explicitly test for hybrid origins using several analytical approaches. Our results indicate recurrent hybrid origins for the populations of *C. salina* and *C. ramenskii* sampled in our study and show that these populations are characterized by high interspecific heterozygosity. Our results support the idea that hybridization may indeed play an important role in the diversification of lineages of *Carex* and highlight the important role that clonal propagation might play in maintaining hybrid populations. Future studies focusing on a broader geographical sampling would be needed to assess if the genetic structuring in these Nordic populations reflects range-wide patterns in these hybrid lineages.

ADDITIONAL KEYWORDS: *Carex* – clonal propagation – Cyperaceae – ddRAD genotyping – genetic diversity – hybridization.

## INTRODUCTION

The formation of hybrids resulting from reproduction between two species is a relatively common phenomenon in plants (Stebbins, 1950; Abbott, Barton & Good, 2016), and estimates of the frequency of interspecific hybridization suggest that 25% or more of all plant species routinely produce hybrid offspring with other species of varying degrees of relatedness (Rieseberg, Wood & Baack, 2006; Mallet, 2005, 2007). Despite the relatively high rate of interspecific hybridization, there is considerable debate regarding the evolutionary importance of this phenomenon because the ultimate outcomes of hybridization are known to vary considerably (Abbott *et al.*, 2013). In some cases, hybridization may primarily result in the introgression of presumably advantageous alleles between parental species (Harrison & Larson, 2014), a process thought to be particularly important in hybrid zones (Lexer *et al.*, 2010). A second potential outcome of interspecific hybridization in plants is the establishment of a polyploid lineage (i.e. allopolyploidy) resulting from meiotic dysfunction in the F1 hybrid (e.g. the production of unreduced gametes; Mason & Pires, 2015). A third possible outcome of interspecific hybridization is the generation of a new homoploid hybrid species, which is reproductively isolated from the parental species due to ecological niche shifts or the evolution of other pre- or post-zygotic isolating mechanisms (Rieseberg, 1997; Buerkle *et al.*, 2000; Gross & Rieseberg, 2005; Abbott *et al.*, 2010).

The earliest stages of homoploid hybrid speciation are most likely to occur in close geographical proximity to one or both of the parental species (i.e. in sympatry; Buerkle *et al.*, 2000), and a newly formed hybrid lineage may be transient if gene flow with parental populations is persistent. Therefore, pre- and/or post-zygotic reproductive barriers must evolve rapidly for

*Corresponding author. E-mail: michaeldnowak@gmail.com
#These authors contributed equally

the hybrid lineage to have any chance of founding a lineage with an evolutionary trajectory independent of its parents (Rieseberg, 1997). A first-generation hybrid (F1) between two species is characterized by genome-wide patterns of interspecific heterozygosity, but if sufficient reproductive barriers to the parental species exist, and an F1 hybrid predominantly engages in self-fertilization or mating with other hybrid individuals, alleles from both parental species will become fixed throughout the genome of the hybrid lineage. This process has been referred to as 'genomic stabilization' (Buerkle & Rieseberg, 2008), and the outcome of such stabilization (e.g. which parental alleles become fixed) is expected to vary throughout the genome, leading to a mosaic of hybrid ancestry that can be shaped both by the random process of genetic drift and by the more deterministic process of natural selection and genetic linkage (Payseur & Rieseberg, 2016; Elgvin *et al.*, 2017).

Numerous homoploid hybrid species have been proposed by plant taxonomists (Nieto Feliner *et al.*, 2017), but such taxonomic hypotheses must be tested using a sufficiently large number of nuclear loci to evaluate patterns of genomic stabilization and variation in hybrid ancestry in several different populations (Schumer, Rosenthal & Andolfatto, 2014; Schumer *et al.*, 2016). Here, we apply genome-scale reduced representation library sequencing (ddRAD) to test taxonomic hypotheses of hybrid origins in two species of the cosmopolitan monocot plant genus *Carex* L. (Cyperaceae).

*Carex* is one of the most species-rich groups of vascular plants, consisting of perennial, rhizomatous herbs that form tussocks or mats, with representatives found in nearly all biomes, but with the greatest species diversity in Arctic and boreal wetlands (Reznicek, 1990; Ball & Reznicek, 2002). The importance of sexual vs. asexual (e.g. clonal) reproduction probably varies significantly among the different sections of the genus, but, to our knowledge, a systematic analysis of this trait has yet to be conducted. In the Arctic, species of *Carex* section *Phacocystis* Dumort. propagate extensively via clonal rhizomatous growth (Standley, 1990; Volkova *et al.*, 2008), often forming large mats that become fragmented over time, whereas species of *Carex* section *Ceratocystis* Dumort. tend to produce short rhizomes and thus probably colonize habitats primarily through sexual reproduction and seed dispersal. *Carex* spp. are also known as classic examples of intraspecific karyotypic diversity, and several species exhibit long aneuploid series of haploid chromosome numbers (e.g. base chromosome numbers ranging from $N = 6$ to $N = 56$; Hipp 2007; Roalson, 2008; Hipp, Rothrock & Roalson, 2009). This karyotypic diversity is probably caused by unlocalized centromeric activity of the holocentric chromosomes that are a key characteristic of the genus (Cayouette & Morriset, 1986; Kukkonen & Toivonen, 1988).

Taxonomists have proposed many putative hybrid species in *Carex* based on intermediate morphology. For example, in the North American sedge flora, which contains *c*. 420 named taxa, Cayouette & Catling (1992) reported 300 putative hybrid taxa of *Carex* and found that the vast majority of these hybrids occur in recently glaciated areas. Recent empirical studies of putative hybrid lineages in *Carex* suggest that hypotheses of hybrid origin may be supported in some lineages, but refuted in others. For example, by analysing data from 15 microsatellite loci and pollen fertility measurements, Pedersen *et al.* (2016) confirmed taxonomic hypotheses of the hybrid origins in *C. rostrata* Stokes var. *borealis* (Hartm.) Kük. and *C. stenolepis* Less. in *Carex* section *Vesicariae* (Heuff.) J.Carey. In contrast, Escudero *et al.* (2014) employed genotyping-by-sequencing data to reject a taxonomic hypothesis of hybrid origin for *C. waponahkikensis* Lovit & A.Haines. These two case studies highlight the important role that genetic and genomic studies can play in testing hypotheses of hybrid origins in *Carex*, which is a fundamental component in understanding the evolutionary history and diversification of this taxonomically complicated group of plants.

Taxonomic hypotheses of hybridization in *Carex* appear to be most common in certain sections of the genus (Kukkonen & Toivonen, 1988; Cayouette & Catling, 1992), and there is some evidence that hybridization may have been especially frequent in Arctic lineages (Toivonen, 1974) and among estuarine and palustrine species of the large *Carex* section *Phacocystis* Dumort. (Cayouette, 1987; Standley, 1990). One clade supported by ITS, ETS and *matK* sequence data in *Carex* section *Phacocystis* (Jiménez-Mejias *et al.*, 2016) contains five taxa known to play a dominant and ecologically important role in circumpolar Arctic and subarctic coastal ecosystems (Volkova *et al.*, 2008): *Carex lyngbyei* Hornem., *C. paleacea* Schreb. ex Wahlenb., *C. subspathacea* Wormsk. and two putative hybrid taxa, *C. salina* Wahlenb. and *C. ramenskii* Kom. (Kristinsson, 2010; Elven *et al.*, 2011); *C. ramenskii* has been previously referred to as *C. salina* (Mossberg & Stenberg, 2003). *Carex paleacea* and *C. salina* have distributions along the North Atlantic coasts of Scandinavia, north-western Russia and Canada. *Carex subspathacea*, *C. lyngbyei* and *C. ramenskii* are more widely distributed in the North Atlantic (including Iceland) and throughout the northern Pacific coasts of North America and Russia (see Supporting Information, Fig. S1 for range maps). Several taxonomists have suggested hybrid origins for *C. salina* and *C. ramenskii* because *C. salina* appears to be both morphologically and ecologically intermediate between *C. paleacea*

and *C. subspathacea*, and *C. ramenskii* appears to be morphologically and ecologically intermediate between *C. lyngbyei* and *C. subspathacea* (Cayouette & Morriset, 1985, 1986; Standley, Cayouette & Bruederle, 2002; Kristinsson, 2010; Elven *et al.*, 2011). In the current study, we use genomic data from Norwegian and Icelandic populations of *C. salina* and *C. ramenskii* to test if these species exhibit genetic compositions consistent with origins via interspecific hybridization. If these two species appear to have hybrid origins, they could represent examples of homoploid hybrid speciation given that previous cytological work shows no evidence of polyploidy (see Elven *et al.*, 2011, for a review). In the current study, we specifically aim to test taxonomic hypotheses of the hybrid origins of *C. salina* and *C. ramenskii*.

Two previous empirical studies have attempted to test the hypothesis of a hybrid origin of *C. salina* using isozymes (Standley, 1990) and AFLPs (Volkova *et al.*, 2008). In a study examining isozyme polymorphism in Canadian populations of *C. salina*, Standley (1990) hypothesized a hybrid origin of *C. salina* from the parental species *C. paleacea* and *C. subspathacea*. Volkova *et al.* (2008) employed dominant molecular markers (AFLPs in Volkova *et al.*, 2008), which make it impossible to identify loci exhibiting interspecific heterozygosity. Current approaches in high-throughput reduced representation library sequencing allow the genotyping of thousands of loci randomly distributed throughout the genome in non-model species (Andrews *et al.*, 2016). Such data sets are well suited for studies of hybridization and introgression when analysed using modern analytical approaches capable of distinguishing between introgression and incomplete sorting of ancestral polymorphism (Eaton & Ree, 2013; Escudero *et al.*, 2014; Rheindt *et al.*, 2014; Streicher *et al.*, 2014; Zinenko *et al.*, 2016). In this study, we apply high-throughput genotyping data (ddRAD sequencing) from populations representing a subset of the geographical range of each species to show that the putative hybrid species *C. salina* and *C. ramenskii* indeed originated through interspecific hybridization between *C. subspathacea* and either *C. paleacea* (*C. salina*) or *C. lyngbyei* (*C. ramenskii*). Moreover, we use these data to investigate whether the hybrids appear to have formed once or multiple times and whether their hybrid origins appear to be relatively recent or ancient (e.g. followed by genomic stabilization) thereby clarifying the evolutionary history of this clade.

## MATERIAL AND METHODS

### Taxon sampling

We sampled 44 (12 *C. paleacea*, nine *C. salina*, seven *C. subspathacea*, nine *C. ramenskii* and seven *C. lyngbyei*) individuals for genetic analysis. Sampling was conducted in the summer of 2013 in Norwegian and Icelandic populations of the five target species (Fig. 1, Supporting Information, Table S1). We were unable to sample populations representing the entire distributions of the five target species (e.g. see Fig. S1). Only a single individual was sampled per population because our field observations indicated that there is a high probability that a single genet that has propagated itself clonally dominates each population (A.T.M. Pedersen, C.S. Bjorå & R. Elven, pers. obs.), similarly to our previous observations in species of *Carex* section *Vesicariae* (Pedersen *et al.*, 2016). Consistent with this assumption, we found that genotypes from individuals of the same population were always identical in a small pilot experiment using microsatellites (Pedersen and Bjorå, unpublished data). Small amounts of fresh leaf material from each shoot were dried in silica gel and used for molecular analyses and the shoot was pressed and used as the voucher specimen. Leaf tissue samples and associated vouchers are deposited in the herbarium of the Natural History Museum, University of Oslo (O; Supporting Information, Table S1).

### DNA extraction, ddRAD library preparation and sequencing

DNA was extracted from 10–20 mg of dry leaf tissue using the Qiagen DNeasy Plant mini kit following the manufacturer's protocol. Genomic DNA was quantified using a Qubit 2.0 fluorometer. Double digest restriction site associated DNA (ddRAD) sequencing was performed using modifications to the protocol described in Pukk *et al.* (2015) and Vivian-Smith & Sønstebø (2017). Genomic DNA digestion and ligation of 150–300 ng gDNA per sample was performed in 40 µl total volume consisting additionally of 1× NEB buffer 4, 10 U *PstI*, 10 U *NdeI*, 250 µM rATP, 0.5 µM P1 adapter, 0.5 µM barcoded A adapter and 400 U T4 ligase in a thermal cycler set to 1 h at 37 °C, 10 min at 65 °C and a slow cool to 4 °C. At this point, the barcoded samples were pooled and fragments from 390–490 bp were isolated using a Blue Pippin (Sage Science). The libraries were amplified in 100 µl using 10 U of Q5 HiFi polymerase and primers designed to anneal to the A and P1 adapters under the following cycling conditions: 98 °C for 30 s (98 °C for 10 s, 60 °C for 15 s, 72 °C for 15 s) × 14, 72 °C for 2 min. After each step in the protocol, the products were cleaned once using 0.8 volumes of Ampure XP beads. Final libraries were cleaned twice with 0.65 volumes of Ampure XP beads to remove all short molecules before sequencing. Library quantification was performed on a Fragment Analyzer (Advanced Analytical) and sequenced on an ION Torrent PGM (Thermo Fisher) using Ion 316 chips, multiplexing between 11 and 14 samples per chip.

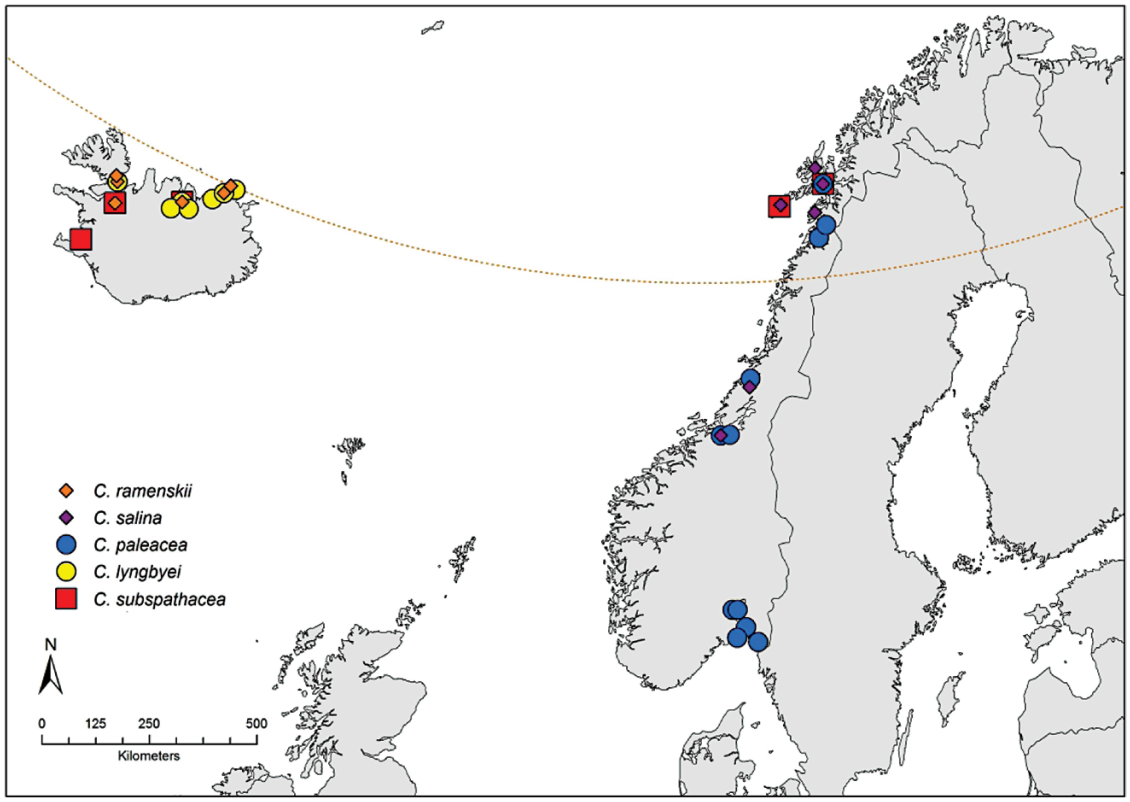


**Figure 1.** Sampling localities for all specimens of the five species of *Carex* section *Phacocystis* included in this study. Collection information for all specimens can be found in Supporting information, Table S1. Range maps for each species are shown in Supporting information, Fig. S1.

### Processing of raw data, filtration and variant calling

The raw reads were demultiplexed using the Torrent Suite Software v.5.0.5 and exported in fastq format for further processing. The raw fastq files were processed to trim adapter sequences and to identify valid ddRAD loci (i.e. those fragments containing the 5′ restriction site overhang sequence). The ddRAD library prep employed *PstI* as the rare-cutting enzyme and thus each valid ddRAD locus is expected to begin with the partial *PstI* motif TGCAG. Since the Ion Torrent platform produces variable read lengths, the 3′ partial restriction site overhang of *NdeI* (GTAT) was only used to identify trailing adapter sequence. Allowing for a single nucleotide error in the *PstI* motif, all reads starting with any sequence other than this motif were removed using cutadapt v.1.4.1 (Martin, 2011), which was also used to remove any contaminating adapter sequences and trailing sequence from each read, discarding any reads < 100 nucleotides long following adapter trimming. To facilitate further analyses in the absence of a reference genome, we produced a chimaeric pseudo-reference (CPR) genome of ddRAD loci for this study by conducting a *de novo* assembly of all of the cleaned sequence data with the mira v.4.0.2 assembler (Chevreux, Wetter & Suhai, 1999) using default settings for Ion Torrent data. Reads were mapped to the CPR using the bwa mem algorithm (Li & Durbin, 2010) with default settings, and variants were called using samtools mpileup v.1.3.1 (Li *et al.*, 2009; Li, 2011) and VarScan v.2.4.2 (Koboldt *et al.*, 2012) with a minimum coverage of 10× required for a valid SNP call. Using vcftools v.0.1.13 (Danecek *et al.*, 2011), the data set was filtered to include only biallelic SNPs, and any individual with > 50% missing data was removed (samples T171_2, T192_4, T227_4, T286_5 and T355_4). At this point the data set including all five taxa comprised a total of 39 (11 *C. paleacea*, eight *C. salina*, five *C. subspathacea*, eight *C. ramenskii* and seven *C. lyngbyei*) individual samples genotyped at 10 332 SNPs for 2145 ddRAD loci.

### Population genetic parameters and population structure

To reduce the potential for including singleton SNPs that may be due to sequencing errors, our data set was further filtered to include only biallelic SNPs with a

maximum of 20% missing data, and only SNPs with a minor allele count > 4 were retained. Population genetic statistics and $F_{ST}$ values were then calculated in GenAlEx v.6.502 (Peakall & Smouse, 2006, 2012) for 1659 SNPs at 264 loci in 39 samples. Prior to running Bayesian cluster analyses using STRUCTURE v.2.3.3 (Pritchard, Stephens & Donnelly, 2000) to assess genetic admixture, we selected one random SNP per ddRAD locus (264 loci in total) to reduce the impact of linkage disequilibrium. We ran five replicate runs for each value of $K$ from 1 to 10, with each run having a burn-in of 200 000 and 1 000 000 Markov chain Monte Carlo (MCMC) iterations, using the admixture model and correlated allele frequencies settings. The optimal value for $K$ was selected using STRUCTURE Harvester (Earl & vonHoldt, 2012), and cluster assignments were further inspected and visualized using CLUMPAK (Kopelman *et al.*, 2015). Genetic differentiation of the 39 samples was also evaluated using principal components analysis (PCA) in NTSYSpc v.2.11a (Rohlf, 2000) using the same data from 1659 SNPs at 264 ddRAD loci.

### Quantifying admixture and testing evolutionary hypotheses

To explore patterns of admixture in the putative hybrids, the data set was divided into two parts following the groupings implied by the STRUCTURE analysis. The '*C. salina*' data set contained all individuals of *C. salina* (eight), and its putative parental species *C. paleacea* (11) and *C. subspathacea* (five) and the '*C. ramenskii*' data set contained all individuals of *C. ramenskii* (eight) and its putative parental species *C. lyngbyei* (seven) and *C. subspathacea* (the same five individuals as above). Once again, only biallelic SNPs with a maximum of 20% missing data and a minor allele count > 4 were retained. The resulting *C. salina* data set consisted of 437 SNPs for 126 ddRAD loci genotyped in 24 individuals, and the *C. ramenskii* data set consisted of 1295 SNPs for 311 ddRAD loci genotyped in 20 individuals.

Each of these data sets was analysed using both STRUCTURE and BAPS v.6.0 (Corander & Marttinen, 2006; Corander *et al.*, 2008) to examine the genomic composition of the putative hybrids relative to their parental species (i.e. admixture coefficients, $Q$). As before, a single SNP per ddRAD locus was selected for each data set to reduce the impact of linkage disequilibrium (126 SNPs for the *C. salina* data set; 311 SNPs for the *C. ramenskii* data set). STRUCTURE runs were performed on these two data sets using the same settings as described above. The BAPS analysis was performed to estimate admixture proportions of the putative hybrids based on pre-defined clustering and allowing for two genetic clusters (i.e. $K$ = 2) with the results based on 500 simulations from the posterior allele frequencies. Analyses with larger values of $K$ were performed in the absence of pre-defined clusters, but clustering based on $K$ = 2 was consistently the best fit to the data (results not shown). Genetic differentiation was also evaluated using PCA separately on the two hybrid taxa with their respective presumed parents (using the same SNP-reduced data sets as for STRUCTURE and BAPS). To further explore genomic patterns of admixture in the putative hybrids, we estimated interspecific heterozygosity and hybrid index for both data sets (*C. salina* = 437 SNPs; *C. ramenskii* = 1295 SNPs) using maximum likelihood with the R package INTROGRESS v.1.22 (Gompert & Buerkle, 2009, 2010). These results were then compared to 1000 synthetic F1 hybrid genotypes for each putative hybrid simulated by sampling alleles from the respective putative parental species using HybridLab v.1.1 (Nielsen, Bach & Kotlicki, 2006).

The historical relationships of the study species, treated here as populations (i.e. each taxon = one population), were examined using two types of analyses using TreeMix v.1.12 (Pickrell & Pritchard, 2012). First, the 'three population' test of Reich *et al.* (2009) was conducted to estimate the 'treeness' of three population trees. This test estimates an $f_3$ statistic for each of the three species, and a significantly negative value implies a history of admixture. These tests were performed multiple times using three SNP block sizes (1, 10, 100). TreeMix v.1.12 was then used to estimate maximum-likelihood population graphs based on allele frequencies. For these analyses, we produced data sets that included samples from an outgroup species for rooting. The resulting *C. salina* data set included 1659 SNP genotypes (distributed among 264 ddRAD loci, no loci had > 20% missing data) for all of the *C. salina* (eight), *C. paleacea* (11) and *C. subspathacea* (five) samples, including the seven *C. lyngbyei* samples as an outgroup. The *C. ramenskii* data set included the same 1659 SNPs genotyped for all of the *C. ramenskii* (eight), *C. lyngbyei* (seven) and *C. subspathacea* (five) samples, including the 11 *C. paleacea* samples as an outgroup. The TreeMix analyses were conducted multiple times with three SNP block sizes (1, 10, 100) using sample size correction and allowing for a single migration event. Population graph confidence was quantified by performing 1000 bootstrap replicates.

To test for a hybrid origin of *C. salina* and *C. ramenskii*, three evolutionary scenarios were compared in a coalescence framework using approximate Bayesian computation with the DIYABC v.2.1 package (Cornuet *et al.*, 2014). The analyses of the *C. salina* and *C. ramenskii* data sets were conducted similarly, with only minor deviations as indicated,

using the versions of the data sets containing one randomly selected SNP per ddRAD locus (i.e. 126 SNPs for the *C. salina* data set and 311 SNPs for the *C. ramenskii* data set). In the first scenario (Scenario 1, Supporting Information, Fig. S2) the putative hybrid lineages (i.e. *C. salina* and *C. ramenskii*, respectively) were the product of admixture between the two parental species (i.e. *C. subspathacea* and either *C. paleacea* or *C. lyngbyei*). The second scenario (Scenario 2, Supporting Information, Fig. S2) modelled the putative hybrid lineages as splitting more recently from *C. subspathacea*, and the third scenario (Scenario 3, Supporting Information, Fig. S2) modelled the putative hybrid lineages splitting more recently from *C. paleacea* or *C. lyngbyei*, respectively. Uniform prior distributions were placed on the time of origin for the putative hybrids (100–50 000 generations), the age of the most recent common ancestor of all samples (1000–100 000 generations) and the effective population size of each species (e.g. N1, N2 and N3 in Supporting Information, Fig. S2; prior uniformly distributed between 100 and 30 000 for all effective population sizes). Three million data sets were simulated (one million data sets for each of the three scenarios). The three scenarios were compared both by direct estimation approach (e.g. counting scenario frequencies among the simulated data sets that are most similar to the observed parameters; Miller *et al.*, 2005) and through logistic regression of the probability of each scenario for the most similar simulated data sets on the deviations between simulated and observed summary statistics (Fagundes *et al.*, 2007; see Supporting Information, Fig. S3). Following the guidelines outlined in the manual for the direct estimation approach, 0.1% of the simulated data sets closest to the observed values were used; for the logistic regression, 1% of the closest simulated data sets were used. Simulating 1000 pseudo-replicates drawn from the prior distributions of parameters was used to assess confidence in scenario choice. The summary statistics of these pseudo-replicates were replaced by discriminant scores of a linear discriminant analysis (Estoup *et al.*, 2012) for the two alternative scenarios relative to the scenario of hybrid origin (Scenario 1). The proportion of pseudo-replicates in which the scenario of hybrid origin had the highest posterior probability served as an estimate of type II error.

The posterior probability of hybrid category group membership (i.e. pure, F1, F2 etc.) was estimated through MCMC simulation using NewHybrids v.1.1 (Anderson & Thompson, 2002; Anderson, 2008). This analysis was also conducted with the data sets containing one randomly selected SNP per ddRAD locus (i.e. 126 SNPs for the *C. salina* data set and 311 SNPs for the *C. ramenskii* data set), and samples from the putative parental species were identified as representing 'pure' samples from the parental allele frequency distributions and were thus not considered part of the mixture for estimating the π parameter (the vector of mixing proportions; Anderson & Thompson, 2002). The only exception to this was in the *C. salina* data set, in which a single sample of *C. paleacea* (T243_1) was included in the mixture and no prior was placed on group membership because the results of the STRUCTURE analyses and hybrid index estimates suggested an intermediate genotype for this individual. Jeffrey's priors were placed on π and θ (a parameter characterizing the multilocus allele frequencies), and duplicate runs using uniform priors were also performed, but this had no impact on the results. Following 100 000 generations of burn-in, the Markov chain was run for one million generations with a sample drawn once every 1000 generations from the posterior distribution. The trace of the π parameter was visually examined to ensure good mixing throughout the run.

## RESULTS

For the 39 individuals that we analysed in this study, we obtained a total of 8 469 920 raw IonTorrent sequence reads (Supporting Information, Table S2). The number of raw reads for each sample ranged from 36 538 to 586 567 (average number of reads 217 177), and the mean read length ranged from 250 to 301 bp (with an average of 277 bp). After filtering and mapping these reads to the pseudo-reference sequence (see Material and Methods), the total number of informative SNPs varied from 126 to 1659 depending on the data set (e.g. including all individuals vs. data sets that only included each putative hybrid with its respective parents) and whether one or more SNP per ddRAD locus was included (see Material and Methods).

### Population genetic parameters and population structure

Population genetic parameters were estimated using the data set including all 39 individuals genotyped at 1659 SNPs for 264 ddRAD loci (Table 1). Observed heterozygosity ($H_o$) in the putative parental species (*C. paleacea*, *C. subspathacea* and *C. lyngbyei*) is consistently close to or slightly higher than expected heterozygosity ($H_e$); in contrast, the putative hybrid species (*C. salina* and *C. ramenskii*) exhibit a consistently higher observed heterozygosity relative to expected heterozygosity. The number of private alleles observed in each species ranged from 40 to 116, which represents a relatively small proportion of the 1659 SNPs in the data set (e.g. 1.2–3.5%). Pairwise $F_{ST}$ estimates between the five species were

**Table 1.** Population genetic parameters estimated using the data set including all 39 individuals genotyped at 1659 SNPs for 264 ddRAD loci. Key: $H_e$ (expected heterozygosity), $H_o$ (observed heterozygosity), Sites (total number of polymorphic loci genotyped within each taxon, e.g. not including missing data), NewHybrids class membership and private alleles within each taxon (also presented as a percentage of all 1659 SNPs).

| Taxon | Individual | $H_e$ | $H_o$ | Sites | NewHybrids class membership (Scaled posterior probability) | Private alleles (% of alleles) |
|---|---|---|---|---|---|---|
| *C. paleacea* | | | | | | 116 (3.5%) |
| | T170_1 | 127.4 | 134 | 455 | Pure *C. paleacea* (1.0) | |
| | T173_5 | 126.6 | 149 | 449 | Pure *C. paleacea* (1.0) | |
| | T196_5 | 123.5 | 94 | 440 | Pure *C. paleacea* (1.0) | |
| | T204_3 | 110.5 | 114 | 403 | Pure *C. paleacea* (1.0) | |
| | T208_2 | 104.2 | 124 | 366 | Pure *C. paleacea* (1.0) | |
| | T215_1 | 127.5 | 149 | 455 | Pure *C. paleacea* (1.0) | |
| | T220_3 | 113.9 | 140 | 381 | Pure *C. paleacea* (1.0) | |
| | T237_4 | 126.1 | 116 | 454 | Pure *C. paleacea* (1.0) | |
| | T246_3 | 118.1 | 143 | 426 | Pure *C. paleacea* (1.0) | |
| | T259_5 | 120.2 | 144 | 438 | Pure *C. paleacea* (1.0) | |
| | T243_1 | 126.8 | 323 | 456 | Backcross between *C. paleacea* × *C. subspathacea* (F1) and pure *C. paleacea* (1.0) | |
| *C. salina* | | | | | | 40 (1.2%) |
| | T172_5 | 278.7 | 404 | 632 | F1 between *C. paleacea* and *C. subspathacea* (1.0) | |
| | T185_4 | 308.6 | 515 | 699 | F1 between *C. paleacea* and *C. subspathacea* (1.0) | |
| | T250_5 | 312.8 | 479 | 707 | F1 between *C. paleacea* and *C. subspathacea* (1.0) | |
| | T252_1 | 312.6 | 569 | 706 | F1 between *C. paleacea* and *C. subspathacea* (1.0) | |
| | T260_1 | 301.1 | 475 | 683 | F1 between *C. paleacea* and *C. subspathacea* (1.0) | |
| | T277_4 | 287.5 | 477 | 658 | F1 between *C. paleacea* and *C. subspathacea* (1.0) | |
| | T278_3 | 306.1 | 453 | 689 | F1 between *C. paleacea* and *C. subspathacea* (1.0) | |
| | T279_1 | 283 | 404 | 641 | F1 between *C. paleacea* and *C. subspathacea* (1.0) | |
| *C. subspathacea* | | | | | | 43 (1.3%) |
| | T350_5 | 121.1 | 265 | 382 | Pure *C. subspathacea* (1.0) | |
| | C1360 | 89.9 | 73 | 307 | Pure *C. subspathacea* (1.0) | |
| | T258_5 | 122.4 | 95 | 383 | Pure *C. subspathacea* (1.0) | |
| | T276_1 | 118 | 95 | 375 | Pure *C. subspathacea* (1.0) | |
| | T345_5 | 120.6 | 84 | 379 | Pure *C. subspathacea* (1.0) | |
| *C. ramenskii* | | | | | | 52 (1.6%) |
| | T331_3 | 441.5 | 550 | 1238 | F1 between *C. subspathacea* and *C. lyngbyei* (1.0) | |
| | T333_5 | 428.3 | 589 | 1206 | F1 between *C. subspathacea* and *C. lyngbyei* (1.0) | |
| | T339_5 | 441.6 | 678 | 1238 | F1 between *C. subspathacea* and *C. lyngbyei* (1.0) | |
| | T340_3 | 436.2 | 516 | 1208 | F1 between *C. subspathacea* and *C. lyngbyei* (1.0) | |
| | T347_4 | 432.3 | 591 | 1205 | F1 between *C. subspathacea* and *C. lyngbyei* (1.0) | |
| | T349_4 | 437.6 | 588 | 1229 | F1 between *C. subspathacea* and *C. lyngbyei* (1.0) | |
| | T354_5 | 430 | 536 | 1208 | F1 between *C. subspathacea* and *C. lyngbyei* (1.0) | |

**Table 1.** Continued

| Taxon | Individual | $H_e$ | $H_o$ | Sites | NewHybrids class membership (Scaled posterior probability) | Private alleles (% of alleles) |
|---|---|---|---|---|---|---|
| | T356_2 | 433.5 | 602 | 1209 | F1 between *C. subspathacea* and *C. lyngbyei* (1.0) | |
| *C. lyngbyei* | | | | | | 76 (2.3%) |
| | C1279 | 313.8 | 255 | 895 | Pure *C. lyngbyei* (1.0) | |
| | T321_2 | 322.3 | 410 | 915 | Pure *C. lyngbyei* (1.0) | |
| | T324_1 | 319.2 | 271 | 912 | Pure *C. lyngbyei* (1.0) | |
| | T329_3 | 334.9 | 353 | 942 | Pure *C. lyngbyei* (1.0) | |
| | T337_2 | 332.1 | 399 | 932 | Pure *C. lyngbyei* (1.0) | |
| | T338_1 | 326.7 | 321 | 930 | Pure *C. lyngbyei* (1.0) | |
| | T353_5 | 321.9 | 383 | 912 | Pure *C. lyngbyei* (1.0) | |

generally low (i.e. < 0.25) and ranged from 0.068 to 0.234. In the STRUCTURE analysis including all five taxa (39 samples, 264 SNPs; one SNP per ddRAD locus) the optimal number of genetic clusters was found to be three, with each of the parental species (*C. paleacea*, *C. subspathacea* and *C. lyngbyei*) forming distinct clusters and *C. salina* and *C. ramenskii* each combining a roughly equal number of alleles from each of their respective assumed parents (Fig. 2B). Two samples of the parental species (one *C. paleacea* and one *C. subspathacea*) seemingly combined genetic material from more than one cluster and are discussed below. The PCA analysis of the same data set showed a similar pattern, as the parental species were well separated along the first and second principal components (explaining 36.3 and 25.3% of the total variation, respectively) and both putative hybrid lineages appeared to be intermediate between their respective parents (Fig. 2A).

### Quantifying admixture and testing evolutionary hypotheses

In the results of the separate STRUCTURE analyses for the two hybrids and their respective parents (24 samples and 126 SNPs included in the *C. paleacea*, *C. salina* and *C. subspathacea* analysis, 20 samples and 311 SNPs included in the *C. lyngbyei*, *C. ramenskii* and *C. subspathacea* analysis), the optimum number of clusters in both analyses was found to be two. Furthermore, the parental species formed distinct clusters, whereas the hybrid taxa were composed of an approximately equal contribution of alleles from their respective parental gene pools (Figs 3A, 4A). Clustering based on $K = 2$ in BAPS revealed similar patterns to the genetic clusters observed from the Bayesian clustering analysis (Figs 3C, 4C). These results were also consistent with the PCA analyses of the *C. salina* and *C. ramenskii* data sets (Figs 3A, 4A). In the *C. salina* data set, the two parental taxa and their hybrid were all well separated along the first principal component axis (accounting for 57.4% of the total variation), with *C. salina* occupying an intermediate position between its parents. *Carex salina* was also well separated from its parents along the second principal component (accounting for 18.6% of the total variation). The same pattern occurred in the analysis of the *C. ramenskii* data set, in which the first and second principal components accounted for 48.3 and 15.3% of the total variation, respectively. The results of the STRUCTURE analysis of all five taxa showed that one sample of *C. paleacea* appears to be admixed between *C. paleacea* and *C. subspathacea*, and this individual (sample number T243_1) also occupied an intermediate position in the PCA. These results indicate that this individual may be

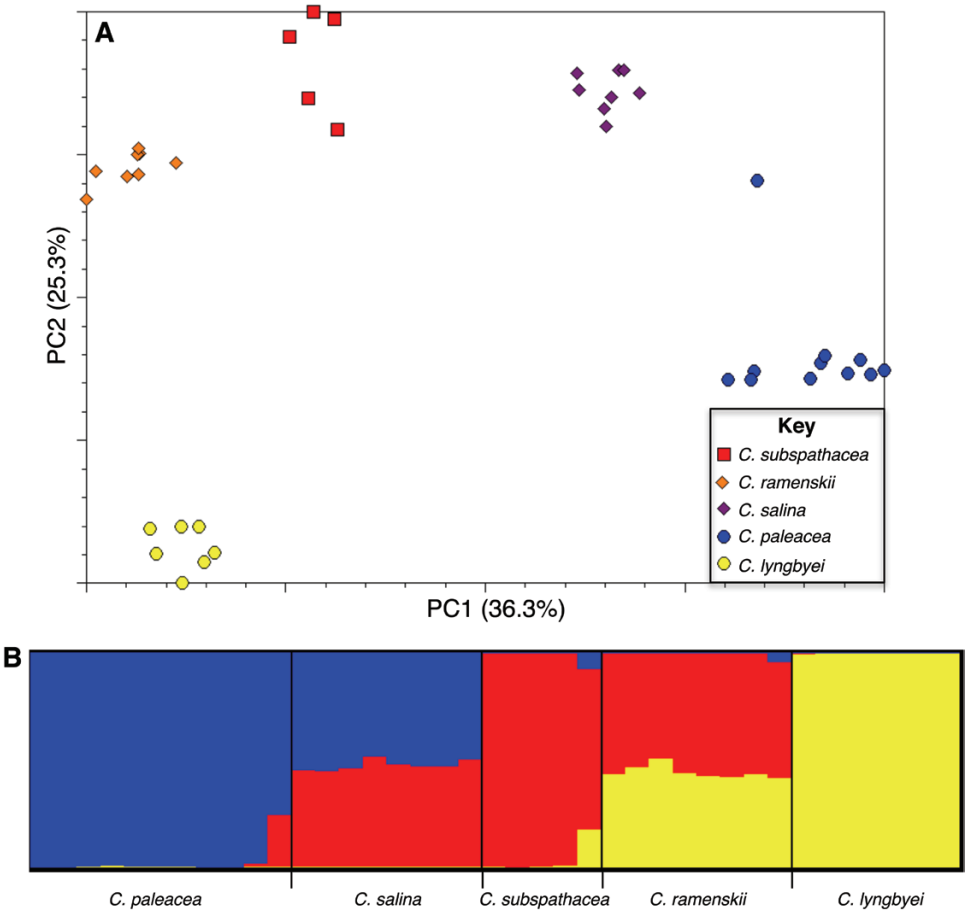


**Figure 2.** Genetic structure of 39 samples of *Carex* section *Phacocystis* sampled in this study. A, The first two axes of a principal components analysis (PCA) showing genetic differentiation for 264 SNPs on the same number of ddRAD loci. B, Graphical representation of the cluster assignment pattern for the best-fit model of genetic structure ($K$ = 3) based on a STRUCTURE analysis of the same 264 SNP loci. These results show that *C. salina* may be a hybrid between *C. subspathacea* and *C. paleacea*, whereas *C. ramenskii* may be a hybrid between *C. subspathacea* and *C. lyngbyei*.

a backcross between *C. salina* and one of its parent species. A careful morphological re-examination of the individual in question confirmed that it displayed characters from both *C. salina* and *C. paleacea*. This result was confirmed in the STRUCTURE and PCA analyses on the *C. salina* data set (Fig. 3A–E). One *C. subspathacea* individual also appeared to exhibit some admixture in the results of the STRUCTURE analysis of all five taxa, but this apparent admixture is not well supported based on PCA analysis of the five taxa data set. The apparent admixture in this individual is also not well supported in the BAPS and PCA analyses of the *C. salina* or *C. ramenskii* data sets. This individual contains slightly more missing data than other samples in our data set, and it does not deviate from the classic *C. subspathacea* morphology, so we do not suspect it to represent a backcross individual.

The results of INTROGRESS analyses for each hybrid taxon and its putative parents were consistent with the results of the STRUCTURE analyses and indicated extensive interspecific heterozygosity and considerable multilocus genotypic variability between samples of both *C. salina* (Fig. 3A) and *C. ramenskii* (Fig. 4A). Estimates of hybrid index for samples of *C. salina* ranged from 0.48 to 0.54, whereas the hybrid index estimated for samples of *C. ramenskii* ranged from 0.43 to 0.52. For comparison, the admixture coefficients from the STRUCTURE analyses showed that the genetic proportions estimated to have been contributed by *C. paleacea* ranged from 0.50 to 0.55 in the *C. salina* individuals, whereas the admixture proportions attributed to *C. lyngbyei* ranged from 0.44 to 0.53 in C. *ramenskii*. Admixture coefficients from *C. paleacea* estimated from the BAPS analyses ranged from 0.49 to 0.56 in *C. salina*, whereas the admixture

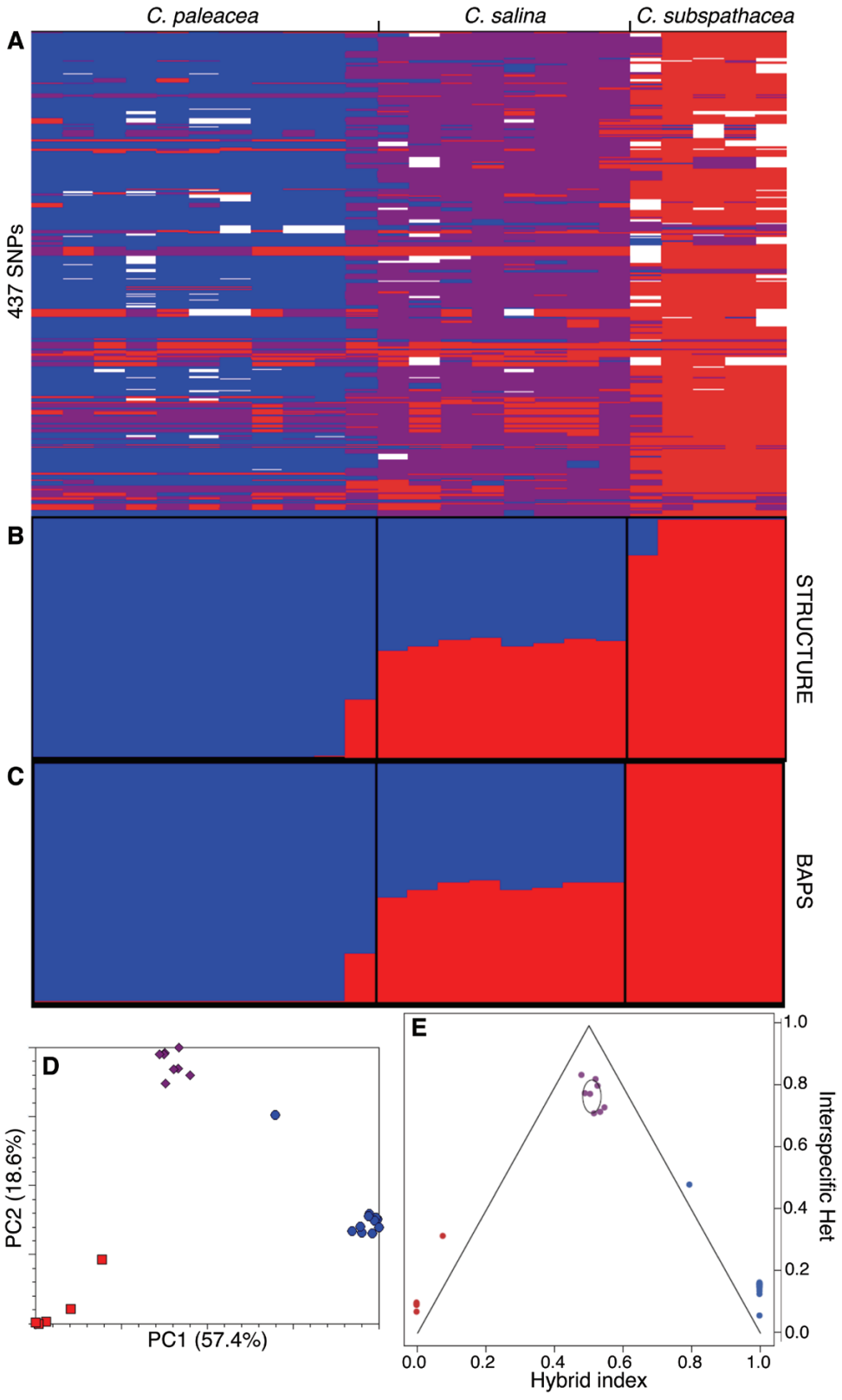


**Figure 3.** Genetic structure of the putative hybrid species *C. salina* and the parental species *C. subspathacea* and *C. paleacea*. A, Graphical representation of genotypes at 437 SNPs in the three species (24 individuals total). Each genotype

coefficients attributed to *C. lyngbyei* ranged from 0.40 to 0.49 in *C. ramenskii*. By comparing interspecific heterozygosity and hybrid index in the *C. salina* and *C. ramenskii* data sets with the 1000 simulated F1 hybrids for each taxon, it can be seen that the observed *C. salina* genotypes overlapped significantly with the synthetic *C. paleacea* × *C. subspathacea* hybrids (Fig. 3E), whereas the observed *C. ramenskii* genotypes deviated only slightly from hybrids simulated between *C. lyngbyei* and *C. subspathacea* (Fig. 4E). These results were consistent with those obtained using NewHybrids, where all samples assigned to *C. salina* and *C. ramenskii* had an F1 posterior probability of 1 (Table 1). NewHybrids also showed that sample T243_1, thought to be an intermediate genotype based on STRUCTURE, PCA, and hybrid index estimates, was estimated to be a backcross between a *C. paleacea* × *C. subspathacea* (i.e. *C. salina*) F1 and a pure *C. paleacea* individual by NewHybrids (posterior probability = 1).

The maximum-likelihood population trees generated in TreeMix were also consistent with hypotheses of hybrid origins for *C. salina* and *C. ramenskii*. In the analysis that included *C. paleacea*, *C. salina* and *C. subspathacea*, with *C. lyngbyei* as an outgroup (Fig. 5A), *C. salina* was grouped together with *C. paleacea*, but with significant gene flow from *C. subspathacea* to *C. salina* (indicated by a migration weight of 0.5), suggesting a hybridization event. The tree including *C. lyngbyei*, *C. ramenskii* and *C. subspathacea* with *C. paleacea* as an outgroup likewise grouped *C. ramenskii* with *C. subspathacea*, but also indicated significant gene flow from *C. lyngbyei* to *C. ramenskii* (as indicated by a migration weight of 0.5; Fig. 5B).

Using approximate Bayesian computation in the DIYABC software suite, it was found that a scenario in which the putative hybrid lineages were the product of admixture between the two parental species consistently exhibited the highest posterior probability for both putative hybrid species. The posterior probability of a scenario of hybrid origin (Scenario 1, Supporting Information, Fig. S2) for *C. salina* was > 0.99 (95% confidence intervals 0.96–1.00) by direct estimation and > 0.80 (95% confidence intervals 0.78–0.81) by logistic regression, whereas the posterior probability of a hybrid origin for *C. ramenskii* was > 0.72 (95% confidence intervals 0.33–1.00) by direct estimation and > 0.99 (95% confidence intervals 0.99–1.0) by logistic regression (Supporting Information, Fig. S3). Posterior predictive error estimates (i.e. the probability of selecting the wrong scenario) indicate high confidence in scenario choice for both *C. salina* (direct estimate: 0.007; logistic regression: 0.005) and *C. ramenskii* (direct estimate: 0.1; logistic regression: 0.043).

## DISCUSSION

Our results consistently support a scenario in which both *C. salina* and *C. ramenskii* originated through interspecific hybridization. In accordance with previous taxonomic hypotheses (Cayouette & Morriset, 1985, 1986; Standley *et al.*, 2002; Kristinsson, 2010; Elven *et al.*, 2011), our results indicate that Norwegian *C. salina* probably originated through hybridization between *C. paleacea* and *C. subspathacea*, and Icelandic *C. ramenskii* probably originated through hybridization between *C. lyngbyei* and *C. subspathacea*. Given the large variation observed in multilocus genotypes between individual samples (Figs 3, 4), it is most likely that both of these hybrid lineages formed multiple times independently, and it is possible that each of the populations sampled from *C. salina* and *C. ramenskii* may represent unique instances of hybridization between the respective parents. The phenomenon of multiple hybrid origins has been documented previously for both homoploid (Schwarzbach & Rieseberg, 2002) and polyploid (Dillenberger *et al.*, 2018) hybrid plant lineages, but to our knowledge, this is the first documentation in *Carex*.

Several previous studies have attempted to test the hypothesis of a hybrid origin of *C. salina* using molecular markers and data on pollen fertility. These studies have reported low pollen fertility and seed set and disturbed meiotic pairing in *C. salina* (Cayouette & Morriset, 1985, 1986; Dean *et al.*, 2008), and this feature has been used as evidence for a hybrid origin of this species. Standley (1990) examined polymorphism at three allozyme loci in several Canadian populations

is assigned a colour: homozygous sites for *C. paleacea* alleles are shown in blue, homozygous sites for *C. subspathacea* alleles are shown in red, and sites exhibiting interspecific heterozygosity are shown in purple. Estimation of allelic origin is based on allele frequency (Gompert & Buerkle, 2010). Missing genotypes are shown in white. B, C, Graphical representations of the cluster assignment patterns for $K = 2$ based on STRUCTURE (B) and BAPS (C) analyses of 126 SNPs. D, The first two axes of a PCA showing genetic differentiation at 126 SNP loci. E, Interspecific heterozygosity (Het) plotted against hybrid index based on genotypes at 1295 SNPs. The ellipse represents the range of parameters extracted from 1000 synthetic F1 hybrids simulated by sampling alleles from the respective parental species using HybridLab v.1.1 (see Methods). Key for D and E: blue = *C. paleacea*; red = *C. subspathacea*; purple = *C. salina*.

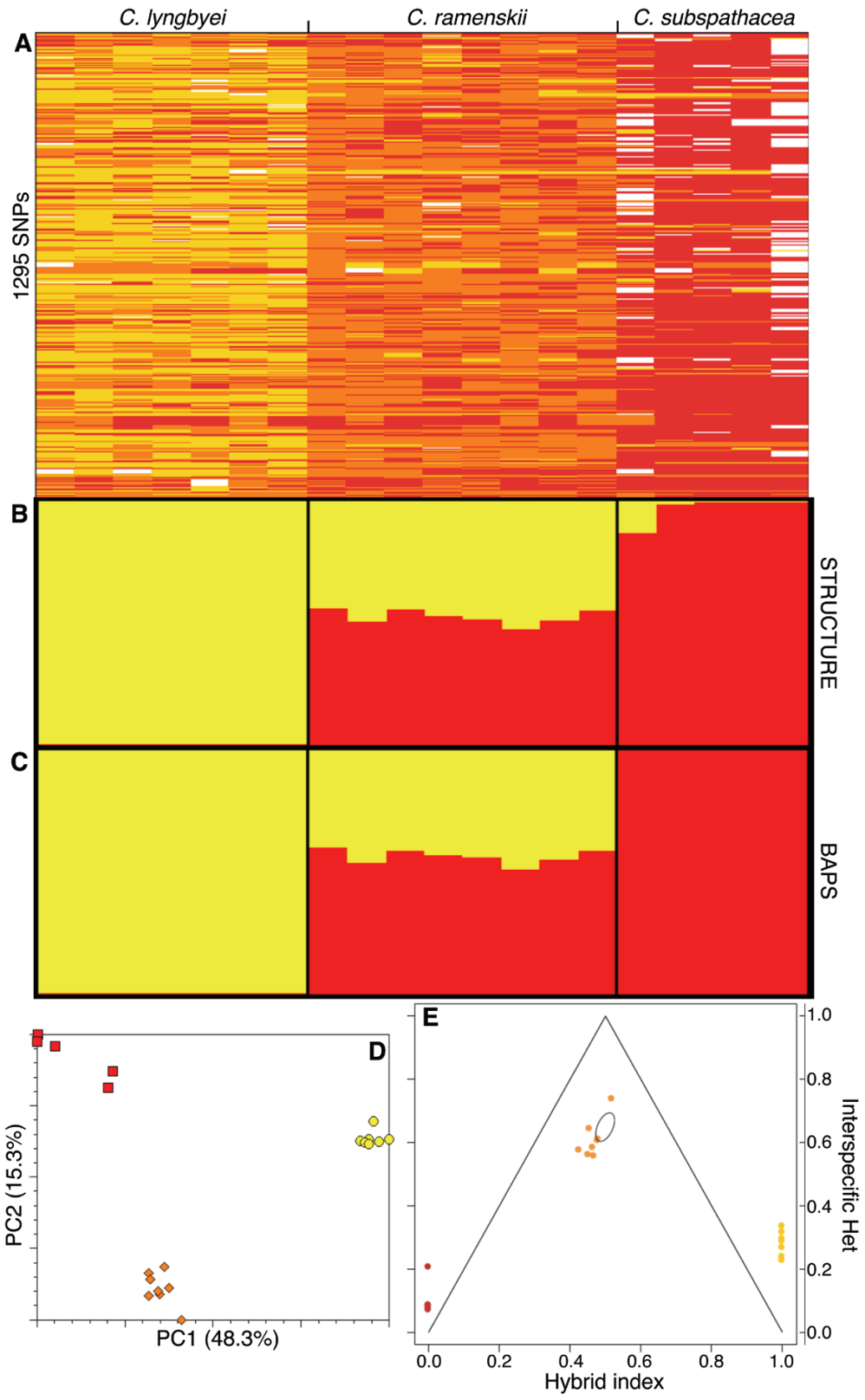


**Figure 4.** Genetic structure of the putative hybrid species *C. ramenskii* and the parental species *C. subspathacea* and *C. lyngbyei*. A, Graphical representation of genotypes at 1295 SNPs in the three species (20 individuals total). Each genotype

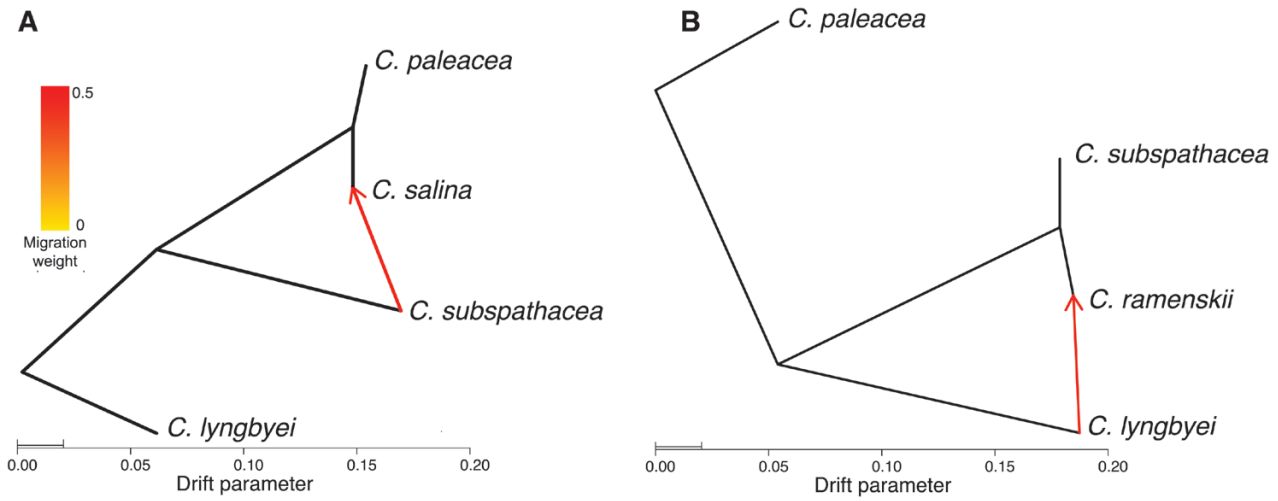


**Figure 5.** Maximum-likelihood population trees based on allele frequencies using TreeMix v.1.12. A, Population tree of *C. salina*, *C. subspathacea* and *C. paleacea*, rooted with *C. lyngbyei*. The results are based on genotypes at 1659 SNPs, and 1000 bootstrap replicates. A single migration event was modelled, and the results indicate that *C. salina* is composed of a contribution of *c*. 50% of alleles from each of *C. paleacea* and *C. subspathacea*. B, Population tree of *C. ramenskii*, *C. subspathacea* and *C. lyngbyei*, rooted with *C. paleacea*. The results are based on genotypes of the same 1659 SNPs, and 1000 bootstrap replicates. A single migration event was modelled, and the results indicate that *C. ramenskii* is composed of a contribution of *c*. 50% of alleles from each of *C. lyngbyei* and *C. subspathacea*.

and, consistent with our results, hypothesized a hybrid origin of *C. salina* from the parental species *C. paleacea* and *C. subspathacea*. Contrary to this, Volkova *et al.* (2008) used AFLP markers and hypothesized that *C. salina* represents a hybrid swarm between *C. subspathacea* and *C. recta* Boott (*s.s.*). However, as AFLP are dominant markers, these data are not well suited to evaluating patterns of admixture in early generation hybrids. No similar studies have previously addressed the question of hybrid origins in *C. ramenskii*.

The consistently high interspecific heterozygosity observed in all samples from *C. salina* and *C. ramenskii* in this study indicates that the sampled hybrid genotypes probably represent first-generation (F1) hybrids (Table 1). These results suggest that there has been either insufficient time or capacity for sexual reproduction to stabilize the genomes of these hybrid lineages. Given that several of the hybrid populations sampled in this study are clearly allopatric with respect to populations of their parental species (Fig. 1, Supporting Information, Table S1) and each putative hybrid species contains private allelic diversity (Table 1), it is unlikely that all of these populations represent recent instances of interspecific hybridization. Instead, our results are more likely the product of the combined effects of clonal propagation, partial hybrid sterility and pre- and post-zygotic isolating barriers between the hybrid lineages and their respective parental species. In our current study, we have not directly tested the relative impacts of these processes, but next we discuss aspects of the ecology and life history of these species that can help to interpret the genetic structure and composition of the hybrid lineages that we sampled.

All of the taxa examined in this study (including the parental species) predominantly reproduce asexually via clonal propagation. If clonal propagation, rather than seed dispersal following sexual reproduction, is the predominant mode of stand expansion in *C. salina*

is assigned a colour: homozygous sites for *C. lyngbyei* alleles are shown in yellow, homozygous sites for *C. subspathacea* alleles are shown in red, and sites exhibiting interspecific heterozygosity are shown in orange. Estimation of allelic origin is based on allele frequency (Gompert & Buerkle, 2010). Missing genotypes are shown in white. Graphical representations of the cluster assignment patterns for $K = 2$ based on (B) STRUCTURE and (C) BAPS analyses of 311 SNPs. D, The first two axes of a PCA showing genetic differentiation at 311 SNP loci. E, Interspecific heterozygosity (Het) plotted against hybrid index based on genotypes at 1295 SNPs. The ellipse represents the range of parameters extracted from 1000 synthetic F1 hybrids simulated by sampling alleles from the respective parental species using HybridLab v.1.1 (see Methods). Key for D and E: yellow = *C. lyngbyei*; red = *C. subspathacea*; orange = *C. ramenskii*.

**Table 2.** Pairwise $F_{ST}$ estimates between the five species of *Carex* section *Phacocystis* included in this study based on the analysis of 1659 SNPs for 264 ddRAD loci.

| | *C. paleacea* | *C. salina* | *C. subspathacea* | *C. ramenskii* |
|---|---|---|---|---|
| *C. salina* | 0.118 | | | |
| *C. subspathacea* | 0.130 | 0.234 | | |
| *C. ramenskii* | 0.120 | 0.101 | 0.098 | |
| *C. lyngbyei* | 0.219 | 0.220 | 0.126 | 0.068 |

and *C. ramenskii*, it is reasonable to assume that F1 genotypes could persist in perpetuity (e.g. Jónsdóttir *et al.*, 2000). Clonal reproduction may explain why samples of these putative hybrid species collected at sites where one or both parental species are missing still appear to carry genotypes consistent with F1 hybrids.

A second process capable of contributing directly to high interspecific heterozygosity of the putative hybrid *Carex* spp. in our study is hybrid sterility resulting from karyotypic differences between the parental species (Renaut *et al.*, 2014). Chromosome evolution has indeed been suggested to be an important driver of lineage diversification and radiation in *Carex* with its holocentric chromosomes (Hipp, 2007; Hipp *et al.*, 2009; Escudero, Hipp & Luceño, 2010; Escudero *et al.*, 2012; see Escudero *et al.*, 2016 for further references). A recent study by Escudero *et al.* (2016) found that chromosomal rearrangements in *Carex* can play a substantial role in creating post-zygotic reproductive isolation through F1 inviability and sterility in hybrids resulting from the crosses of two populations of *C. scoparia* Schkuhr that differ in chromosome numbers. As the three parental species examined here are diploid but modestly differ from one another in base chromosome number (for *C. lyngbyei* $2n$ = 68–78, for *C. paleacea* $2n$ = 71–73 and for *C. subspathacea*, $2n$ = 78–83; see Elven *et al.*, 2011, for citations of various reports), it is possible that these karyotypic differences could lead to meiotic dysfunction when the putative hybrid species *C. salina* ($2n$ = 77– 79; Cayouette & Morisset, 1985; Cayouette, 1986) and *C. ramenskii* ($2n$ = 80; Zhukova & Petrovsky, 1987) produce gametes, and this could manifest in the production of infertile pollen/ovules or inviable offspring. Nonetheless, future studies on the cytogenetics and meiotic stability of these putative hybrid *Carex* spp. are needed to experimentally test for the existence of karyotypic incompatibilities that could reduce hybrid fertility.

The observed high interspecific heterozygosity in *C. salina* and *C. ramenskii* could also be a product of the evolution of pre- and/or post-zygotic reproductive barriers relative to parental species. Given that our study has identified a likely backcross between *C. salina* and its parental species, it is likely that this hybrid lineage is not completely reproductively isolated from its parents. However, a larger sample size within *C. salina* and *C. ramenskii* would be required to more accurately assess the degree of interspecific mating between these species and their respective parental species. Pre-zygotic barriers may play an important role in maintaining the genomic integrity of the putative hybrid species because *Carex* spp. frequently exhibit strong ecological and phenological divergence (Whitkus, 1988; Standley, 1990). Our personal experience with these species indicates no significant phenological divergence as their periods of anthesis overlap considerably in wild populations (A.T.M. Pedersen, C.S. Bjorå & R. Elven, pers. obs.). However, the species in our study do exhibit consistent microhabitat preferences, specifically reflected in salinity tolerance and the frequency and duration of seawater inundation (Halvorsen *et al.*, 2015). *Carex salina* and *C. ramenskii* both occupy intermediate habitats where their parents appear to be at the margin of their ranges. *Carex salina* intersperses between the more saline and more frequently inundated range of *C. subspathacea* and the less saline, more freshwater-influenced range of *C. paleacea* in Scandinavia, whereas *C. ramenskii* does the same between *C. subspathacea* and *C. lyngbyei* in Iceland (and along the northern Pacific coast of North America and north-eastern Asia). This is a good example of *Carex* spp. displaying strong local habitat specificity that might isolate them from one another despite being hypothetically interfertile (Standley, 1985; Cayouette & Catling, 1992). The habitat preferences of both *C. salina* and *C. ramenskii* suggest that these lineages are ecologically divergent from their parents and probably have different optimal niche requirements along the major ecological gradients on the seashore, but future studies utilizing niche models would be valuable to quantify this observed ecological divergence.

Our results show strong genetic evidence for a hybrid origin of both *C. salina* and *C. ramenskii* populations in Norway and Iceland, but further studies utilizing a broader geographical sampling would be needed to assess if the genetic structuring in these Nordic populations reflects range-wide patterns in these species. Given that there is no cytological or genetic

evidence for polyploidy in these species (Elven *et al.*, 2011), our results are consistent with a homoploid hybrid origin of *C. salina* and *C. ramenskii*, but further experiments to measure pre- and post-zygotic reproductive barriers between the parental and putative hybrid species would be required to specifically test whether these species represent hybrid species or persistent hybrid populations. In a recent perspective piece, Schumer *et al.* (2014) define hybrid speciation as a speciation event where hybridization plays a central role in the formation of reproductive barriers between the newly formed hybrid species and its parents, which can manifest as both pre-zygotic (e.g. ecological niche shifts or phenological divergence) and post-zygotic (e.g. hybrid incompatibility or inviability) barriers (see also Gross & Rieseberg, 2005; Mallet, 2007; Abbott *et al.*, 2013, and references therein). Alternatively, Nieto Feliner *et al.* (2017) defined hybrid speciation as hybridization that results in novel diversity in the form of morphologically and ecologically distinct hybrid lineages that are both established and persistent in a specific ecological niche. Regardless of the definition of hybrid speciation applied, further data quantifying the ecological niches and barriers to gene flow between the hybrid and parental lineages are needed to determine if *C. ramenskii* and *C. salina* should be characterized as homoploid hybrid species or simply persistent hybrid populations between the parental species. Consistent with previous taxonomic hypotheses, our results clearly show that North Atlantic populations of these species probably originated via recurrent interspecific hybridization.

## ACKNOWLEDGEMENTS

The authors thank Hörður Kristinsson for providing excellent locations of *Carex ramenskii*, Adam Vivian-Smith and colleagues for sharing their modifications to the ddRAD protocol with us and Oddvar Pedersen for creating maps. Computational work was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian Metacenter for High Performance Computing (NOTUR) and operated by the Department for Research Computing at USIT, the University of Oslo IT-department (http://www.hpc.uio.no/).

## FUNDING

This work was supported by the University of Oslo. M.D.N. was supported by Norges forskningsråd (Grant Number 240223).

## REFERENCES

**Abbott RJ**, **Albach D**, **Ansell S**, **Arntzen JW**, **Baird SJE**, **Bierne N**, **Boughman J**, **Brelsford A**, **Buerkle CA**, **Buggs R**, **Butlin RK**, **Dieckmann U**, **Eroukhmanoff F**, **Grill A**, **Cahan SH**, **Hermansen JS**, **Hewitt G**, **Hudson AG**, **Jiggins C**, **Jones J**, **Keller B**, **Marczewski T**, **Mallet J**, **Martinez-Rodriguez P**, **Möst M**, **Mullen S**, **Nichols R**, **Nolte AW**, **Parisod C**, **Pfennig K**, **Rice AM**, **Ritchie MG**, **Seifert B**, **Smadja CM**, **Stelkens R**, **Szymura JM**, **Väinölä R**, **Wolf JBW**, **Zinner D. 2013.** Hybridization and speciation. *Journal of Evolutionary Biology* **26:** 229–246.

**Abbott RJ**, **Barton NH**, **Good JM. 2016.** Genomics of hybridization and its evolutionary consequences. *Molecular Ecology* **25:** 2325–2332.

**Abbott RJ**, **Hegarty MJ**, **Hiscock SJ**, **Brennan, AC. 2010.** Homoploid hybrid speciation in action. *Taxon* **59:** 1375–1386.

**Anderson EC. 2008.** Bayesian inference of species hybrids using multilocus dominant genetic markers. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **363:** 2841–2850.

**Anderson EC**, **Thompson EA. 2002.** A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160:** 1217–1229.

**Andrews KR**, **Good JM**, **Miller MR**, **Luikart G**, **Hohenlohe PA. 2016.** Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews. Genetics* **17:** 81–92.

**Ball PW**, **Reznicek AA. 2002.** *Carex* Linnaeus. In: Flora of North America Editorial Committee, eds. *Flora of North America North of Mexico. Vol. 23, Magnoliophyta: Commelinidae (in part): Cyperaceae*. New York: Oxford University Press, 254–572.

**Buerkle CA**, **Morris RJ**, **Asmussen MA**, **Rieseberg LH. 2000.** The likelihood of homoploid hybrid speciation. *Heredity* **84 (Pt 4):** 441–451.

**Buerkle CA**, **Rieseberg LH. 2008.** The rate of genome stabilization in homoploid hybrid species. *Evolution* **62:** 266–275.

**Cayouette J. 1987.** *Carex lyngbyei* excluded from the flora of eastern North America, and taxonomic notes on related species and hybrids. *Canadian Journal of Botany* **65:** 1187–1198.

**Cayouette J**, **Catling PM. 1992.** Hybridization in the genus *Carex* with special reference to North America. *Botanical Review* **58:** 351–438.

**Cayouette J**, **Morriset P. 1985.** Chromosome studies on natural hybrids between maritime species of *Carex* (sections *Phacocystis* and *Cryptocarpae*) in northeastern North America, and their taxonomic implications. *Canadian Journal of Botany* **63:** 1957–1982.

**Cayouette J**, **Morriset P. 1986.** Chromosome studies on the *Carex salina* complex (Cyperaceae, section *Cryptocarpae*) in northeastern North America. *Cytologia* **51:** 817–856.

**Chevreux B**, **Wetter T**, **Suhai S. 1999.** Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology* **99:** 45–56.

**Corander J**, **Marttinen P. 2006.** Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology* **15:** 2833–2843.

**Corander J**, **Marttinen P**, **Sirén J**, **Tang J. 2008.** Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9:** 539.

**Cornuet JM**, **Pudlo P**, **Veyssier J**, **Dehne-Garcia A**, **Gautier M**, **Leblois R**, **Marin JM**, **Estoup A. 2014.** DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30:** 1187–1189.

**Danecek P**, **Auton A**, **Abecasis G**, **Albers CA**, **Banks E**, **DePristo MA**, **Handsaker RE**, **Lunter G**, **Marth GT**, **Sherry ST**, **McVean G**, **Durbin R**; **1000 Genomes Project Analysis Group**. **2011.** The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158.

**Dean M**, **Ashton PA**, **Hutcheon K**, **Jermy AC**, **Cayouette J. 2008.** Description, ecology and establishment of *Carex salina* Wahlenb. (saltmarsh sedge) – a new British species. *Watsonia* **27:** 51–57.

**Dillenberger MS**, **Wei N**, **Tennessen JA**, **Ashman T-L**, **Liston A. 2018.** Plastid genomes reveal recurrent formation of allopolyploid *Fragaria*. *American Journal of Botany* **105:** 862–874.

**Earl DA**, **vonHoldt BM. 2012.** STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:** 359–361.

**Eaton DA**, **Ree RH. 2013.** Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* **62:** 689–706.

**Elgvin TO**, **Trier CN**, **Tørresen OK**, **Hagen IJ**, **Lien S**, **Nederbragt AJ**, **Ravinet M**, **Jensen H**, **Sætre GP. 2017.** The genomic mosaicism of hybrid speciation. *Science Advances* **3:** e1602996.

**Elven R**, **Murray DF**, **Razzhivin VY**, **Yurtsev BA. 2011.** *Annotated checklist of the Panarctic Flora (PAF)*. Available at: http://panarcticflora.org/. Accessed 15 September 2017.

**Escudero M**, **Eaton DA**, **Hahn M**, **Hipp AL. 2014.** Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: a case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution* **79:** 359–367.

**Escudero M**, **Hahn M**, **Brown BH**, **Lueders K**, **Hipp AL. 2016.** Chromosomal rearrangements in holocentric organisms lead to reproductive isolation by hybrid dysfunction: the correlation between karyotype rearrangements and germination rates in sedges. *American Journal of Botany* **103:** 1529–1536.

**Escudero M**, **Hipp AL**, **Luceño M. 2010.** Karyotype stability and predictors of chromosome number variation in sedges: a study in *Carex* section *Spirostachyae* (Cyperaceae). *Molecular Phylogenetics and Evolution* **57:** 353–363.

**Escudero M**, **Hipp AL**, **Waterway MJ**, **Valente LM. 2012.** Diversification rates and chromosome evolution in the most diverse angiosperm genus of the temperate zone (*Carex*, Cyperaceae). *Molecular Phylogenetics and Evolution* **63:** 650–655.

**Estoup A**, **Lombaert E**, **Marin JM**, **Guillemaud T**, **Pudlo P**, **Robert CP**, **Cornuet JM. 2012.** Estimation of demo-genetic model probabilities with approximate Bayesian computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources* **12:** 846–855.

**Fagundes NJ**, **Ray N**, **Beaumont M**, **Neuenschwander S**, **Salzano FM**, **Bonatto SL**, **Excoffier L. 2007.** Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 17614–17619.

**Gompert Z**, **Buerkle CA. 2009.** A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology* **18:** 1207–1224.

**Gompert Z**, **Buerkle CA. 2010.** introgress: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10:** 378–384.

**Gross BL**, **Rieseberg LH. 2005.** The ecological genetics of homoploid hybrid speciation. *The Journal of Heredity* **96:** 241–252.

**Halvorsen R**, **Bryn A**, **Erikstad L**, **Lindgaard A. 2015.** *Natur i Norge - NiN. Versjon 2.0.0*. Trondheim: Artsdatabanken.

**Harrison RG**, **Larson EL. 2014.** Hybridization, introgression, and the nature of species boundaries. *The Journal of Heredity* **105**(Suppl 1): 795–809.

**Hipp AL. 2007.** Nonuniform processes of chromosome evolution in sedges (*Carex*: Cyperaceae). *Evolution* **61:** 2175–2194.

**Hipp AL**, **Rothrock PE**, **Roalson EH. 2009.** The evolution of chromosome arrangements in *Carex* (Cyperaceae). *The Botanical Review* **75:** 96–109.

**Hultén E. 1968.** *Flora of Alaska and neighboring territories*. Stanford: Stanford University Press.

**Hultén E**, **Fries M. 1986.** *Atlas of North European vascular plants north of the Tropic of Cancer*. Königstein: Koeltz Scientific Books.

**Jiménez-Mejías P**, **Hahn M**, **Lueders K**, **Starr JR**, **Brown BH**, **Chouinard BN**, **Chung K-S**, **Escudero M**, **Ford BA**, **Ford KA**, **Gebauer S**, **Gehrke B**, **Hoffmann MH**, **Jin X-F**, **Jung J**, **Kim S**, **Luceño M**, **Maguilla E**, **Martín-Bravo S**, **Míguez M**, **Molina A**, **Naczi RFC**, **Pender JE**, **Reznicek AA**, **Villaverde T**, **Waterway MJ**, **Wilson KL**, **Yang J-C**, **Zhang S**, **Hipp AL**, **Roalson EH. 2016.** Megaphylogenetic specimen-level approaches to the *Carex* (Cyperaceae) phylogeny using ITS, ETS, and *matK* sequences: implications for classification. *Systematic Botany* **41:** 500–518.

**Jónsdóttir IS**, **Augner M**, **Fagerström T**, **Persson H**, **Stenström A. 2000.** Genet age in marginal populations of two clonal *Carex* species in the Siberian Arctic. *Ecography* **23:** 402–412.

**Koboldt DC**, **Zhang Q**, **Larson DE**, **Shen D**, **McLellan MD**, **Lin L**, **Miller CA**, **Mardis ER**, **Ding L**, **Wilson RK. 2012.** VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* **22:** 568–576.

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

**Kristinsson H. 2010.** *Flowering plants and ferns of Iceland*, *3rd edn*. Reykjavik: Mál og menning/Forlagid.

**Kukkonen I**, **Toivonen H. 1988.** Taxonomy of wetland carices. *Aquatic Botany* **30:** 5–22.

**Lexer C**, **Joseph JA**, **van Loo M**, **Barbará T**, **Heinze B**, **Bartha D**, **Castiglione S**, **Fay MF**, **Buerkle CA. 2010.** Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics* **186:** 699–712.

**Li H. 2011.** Improving SNP discovery by base alignment quality. *Bioinformatics* **27:** 1157–1158.

**Li H**, **Durbin R. 2010.** Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26:** 589–595.

**Li H**, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**; **1000 Genome Project Data Processing Subgroup**. **2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

**Martin M. 2011.** Cutadapt removes adapter sequences from high–throughput sequencing reads. *EMBnet.Journal* **17:** 10–12.

**Mallet J. 2005.** Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20:** 229–237.

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

**Mason AS**, **Pires JC. 2015.** Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends in Genetics* **31:** 5–10.

**Miller N**, **Estoup A**, **Toepfer S**, **Bourguet D**, **Lapchin L**, **Derridj S**, **Kim KS**, **Reynaud P**, **Furlan L**, **Guillemaud T. 2005.** Multiple transatlantic introductions of the western corn rootworm. *Science* **310:** 992.

**Mossberg B**, **Stenberg L. 2003.** *Den nya nordiska floran*. Stockholm: Wahlström and Widstrand.

**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.

**Nieto Feliner G**, **Álvarez I**, **Fuertes-Aguilar J**, **Heuertz M**, **Marques I**, **Moharrek F**, **Piñeiro R**, **Riina R**, **Rosselló JA**, **Soltis PS**, **Villa-Machío I. 2017.** Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity* **118:** 513–516.

**Payseur BA**, **Rieseberg LH. 2016.** A genomic perspective on hybridization and speciation. *Molecular Ecology* **25:** 2337–2360.

**Peakall R**, **Smouse PE. 2006.** GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6:** 288–295.

**Peakall R**, **Smouse PE. 2012.** GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics* **28:** 2537–2539.

**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* **12:** e0171398.

**Pickrell JK**, **Pritchard JK. 2012.** Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8:** e1002967.

**Pritchard JK**, **Stephens M**, **Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

**Pukk L**, **Ahmad F**, **Hasan S**, **Kisand V**, **Gross R**, **Vasemägi A. 2015.** Less is more: extreme genome complexity reduction with ddRAD using Ion Torrent semiconductor technology. *Molecular Ecology Resources* **15:** 1145–1152.

**Renaut S**, **Rowe HC**, **Ungerer MC**, **Rieseberg LH. 2014.** Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. *Philosophical Transactions of the Royal Society, B, Biological Sciences* **369:** 20130345.

**Reznicek AA. 1990.** Evolution in sedges (*Carex*, Cyperaceae). *Canadian Journal of Botany* **68:** 1409–1432.

**Reich D**, **Thangaraj K**, **Patterson N**, **Price AL**, **Singh L. 2009.** Reconstructing Indian population history. *Nature* **461:** 489–494.

**Rheindt FE**, **Fujita MK**, **Wilton PR**, **Edwards SV. 2014.** Introgression and phenotypic assimilation in *Zimmerius flycatchers* (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. *Systematic Biology* **63:** 134–152.

**Rieseberg LH. 1997.** Hybrid origins of plant species. *Annual Review of Ecology, Evolution, and Systematics* **28:** 359–389.

**Rieseberg LH**, **Wood TE**, **Baack EJ. 2006.** The nature of plant species. *Nature* **440:** 524–527.

**Roalson EH. 2008.** A synopsis of chromosome number variation in the Cyperaceae. *The Botanical Review* **74:** 209–393.

**Rohlf F. 2000.** *NTSYSpc: numerical taxonomy and multivariate analysis system. Version 2.11a*. Setauket: Exeter Software.

**Schumer M**, **Cui R**, **Powell DL**, **Rosenthal GG**, **Andolfatto P. 2016.** Ancient hybridization and genomic stabilization in a swordtail fish. *Molecular Ecology* **25:** 2661–2679.

**Schumer M**, **Rosenthal GG**, **Andolfatto P. 2014.** How common is homoploid hybrid speciation? *Evolution* **68:** 1553–1560.

**Schwarzbach AE**, **Rieseberg LH. 2002.** Likely multiple origins of a diploid hybrid sunflower species. *Molecular Ecology* **11:** 1703–1715.

**Standley LA. 1985.** Systematics of the *Acutae* group of *Carex* (Cyperaceae) in the Pacific Northwest. *Systematic Botany Monographs* **7:** 1–106.

**Standley LA. 1990.** Allozyme evidence for the hybrid origin of the maritime species *Carex salina* and *Carex recta* (Cyperaceae) in Eastern North America. *Systematic Botany* **15:** 182–191.

**Standley LA**, **Cayouette J**, **Bruederle L. 2002.** *Carex* Linnaeus section *Phacocystis*. In: Flora of North America Editorial Committee, eds. *Flora of North America north of Mexico. Volume 23. Magnoliophyta: Commelinidae (in part): Cyperaceae*. New York: Oxford University Press.

**Stebbins GL. 1950.** *Variation and evolution in plants*. New York: Columbia University Press.

**Streicher JW**, **Devitt TJ**, **Goldberg CS**, **Malone JH**, **Blackmon H**, **Fujita MK. 2014.** Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. *Molecular Ecology* **23:** 3273–3291.

**Toivonen H. 1974.** Chromatographic comparison of the species of *Carex* section *Heleonastes* and some *Carex canescens* hybrids in eastern Fennoscandia. *Annales Botanici Fennici* **11:** 225–230.

**Vivian-Smith A**, **Sønstebø JH. 2017.** *A streamlined ddRAD tag protocol for use with the Ion Torrent sequencer, as a versatile probe for populations, genetics and genomics*. Available at: https://www.protocols.io/view/a-streamlined-ddrad-tag-protocol-for-use-with-the-khuct6w. Accessed 3 October 2018.

**Volkova PA**, **Shipunov AB**, **Elven R**, **Brochmann C. 2008.** The seashore sedges of the Russian Kola Peninsula: how many species? *Flora* **203:** 523–533.

**Whitkus R. 1988.** Experimental hybridizations among chromosome races of *Carex pachystachya* and the related species *C. macloviana* and *C. preslii* (Cyperaceae). *Systematic Botany* **13:** 146–153.

**Zinenko O**, **Sovic M**, **Joger U**, **Gibbs HL. 2016.** Hybrid origin of European vipers (*Vipera magnifica* and *Vipera orlovi*) from the Caucasus determined using genomic scale DNA markers. *BMC Evolutionary Biology* **16:** 76.

**Zhukova PG**, **Petrovsky VV. 1987.** Chromosome numbers and taxonomy of some plant species from the northern Asia. *Botanicheskii Zhurnal* **72:** 1617–1624 [In Russian].

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Figure S1** A–E, Geographic ranges of the five *Carex* taxa of section *Phacocystis* included in this study, modified from Hultén & Fries (1986) and Hultén (1968). Dubious reports are denoted with question marks.

**Figure S2.** The three scenarios of historical relationships tested against one another for DIYABC analyses. The parameters *t1* and *t2* refer to time; *t1* is the time of the most recent population divergence and *t2* is the time of the most ancient population (Pop) divergence in each model. The branches of each tree are coloured to indicate unique effective population sizes (*N1*, *N2* and *N3*) according to the included key. In Scenario 1, the additional parameter *ra* is included to indicate the proportion of population 1 (*ra*) and population 3 (1-*ra*) alleles that constitute the admixture represented by population 2. See Cornuet *et al.*, (2014) for more details about DIYABC model parameterization.

**Figure S3.** A–D, Scenario selection estimated by direct estimation (A, C) and logistic regression (B, D) for both the *C. salina* (A, B) and *C. ramenskii* (C, D) data sets. Deviations between simulated and observed summary statistics are plotted for ten distinct categories of simulated data sets closest to the observed data ranging from 0.01 to 0.1% and from 0.1 to 1% in direct and logistic approach, respectively. Scenario 1 = green, scenario 2 = red, scenario 3 = blue.

**Table S1.** For all lines where the "Taxon and sample ID" starts with "T" the column "Collected by" should start with "A. T. M. Pedersen".

**Table S2.** Read number and mean length of raw reads obtained from ddRAD sequencing on the Ion Torrent platform.