NTNU
Norwegian University of
Science and Technology

Population Genetic Study of *Carex jemtlandica* and *C. lepidocarpa* sect. *Ceratocystis* (Cyperaceae), using RAD-seq

Malene Østreng Nygård

MSc in Biology
Submission date: August 2016
Supervisor: Mika Bendiksby, IBI
Co-supervisor: Hans K. Stenøien, IBI
Heidi Solstad, Tiril Myhre Pedersen , NHM
Narjes Yousefi, BIOLOGI

Norwegian University of Science and Technology
Department of Biology

Preface

This thesis is part of the requirements for a Master of Science degree in the field of Biology at the Norwegian University of Science and Technology (NTNU), and was carried out at NTNU University Museum (NTNU VM). I really appreciate the opportunity I was given to study the fascinating and intriguing *Carex* genus, and being a part of the "RAD-seq movement". It has been a wonderful and educational journey, but also challenging. At some stages of the project the goal almost seemed unreachable due to technical challenges. However, I made it! In this context, I want to thank my main supervisor Mika Bendiksby (NTNU VM) for her support, motivation, excellent collaboration (during project design, field work, and DNA laboratory work), and her ability to help me get in contact with people exhibiting particular knowledge outside her expertise. I also want to thank my co-supervisors: (1) Hans K. Stenøien (NTNU VM) for suggestions and knowledge regarding population genetics, (2) Narjes Yousefi (NTNU VM) for help at the DNA laboratory at any hour of the day, guidance when processing RAD-data, and permission to copy "Preparing RAD-tag libraries" (3) Heidi Solstad (NHM, UiO) for initiating the project, planning and helping during fieldwork, and fun times travelling with her family, and (4) Tiril Myhre Pedersen (NHM, UiO) for comparing results regarding RAD-data on *Carex* species.

I really appreciated the large number of people that wanted to support this thesis and help during different parts of the project. Thanks to Kjell-Ivar Flatberg, Leif Galten, and Reidar Elven for the exchanging of ideas about hypotheses, and assistance with species identification in the field. Thanks to Leif Galten, Reidar Elven and Reidar Haugan for hosting me, good food, and sightseen at Engerdal, Drammen and Storfosna. A scientific thanks goes to the bio technicians at NTNU Genomic Core Facility for sequencing and helping solve barcode problems. A huge thanks also goes to Petri Kemppainen for help and patients with data analyses. In addition, thanks to Hans K. Stenøien, Heidi Solstad, Mika Bendiksby, Petri Kemppainen, and Reidar Elven for comments on my thesis.

Lastly, I want to thank my family and friends for their support and encouragement.

Sammendrag

Carex L. seksj. *Ceratocystis* Dumort er en liten og nylig evolvert, distinkt evolusjonær enhet i starrfamilien (Cyperaceae Juss.) som innad er kjent for sin problematiske taksonomi og komplekse morfologi. Det er for eksempel minst fire ulike taksonomiske behandlinger av de to mest iøynefallende like artene, *C. jemtlandica* (Palmgren) Palmgren og *C. lepidocarpa* Tausch, henholdsvis også kjent som jemtlandstarr og nebbstarr. *Carex jemtlandica* og *C. lepidocarpa* har svært liknende morfologi og økologi, men er geografisk begrenset til henholdsvis kontinentale og kystlige områder. Følgende kan de to artene mulig representere ekstremer i et kontinuum av intraspesifikk klinal variasjon. Jeg samlet 156 individer av *C. jemtlandica*, *C. lepidocarpa* og antatte hybrider i Norge og Sør-Sverige. ”Restriction site associated DNA sequencing” (RAD-seq) ble anvendt for å studere populasjonsgenetikk i, og hybridisering og tilbakekrysning mellom, disse artene. Analysene av 493 SNPs indikerte tilstedeværelsen av to distinkte ”gene pools”, som i stor grad samsvarte med morfologisk tildeling. *Carex jemtlandica* og *C. lepidocarpa* forble genetisk distinkte også i sympatriske populasjoner, mest sannsynlig grunnet redusert fertilitet hos hybrider som følge av intrinsiske post-zygotiske inkompatibiliteter. Dermed kunne klin-hypotesen forkastes. Lav genetisk diversitet ble observer innad *C. jemtlandica* sammenliknet med *C. lepidocarpa*, og støtter opp om opprinnelse fra *C. lepidocarpa* (eller en nær stamfar) som et resultat av ”founder effect” eller ved geografisk ekspansjon inn i Fennoskandia i en tidlig postglasial periode. *Carex jemtlandica* og *C. lepidocarpa* er mulige taksonomiske enheter, men taksonomisk nivå ble ikke foreslått grunnet begrenset samlings område og få taxa inkludert fra *Ceratocystis.*

Abstract

Carex L. sect. *Ceratocystis* Dumort. is a small, presumably young, distinct evolutionary entity of the sedge family (Cyperaceae Juss.) known for its internally problematic taxonomy and complex morphology. For instance, there are at least four different taxonomical treatments of the two seemingly most similar species, *C. jemtlandica* (Palmgren) Palmgren and *C. lepidocarpa* Tausch. *Carex jemtlandica* and *C. lepidocarpa* have high morphological and ecological resemblance, but are geographically constrained to continental and costal areas, respectively. Therefore, these two species could represent the extremes in a continuum of intraspecific clinal variation. I sampled 156 individuals of *C. jemtlandica, C. lepidocarpa,* and putative hybrids from Norway and southern Sweden. Restriction site associated DNA sequencing (RAD-seq) was used to study population genetic structure in, and hybridization and introgression between, these species. Analyses of 493 SNPs indicated the presence of two distinct gene pools, largely in agreement with morphological assignment. *Carex jemtlandica* and *C. lepidocarpa* remained genetically distinct also in sympatric populations, most likely due to reduced fertility in hybrids as a result of intrinsic post-zygotic incompatibilities. The cline hypothesis can therefore be rejected. Low genetic diversity was observed within *C. jemtlandica* compared to *C. lepidocarpa*, lending support for the origin from *C. lepidocarpa* (or a near ancestor) as a result of founder effect or range expansion into Fennoscandia during the early postglacial period. *Carex jemtlandica* and *C. lepidocarpa* are possible taxonomical units, but a taxonomical rank was not suggested due to restricted sampling area and limited taxa included from *Ceratocystis.*

Contents

1. Introduction

The genus *Carex* L., of the sedge family (Cyperaceae Juss.), is one of the largest genera of angiosperms (Schmid, 1983) comprising approximately 2000 species (Reznicek, 1990; Egorova, 1999). *Carex* section *Ceratocystis* Dumort is a distinct evolutionary entity, both genetically and morphologically (Palmgren, 1956; Crins & Ball, 1988a; Escudero *et al.,* 2008; Jiménez-Mejías *et al*., 2012). Internally, however, the variation in morphology is complex and taxonomy problematic (Schmid, 1983; Hedrén, 2004; Jiménez-Mejías *et al*., 2014). Depending on circumscription, the section consists of c. 7-15 species (Egorova, 1999; see Jiménez-Mejías *et al*., 2012, and references therein; Derieg *et al*., 2013), predominantly distributed in temperate Eurasia, North Africa and North America (Egorova, 1999). Hybridization appears to be common between the taxa (Schmid, 1982; see Cayouette & Catling, 1992, and references therein), and introgression has also been documented (Blackstock & Ashton, 2010; Więcław & Wilhelm, 2014). Although the members of section *Ceratocystis* are united by several morphological traits (Palmgren, 1956; Schmid, 1983; Egorova, 1999), there is less consensus concerning species delimitation within the group, and the number species accepted by different authors vary greatly. Hedrén (2002) proposed a "traditional" total-evidence phylogenetic hypothesis of the North and Central European taxa of the section (see Figure 1.1; modified from Hedrén, 2002).

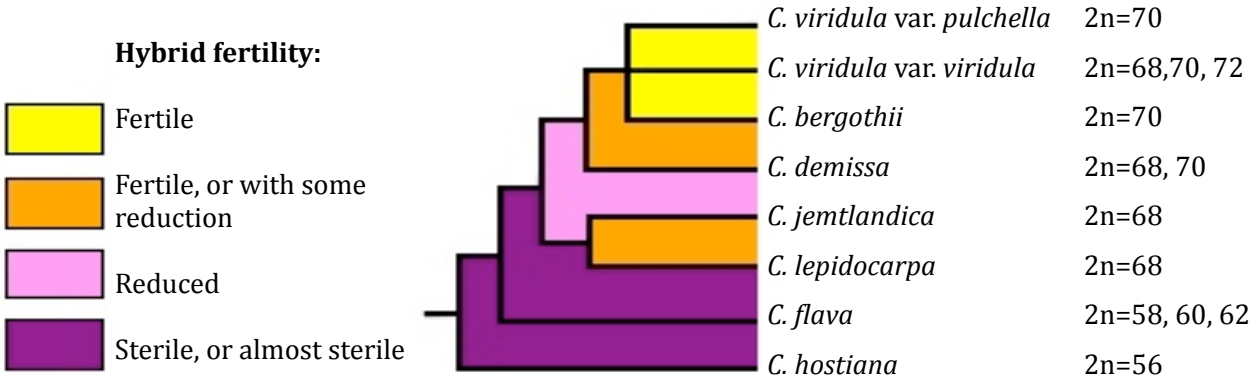

Figure 1.1. Phylogenetic hypothesis of, and hybrid fertility within, taxa of *Ceratocystis* based on available scientific literature (see Hedrén, 2002 and references therein). The position of *C. demissa* (and whether it produced more or less fertile hybrids with any other species) is debated. Figure modified from Hedrén (2002).

However, overall there is little consensus regarding the distinction and taxonomy of the two seemingly most similar species, *C. jemtlandica* (Palmgren) Palmgren and *C. lepidocarpa* Tausch, the two focal species in this study. There are mainly four treatments of these taxa. Aspects analysed and evaluated in the different treatments include morphology, geographical range, and degree of sterility in presumed hybrids. Palmgren (1959) and several recent regional floras of the Nordic countries treat them as separate species, *C. jemtlandica* and *C. lepidocarpa* (Toivonen, 1986; Elven in Lid & Lid, 2005; Elven in Elven *et al*., 2013). Other authors regard them as subspecies of one species, *C. lepidocarpa* ssp. *lepidocarpa* and *C. lepidocarpa* ssp. *jemtlandica* Palmgren (Palmgren, 1926; Chater, 1980; Pykälä & Toivonen, 1994; Hedrén & Prentice, 1996; Hedrén, 2002; Mossberg & Stenberg, 2003; Koopman, 2011). Both taxa have also been included in *C. viridula* Michx ssp. *brachyrhyncha* (Čelak.) B. Schmid (Schmid, 1983,1984; Crins & Ball, 1988b; Blackstock & Ashton, 2001). However, the latter taxonomical treatment is rarely used today as other well-defined species (e.g. *C. demissa* Hornem.; Chater, 1980; Koopman, 2011) are also considered a subspecies of *C. viridula* in Schmid's treatment.

Both *C. jemtlandica* and *C. lepidocarpa* are calciphile and prefer rich fen mires, found growing in lawns, carpets and partly mud-bottoms, i.e., in the wetter parts of the mires. *Carex lepidocarpa* is more common on open mires, while *C. jemtlandica* seems to prefer mire margin sites more strongly influenced by calcareous spring water (Davies, 1956; Palmgren, 1959; Schmid, 1983; Pykälä & Toivonen, 1994). However, *C. jemtlandica* is mainly confined to the interior (continental) parts of Northeast Europe from Norway through Sweden and Finland to the Baltic States and western European Russia (Fig. 1.2A; Palmgren, 1956; Koopman, 2011). In addition, the Trondheim herbarium (TRH) has collections from Newfoundland, Canada, which seem morphologically identical with Scandinavian material of *C. jemtlandica*. In contrast, *C. lepidocarpa* has a disjunct amphi-Atlantic and suboceanic distribution (Pykälä & Toivonen, 1994), ranging from eastern North America on one side, and from North Africa, through Europe to southern Sweden and coastal northern Norway on the other side (Fig. 1.2B; Palmgren, 1956; Hedrén, 2002; Jiménez-Mejías *et al*., 2012; Koopman, 2011). Therefore, it is also possible that *C. jemtlandica* and *C. lepidocarpa* represents the extremes in a continuum of intraspecific clinal variation rather than two distinct species.

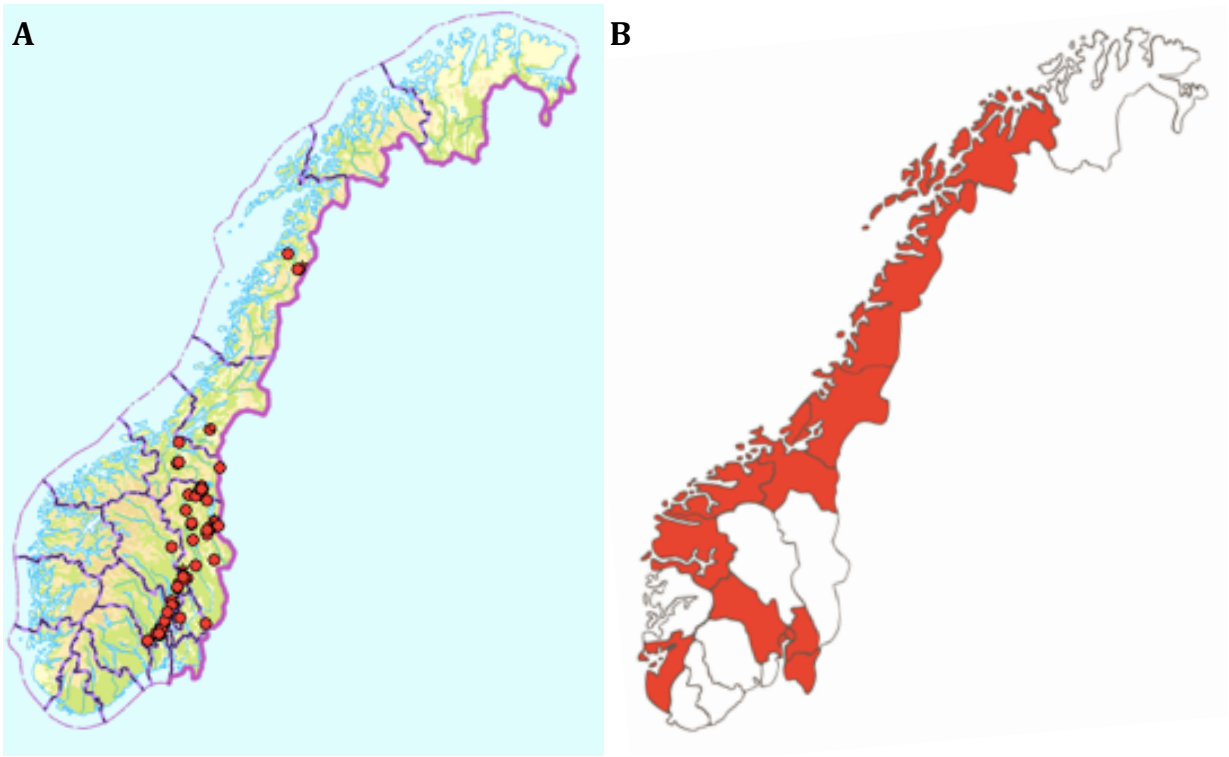

Figure 1.2. Geographical distribution of *C. jemtlandica* and *C. lepidocarpa* in Norway. (A) Red dots represent registered sampling sites of collected material of *C. jemtlandica* in Norway. Downloaded14/07/2016 from http://artskart.artsdatabanken.no. (B) County occurrences of *C. lepidocarpa* in Norway. Downloaded 14/07/2016 from http://data.artsdatabanken.no/Rodliste.

Although closely related and often difficult to distinguish, a number of morphological differences have been reported to separate between *C. jemtlandica* and *C. lepidocarpa* (Palmgren, 1959; Hedrén, 1994; Hedrén, 2002). *Carex jemtlandica* (Fig. 1.3) has more congested globular to ovate upper female spikes with long bracts, and often a smaller lowermost female spike. In comparison, *C. lepidocarpa* (Fig. 1.3) has more or less well separated, equally sized ovate to cylindrical female spikes with shorter bracts. The utricles of *C. jemtlandica* are less crowded and with straight beaks, as opposed to those of *C. lepidocarpa* that are more crowded with downward-pointing beaks. The peduncles of male spikes are often short and erect in *C. jemtlandica*, and longer and oblique in *C. lepidocarpa*. These traits often vary and overlap, and the ratio between the different characters is often more useful to distinguish the two taxa. For example, the ratio between the male spike peduncle length and the length of the uppermost female spike bract is smaller in *C. jemtlandica* than in *C. lepidocarpa*. In addition, the ratio between the leaf length and culm is much greater in *C. jemtlandica* than in *C. lepidocarpa*, i.e., *C. jemtlandica* has leaves nearly as long as the culms, *C. lepidocarpa* has leaves much

shorter than the culm. Morphological intermediates between *C. jemtlandica* and *C. lepidocarpa* can be observed in areas of Scandinavia where they occur in sympatry, and hybridization is thought to take place (Hedrén & Prentice, 1996; Hedrén, 2002).

Figur 1.3: *Carex jemtlandica* (left) and *C. lepidocarpa* (right) from Lier, Buskerud, Norway. Photos taken by Heidi Solstad 2014.

Introduction

Section *Ceratocystis* was recently subjected to a molecular phylogenetic investigation based on DNA sequence data of one nuclear (ITS) and two plastid (rps16 and 5'trnK) genetic regions combined with cytogenetic information (Jiménez-Mejías *et al*., 2012). These are conservative markers that rarely reflect recent evolutionary events, and their study did not resolve the phylogenetic relationships within *Ceratocystis*. This suggests that the sections mainly comprise recently evolved species, which is consistent with a presumed young age of the group (Hedrén, 2002; Derieg *et al.*, 2008). It has been suggested that at least parts of the diversity in the group is less than 10 000 years old, as some of the described species of *Ceratocystis* are restricted to areas covered by the Pleistocene glaciations in North Europe (*C. jemtlandica*, *C. bergrothii* Palmgr., and *C. kotilaini* Palmgr.). Yet, allozyme variation suggests evolutionary lineages more or less consistent with taxonomic treatments (Hedrén & Prentice, 1996; Hedrén, 2002); even clear allele frequency differences between *C. jemtlandica* and *C. lepidocarpa* were observed. This study was based on only 12-15 polymorphic loci. Thus, further studies using larger numbers of loci are needed to resolve the evolutionary relationships within and between species in section *Ceratocystis* (Koopman, 2011).

Several molecular methods that make use of high-throughput sequencing technology are about to become mainstream for studying inter- and intraspecific relationships (Lemmon & Lemmon, 2013). Restriction site associated DNA sequencing (RAD-seq; Miller *et al*., 2007; Baird *et al*., 2008) is one such method that, combined with Illumina sequencing technology, is capable of identifying thousands of genetic markers distributed across the genome. This technique is applicable to any organism without prior knowledge of their genomes (Davey & Blaxter, 2011). In the absence of a reference genome, RAD tags are assembled *de novo* to identify the potential loci, and single nucleotide polymorphisms (SNPs) are then called from orthologous loci.

In this project, I have used RAD-seq to study population genetic structure in, and hybridization and introgression between, natural populations of *C. jemtlandica* and *C. lepidocarpa* in Norway and southern Sweden. I specifically address the following questions: (1) is there evidence for two distinct gene pools among samples of *C. jemtlandica* and *C. lepidocarpa*, (2) do these gene pools correspond to morphological assignment, (3) do hybridization occur between *C. jemtlandica* and *C. lepidocarpa*, and (4), are there indications of backcrossing between F1 hybrids and parental forms of *C. jemtlandica* and *C. lepidocarpa* (introgression). If support for (1) is found, the hypothesis

that genetic variation within these two species represents intraspeciefic clinal variation can be rejected.

Both *Carex jemtlandica* and *C. lepidocarpa* are treated as separate species and listed as, respectively, vulnerable (VU) and near threatened (NT) in the current Norwegian Red List for species from 2015. This is due to reduction in suitable habitats, i.e., ditching of rich fens (Solstad & Elven, 2015). In addition, *C. lepidocarpa* is used as an indicator species for lowland rich fens in Norway; a red listed habitat (A. Moen. & D.I. Øien pers. comm.; Moen & Øien, 2011). Furthermore, it is stated in the Norwegian Biodiversity Act («Naturmangfoldloven") §5 that Norwegian species and their genetic diversity shall be long-term conserved. Thus, properly understanding how genetically distinct *C. jemtlandica* and *C. lepidocarpa* are and how much introgression between them occur in natural populations has large implications for conservation management strategies for these two species.

2. Material and Methods

2.1. Sampling

The data set comprised 156 samples representing three sympatric and 30 allopatric populations. Two of the sympatric population samples comprised *C. jemtlandica*, *C. lepidocarpa* and putative hybrids, but only *C. jemtlandica* and *C. jemtlandica* x *C. lepidocarpa* were morphologically identified in the third sympatric population sample. Of the allopatric population samples, 17 and 13 were morphologically assigned to *Carex jemtlandica* and *C. lepidocarpa*, respectively (Appendix A; Table A.1). The population samples were collected in different parts of the focal species' ranges in Norway. In addition, two population samples were collected in southern Sweden (Fig. 2.1). They were *a priori* thought to only comprise *C. lepidocarpa* as the occurrence of *C. jemtlandica* has never been recorded in this geographic region (Mossberg & Stenberg, 2003). An allopatric population sample consisted of a median sample size of five (range one to eight) individuals of the same assumed taxon collected from the same mire. In sympatric populations up to 15 individuals were collected (Appendix A: Table A.1). Preliminary species identification was made in the field and used as sample ID. However, species identification was later evaluated during workshops at different stages of the project (including evaluation of macroscopic characters). Morphological assignment was mainly based on characters described by Palmgren (1956). Vouchers are deposited in the Herbarium of the Natural History Museum in Oslo (O) and the NTNU University Museum in Trondheim (TRH).

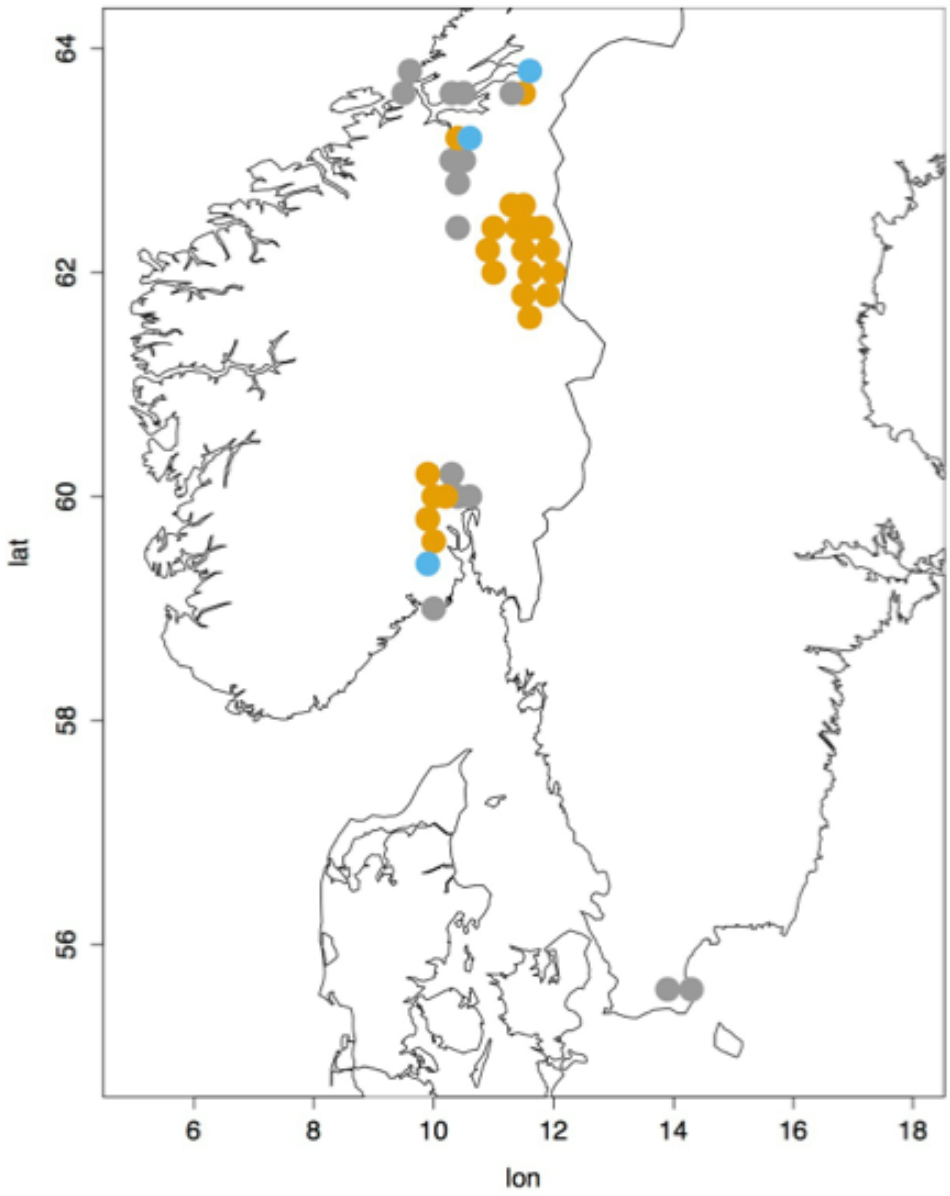

Figure 2.1. Collection site of population samples of *C. jemtlandica*, *C. lepidocarpa* and putative hybrids used in the present project. Points are spread out to prevent overlapping (exact coordinates are given in Appendix A; Table A.1). Colour represents morphologically assigned taxa (orange=*C. jemtlandica*, grey=*C. lepidocarpa*, blue=*C. jemtlandica* x *lepidocarpa*).

2.2. DNA extraction, RAD-seq library preparation and sequencing

Ten mg silica-dried leaf tissue from each sample was grinded into fine powder using 2 x 2mm tungsten beads (in each tube) on a mixer mill (MM301, Retsch GmbH & Co., Haan, Germany). I cut leaf sections larger than 5 mm in length into smaller pieces prior to the crushing step. All samples were crushed with two cycles of 20 oscillations/sec for 120 sec and one cycle of 18 oscillations/sec for 120 sec. Total genomic DNA was extracted using the E.Z.N.A. SP Plant DNA Kit (Omega Bio-tek, Inc., Norcross, GA, USA) following the manufacture's protocol. I eluted samples twice (50 µL Elution Buffer). DNA quality

and quantity was measured (1) on a 1,25 % agarose gel stained with SYBR safe (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA, USA), (2) using a Qubit 2.0 fluorometer (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA, USA), and (3) using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). All extracted DNA samples were quality controlled using (1) and (2), whereas (3) was used mainly for quality control of the RAD-seq libraries.

RAD-seq libraries were prepared using the protocol of Etter et al (2011) with modifications as described in Yousefi *et al.* (in prep.; see Appendix B). In total, I prepared 192 samples (eight libraries of 24 multiplexed samples each), of which 156 samples are included in the present study. Paired-end sequencing of the RAD-seq libraries were conducted over two lanes on an Illumina HiSeq 2500 (Illumina, Inc., San Diego, CA, USA) at the NTNU Genomic Core Facility (https://www.ntnu.edu/dmf/gcf).

2.3. Processing RAD-seq data

I assessed quality, sequence lengths, and base composition of raw forward and reverse reads of each lane using *FastQC* v. 0.11.4 (Bioinformatics, B., 2011). Raw reads were de-multiplexed according to barcodes (Appendix C) using both *Skewer* v. 0.2.1 (Jiang *et al.*, 2014) and *Stacks* v. 3.5 (Catchen *et al.*, 2013; see Appendix D). Quality filtering was conducted using the *process_radtags* program from the *Stacks* pipeline. Raw reads were discarded when: (1) containing one or more bases with a Phred quality score below 10 (reads with a raw Phred score of 10 within a sliding window of 15% of the total read length were retained), (2) including more than one ambiguous nucleotide or primer sequence (inferred by *Minion* v. 15-065; Davis *et al.*, 2013), and (3) missing the complete barcode or SbfI recognition site. Barcode sequences were trimmed away from the target reads, and the de-multiplexed and quality-filtered reads were assembled *de novo* in *ustacks* (*Stacks* pipeline). *Ustacks* aligns identical reads into a "stack" (equivalent of an allele) with a minimum stack depth (-m) of three reads within a sample (secondary reads excluded). The sample-specific stacks were pairwise compared and merged into putative loci (RAD-tags) if they differed by one or less nucleotides (-M). Highly repetitive stacks were discarded (-r) and over-merged RAD-tags resolved (-d). Single nucleotide polymorphisms (SNPs) at each RAD-tag were detected using a maximum likelihood framework implemented in *Stacks*. The *cstacks* program (*Stacks* pipeline) was used for building a catalog, a set of consensus sequences for each RAD-tag. Initially all individuals

were used for building the catalogue. However, to avoid possible erroneous RAD-tags (Catchen *et al.*, 2011) only subsamples of genetically pure (non-admixed) individuals representing distinct genetic clusters (see Results) were used for constructing the reference catalogue (but see Discussion). Sample-specific RAD-tags were merged into homologous RAD-tags if they differed by no more than one nucleotide (-n). Sample-specific RAD-tags that matched to more than one entry (-m) were also included in the catalog (see *Stacks* documentation for details). Finally, sets of stacks from each sample were matched against the catalog using the *sstacks* program (*Stacks* pipeline).

Homologous RAD-tag consensus sequences were then aligned among samples using the *population* program (*Stacks* pipeline), and all bi-allelic SNPs from each RAD-tag with less than 50% missing data (-r) per locus were exported. Lastly, individuals with more than 50% missing genotypes were excluded prior to analysis.

2.4 Genetic clustering and identification of admixed individuals

To study genetic clustering, I used *R* (Team, R. C., 2014) function *find.clusters* ("adegenet"; Jombart & Ahmed, 2011). The number of distinct genetic clusters (*k*) present was first identified by running the function *find.clusters*, in which the function *kmeans* is run sequentially with increasing *k*. Different clustering solutions were compared using the Bayesian information criterion (BIC). Population structure was further tested in *Structure* v. 2.3.4 (Pritchard *et al.*, 2000), which uses a Bayesian model-based clustering algorithm that can account for potential admixed origin of individuals. I applied the "admixture" model with uncorrelated allelic frequencies among populations to obtain estimates of the proportion of ancestry (*q*) of each individual genotype in each of the *k* clusters. *Structure* MCMCs were run for 1,000,000 generations with 100,000 generations burn-in and values of *k* ranging from one to ten. I used five replicates for each *k* and *ln* Pr(D|K) (logarithm of the posterior probability of the data, given the number of clusters) with increasing *k* were visually examined using *Structure Harvester* (Earl & vonHoldt, 2012). The optimal value of *k* was estimated using the Δ*k*-method (Evanno *et al.*, 2005). In figures, colour of individuals was represented by a linear interpolation of *Structure* posterior probability (at *k*=2) going from pure red via pure green to pure blue in RGB colour space (R function *colorRampPallette*). At *k*=2, all individuals were ordered with increasing posterior probability. The Structure posterior probability of each individual was in addition plotted on a map according to sample sites

and using colour as defined above. To avoid overlapping and achieve optimal visualization, points were spread out on the map.

Finally, I conducted a principal components analysis (PCA) implemented in the *R* package "adegenet" with function *dudi.pca*. Prior to this, *scaleGen* was used to scale allele frequencies to mean zero and missing genotypes were replaced by the mean allele frequency among the individuals (see documentation for "adegenet" for details). I tested for correlation between the *Structure* posterior probability and the first principal component using Pearson's product moment correlation coefficient in *R*.

2.5. Isolation by distance

I performed isolation by distance (IBD) analysis to test for population differentiation as a consequence of limited dispersal (Nielsen & Slatkin, 2013). IBD was tested both within and between the focal taxa by comparing matrices of Euclidean genetic distances and Euclidean geographical distances between all pairs of individuals. All pairwise comparisons were coloured by the difference in *Structure* posterior probability at k=2. This value ranged between one and zero, where zero represented pairwise comparisons between individuals belonging to the same genetic cluster, and one represented pairwise comparisons between individuals belonging to different distinct clusters. Colour interpolation was achieved the same way as for posterior probability as described above. A least-square regression line was fitted to all pairwise comparisons (within and between focal taxa). I used Mantel test with 10,000 replicates to assess significance of correlations between genetic and geographical distances using *R* function *mantel.randtest* from package "ade4" (Chessel *et al*., 2004). Pairwise Euclidean genetic distances were also visualized using histograms (mismatch distribution).

2.6 Within population variation and genetic differentiation

Function *summary* from "adegenet" was used to attain expected (H_{exp}) and observed heterozygosities (H_{obs}), and the inbreeding fixation coefficient (F_{IS}) was estimate as 1-(H_{exp}/ H_{obs}). The fixation index F_{ST} (Weir & Cockerham, 1984) between population samples within each taxon and between both taxa was calculated for morphologically and genetically assigned *C. jemtlandica* and *C. lepidocarpa* using function *Fst* from *R* package "pegas" (Paradis, 2010). As many population samples (Appendix A: Table A.1)

contained few individuals (<5), when estimating within species population differentiation, geographical coordinates (both latitude and longitude) for samples sites were rounded to closest whole number to obtain larger sample sizes. F_{ST} was also calculated between pairs of sympatric and allopatric population samples of pure (see Results) *C. jemtlandica* and *C. lepidocarpa*.

Admixture of two distinct gene pools is expected to result in increased proportion of heterozygous loci in individuals that have *Structure* posterior probabilities close to 0.5 (putative F1 hybrids). To study this, the proportion of heterozygous loci per individual were plotted against structure posterior probability (at k=2) and smoothing curve was fitted using a qubic smoothing spline with 95% confidence intervals (*R* function *smooth,spline* with default settings; *spar*=0.99, λ=4.4x10^{-4}, *df*=7.7). The difference in proportion of heterozygous loci per individual between admixed individuals (here defined as individuals with posterior probability >0.4 and < 0.6), pure *C. jemtlandica* and *C. lepidocarpa* (posterior probability <0.05 and >0.95, respectively) were tested using Turkey's Honest Significant Differences test following a single factor ANOVA. In this particular analysis, individuals that did not meet the stated posterior probability requirements were excluded.

3. Results

3.1. RAD-sequencing

The quality-filtered and de-multiplexed data set contained 354 million reads, each 96 bp long. The number of reads per sample varied from 0.73 million to 5.24 million with a median value of 2.24 million. The first draft catalog based on all 156 individuals contained 756 RAD-tags and 1,880 SNPs after final filtering. Construction of a reference catalog based on subsamples of pure individuals reduced these numbers to 335 RAD-tags and 493 SNPs. The median number of SNPs was one in both dataset, but varied from one to 20 (on average 2.5 SNPs per RAD tag) and from one to six (on average 1.5 SNPs per RAD tag) in the draft catalog and reference catalog, respectively. Initial analyses indicated that both data sets gave qualitatively similar results with respect to population genetic parameters (data not shown). As using a reference catalog for SNP calling is recommended (but see Discussion) I here present data from this data set only. Thirteen individuals containing more than 50% missing data were discarded (Appendix A: Table A.2).

3.2. Genetic clustering and identification of admixed individuals

Bayesian information criterion (BIC) decreases steeply from k=1 to k=2 with minimum at k=6, and thereafter increases gradually for subsequent k's (Appendix E: Fig. E.1). At k=2, 63 morphologically assigned *C. jemtlandica* and eight morphologically assigned *C. lepidocarpa* were genetically assigned to one cluster, and 15 of morphologically assigned *C. jemtlandica* and 47 morphologically assigned *C. lepidocarpa* were genetically assigned to the other cluster (Fig. 3.1). At k=3 to k=6 further sub-structuring occurs only in cluster comprising mainly *C. lepidocarpa* (Appendix E: Fig. E.2).

Running *Structure* simulations for k=1 to k=10 lead to an increase in ln Pr(D|k) that reached a maximum at k=6 (Appendix F: Fig. F.1), closely reflecting results from the clustering analysis above. A peak in the second order rate of change with respect to k of the likelihood function (Δk) is identified at k=2 suggesting this as that the optimal numbers of clusters (Appendix F: Fig. F.2). At k=2, the *Structure* model has a support of 5 out of 5 runs, and *Structure* posterior probability is strongly correlated with the first principal component (r_p=0.97, t=47.5, df=141, p<0.01; Fig. 3.3B). *Structure* also identifies

some individuals (n=45) being admixed. These individuals have posterior probabilities between 0.05 and 0.95 (Fig. 3.2) and includes all individuals *a priori* morphologically identified as putative hybrids. At *k*=3 to *k*=6, further sub-structuring occurs only in the *C. lepidocarpa* cluster (Appendix F: Fig. F.3).

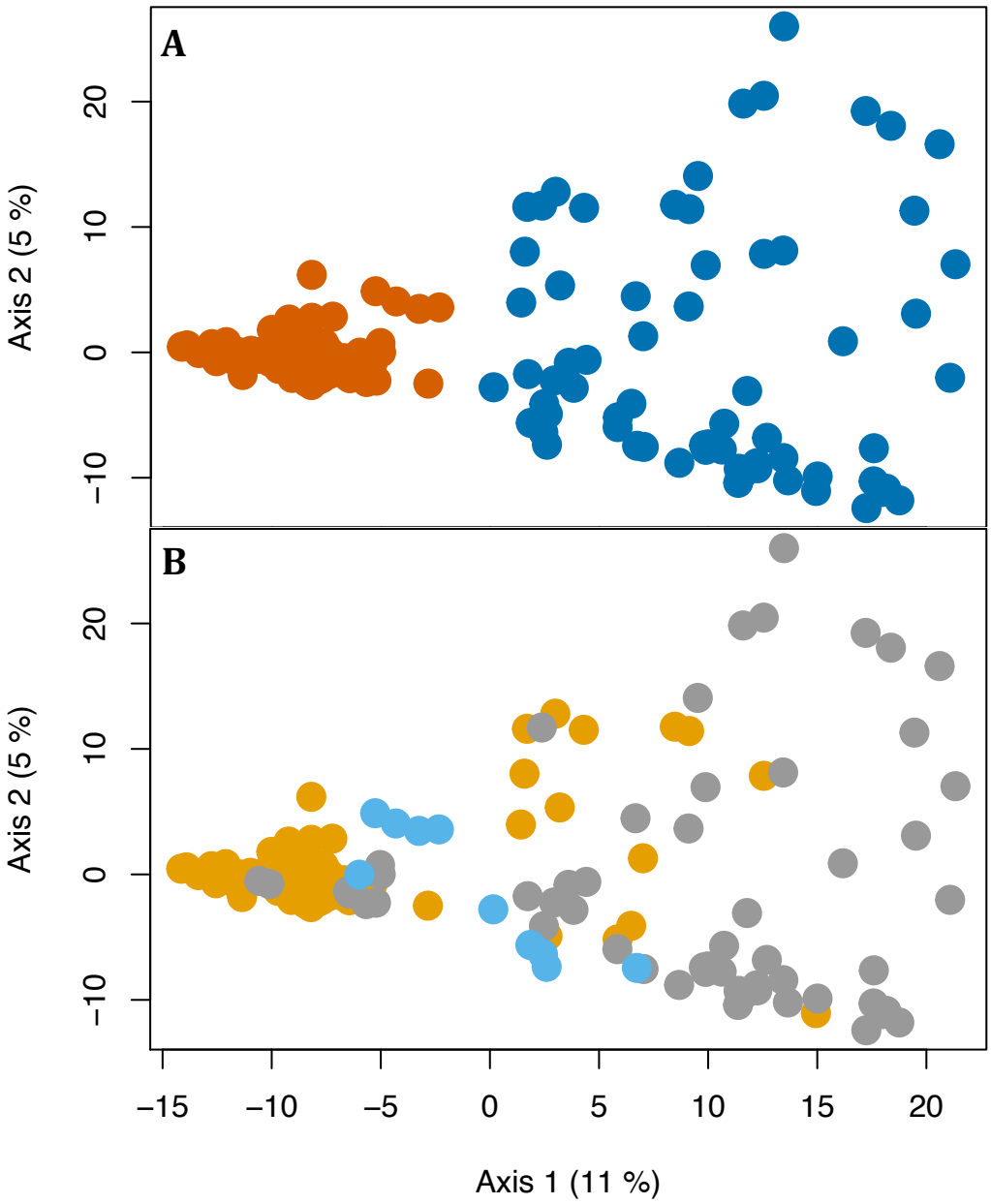

Figure 3.1. Principal components analyses (PCA) of populations of *C. jemtlandica* and *C. lepidocarpa* obtained from the analyses of 493 SNPs. Scatterplot for the first two principal components of the PCA performed on all 143 samples. (A) Individuals assigned to each of the two genetic clusters (*k*=2) are shown in different colours (red=cluster 1, blue=cluster 2). (B) Individuals morphologically assigned to each of the two taxa and to assumed hybrids between them are shown in different colours (orange=*C. jemtlandica*, grey=*C. lepidocarpa*, blue=*C. jemtlandica* x *lepidocarpa*).

Individuals with a posterior probability <0.05 are here considered as genetically pure (non-admixed) *C. jemtlandica* (*n*=64), and individuals with a posterior probability >0.95 as pure *C. lepidocarpa* (*n*=34). According to this criterion only two and four individuals were morphologically misidentified in *C. jemtlandica* and *C. lepidocarpa*, respectively. The admixed individuals are mostly located in geographical proximity to genetically pure *C. jemtlandica* and *C. lepidocarpa* (Fig. 3.3A).

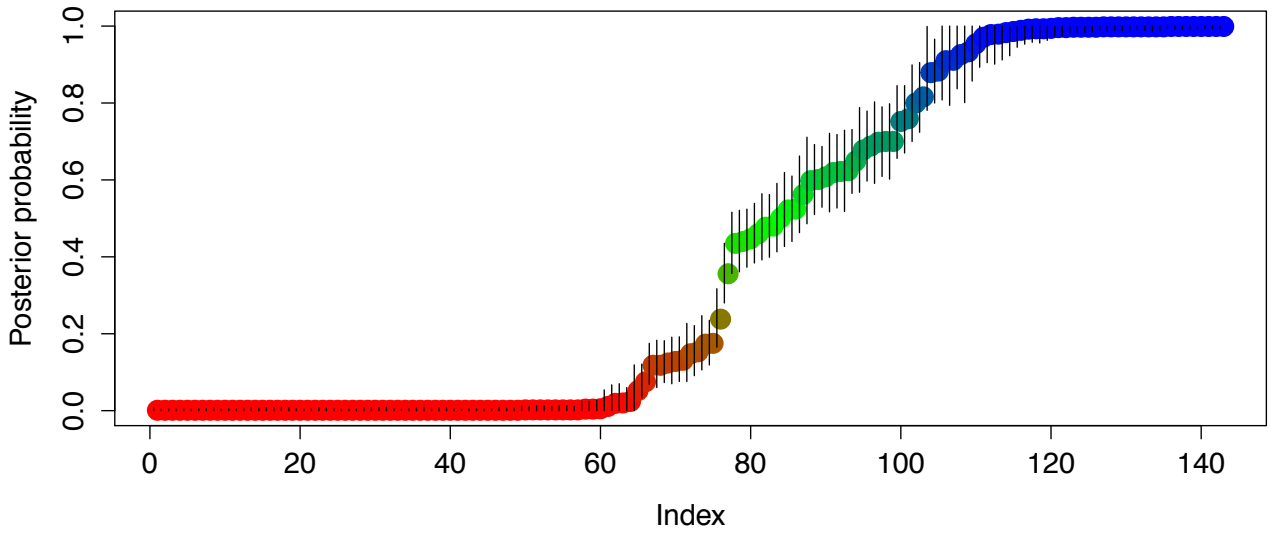

Figure 3.2. Population structure of the studied populations of *C. jemtlandica* and *C. lepidocarpa* at *k*=2 as inferred by *Structure*. All 143 Individuals (indexed) are ordered with increasing posterior probability. Colour represents posterior probability following linear interpolation going from pure red via pure green to pure blue in RGB colour space. Accordingly, red represents genetically pure *C. jemtlandica* (<0.05) blue represents genetically pure *C lepidocarpa* (>0.95). Admixed individuals are represented by green and green/red and green/blue. Vertical lines represent the 90% probability interval.

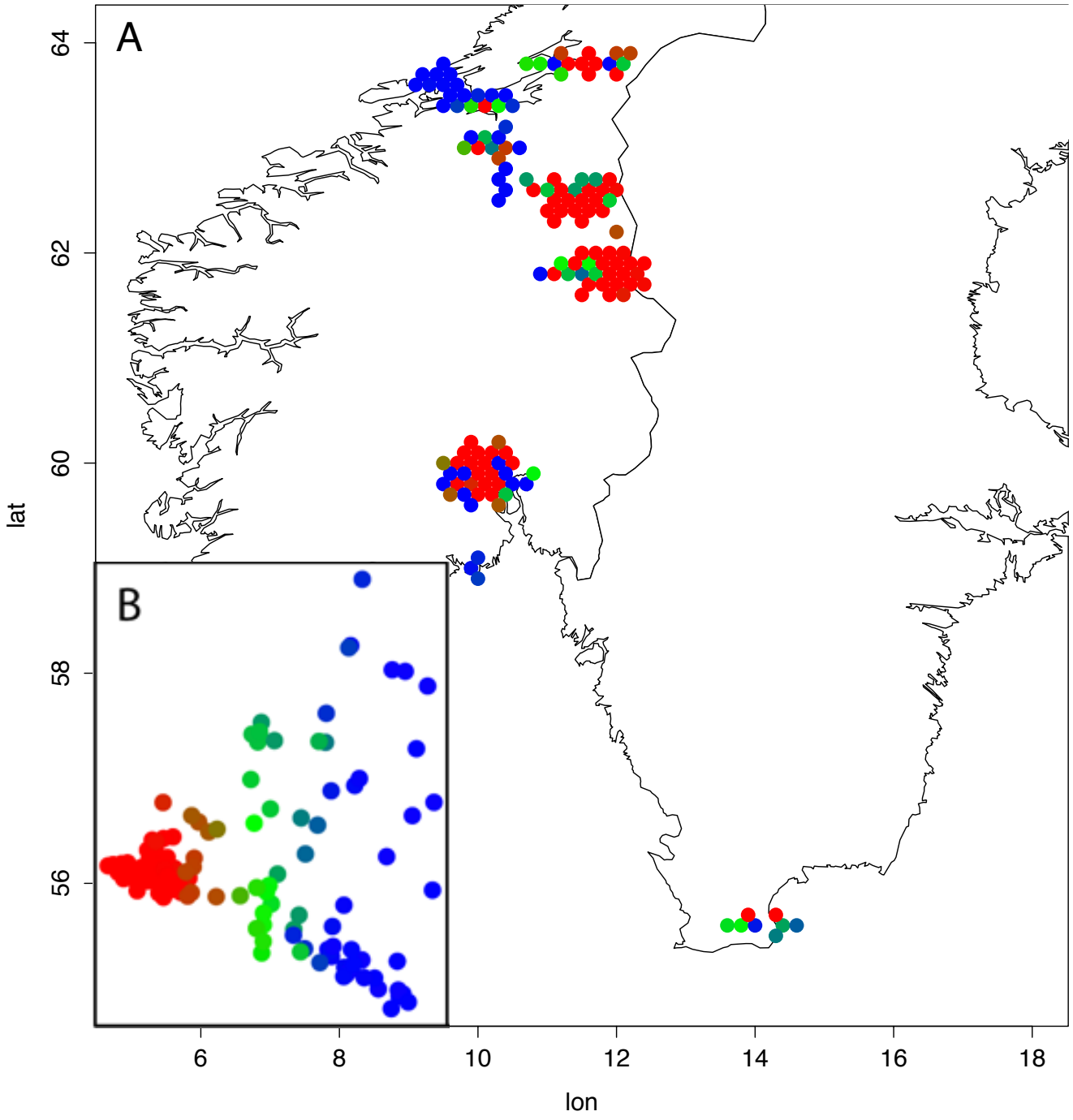

Figure 3.3. Geographic distribution of genetic clusters in the present project. (A) Geographic sample location of all individuals. Each point represents an individual, and points are spread to prevent overlapping. Colours represent posterior probability as in Fig. 3.2. (B) Scatterplot of the first two principal components of a PCA performed on all individuals, coloured as in A.

3.3. Isolation by distance

Isolation by distance analysis are summarized in Fig. 3.4. When considering all samples in the data (Fig. 3.4A), individuals from the same genetic cluster (red dots) and individuals from different genetic clusters (blue dots) are found both in close geographic proximity as well as with large distances apart. Posterior probabilities differed by >0.9 in 4.3% of all pairwise comparisons between individuals within the same sample sites (geographic distance = 0), i.e. these individuals remain genetically distinct in sympatry. Individuals from different genetic clusters (blue dots) are also found when each morphologically assigned species is analysed separately (Fig. 3.4B and C), indicating that some individuals may have been misidentified morphologically. All correlations between genetic and geographic distances are positive and significant (Fig. 3.4A, C and D), except in morphologically assigned *C. jemtlandica* (Fig. 3.4B; Table 3.1). When considering pairwise comparisons between all individuals (Fig. 3.4A) with a difference in posterior probability >0.9 (blue dots in Fig. 3.4A) the correlation between geographic and genetic distance was negative (r_p=-0.04).

Histograms of pairwise genetic distances (mismatch distributions) from Fig. 3.4 are shown in Fig. 3.5. When considering all samples (Fig.3.5A) and morphologically assigned *C. jemtlandica* (Fig. 3.5B), the mismatch distributions are bimodal indicating the existence of two distinct gene pools (one mode representing pairwise comparisons between individuals from the same gene pools and the other representing pairwise comparisons between individuals from different gene pools). For morphologically assigned *C. lepidocarpa* the distribution of pairwise genetic distances is unimodal with mean=25.4 and SE=0.119 (Fig. 3.5C). When considering only pure *C. jemtlamdica* and *C. lepidocarpa* (Fig. 3.5D) the distributions of pairwise genetic distances are unimodal with mean=14.1 (SE=0.048) for *C. jemtlandica* and mean=22.7 (SE=0.17) for *C. lepidocarpa*.

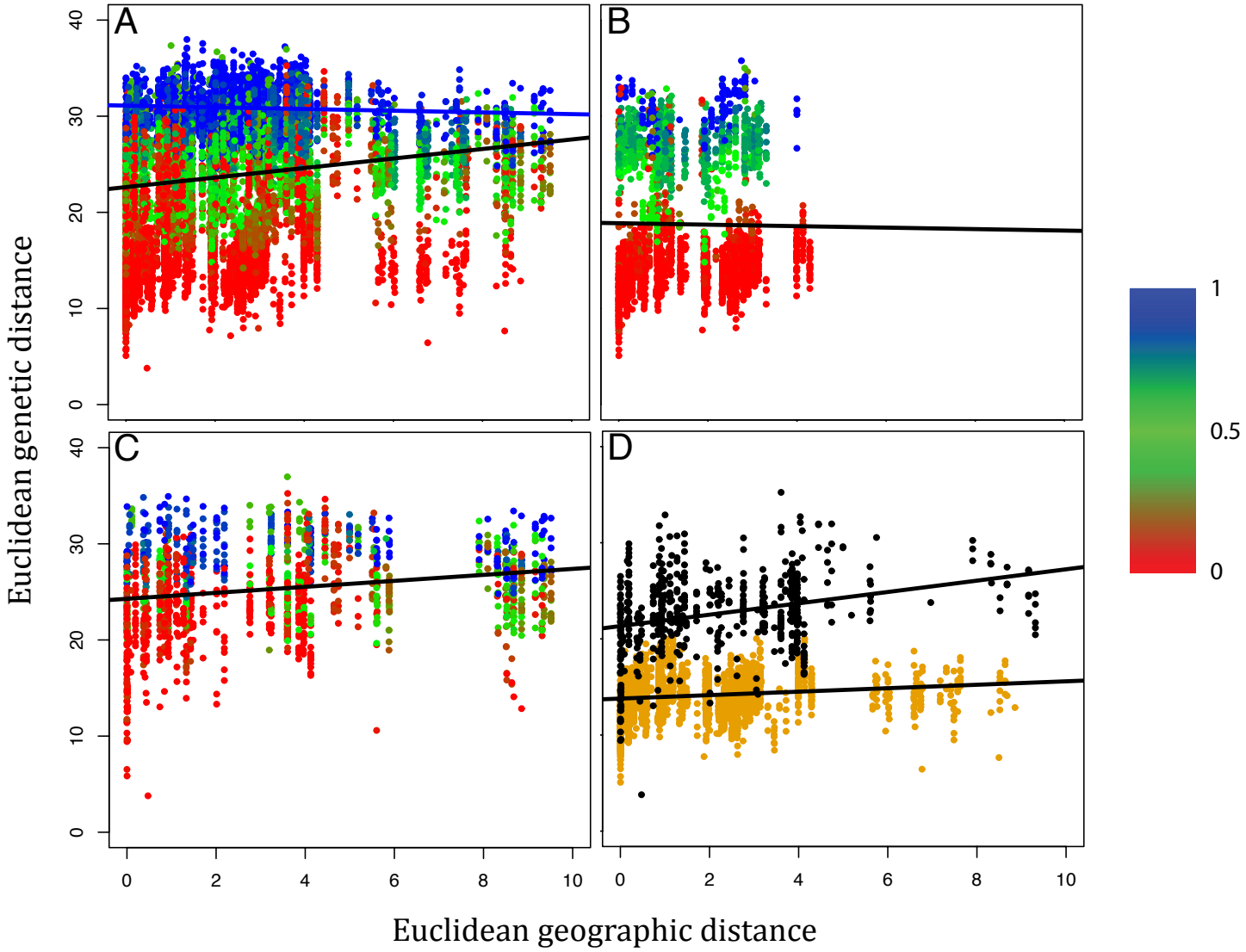

Figure 3.4. Isolation-by-distance analyses of the studied populations of *C. jemtlandica* and *C. lepidocarpa* based on 493 SNPs. Each point represents comparisons of Euclidean geographic and Euclidean genetic distances between pairs of individuals. Lines represent least-square regression. Colour in A to C (see legend) represent difference in posterior probability between pairs of individuals following linear interpolation going from pure red via pure green to pure blue in RGB colour space. Accordingly, red represents difference in posterior probability close to zero, and blue represents difference close to one. (A) Pairwise comparisons between all individuals in this study. Blue line represents regression line for a subset of individuals in this plot representing pairwise comparisons between genetically pure *C. jemtlandica* and *C. lepidocarpa*. (B) Pairwise comparisons between individuals of morphologically assigned *C. jemtlandica*. (C) Pairwise comparisons between individuals of morphologically assigned *C. lepidocarpa*. (D) Pairwise comparisons between all individuals of genetically pure *C. jemtlandica*, and between all individuals of genetically pure *C. lepidocarpa* (orange=*C. jemtlandica*, black=*C. lepidocarpa*).

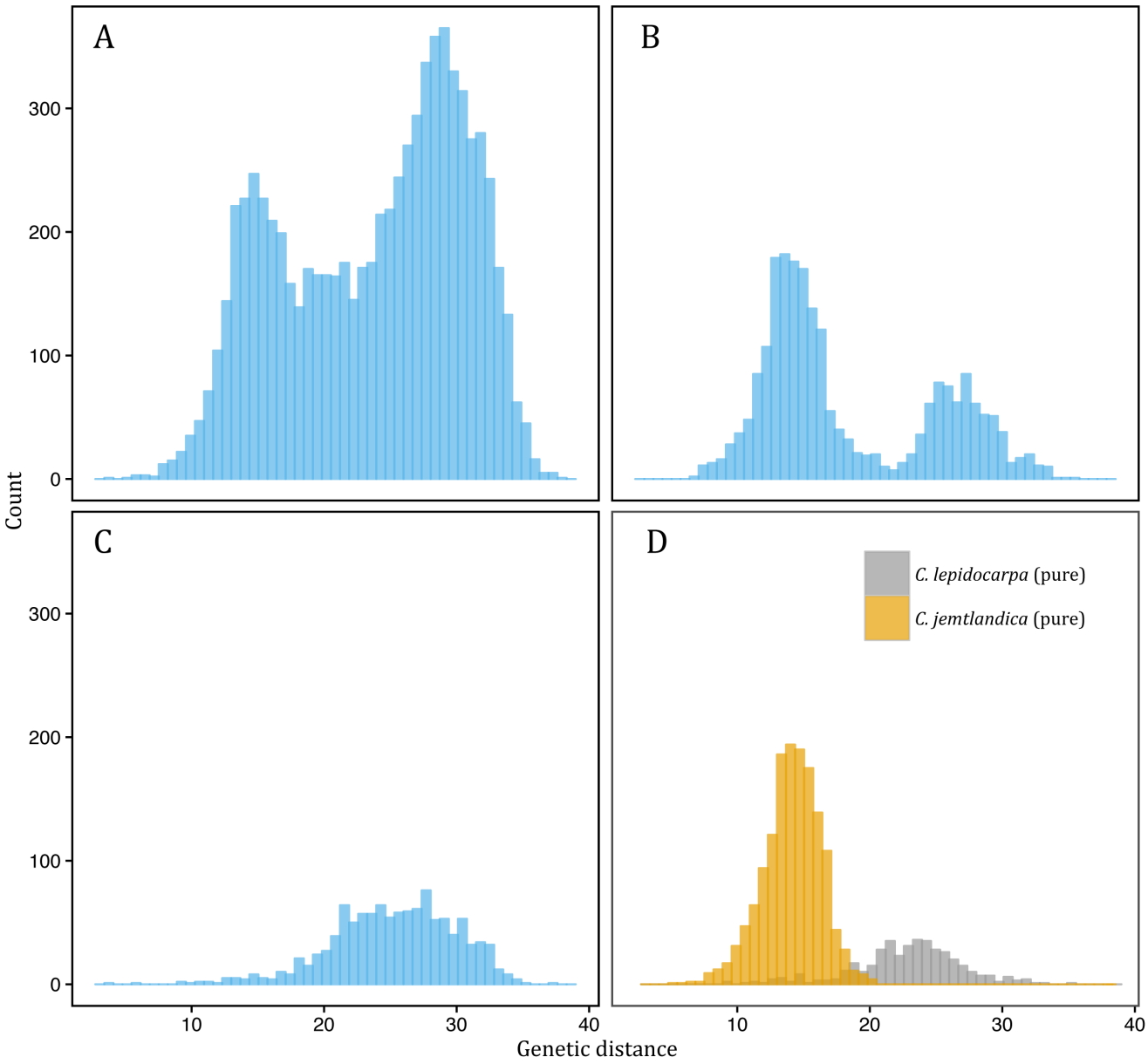

Figure 3.5. Mismatch-distribution of *C. jemtlandica* and *C. lepidocarpa* in present study. Histograms of genetic Euclidean distances between individuals for (A) all individuals, (B) morphologically assigned *C. jemtlandica*, (C) morphologically assigned *C. lepidocarpa*, and (D) genetically pure *C. jemtlandica* (orange) and *C. lepidocarpa* (grey), respectively.

3.4 Within population variation and genetic differentiation

The observed proportion of heterozygous loci per individual in Fig. 3.6 is significantly different between pure *C. jemtlandica*, *C. lepidocarpa* and admixed individuals (p<0.01 for all pairwise comparisons). Individuals with posterior probability between 0.4 and 0.6 have higher proportions of heterozygous loci, consistent with genetic admixture between distinct gene pools. Furthermore, proportion of heterozygous loci per individual (Fig. 3.6) is higher in pure *C. lepidocarpa* compared to pure *C. jemtlandica*.

Expected heterozygosity (Table 3.1) is lower in *C. jemtlandica* compared to *C. lepidocarpa* regardless if morphological assigned or pure individuals are used. For both *C. jemtlandica* and *C. lepidocarpa*, expected heterozygosity is higher when morphological rather than genetic assignment of individuals is used (Table 3.1). All $F_{IS} > 0$, indicating deficiency of heterozygotes compared to what is expected by Hardy-Weinberg equilibrium.

F_{ST} between geographic locations within *C. lepidocarpa* is higher than within *C. jemtlandica* (Table 3.1). In all cases, F_{ST} between geographic locations is higher when species are defined according to genetic clustering rather than morphology (Table 3.1), except within *C. jemtlandica*. F_{ST} was the highest between genetically pure *C. lepidocarpa* and *C. jemtlandica.*

F_{ST} between sympatric populations of *C. jemtlandica* and *C. lepidocarpa* was always higher than F_{ST} between distant populations within the same species (Table 3.2). In addition, F_{ST} between *C. jemtlandica* and *C. lepidocarpa* was higher between allopatric compared to sympatric populations.

Table 3.1. Expected heterozygosity (H_{exp}), observed heterozygosity (H_{obs}), Inbreeding fixation coefficient (F_{IS}), Fixation index (F_{ST}), Parson's *r* for IBD (r_p; Fig. 3.4), and *p*-value from Mantel test (*p*) of the studied populations of *C. jemtlandica* and *C. lepidocarpa* obtained from the analysis of 493 SNPs. Values marked with (*) are calculated based on allele frequencies from all individuals including admixed. Within species, the calculations are performed among populations (see Material & Methods). Abbreviations: jemt=*C. jemtlandica*, lep=*C. lepidocarpa*, morph=morphologically assigned, and pure=genetically pure (non-admixed).

Group	H_{exp}	H_{obs}	F_{IS}	F_{ST}	r_p	*p*
Jemt vs. lep (morph)	0.181*	0.067*	0.626*	0.097	0.15	<0.001
Within jemt (morph)	0.127	0.058	0.539	0.072	-0.01	0.558
Within lep (morph)	0.204	0.074	0.634	0.097	0.20	0.001
Jemt vs. lep (pure)	0.171	0.061	0.642	0.154	-0.04	*na*
Within jemt (pure)	0.098	0.053	0.456	0.071	0.14	0.019
Within lep (pure)	0.193	0.078	0.593	0.121	0.29	0.002

Table 3.2. Differentiation of allopatric and sympatric populations of studied *C. jemtlandica* and *C. lepidocarpa* based on 493 SNPs. Genetic differentiation of allopatric populations is calculated between all combinations of *C. jemtlandica* and *C. lepidocarpa* sampled in mid and eastern Norway; specifically, population j_1413, j_1463, l_1403, and l_1430 (Appendix A; prefixes indicate species, j for *C. jemtlandica* and l for *C. lepidocarpa*). Genetic differentiation of *C. jemtlandica* and *C. lepidocarpa* in sympatric population is calculated based on population j_1446 and l_1448 (Appendix A) that were collected from the same mire.

Population pairs	F_{ST}	Population type
j_1413 vs. l_1403	0.29	Allopatric
j_1463 vs. l_1430	0.24	Allopatric
j_1413 vs. l_1430	0.26	Allopatric
j_1463 vs. l_1403	0.28	Allopatric
j_1413 vs. j_1463	0.049	Allopatric
l_1403 vs. l_1430	0.11	Allopatric
j_1446 vs. l_1448	0.19	Sympatric

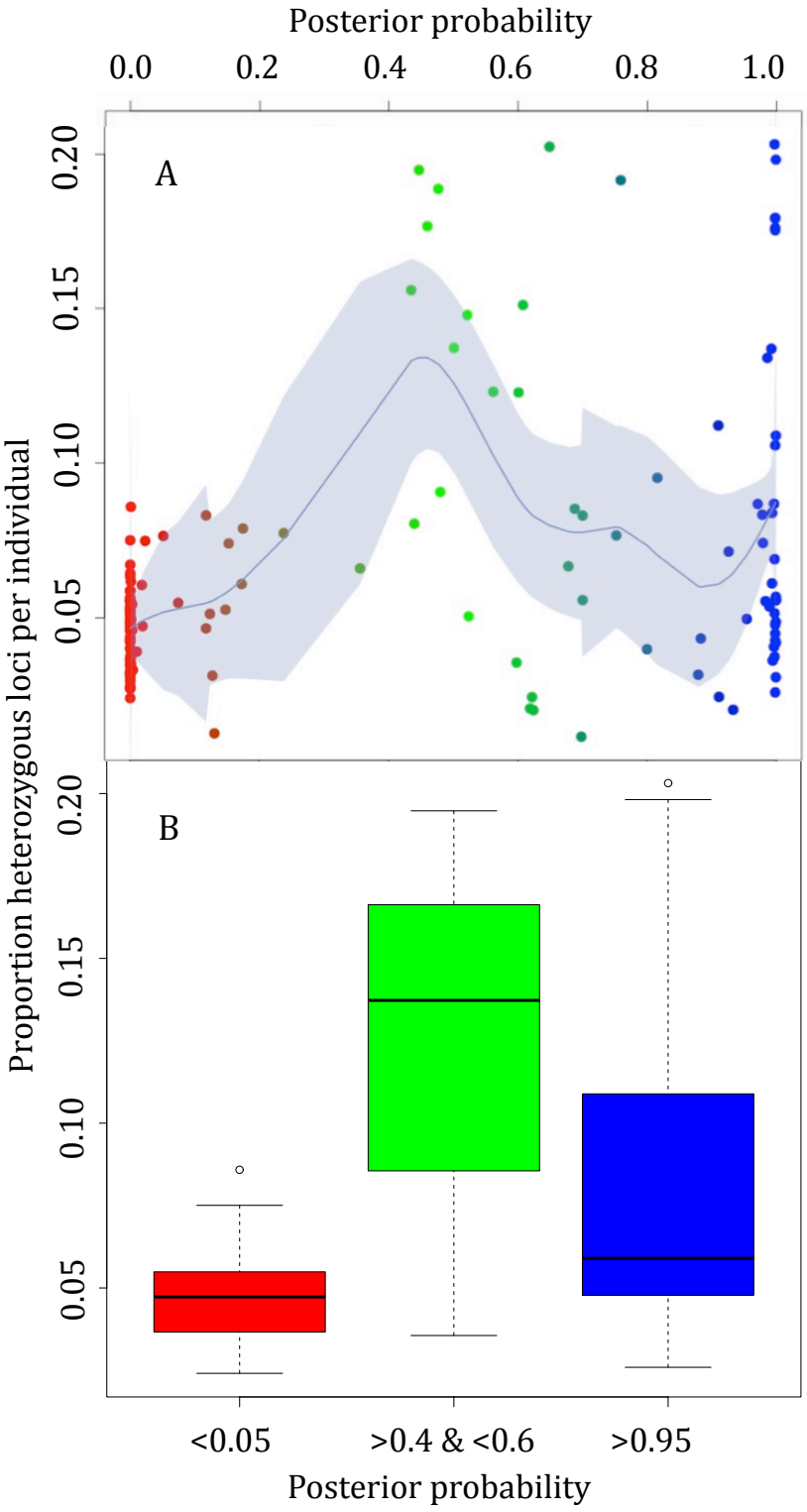

Figure 3.6. Proportion of heterozygous loci per individual as a function of posterior probability. In (A) proportion of heterozygous loci per individual of all individuals is plotted against posterior probability. Line represents a qubic smoothing spline with 95% confidence intervals. (B) Box plot of proportion heterozygous loci per individual for genetically assigned pure *C. jemtlandica* (red; posterior probability <0.05), genetically assigned pure *C. lepidocarpa* (blue; posterior probability>0.95), and admixed (green; 0.4<posterior probability>0.6).

4. Discussion

Here I used a genomic approach (RAD-seq) to study population genetic structure in, and hybridization and introgression between, *C. jemtlandica* and *C. lepidocarpa*, two closely related species in *Carex* section *Ceratocystis*. Even though *Ceratocystis* is one of the most well studied sections of *Carex* (Derieg *et al*., 2013), the taxonomy within the section remains elusive. RAD-seq is capable of identifying thousands of genetic markers distributed across the genome and is thus well suited to further study this recently evolved and seemingly homogeneous group.

4.1 Genetic clustering and introgressive hybridization

Results from all analyses in the present study indicate the presence of two distinct gene pools largely corresponding to the morphologically assigned *C. jemtlandica* and *C. lepidocarpa*. It is therefore possible to reject the hypothesis that *C. jemtlandica* and *C. lepidocarpa* represents the extremes of a continental to coastal clinal variation. This is particularly apparent in the isolation-by-distance plot (IBD; including all individuals; Fig. 3.4A), where pairs of genetically distinct individuals (difference in posterior probability ~1) exist across all geographic distances, including fully sympatric populations. Despite that individuals can remain genetically distinct also in sympatric populations, *Structure* identifies a large proportion of the individuals (31%) as admixed. Admixture proportions (with narrow 90% probability intervals) are continuously distributed between zero and one (Fig. 3.2). This suggesting not only hybridization but also extensive introgression back to each of the parental gene pools via later generation backcrossing. Admixture between two distinct gene pools is further supported by a larger proportion of heterozygous loci per individual in admixed individuals (Fig. 3.6). This is consistent with an earlier allozyme study (Hedrén, 2002), in which two genetically distinct clusters, *C. jemtlandica* versus *C. lepidocarpa*, were connected by genetically admixed individuals, most of which were determined as morphologically intermediate in field.

According to Schmid (1983), introgressive hybridization is common between members of *Ceratocystis* growing in close geographical proximity. Blackstock & Ashton (2010) undertook genetic and morphometric studies of natural hybrid populations of

Ceratocystis from 36 sites in North America, Europa and the British Isles. They found that low-pollen fertility hybrids between *C. flava* and *C. lepidocarpa* were able to backcross with *C. lepidocarpa*. Więcław & Wilhelm (2014) observed the same backcrossing pattern in natural hybrid populations in Poland, based only on morphometric studies. However, *C. jemtlandica* was not included in either of these studies. *Carex lepidocarpa* is considered more distantly related to *C. flava* than to *C. jemtlandica.* It is therefore reasonable to expect introgression between *C. lepidocarpa* and *C. jemtlandica*. In my data, individuals with varying degrees of admixture can be found in all areas were pure forms of both *C. jemtlandica* and *C. lepidocarpa* co-occur. However, admixed individuals are also found in geographic areas with individuals from only one of the genetic clusters. This can be explained either by (1) failure to find and collect individuals of the missing parent species in that particular area (it may be rare or present only in nearby localities), or (2) the missing parent species only recently became extinct from that particular area (Schmid, 1983). In addition, mix-up of sample labels (for instance when de-multiplexing individual barcodes; Appendix E), cannot at this stage entirely be excluded.

My results, that genetically distinct individuals occur in sympatry despite introgressive hybridization, suggests the existence of some level of reproductive barrier between *C. jemtlandica* and *C. lepidocarpa*. Schmid (1982) studied the fertility of cultivated hybrids between species of *Ceratocystis* in Switzerland. He found that seed set and pollen fertility in artificial *C. lepidocarpa* x *C. viridula* hybrids varied between 6-12% and 25-37%, respectively. Although he did not study the fertility of natural hybrids between the two species, Schmid (1982) assumed that they might be less fertile than the experimental ones. He discovered that pollen fertility was as low as 0-2% in natural *C. flava* x *C. lepidocarpa* hybrids and up to 3% in artificial hybrids. It is reasonable to assume that the pollen fertility in *C. jemtlandica* x *C. lepidocarpa* hybrids would be even higher than 25-37% as *C. jemtlandica* and *C. lepidocarpa* are considered to be more closely related to each other than to the other taxa in this group. Reduced pollen fertility (aborted anthers) was clearly observed also in my study and used for sampling of morphologically assumed hybrids in the field. All these individuals were indeed genetically admixed between *C. jemtlandica* and *C. lepidocarpa* with admixture proportions ranging from 0.13 to 0.61 (Appendix A; Table A.1).

Reduced pollen fertility indicates intrinsic post-zygotic gene flow barriers due to the gradual accumulation of genetic incompatibilities (Bateson-Dobzhansky-Muller incompatibilities) in the absence of gene flow (Futuyma, 2013). Based on allozymic similarities and lower genetic diversity within *C. jemtlandica*, Hedrén & Prentice (1996) suggested that *C. jemtlandica* might have evolved from *C. lepidocarpa* (or a near ancestor), either (1) as a result of population fragmentation and isolation in glacial refugia, or (2) during the process of range expansion into Fennoscandia during the early postglacial period. Both processes potentially preclude prolonged or repeated contact between *C. jemtlandica* and *C. lepidocarpa* allowing for mutational differences to accumulate in allopatry and intrinsic post-zygotic incompatibilities to form. Also in my study, genetic diversity was considerably lower in *C. jemtlandica* (Fig. 3.5 D; Table 3.1) compared to *C. lepidocarpa* lending support for this hypothesis. An alternative explanation for the lower genetic diversity in *C. jemtlandica*, but not mutually exclusive, could be a higher degree of self-fertilization in *C. jemtlandica* compared to *C. lepidocarpa* (Hedrén & Prentice, 1996). Indeed, the high F_{IS} for both species (Table 3.1) in my study indicates that some selfing in both *C. jemtlandica* and *C. lepidocarpa* is likely. However, in particular in *C. lepidocarpa* this could also be due to population sub-structuring (Wahlund effect; Nielsen & Slatkin, 2013; but I did not explicitly test this).

4.2 Morphological assignment

Population genetic clustering in the RAD-data agrees well with morphological assignment in the field. However, many more individuals were genetically assigned as admixed, compared to what were *a priori* morphologically assigned as *C. jemtlandica* x *C. lepidocarpa* in the field (Appendix A; Table A.2). There are three potential, non-mutually exclusive explanations for this. First, our criteria for considering an individual non-admixed (pure) were stringent (posterior probability >0.95 or <0.05). Using less strict thresholds (>0.8 and <0.2) reduced the number of admixed individuals from 31% to 18%. Second, backcrossing with parental species over several generations can lead to hybrids with similar morphology as one of the parental species and high pollen fertility (stabilized cryptic backcrossing). This has explicitly been documented in *Ceratocystis* cross-pollination studies (Schmid, 1982). Lastly, many individuals of hybrid origin may have been mistakenly identified as pure *C. jemtlandica* or *C. lepidocarpa* because the

reduced pollen development (anther abortion) was mainly examined only by a hand lense in the field.

The existence of misidentified individuals is most evident from the bimodal distribution in Euclidean genetic distances within morphologically assigned *C. jemtlandica* (Fig. 3.5B) and the existence of genetically distinct pairs of individuals in the IBD plots in this species (Fig. 3.4B). Although morphologically assigned *C. lepidocarpa* also includes genetically distinct pairs of individuals in the IBD plots (Fig. 3.4C), the distribution of Euclidean genetic distances was not bimodal (Fig. 3.5C). This is likely due to a much higher genetic diversity within pure *C. lepidocarpa* than in pure *C. jemtlandica*. The inclusion of *C. jemtlandica* among samples of *C. lepidocarpa* (erroneous morphological assignment) will thus have less effect on the mismatch distribution in *C lepidocarpa* than in *C. jemtlandica*.

Some of the uncertainties in genetic assignment in my study may also be due the use of a reference catalogue for SNP calling. The intention of using a reference catalogue for SNP calling is to exclude possible erroneous RAD-tags in the data. This works well when little uncertainty is associated with morphological assignment of the individuals. Taxa in *Ceratocystis* are morphologically sometimes difficult to distinguish, especially individuals of hybrid origin (see above). Thus, there is a risk that samples in my study also include *C. flava*, *C. viridula*, *C. demissa* or, more likely, hybrid combination between any of these and either of the two focal species (Fig, 1.1). In particular, young individuals of *C. flava* can morphologically resemble *C. jemtlandica* (R. Elven, pers. comm.). However, when a reference catalogue of pure *C. jemtlandica* and pure *C. lepidocarpa* was used for SNP calling, alleles specific to the non-focal taxa will be regarded as sequencing errors. This has the effect that (1) any rare alleles in the focal species will most likely be missed (for instance alleles only present in a local population not having representatives in the reference catalogue), (2) any alleles of hybrid origin that do not involve the focal species are discarded, and (3) individuals with alleles of hybrid origin will contain high proportions of missing genotypes.

Initial analyses on the data where all individuals were used for the catalogue indicated 11 genetic outliers (visually determined from a PCA plot, data not shown). Three of these individuals were later removed due to more than 50% missing data, when using the reference catalogue. All the genetic outliers except four are from populations where hybridization between the focal species (mainly *C. jemtlandica*) and

C. flava/C. demissa is suspected (Appendix A; Table A.2). No outliers were present in the PCA when the reference catalogue was used for SNP calling, strongly suggesting that alleles of non-focal species origin were indeed removed (see above). A more comprehensive view of possible hybridization with non-focal species in my study would be possible by inclusion of other taxa in *Ceratocystis* that potentially hybridize with the focal species not using a reference catalogue. In other words, I do not recommend using a reference catalogue for SNP calling, when the potential to detect introgression from closely related, non-focal, taxa is important.

The two populations from Sweden were *a priori* assumed to comprise only *C. lepidocarpa* as no prior records of *C. jemtlandica* exists from this geographic region (Mossberg & Stenberg, 2003). However, 7 out of 8 of these individuals were not genetically assigned to pure *C. lepidocarpa* but instead to pure *C. jemtlandica* (two) and admixed (five; Appendix A; Table A.2). Three out of eight of these individuals (l_1401_2, l_1402_2 and l_1402_5) contained the P2-barcode (TGCAT) that performed less successfully (increased number of mismatches compared to other barcodes; Appendix E). For these individuals, an alternative de-multiplexing procedure was used (Appendix E) that potentially could have mis-assigned reads to any of the individuals with the problematic P2-barcode (Appedix A; Table A.2). Two of these individual were genetically assigned to pure *C. jemtlandica* and one to pure *C. lepidocarpa.* However, mis-assignment of reads among individuals with the problematic P2-barcode is only expected to cause noise, not completely assign them to a different genetic cluster.

More in depth morphological investigations suggested that two out of the seven potentially misidentified individuals in the two Swedish populations might be of hybrid origin with *C. demissa*. Indeed, *C. demissa* (but also *C. viridula*) was present in these locations. It is therefore possible that introgression from *C. demissa* into *C. lepidocarpa* may have caused *Structure* (at k=2) to force them to appear admixed between *C. jemtlandica* and *C. lepidocarpa*. This is further complicated by the fact that a reference catalogue was used for SNP calling (see above) causing any (rare) introgressed alleles (from non-focal species) to drop out from the final data set. Re-analysis of the data without using a reference catalogue for SNP calling and *Structure* analyses at k>2 is necessary to resolve this. However, it is still difficult to explain why two of the eight individuals were genetically assigned to pure *C. jemtlandica*.

5. Conclusions

The analyses of 493 SNPs from 143 morphologically assigned *C. jemtlandica* and *C. lepidocarpa* indicate the presence of two distinct gene pools among the sampled populations from Norway and southern Sweden. The different gene pools correspond to morphologically assigned *C. jemtlandica* and *C. lepidocarpa* with the exception of very few misidentified individuals. As differentiation of *C. jemtlandica* and *C. lepidocarpa* cannot be explained by isolation by distance, the cline hypothesis can be rejected. Even though introgressive hybridization occurs between *C. jemtlandica* and *C. lepidocarpa*, they remain genetically distinct also in sympatric populations. This is suggestive of reduced fertility in hybrids as a result of intrinsic post-zygotic incompatibilities. However, since the sample area constitutes only a part of the total geographical ranges of *C. jemtlandica* and *C. lepidocarpa*, and because not all of the taxa in *Ceratocystis* are included in this study, I find it preliminary to conclude on taxonomical rank for *C. jemtlandica* and *C. lepidocarpa*.

References

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson, E.A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, 3(10): e3376. DOI:10.1371/journal.pone.0003376

Bioinformatics, B. (2011). *FastQC: a quality control tool for high throughput sequence data*. Cambridge, UK: Babraham Institute. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Blackstock, N. & Ashton, P.A. (2001). A re-assessment of the putative *Carex flava* agg. (Cyperaceae) hybrids at Malham Tarn (v.c. 64): A morphometric analysis. *Watsonia*, 23(4): 505-516.

Blackstock, N. & Ashton, P.A. (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(1): 37-47. DOI: 10.1007/s00606-010-0287-0

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, 22(11): 3124-3140 DOI: 10.1111/mec.12354

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J.H. (2011). *Stacks*: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3* (Bethesda), 1(3): 171-182. DOI: 10.1534/g3.111.000240

Cayouette, J. & Catling, P.M. (1992). Hybridization in the Genus *Carex* with Special Reference to North America. *The Botanical Review*, 58(4): 351-438. DOI: 10.1007/BF02858773

Chater, A.O. (1980). *Carex* L. -In: Tutin, T.G., Heywood, V.H., Burges, N.A., Moore, D.M., Valentine, D.H., Walters, S.M., Webb, D.A, Charter, A.O. & Richardson, I.B.K. (eds). *Flora Europaea* (5th ed): 290-323. Cambridge, Great Britain: Cambridge University Press.

Chessel, D., Dufour, A.B. & Thioulouse, J. (2004). The ade4 package-I: One-table methods. *R News*, 4(1): 5-10.

Crins, W.J. & Ball, P.W. (1988a). Sectional Limits and Phylogenetic Considerations in *Carex* Section *Ceratocystis* (Cyperaceae). *Brittonia*, 40(1): 38-47. DOI: 10.2307/2806873

Crins, W.J. & Ball, P.W. (1988b). Taxonomy of the *Carex flava* complex (Cyperaceae) in North America and northern Eurasia. I. Numerical taxonomy and character analysis. *Canadian Journal of Botany*, 67(4): 1032-1047. DOI: 10.1139/b89-137

Davey, J.W. & Blaxter, M.L. (2011). RADSeq: next-generation population genetics. *Briefings in Functional Genomics*. 9(5-6): 416-423. DOI: 10.1093/bfgp/elq031

Davies E.W. (1956). The Ecology and Distribution of *Carex flava* and its Allies in the British Isles. *Botaniska Notiser*, 109(1): 50-74.

Davis, M.P.A., van Dongen, S., Abreu-Goodger, C., Bartonicek, N., & Enright, A.J. (2013). Kraken: A set of tools for quality control and analysis of high-throughput sequence data. *Methods*, 63(1): 41-49. DOI: 10.1016/j.ymeth.2013.06.027. Available for download at: http://www.ebi.ac.uk/research/enright/software/kraken

Derieg, N.J., Sangaumphai, A. & Bruederle, L.P. (2008). Genetic diversity and endemism in North American *Carex* section *Ceratocystis* (Cyperaceae)1. *American Journal of Botany*, 95(10): 1287-1296. DOI: 10.3732/ajb.2007069

Derieg, N.J., Weil, S.J., Reznicek, A.A. & Bruederle, L.P. (2013). *Carex viridistellata* sp. nov. (Cyperaceae), a New Cryptic Species from Prairie Fens of the Eastern United States. *Systematic Botany*, 38(1): 82-91. DOI: 10.1600/036364413X661971

Earl, D.A. & vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*(2): 359-361. DOI: 10.1007/s12686-011-9548-7. Available online at: http://taylor0.biology.ucla.edu/structureHarvester/

Egorova, T.V. (1999). *The sedges (Carex L.) of Russia and adjacent states*. St. Petersburg, Russia: St. Petersburg State Chemical–Pharmaceutical Academy, and St. Louis, Missouri, USA: Missouri Botanical Garden Press.

Elven, R. (2005). Seksjon *Ceratocystis*. -In: Lid, J. & Lid, D.T. (eds), 2005: p. 1006. *Norsk flora* (7th ed). Oslo, Norway: Det Norske Samlag.

Elven, R. (2013). *Carex jemtlandica (Palmgr.) Palmgr*. -In: Elven, R., Fremstad, E. & Pedersen, O. (eds) 2013: p. 123. *Distribution maps of Norwegian vascular plants. IV. The eastern and northeastern elements*. Trondheim, Norway: Akademika Publishing.

Escudero, M., Valcárcel, V., Vargas, P. & Luceño, M. (2008). Evolution in *Carex* L. sect. *Spirostachyae* (Cyperaceae): A molecular and cytogenetic approach. *Organisms, Diversity & Evolution*, 7(4): 271-291. DOI: 10.1016/j.ode.2006.08.006

Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., & Cresko, W.A. (2011). SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing. *Molecular Methods for Evolutionary Genetics*, Methods in Molecular Biology, 772: 157-178. DOI: 10.1007/978-1-61779-228-1_9

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, 14(8): 2611-2620. DOI: 10.1111/j.1365-294X.2005.02553.x

Futuyma, D. (2013). *Evolution* (3rd ed). Sunderland, MA, USA: Sinauer Associates, Inc.

Hedrén, M. (1994). Morfologisk variation hos näbbstarr och jämtstarr I Sverige. *Svensk Botanisk Tidskrift*, 88(3): 129-141.

Hedrén, M. (2002). Patterns of allozyme and morphological differentiation in the *Carex flava* complex (Cyperaceae) in Fennoscandia. *Nordic Journal of Botany*, 22(3): 257-301. DOI: 10.1111/j.1756-1051.2002.tb01373.x

Hedrén, M. (2004). Species delimitation and the partitioning of genetic diversity - an example from the *Carex flava* complex (Cyperaceae). *Biodiversity and Conservation*, 13(2): 293-316. DOI: 10.1023/B:BIOC.0000006499.20169.f9

Hedrén, M. & Prentice, H.C. (1996). Allozyme variation and racial differentiation in Swedish *Carex lepidocarpa s.l.* (Cyperaceae). *Biological Journal of the Linnean Society*, 59(2): 179-200. DOI: 10.1111/j.1095-8312.1996.tb01460.x

Jiang, H., Lei, R., Ding, S.W. & Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BioMed Central Bioinformatics*, 15(1): 182. DOI: 10.1186/1471-2105-15-182. Available for download at: https://sourceforge.net/projects/skewer/

Jiménez-Mejías, P., Luceño, M. & Matrín-Bravo, S. (2014). Species Boundaries within Southwest Old World Populations of the *Carex flava* Group (Cyperaceae). *Systematic Botany*, 39(1): 117-131. DOI: 10.1600/036364414X677973

Jiménez-Mejías, P., Martín-Bravo, S. & Luceño, M. (2012). Systematic and Taxonomy of *Carex* sect. *Ceratocystis* (Cyperaceae) in Europe: A Molecular and Cytogenetic Approach. *Systematic Botany*, 37(2): 382-398. DOI: 10.1600/036364412X635449

References

Jombart, T. & Ahmed, I. (2011). *adegenet 1.3-1*: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21): 3070-3071. DOI: 10.1093/bioinformatics/btr521

Koopman, J. (2011). *Carex Europaea. The Genus Carex L. (Cyperaceae) in Europe, 1. Accepted Names, Hybrids, Synonyms, Distribution, Chromosome Numbers*. Weikersheim, Germany: Margraf Publishers.

Lemmon, E.M., & Lemmon, A.R. (2013). High-Throughput Genomic Data in Systematics and Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, 44: 99-121. DOI: 10.1146/annurev-ecolsys-110512-135822

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A. & Johnson, E.A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17(2): 240-248. DOI: 10.1101/gr.5681207

Moen, A. & Øien, D.I. (2011). *Våtmarkssystemer.* – In: Lindgaard, A. & Henriksen, S. (eds), 2011. *Norsk rødliste for naturtyper 2011*. Trondheim, Norway: Artsdatabanken. Downloaded 27/06/2016 from http://www.artsdatabanken.no/Article/Article/133431.

Mossberg, B. & Stenberg, L. (2003). *Den nya nordiska Floran*. Tangen, Norway:
Wahlström & Widstrand.

Nielsen, R. & Slatkin, M. (2013). *An Introduction to Population Genetics: Theory and Applications*. Sunderland, MA, USA: Sinauer Associates, Inc.

Palmgren, A. (1926). *C. lepidocarpa* subsp. *jemtlandica* -In: Lindman (eds), C. A. M., *Svensk fanerogamflora* (2nd ed): p. 153. Stockholm, Sweden: Nordstedt & Söner.

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

Pritchard, J.K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *GENETICS*, 155(2): 945-959. Available for download at: http://pritchardlab.stanford.edu/structure.html

Pykälä, J. & Toivonen, H. (1994). Taxonomy of the *Carex flava* complex (Cyperaceae) in Finland. *Nordic Journal of Botany*, 14(2): 173-191. DOI: 10.1111/j.1756-1051.1994.tb00583.x

Reznicek, A.A. (1990). Evolution in sedges (*Carex*, Cyperaceae). *Canadian Journal of Botany*, 68(7): 1409-1432. DOI: 10.1139/b90-180

Schmid, B. (1982). Karyology and hybridization in the *Carex flava* complex in Switzerland. *Feddes Repertorium*, 93(1-2): 23-59. DOI: 10.1002/fedr.19820930103

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): 1-5. DOI: 10.1007/BF00379777

Solstad, H. & Elven, R. (2015). *Karplanter (Pteriodophyta, Pinophyta og Magnoliophyta)*. - In: Henriksen, S. & Hilmo, O. (eds) 2015. *Norsk rødliste for arter 2015*. Trondheim, Norway: Artsdatabanken. Downloaded 27/06/2016 from http://www.artsdatabanken.no/Rodliste/Artsgruppene/Karplanter

Team, R.C. (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available for download at: http://www.R-project.org/

Toivonen, H. (1986). *Carex*. – In: Hämet-Ahti, L., Suominen, J., Ulvinen, T., Uotila, P. & Vuokko, S. (eds), *Retkeilykasvio* (3rd ed): p. 492-526. Helsinki, Finland: Suomen Luonnonsuojelun Tuki Oy.

Weir, B.S. & Cockerham, C C. (1984). Estimating F-statistics for the Analysis of Population Structure. *Evolution*, 38(6): 1358–1370. DOI: 10.2307/2408641

Więcław, H. & Wilhelm, M. (2014). Natural Hybridization within the *Carex flava* Complex (Cyperaceae) in Poland: Morphometric Studies. *Annales Botanici Fennici*, 51(3): 129-147. DOI: 10.5735/085.053.0101

Appendix

Appendix A: Specimens

Table A.1. Collection information regarding samples used in present study. Species identification was made in field and later re-evaluated. Re-assignments are indicated within parenthesis, and slash (/) is used to separate possible re-assignments when more than one was suggested. Sympatric populations are coloured in blue. Abbreviations: dem=*C. demissa*, flava=*C. flava*, jemt=*C. jemtlandica*, lep=*C. lepidocarpa*, pul=*C. viridula* ssp. *pulchella*, NOR=Norway, SWE=Sweden, AKR=Akershus, BUS=Buskerud, HED=Hedmark, NTR=Nord-Trøndelag, STR=Sør-Trøndelag, VFO=Vestfold, AJ=Arne Jakobsen, HS=Heidi Solstad , KIF=Kjell-Ivar Flatberg, LG=Leif Galten, MB=Mika Bendiksby, MØN=Malene Østreng Nygård, RE=Reidar Elven, RH=Reidar Haugan, O=Herbarium of the Natural History Museum in Oslo, TRH=Herbarium of the NTNU University Museum in Trondheim

Sample ID	Species identification	Pop. ID	Nr. ind. in pop.	Country; State province; County	Locality	Coord.	Coll.	Date	Voucher dep. site
l_1401_1	*C. lepidocarpa*								
l_1401_2	*C. lepidocarpa* (lep x dem)								
l_1401_3	*C. lepidocarpa* (lep x dem)								
l_1401_4	*C. lepidocarpa*	1401	4	SWE; Skåne; Simrishamns	Strandäng	55.5365N 14.3254E	HS & RE	1.6.14	O
l_1402_1	*C. lepidocarpa*								
l_1402_2	*C. lepidocarpa*								
l_1402_4	*C. lepidocarpa*								
l_1402_5	*C. lepidocarpa*	1402	4	SWE; Skåne; Tomelilla	Benestad backar	55.5263N 13.8997E	RE & HS	4.6.14	O
l_1403_1	*C. lepidocarpa*								
l_1403_2	*C. lepidocarpa*								
l_1403_3	*C. lepidocarpa*								
l_1403_4	*C. lepidocarpa*								
l_1403_5	*C. lepidocarpa*	1403	5	NOR; BUS; Lier	Haugerudmyra	59.8156N 10.2789E	MØN, RE & HS	18.6.14	O
l_1404_1	*C. lepidocarpa*								
l_1404_2	*C. lepidocarpa*	1404	5	NOR; BUS; Lier	Gjellebekkmyra	59.8115N 10.2921E	MØN, RE & HS	18.6.14	O

l_1404_3	*C. lepidocarpa*							
l_1404_4	*C. lepidocarpa* (lep x dem)							
l_1404_5	*C. lepidocarpa* (lep x dem)							
l_1405_1	*C. lepidocarpa*	1405	1	NOR; AKR; Asker	Dikemark, Lille Oppsjøe	59.8167N 10.3929E	MØN, RE & HS	18.6.14 O
j_1406_1	*C. jemtlandica*							
j_1406_2	*C. jemtlandica* (lep/dem/lep x jemt/flava)							
j_1406_3	*C. jemtlandica*							
j_1406_4	*C. jemtlandica* (lep/jemt/dem)					59.8485N		
j_1406_5	*C. jemtlandica*				Haskollsetermyra	10.1439E		
j_1407_1	*C. jemtlandica*	1406	6	NOR; BUS; Lier;	Haskollsetermyra E, hillside	59.8487N 10.1459E	MØN, RE & HS	18.6.14 O
j_1412_1	*C. jemtlandica*							
j_1412_3	*C. jemtlandica*							
j_1412_4	*C. jemtlandica*					59.9585N	MØN, RE	
j_1412_5	*C. jemtlandica*	1412	4	NOR; BUS; Lier	Hornsetra W, mire	10.2425E	& HS	19.6.14 O
j_1413_1	*C. jemtlandica*							
j_1413_2	*C. jemtlandica*							
j_1413_3	*C. jemtlandica*							
j_1413_4	*C. jemtlandica*					59.9677N	MØN, RE	
j_1413_5	*C. jemtlandica*	1413	5	NOR; BUS; Lier	Ringemyr	10.2328E	& HS	19.6.14 O
j_1424_1	*C. jemtlandica*	1424	1	NOR; BUS; Nedre Eiker	Brensetjønn S; Bremsa	59.7136N 10.0662E	MØN, RE & HS	20.6.14 O
j_1425_1	*C. jemtlandica*							
j_1425_2	*C. jemtlandica*							
j_1425_3	*C. jemtlandica*			NOR; BUS; Nedre	Rismyr naturereserve	59.7056N	MØN, RE	
j_1425_4	*C. jemtlandica*	1425	4	Eiker	NE	10.0794E	& HS	20.6.14 O
lj_1426_1	*C. jemtlandica* x *C. lepidocarpa* (jemt/jemt x lep)							
lj_1426_2	*C. jemtlandica* x *C. lepidocarpa* (jemt)							
lj_1426_3	*C. jemtlandica* x *C. lepidocarpa* (jemt)			NOR; BUS; Nedre	Rismyr naturereserve	59.7050N	MØN, RE	
lj_1426_4	*C. jemtlandica* x *C. lepidocarpa* (jemt/jemt x lep)	1426	4	Eiker	SE	10.0812E	& HS	20.6.14 O
l_1429_1	*C.* cf. *lepidocarpa* (dem/dem x lep)	1429	3	NOR; VFO; Larvik	Rakke	58.9783N	MØN &	20.6.14 O

l_1429_2	*C.* cf. *lepidocarpa* (dem x lep)					10.0305E	RE		
l_1429_3	*C.* cf. *lepidocarpa* (dem x lep)								
l_1430_1	*C. lepidocarpa*								
l_1430_2	*C. lepidocarpa*								
l_1430_3	*C. lepidocarpa*								
l_1430_4	*C. lepidocarpa* (dem x lep)				Dalehamna W,	63.678N	MØN, HS		
l_1430_5	*C. lepidocarpa*	1430	5	NOR; STR; Ørland	Storfosna	9.399E	& RH	13.7.14	TRH
j_1433_1	*C. jemtlandica* (jemt x flava/flava)					62.6202°			
j_1433_4	*C. jemtlandica* (jemt x flava/flava)					N	MØN, LG		
j_1433_5	*C. jemtlandica* (jemt x flava/flava)	1433	3	NOR; STR; Røros	Jamtbekken	11.3727°E	& MB	15.7.14	TRH
j_1435_1	*C. jemtlandica* (dem x jemt)								
j_1435_2	*C. jemtlandica*								
j_1435_3	*C. jemtlandica* (dem x jemt)								
j_1435_4	*C. jemtlandica* (dem x jemt)					62.5658°			
j_1435_5	*C. jemtlandica* (dem x jemt)				Nedre Hånesset,	N	MØN, LG		
j_1435_6	*C. jemtlandica* (dem x jemt)	1435	6	NOR; STR; Røros	Joesvika	11.3617°E	& MB	15.7.14	TRH
j_1436_1	*C. jemtlandica* (dem/pul)								
j_1436_2	*C. jemtlandica* (jemt/flava)								
j_1436_3	*C. jemtlandica* (flava)					62.5652°			
j_1436_4	*C. jemtlandica* (jemt/flava)				Hånesset, 1 km south	N	MØN, LG		
j_1436_5	*C. jemtlandica* (jemt/flava)	1436	5	NOR; STR; Røros	of Sjåfram	11.332°E	& MB	15.7.14	TRH
j_1438_2	*C. jemtlandica*								
j_1438_3	*C. jemtlandica*					62.3337°			
j_1438_4	*C. jemtlandica*					N	MØN, LG		
j_1438_5	*C. jemtlandica*	1438	4	NOR; HED; Os	Siksjølia	11.6295°E	& MB	15.7.14	TRH
j_1439_1	*C. jemtlandica*								
j_1439_2	*C. jemtlandica*								
j_1439_3	*C. jemtlandica*					61.7712°			
j_1439_4	*C. jemtlandica*			NOR; HED;	Bjørbekkåsen,	N	MØN, LG		
j_1439_5	*C. jemtlandica*	1439	5	Engerdal	Stormyra	11.6905°E	& MB	16.7.14	TRH
j_1441_2	*C. jemtlandica*					61.7485°			
j_1441_3	*C. jemtlandica*			NOR; HED;		N	MØN, LG		
j_1441_4	*C. jemtlandica*	1441	8	Engerdal	Ågård W; Granberget	11.7353°E	& MB	16.7.14	TRH

j_1441_5	*C. jemtlandica*								
j_1442_1	*C. jemtlandica*								
j_1442_2	*C. jemtlandica*					61.7482°			
j_1442_3	*C. jemtlandica*				Ågård WW,	N			
j_1442_5	*C. jemtlandica*				Granberget	11.7367°E			
j_1443_1	*C. jemtlandica*								
j_1443_2	*C. jemtlandica*					61.8242°			
j_1443_3	*C. jemtlandica*			NOR; HED;		N	MØN, LG		
j_1443_4	*C. jemtlandica*	1443	4	Engerdal	Sundsetra	11.7997°E	& MB	16.7.14	TRH
j_1444_1	*C. jemtlandica*								
j_1444_2	*C. jemtlandica*								
j_1444_3	*C. jemtlandica*					61.8272°			
j_1444_4	*C. jemtlandica*			NOR; HED;	Engerdalssetra N,	N	MØN, LG		
j_1444_5	*C. jemtlandica*	1444	5	Engerdal	Lillevold	11.9002°E	& MB	16.7.14	TRH
j_1445_1	*C. jemtlandica*								
j_1445_2	*C. jemtlandica* (jemt/flava)					61.8773°			
j_1445_3	*C. jemtlandica*			NOR; HED;		N	MØN, LG		
j_1445_4	*C. jemtlandica* (jemt/flava)	1445	4	Engerdal	Lilleborg, Myrneset	12.1487°E	& MB	16.7.14	TRH
j_1446_1	*C. jemtlandica*								
j_1446_2	*C. jemtlandica*								
j_1446_3	*C. jemtlandica*					63.7213°	MØN,		
j_1446_4	*C. jemtlandica*					N	KIF &		
j_1446_5	*C. jemtlandica*	1446	5	NOR; NTR; Verdal	Kaldvassmyra	11.586°E	MB	17.7.14	TRH
lj_1447_1	*C. jemtlandica* x *C. lepidocarpa*								
lj_1447_2	*C. jemtlandica* x *C. lepidocarpa*								
lj_1447_3	*C. jemtlandica* x *C. lepidocarpa*					63.7232°	MØN,		
lj_1447_4	*C. jemtlandica* x *C. lepidocarpa*					N	KIF &		
lj_1447_5	*C. jemtlandica* x *C. lepidocarpa*	1447	5	NOR; NTR; Verdal	Kaldvassmyra W	11.5803°E	MB	17.7.14	TRH
l_1448_1	*C. lepidocarpa*								
l_1448_2	*C. lepidocarpa*								
l_1448_3	*C. lepidocarpa* (lep/lep x jemt)					63.7252°	MØN,		
l_1448_4	*C. lepidocarpa* (lep/lep x jemt)					N	KIF &		
l_1448_5	*C. lepidocarpa* (lep/lep x jemt)	1448	5	NOR; NTR; Verdal	Kaldvassmyra N	11.5818°E	MB	17.7.14	TRH

j_1461_1	*C. jemtlandica* (jemt/jemt x flava)								
j_1461_3	*C. jemtlandica* (jemt/jemt x flava)								
j_1461_4	*C. jemtlandica*				Risvollen E, mire				
j_1462_2	*C. jemtlandica* (jemt/jemt x flava)					62.5657N			
j_1462_4	*C. jemtlandica*	1461	5	NOR; STR; Røros	Risvollen E, mire	11.3623E	HS & RE	15.7.14	TRH
j_1463_1	*C. jemtlandica*								
j_1463_2	*C. jemtlandica*								
j_1463_3	*C. jemtlandica*								
j_1463_4	*C. jemtlandica*								
j_1463_6	*C. jemtlandica*					62.4058N			
j_1463_7	*C. jemtlandica*	1463	6	NOR; HED; Tolga	Knappåssetra NW	11.1446E	HS & RE	16.7.14	TRH
l_1482_2	*C. lepidocarpa* (jemt x lep/lep/dem x lep)								
l_1482_3	*C. lepidocarpa* (dem x lep)								
l_1482_4	*C. lepidocarpa*			NOR; STR; Midtre	Røssvatnet S, rich area	63.0470N			
l_1482_5	*C. lepidocarpa* (flava x jemt/lep x dem/lep)	1482	4	Gauldal	beside stream	10.3252E	HS	30.7.14	TRH
l_1484_1	*C. lepidocarpa*								
l_1484_2	*C. lepidocarpa*								
l_1484_3	*C. lepidocarpa* (lep/lep x jemt)								
l_1484_4	*C. lepidocarpa* (lep/lep x jemt)			NOR; STR;	Fjellsetermyrene I,	63.4161N			
l_1484_5	*C. lepidocarpa*	1484	5	Trondheim	rich mire	10.2878E	KIF & HS	31.7.14	TRH
l_1485_1	*C. lepidocarpa*								
l_1485_2	*C. lepidocarpa*								
l_1485_3	*C. lepidocarpa*								
l_1485_4	*C. lepidocarpa*			NOR; STR;	Fjellsetermyrene II,	63.4138N			
l_1485_5	*C. lepidocarpa*	1485	5	Trondheim	rich mire	10.2877E	KIF & HS	31.7.14	TRH
l_1501_1	*C. lepidocarpa*								
l_1501_2	*C. lepidocarpa*								
l_1501_3	*C. lepidocarpa*					63,60063			
l_1501_4	*C. lepidocarpa*					N			
l_1501_5	*C. lepidocarpa*	1501	5	NOR; STR; Agdenes	Eidemstjørna	9,57435E	MØN	27.6.15	TRH
j_B_2	*C. jemtlandica*	B	2	NOR; HED; Rendalen	Harrsjøvollen N, Raudbekken	61.8801° N 11.0457°E	LG	22.8.14	TRH

j_Bbb_1	*C. jemtlandica*				Harrsjøvollen N	61.875°N 11.0387°E			
j_E_1	*C. jemtlandica*								
j_E_2	*C. jemtlandica*								
j_E_3	*C. jemtlandica*								
j_E_4	*C. jemtlandica*			NOR; STR; Midtre	Støren, Enamyra, 2 km	63.0094° N			
j_E_5	*C. jemtlandica*	E	5	Gauldal	S of Støren church	10.2915°E	LG	9.8.14	TRH
j_Ee_5	*C. jemtlandica*								
l_Ee_1	*C. lepidocarpa*					63.0088°			
l_Ee_2	*C. lepidocarpa*				Støren, Enamyra, 2 km	N			
lj_Ee_4	*C. jemtlandica* x *C. lepidocarpa*				S of Støren church	10.2816°E		12.8.14	
l_Aa_1	*C. lepidocarpa*	Ee	5	NOR; STR; Midtre Gauldal	Støren, Enamyra S, Skårvollmarka	63.0081° N 10.2824°E	LG	25.6.14	TRH
l_K_1	*C. lepidocarpa*					62.5723°			
l_K_2	*C. lepidocarpa*					N			
l_K_5	*C. lepidocarpa*	K	3	NOR; HED; Tynset	Kvukne, Sverja S	10.3211°E	LG	10.8.14	TRH
l_R_4	*C. lepidocarpa*	R	1	NOR; STR; Midtre Gauldal	Åsmotet E, Refsåsen, 1,5 km E of residental area. Frøset in Støren.	63.0442° N 10.3621°E	LG, KIF & AJ	12.8.14	TRH

Table A.2. Sample information in the present study. Species identification was made in field and later re-evaluated. Re-assignments are indicated within parenthesis, and slash (/) is used to separate possible re-assignments when more than one were suggested. Sympatric populations are coloured in blue. Samples with more than 50% missing data were discarded form the data set. Pure *C. jemtlandica*, *C. lepidocarpa* and admixed between these two are defined according to *Structure* posterior probability <0.05, >0.95, and between 0.05 and 0.95, respectively. Individuals identified as outliers in data set created without a reference catalogue are marked with x under heading Outliers. Samples containing the least successful barcode (containing many mismatches; Appendix D) are marked with x under heading P2-barcode TGCAT. Abbreviations: dem=*C. demissa*, flava=*C. flava*, jemt=*C. jemtlandica*, lep=*C. lepidocarpa*, pul=*C. viridula* ssp. *pulchella*.

Sample ID	Species identification	Missing data (%)	*Structure* posterior probability				Outliers	P2-barcode TGCAT
			Pure jemt	Admixed	Pure lep	90% prob. Interval		
l_1401_1	*C. lepidocarpa*	43		0.678		(0.567,0.788)		
l_1401_2	*C. lepidocarpa* (lep x dem)	33	0.002			(0.000,0.011)		x
l_1401_3	*C. lepidocarpa* (lep x dem)	40		0.751		(0.656,0.845)		
l_1401_4	*C. lepidocarpa*	34		0.817		(0.724,0.905)		
l_1402_1	*C. lepidocarpa*	37		0.525		(0.440,0.610)		
l_1402_2	*C. lepidocarpa*	41			0.954	(0.857,1.000)		x
l_1402_4	*C. lepidocarpa*	47		0.562		(0.463,0.663)		
l_1402_5	*C. lepidocarpa*	31	0.001			(0.000,0.009)		x
l_1403_1	*C. lepidocarpa*	20			0.999	(0.993,1.000)		x
l_1403_2	*C. lepidocarpa*	>50						x
l_1403_3	*C. lepidocarpa*	25			0.999	(0.992,1.000)		x
l_1403_4	*C. lepidocarpa*	27			0.997	(0.982,1.000)		
l_1403_5	*C. lepidocarpa*	26			0.999	(0.991,1.000)		x
l_1404_1	*C. lepidocarpa*	28			0.999	(0.993,1.000)		x
l_1404_2	*C. lepidocarpa*	36			0.998	(0.991,1.000)		
l_1404_3	*C. lepidocarpa*	29			0.999	(0.993,1.000)		x
l_1404_4	*C. lepidocarpa* (lep x dem)	40		0.522		(0.427,0.619)		x
l_1404_5	*C. lepidocarpa* (lep x dem)	34			0.997	(0.983,1.000)		x

l_1405_1	*C. lepidocarpa*	48		0.621		(0.526,0.717)	x	x
j_1406_1	*C. jemtlandica*	31	0.004			(0.000,0.025)		x
j_1406_2	*C. jemtlandica* (lep/dem/lep x jemt/flava)	>50						
j_1406_3	*C. jemtlandica*	37	0.005			(0.000,0.028)		
j_1406_4	*C. jemtlandica* (lep/jemt/dem)	17		0.075		(0.034,0.121)		x
j_1406_5	*C. jemtlandica*	26	0.002			(0.000,0.012)		x
j_1407_1	*C. jemtlandica*	>50						
j_1412_1	*C. jemtlandica*	38	0.001			(0.000,0.007)		
j_1412_3	*C. jemtlandica*	27	0.001			(0.000,0.004)		
j_1412_4	*C. jemtlandica*	23	0.002			(0.000,0.010)		x
j_1412_5	*C. jemtlandica*	37			0.998	(0.991,1.000)		
j_1413_1	*C. jemtlandica*	16	0.001			(0.000,0.004)		x
j_1413_2	*C. jemtlandica*	17	0.001			(0.000,0.004)		x
j_1413_3	*C. jemtlandica*	19	0.001			(0.000,0.005)		x
j_1413_4	*C. jemtlandica*	15	0.001			(0.000,0.004)		x
j_1413_5	*C. jemtlandica*	18	0.001			(0.000,0.005)		x
j_1424_1	*C. jemtlandica*	25	0.001			(0.000,0.005)		
j_1425_1	*C. jemtlandica*	40	0.001			(0.000,0.007)		
j_1425_2	*C. jemtlandica*	31	0.001			(0.000,0.004)		
j_1425_3	*C. jemtlandica*	38	0.001			(0.000,0.005)		
j_1425_4	*C. jemtlandica*	29	0.001			(0.000,0.004)		
lj_1426_1	*C. jemtlandica* x *C. lepidocarpa* (jemt/jemt x lep)	39		0.153		(0.091,0.221)		
lj_1426_2	*C. jemtlandica* x *C. lepidocarpa* (jemt)	26		0.175		(0.119,0.236)		
lj_1426_3	*C. jemtlandica* x *C. lepidocarpa* (jemt)	42		0.173		(0.106,0.247)		
lj_1426_4	*C. jemtlandica* x *C. lepidocarpa* (jemt/jemt x lep)	33		0.238		(0.165,0.317)		
l_1429_1	*C.* cf. *lepidocarpa* (dem/dem x lep)	49			0.971	(0.893,1.000)		
l_1429_2	*C.* cf. *lepidocarpa* (dem x lep)	39		0.926		(0.837,1.000)		
l_1429_3	*C.* cf. *lepidocarpa* (dem x lep)	42		0.883		(0.800,0.967)		
l_1430_1	*C. lepidocarpa*	32			0.978	(0.905,1.000)		

l_1430_2	*C. lepidocarpa*	34			0.992	(0.954,1.000)	
l_1430_3	*C. lepidocarpa*	38			0.993	(0.957,1.000)	
l_1430_4	*C. lepidocarpa* (dem x lep)	36			0.998	(0.986,1.000)	
l_1430_5	*C. lepidocarpa*	31			0.998	(0.989,1.000)	
j_1433_1	*C. jemtlandica* (jemt x flava/flava)	>50					
j_1433_4	*C. jemtlandica* (jemt x flava/flava)	46		0.698		(0.591,0.805)	x
j_1433_5	*C. jemtlandica* (jemt x flava/flava)	>50					
j_1435_1	*C. jemtlandica* (dem x jemt)	35	0.001			(0.000,0.006)	
j_1435_2	*C. jemtlandica*	31	0.001			(0.000,0.007)	
j_1435_3	*C. jemtlandica* (dem x jemt)	36		0.688		(0.597,0.779)	
j_1435_4	*C. jemtlandica* (dem x jemt)	35	0.001			(0.000,0.006)	
j_1435_5	*C. jemtlandica* (dem x jemt)	37	0.001			(0.000,0.008)	
j_1435_6	*C. jemtlandica* (dem x jemt)	34	0.001			(0.000,0.008)	
j_1436_1	*C. jemtlandica* (dem/pul)	35		0.699		(0.608,0.789)	
j_1436_2	*C. jemtlandica* (jemt/flava)	>50					x
j_1436_3	*C. jemtlandica* (flava)	>50					x
j_1436_4	*C. jemtlandica* (jemt/flava)	50		0.619		(0.517,0.720)	x
j_1436_5	*C. jemtlandica* (jemt/flava)	>50					
j_1438_2	*C. jemtlandica*	33	0.001			(0.000,0.009)	
j_1438_3	*C. jemtlandica*	34	0.001			(0.000,0.008)	
j_1438_4	*C. jemtlandica*	20	0.001			(0.000,0.005)	
j_1438_5	*C. jemtlandica*	31	0.001			(0.000,0.004)	
j_1439_1	*C. jemtlandica*	37		0.8		(0.700,0.900)	
j_1439_2	*C. jemtlandica*	46		0.501		(0.414,0.590)	x
j_1439_3	*C. jemtlandica*	31	0.001			(0.000,0.005)	
j_1439_4	*C. jemtlandica*	41		0.601		(0.511,0.693)	x
j_1439_5	*C. jemtlandica*	48		0.624		(0.519,0.730)	x
j_1441_2	*C. jemtlandica*	30	0.001			(0.000,0.008)	
j_1441_3	*C. jemtlandica*	29	0.002			(0.000,0.013)	

j_1441_4	*C. jemtlandica*	32	0.001		(0.000,0.006)
j_1441_5	*C. jemtlandica*	22	0.001		(0.000,0.005)
j_1442_1	*C. jemtlandica*	28	0.001		(0.000,0.004)
j_1442_2	*C. jemtlandica*	32	0.001		(0.000,0.004)
j_1442_3	*C. jemtlandica*	24	0.001		(0.000,0.004)
j_1442_5	*C. jemtlandica*	44	0.001		(0.000,0.007)
j_1443_1	*C. jemtlandica*	25	0.001		(0.000,0.008)
j_1443_2	*C. jemtlandica*	22	0.024		(0.000,0.061)
j_1443_3	*C. jemtlandica*	25	0.001		(0.000,0.007)
j_1443_4	*C. jemtlandica*	22	0.001		(0.000,0.009)
j_1444_1	*C. jemtlandica*	19	0.001		(0.000,0.005)
j_1444_2	*C. jemtlandica*	29	0.001		(0.000,0.005)
j_1444_3	*C. jemtlandica*	23	0.001		(0.000,0.005)
j_1444_4	*C. jemtlandica*	36	0.001		(0.000,0.007)
j_1444_5	*C. jemtlandica*	33	0.002		(0.000,0.013)
j_1445_1	*C. jemtlandica*	37	0.011		(0.000,0.054)
j_1445_2	*C. jemtlandica* (jemt/flava)	29		0.052	(0.000,0.120)
j_1445_3	*C. jemtlandica*	44		0.149	(0.076,0.228)
j_1445_4	*C. jemtlandica* (jemt/flava)	30	0.02		(0.000,0.070)
j_1446_1	*C. jemtlandica*	32	0.001		(0.000,0.009)
j_1446_2	*C. jemtlandica*	28	0.002		(0.000,0.011)
j_1446_3	*C. jemtlandica*	30	0.001		(0.000,0.007)
j_1446_4	*C. jemtlandica*	32	0.001		(0.000,0.007)
j_1446_5	*C. jemtlandica*	31	0.001		(0.000,0.007)
lj_1447_1	*C. jemtlandica* x *C. lepidocarpa*	22		0.435	(0.357,0.516)
lj_1447_2	*C. jemtlandica* x *C. lepidocarpa*	24		0.608	(0.528,0.687)
lj_1447_3	*C. jemtlandica* x *C. lepidocarpa*	24		0.46	(0.383,0.539)
lj_1447_4	*C. jemtlandica* x *C. lepidocarpa*	34		0.131	(0.077,0.192)
lj_1447_5	*C. jemtlandica* x *C. lepidocarpa*	29		0.447	(0.372,0.524)

l_1448_1	*C. lepidocarpa*	42			0.994	(0.962,1.000)		
l_1448_2	*C. lepidocarpa*	43			0.998	(0.988,1.000)		
l_1448_3	*C. lepidocarpa* (lep/lep x jemt)	27		0.124		(0.073,0.182)		
l_1448_4	*C. lepidocarpa* (lep/lep x jemt)	20		0.118		(0.068,0.175)		x
l_1448_5	*C. lepidocarpa* (lep/lep x jemt)	18	0.001			(0.000,0.006)		x
j_1461_1	*C. jemtlandica* (jemt/jemt x flava)	32	0.019			(0.000,0.067)		
j_1461_3	*C. jemtlandica* (jemt/jemt x flava)	32	0.002			(0.000,0.014)		
j_1461_4	*C. jemtlandica*	32	0.004			(0.000,0.024)		
j_1462_2	*C. jemtlandica* (jemt/jemt x flava)	40		0.7		(0.602,0.798)	x	
j_1462_4	*C. jemtlandica*	47		0.598		(0.487,0.711)	x	
j_1463_1	*C. jemtlandica*	26	0.001			(0.000,0.005)		
j_1463_2	*C. jemtlandica*	21	0.001			(0.000,0.004)		
j_1463_3	*C. jemtlandica*	19	0.001			(0.000,0.004)		
j_1463_4	*C. jemtlandica*	23	0.001			(0.000,0.005)		
j_1463_6	*C. jemtlandica*	22	0.001			(0.000,0.004)		
j_1463_7	*C. jemtlandica*	18	0.001			(0.000,0.005)		
l_1482_2	*C. lepidocarpa* (jemt x lep/lep/dem x lep)	42		0.118		(0.060,0.183)		
l_1482_3	*C. lepidocarpa* (dem x lep)	>50						
l_1482_4	*C. lepidocarpa*	24			0.983	(0.911,1.000)		
l_1482_5	*C. lepidocarpa* (flava x jemt/lep x dem/lep)	34		0.128		(0.070,0.192)		
l_1484_1	*C. lepidocarpa*	25	0.001			(0.000,0.007)		x
l_1484_2	*C. lepidocarpa*	36			0.979	(0.900,1.000)		x
l_1484_3	*C. lepidocarpa* (lep/lep x jemt)	23		0.439		(0.359,0.522)		x
l_1484_4	*C. lepidocarpa* (lep/lep x jemt)	29		0.48		(0.399,0.563)		x
l_1484_5	*C. lepidocarpa*	49		0.934		(0.802,1.000)		x
l_1485_1	*C. lepidocarpa*	26			0.998	(0.990,1.000)		
l_1485_2	*C. lepidocarpa*	34		0.879		(0.781,0.999)		x
l_1485_3	*C. lepidocarpa*	40		0.91		(0.795,1.000)		x
l_1485_4	*C. lepidocarpa*	>50						x

l_1485_5	*C. lepidocarpa*	39			0.999	(0.994,1.000)		x
l_1501_1	*C. lepidocarpa*	47			0.998	(0.990,1.000)		
l_1501_2	*C. lepidocarpa*	40			0.998	(0.989,1.000)		
l_1501_3	*C. lepidocarpa*	41			0.998	(0.988,1.000)		
l_1501_4	*C. lepidocarpa*	32			0.999	(0.993,1.000)		
l_1501_5	*C. lepidocarpa*	30			0.996	(0.975,1.000)		
j_B_2	*C. jemtlandica*	40			0.986	(0.923,1.000)		x
j_Bbb_1	*C. jemtlandica*	34		0.477		(0.392,0.564)		x
j_E_1	*C. jemtlandica*	>50						
j_E_2	*C. jemtlandica*	42		0.759		(0.670,0.845)		
j_E_3	*C. jemtlandica*	29		0.649		(0.566,0.731)		
j_E_4	*C. jemtlandica*	>50						
j_E_5	*C. jemtlandica*	>50						
j_Ee_5	*C. jemtlandica*	35	0.002			(0.000,0.014)		x
l_Ee_1	*C. lepidocarpa*	43			0.998	(0.989,1.000)		
l_Ee_2	*C. lepidocarpa*	43			0.997	(0.984,1.000)		
lj_Ee_4	*C. jemtlandica* x *C. lepidocarpa*	18		0.356		(0.281,0.436)		
l_Aa_1	*C. lepidocarpa*	33		0.911		(0.809,1.000)	x	x
l_K_1	*C. lepidocarpa*	38			0.989	(0.944,1.000)		
l_K_2	*C. lepidocarpa*	42			0.993	(0.957,1.000)		
l_K_5	*C. lepidocarpa*	39			0.996	(0.975,1.000)		
l_R_4	*C. lepidocarpa*	44			0.998	(0.985,1.000)		

Appendix B: Preparing RAD-tag libraries

Copied with permission from Yousefi et al. (in prep.)

RAD-tag libraries were constructed as in Etter *et al.* (2011) with small modifications. As the first step, DNA samples were normalized to 300ng in the strips. To digest the DNA, 5µl cut smart buffer and 0.5µl sbfI restriction enzyme was added to each sample and vortexed well. Samples were incubated at 37˚C for 90 min and the enzyme was inactivated at 80˚C for 20 min. Samples were then cooled down in room temperature (RT). In order to ligate P1 adaptors to the digested DNA, a master mix of 1µl cut smart, 0.6µl rATP, 3.9µl water and 0.5µl T4 DNA ligase, each per sample, was prepared. 4µl of P1 adaptors with unique barcodes were added to each sample. Then, 6µl of the master mix was added to each sample. Samples were incubated at 16˚C overnight and the enzyme was inactivated at 65˚C for 20 min. Each 24 barcoded samples were pooled in a 2 ml microtube. Half of the volume was kept as back up and the other half was used for shearing. Samples were divided in 7-8 sonicator microtbes (c 100µl in each), and shearing was performed using a Bioruptor Pico sonicator with 8 cycles, 15second on, 90 seconds off. For degraded samples, only 4 cycles was used. A test 1.25 % agarose gel with 3µl Sybr safe was prepared and run with 10µl of sheared samples and 2µl orange loading dye. 100bp and 50bp DNA ladders were used on each side of the samples. The gel was run for 1 hour with 85 V. Samples were then concentrated using Qiaquick PCR purification kit, following the instruction, and 21µl elution buffer was added at the end. 4µl orange loading dye was added to the 21µl concentrated libraries and run on a gel, similar to previous step. The majority of library is expected to be between 250-500bp. This bright part was cut using a sharp razor blade and put in a 2ml microtube. The libraries were extracted from the gel using MinElute gel extraction kit (Qiagen), following the instruction except that after adding the QG buffer, samples were incubated at 22°C for 20-25 min with 500-750 rpm to dissolve the gel slice. At the end, 20µl elution buffer was added. To end-repair the fragments, blunt enzyme mix was used. 2.5µl blunting buffer, 2.5µl dNTP and 1µl blunt enzyme was added. Libraries were incubated at 25°C for 30 min and then cleaned using Ampure beads (Agencourt AMPure XP) with the ratio of 0.75 : 1 (beads : sample) following the instruction. This ratio and instruction is used for all of the following cleaning steps. Then 43 µl elution buffer was added. The next step was to add a 3’ –dA overhang to the end of fragments. 5µl NEB2 buffer, 3µl

Klenow (exo) and 1µl dATP was added to each library and incubated at 37°C for 30 min and cooled down at RT for 15 min. Libraries were then cleaned using Ampure beads and then 45µl elution buffer was added. In order to ligate the P2 adaptors (barcoded), 5µl NEB2 buffer, 1µl P2 adapter, 0.5µl rATP and 0.5µl T4 DNA Ligase was added to each library and incubated at 25°C for 30 min, then cleaned using Ampure beads. Finally, 52µl elution buffer was added. A preliminary PCR was performed using 6.5µl water, 12.5 µl Phusion High-Fidelity Master Mix, 1µl of each forward and reverse Solexa primer mix (10 µl) and 4µl of template library. The program 98^0C, 30 sec; 20 * (98^0C, 10 sec; 65^0C 30 sec; 72^0C, 30 sec); 72^0C 10 min was used. A test gel was run using 5µl of PCR product and 1µl of library template each mixed with 1µl orange loading dye. The PCR products were at least twice as bright as template. Then a massive PCR was run using 26µl water, 50 µl Phusion High-Fidelity Master Mix, 4µl of each forward and reverse Solexa primer mix (10 µl) and 16µl of template library. The samples were divided into 8 strips to minimize the PCR bias. The PCR program 98^0C, 30 sec; 16 * (98^0C, 10 sec; 65^0C 30 sec; 72^0C, 30 sec); 72^0C 10 min was used. The PCR products were cleaned using Ampure beads, and in order to remove all of the fragments bellow 250bp, including free adaptors, the cleaning was repeated once more. Finally, 20µl elution buffer was added. The quality was measured using Qubit v. 2.0 and submitted for sequencing.

Reference

Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., & Cresko, W.A. (2011). SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing. *Molecular Methods for Evolutionary Genetics*, Methods in Molecular Biology, 772: 157-178. DOI: 10.1007/978-1-61779-228-1_9

Appendix C: Barcodes

Table C.1. P1-barcodes, differing with at least two nucleotides.

Index	P1-barcode
1	AAAAA
2	AACCC
3	AATTT
4	ACGTA
5	AGGAC
6	ATTAG
7	CCAAC
8	CCCCA
9	CCGGT
10	CCTTG
11	CGGCG
12	CTAGG
13	GAAGC
14	GACTA
15	GGAAG
16	GGCCT
17	GGGGA
18	GTTGT
19	TAATG
20	TCTCT
21	TGCAA
22	TTCCG
23	TTGGC
24	TTTTA

Table C.2. P2-barcodes, differing at all nucleotides.

Index	P2-barcode
1	ACTAG
2	TGATC
3	CACGT
4	GTGCA

Appendix D: De-multiplexing

The last nucleotide (T) of the Illumina PCR Reverse sequencing primer was missing due to design error. As a consequence, the sequencing did not initiate until the second P2-barcode nucleotide in all reverse reads (i.e. the first P2-barcode nucleotide was never recorded). To account for this, I included the T-overhang as the fifth P2-barcode nucleotide (Table D.1). The sequencing primer design error did not have any apparent negative impact on sequencing success. However, I observed an increased number of mismatches in one of the P2-barcodes (marked with asterisk in Table D.1). This barcode comprised only ~1% of the total retained read count (instead of the expected ~25%) in both lanes when attempting to de-multiplex according to barcodes (one nucleotide mismatch allowed) using the *process_radtags* program in *Stacks* v. 3.5 (Catchen *et al*., 2013).

Table D.1. Original and adjusted P2-barcodes used in this study. First nucleotide of original barcode is missing in all reverse reads due to Illumina PCR Reverse sequencing primer design error. Thus, a T-overhang is used as fifth barcode base. The least successful P2-barcode (containing more mismatches compared to other P2-barcodes) is marked with an asterisk (*).

Original P2-barcode	ACTAG	TGATC	CACGT	GTGCA*
Adjusted P2-barcode	CTAGT	GATCT	ACGTT	TGCAT*
***Skewer* P2-barcodes**	CTAGT	GATCT	ACGTT	TNNNN

The *process_radtags* output provides a list of recorded reads and their barcodes, in addition to those not recorded (containing more nucleotide mismatches in the barcode than allowed when de-multiplexing). I observed that most of the reads not recorded contained: (1) a P2-barcode with a T in first nucleotide position, (2) two or more mismatches between second to fifth P2-barcode nucleotide position compared to the least successful P2-barcode TGCAT, and (3) a P1-barcode with no nucleotide mismatches. On the other hand, reads not recorded with P2-barcode more similar to the other P2-barcodes (CTAGT, GATCT, and AGCTT) also contained nucleotide mismatches in the P1-barcode. Thus, all of the reads not recorded that most likely did not belong to the least successful P2-barcode (TGCAT) would probably also have nucleotide mismatches in the P1-barcode.

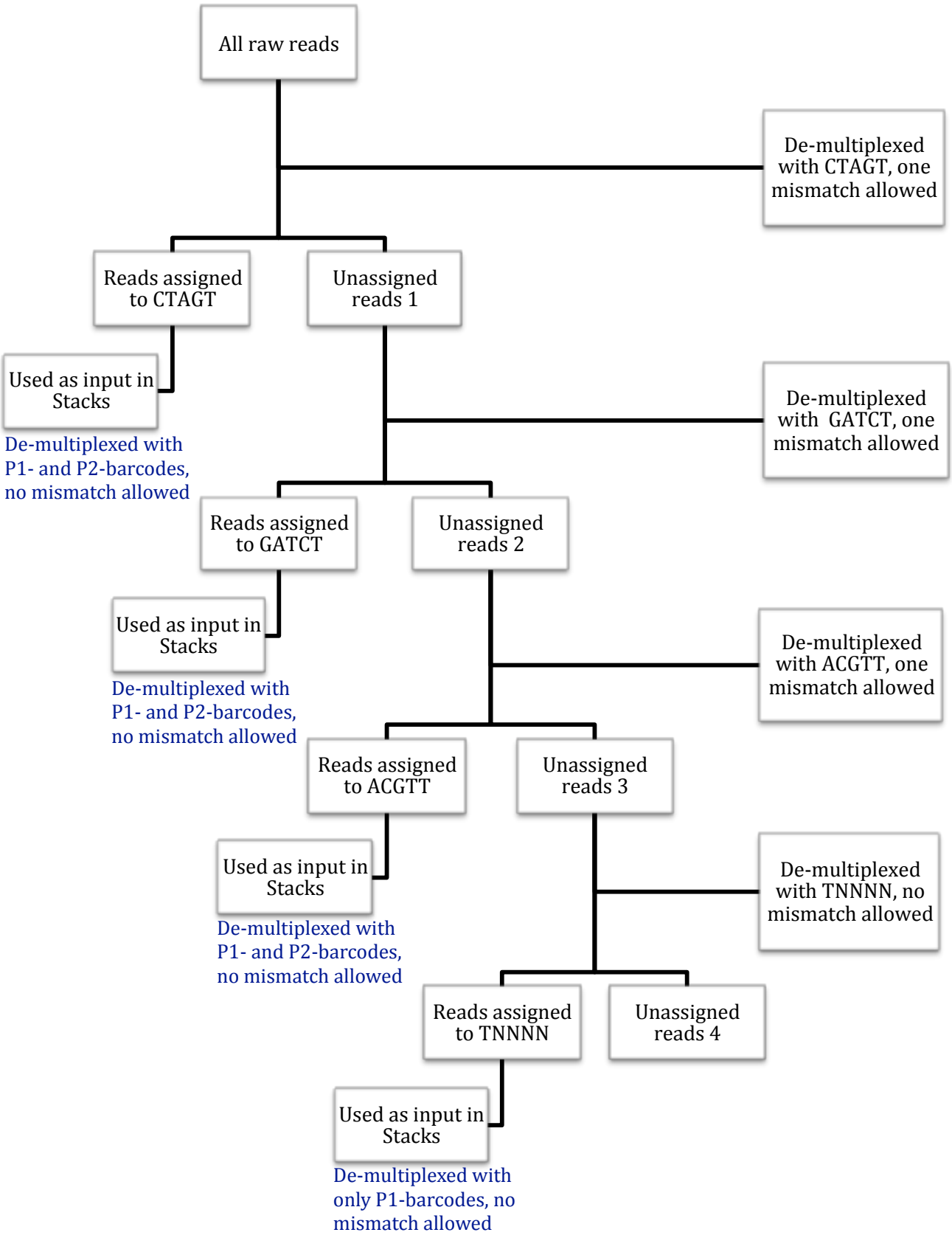

Figure D.1. De-multiplexing procedure using both *Skewer* v. 0.2.1 (Jiang *et al*., 2014) and *Stacks* v. 3.5 (Catchen *et al*., 2013). The process of elimination was utilized to sort out reads that most likely contained the P2-barcode TGCAT. Steps marked in black and blue was conducted in *Skewer* and *Stacks*, respectively.

Compared to *Stacks*, the software *Skewer* v. 0.2.1 (Jiang *et al.*, 2014) has more options with regard to de-multiplexing. For instance, barcode sequences can contain N's, and it can produce three output files: (1) assigned (de-multiplexed) reads (2) excluded (low quality) reads, and (3) unassigned reads (did not fulfil the barcode requirements). By utilizing these options and the process of elimination I managed to sort out reads that most probably contain the least successful P2-barcode (Table D.1), as explained in Figure D.1.

The first three de-multiplexing steps excluded all reads containing the P2-barcodes CTAGT, GATCT, and ACGTT with no or one nucleotide mismatch from the "pool" of raw reads. By doing so, I excluded all reads with a potential T in the first P2-barcode nucleotide position that belonged to another P2-barcode than TGCAT. The last Skewer de-multiplexing step separated the remaining reads into two groups, those that did (assigned) and did not (unassigned) contain a P2-barcode with a T in first nucleotide position. The "assigned to TNNNN" group comprised all remaining reads containing a P2-barcode with a T in first nucleotide position. In other words, the last de-multiplexed group comprised: (1) reads containing the P2-barcode TGCAT, (2) reads containing the P2-barcodes TGCAT with nucleotide mismatches in second to fifth nucleotide position, (3) reads containing other P2-barcodes (CTAGT, GATCT, and AGCTT) with a reading error (T) in the first barcode nucleotide position and at least one more nucleotide mismatch in the P2-barcode sequence, and (4) adapter sequences with a reading error (T) in first nucleotide position. To exclude reads that most likely belonged to another P2-barcode than TGCAT in the "assigned to TNNNN" group, I de-multiplexed further in Stacks according to P1-barcodes allowing no nucleotide mismatches. Because, as explained above, reads not recorded with P2-barcodes more similar to the other P2-barcodes (CTAGT, GATCT, and AGCTT) also contained nucleotide mismatches in the P1-barcode. Finally, adapter sequences were filtered out (see Material and Method). Same procedure was conducted for the second lane.

De-multiplexing using both *Skewer* and *Stacks* made it possible to rescue reads containing the less successful barcode (Table D.1) as shown in Table D.2.

Table D.2. Differences in total retained reads (in percentage) of P2-barcodes in each lane when de-multiplexing using only *Stacks* v. 3.5 (Catchen *et al*., 2013) and using both *Skewer* v. 0.2.1 (Jiang *et al*., 2014) and *Stacks* combined. Excepted percentage of total retained read count for each barcode is ~25.

Percentage of total retained read count after de-multiplexing using:	**Lane**	**P2- barcode**			
		CTAGT	**GATCT**	**ACGTT**	**TGCAT**
Stacks	1	35.10	33.29	30.60	1.01
	2	28.96	32.85	37.44	0.75
Skewer* and *Stacks	1	22.39	24.46	28.62	24.53
	2	29.23	26.76	25.01	19.00

In order to detect any affect of the de-multiplexing method described above I also created a reduced data set were samples containing the least successful P2-barcode TGCAT were excluded. In the reduced data set, raw reads were de-multiplexed based on P1- and P2-barcodes not allowing any nucleotide mismatches in barcode sequences using only *process_radtags* program in *Stacks*. Preliminary analyses of the separate data sets did not display any apparent qualitative differences (data not shown). Since all my collected material were of interest, the data set de-multiplexed using both *Stacks* and *Skewer* were used in all analyses (see Results).

References

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, 22(11): 3124-3140 DOI: 10.1111/mec.12354

Jiang, H., Lei, R., Ding, S.W. & Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BioMed Central Bioinformatics*, 15(1): 182. DOI: 10.1186/1471-2105-15-182. Available for download at: https://sourceforge.net/projects/skewer/

Appendix E: Principal component analysis

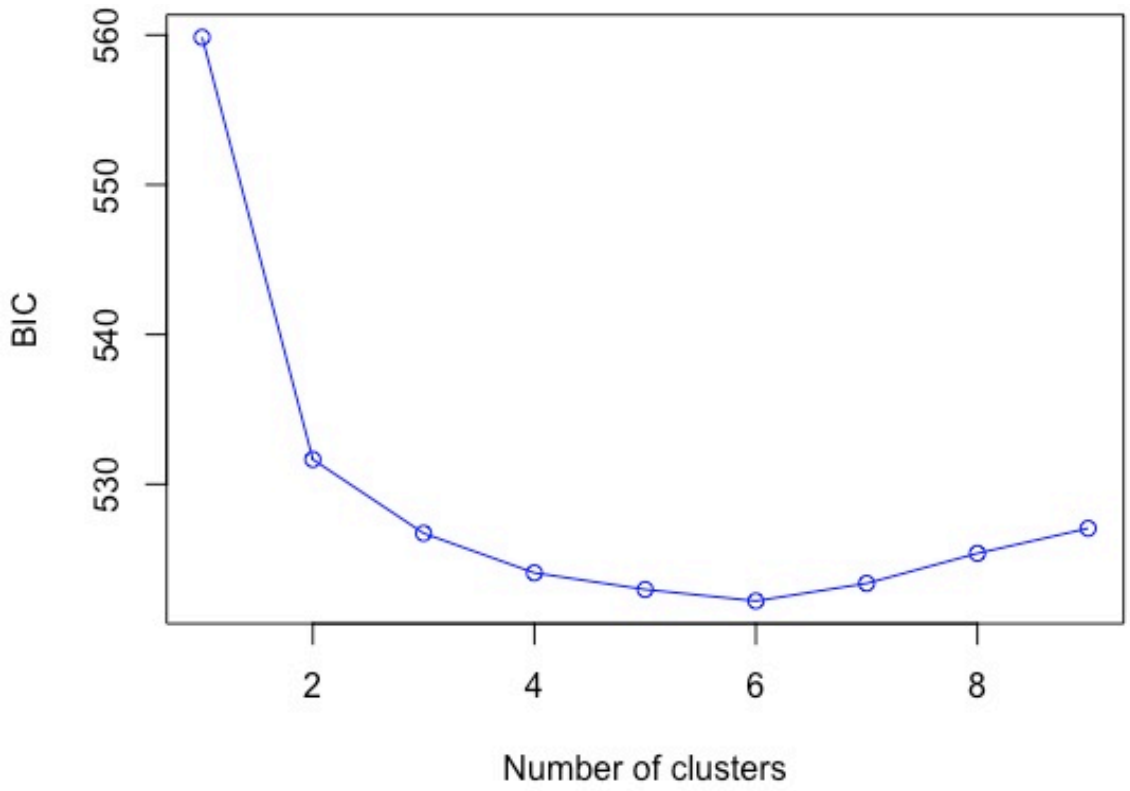

Figure E.1. Bayesian information criterion (BIC) scores as a function of number of clusters (*k*) for successive *k*-means clustering on all samples.

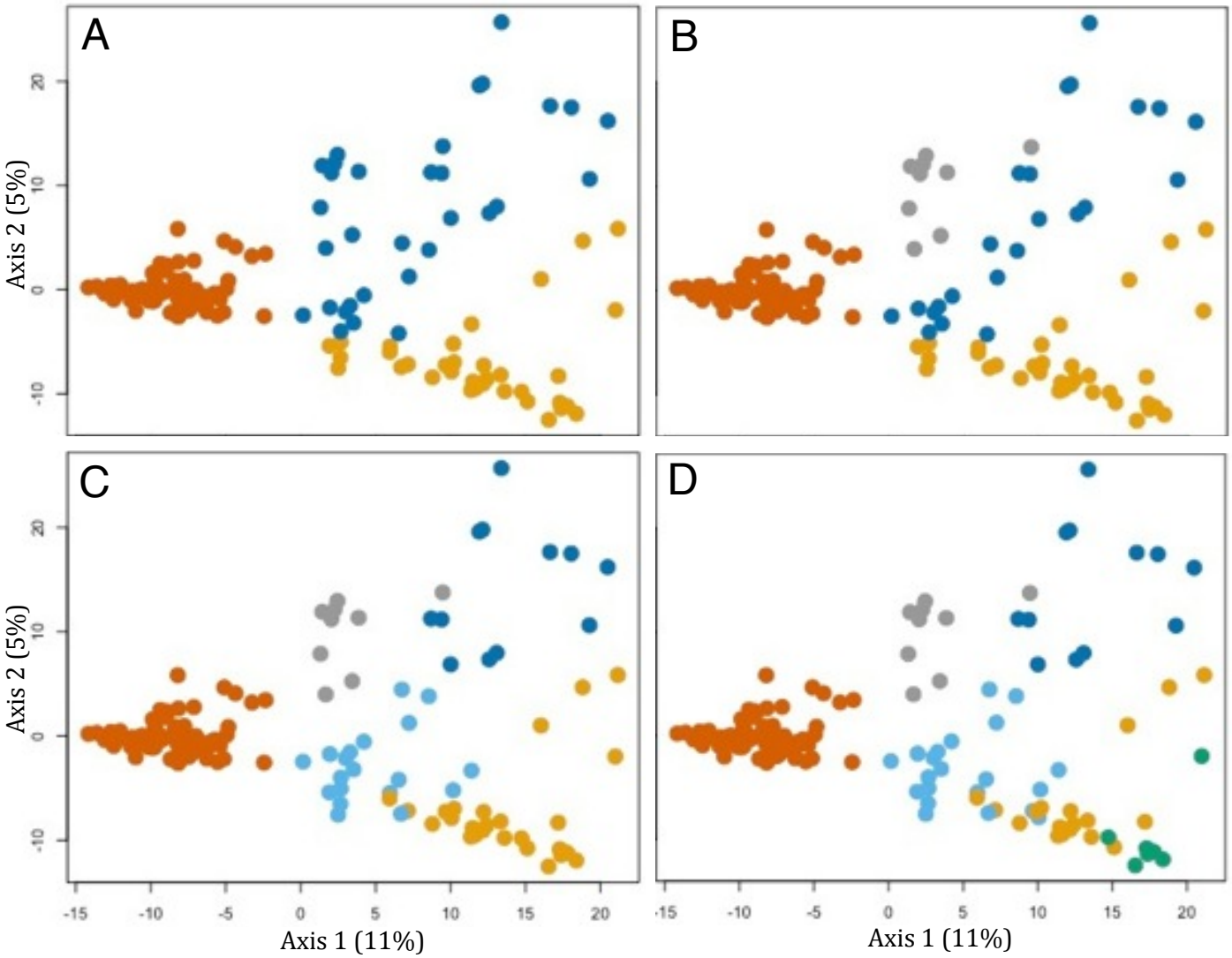

Figure E.2. Principal components analysis (PCA) and *k*-means cluster analysis of all 143 sampled individuals of *C. jemtlandica* and *C. lepidocarpa*. Individuals provisionally assigned to each of the *k* genetic clusters are displayed with different colours at (A) *k*=3 (red=*C. jemtlandica*, blue=*C. lepidocarpa* #1, orange=*C. lepidocarpa* #2), (B) *k*=4 (red=*C. jemtlandica*, blue=*C. lepidocarpa* #1, orange=*C. lepidocarpa* #2, grey=*C. jemtlandica x C. lepidocarpa*), (C) *k*=5 (red=*C. jemtlandica,* blue=*C. lepidocarpa* #1, orange=*C. lepidocarpa* #2, grey=*C. jemtlandica* x *C. lepidocarpa* #1, light blue=*C. jemtlandica* x *C. lepidocarpa* #2), and (D) *k*=6 (red=*C. jemtlandica*, blue=*C. lepidocarpa* #1, orange=*C. lepidocarpa* #2, green=*C. lepidocarpa* #3, grey=*C. jemtlandica* x *C. lepidocarpa* #1, light blue=*C. jemtlandica* x *C. lepidocarpa* #2).

Appendix F: Structure

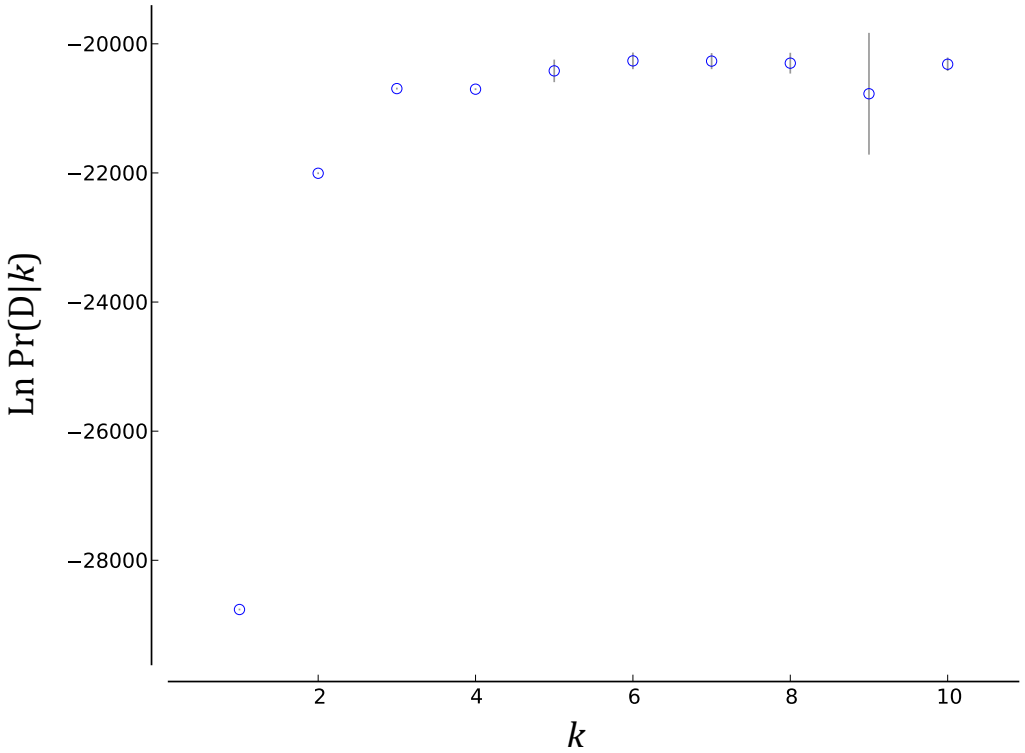

Figure F.2. Mean of estimated likelihood of data given k clusters +- standard deviation (SD), visualised by *Structure Harvester* v0.6.94 (Earl & vonHoldt, 2012).

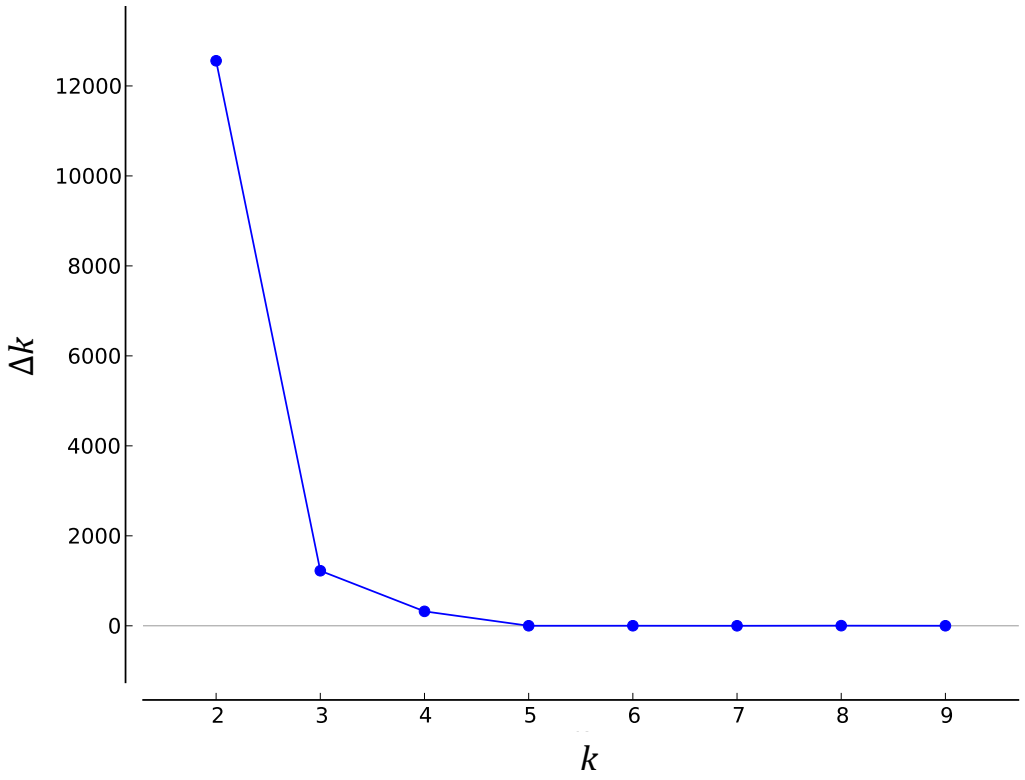

Figure F.3. Delta *k* for successive number of *k* estimated by *Structure Harvester* v0.6.94 (Earl & vonHoldt, 2012). Delta *k* is the absolute value of the second order rate of change of the likelihood function (mean) divided by the standard deviation (SD) of the likelihood of the model.

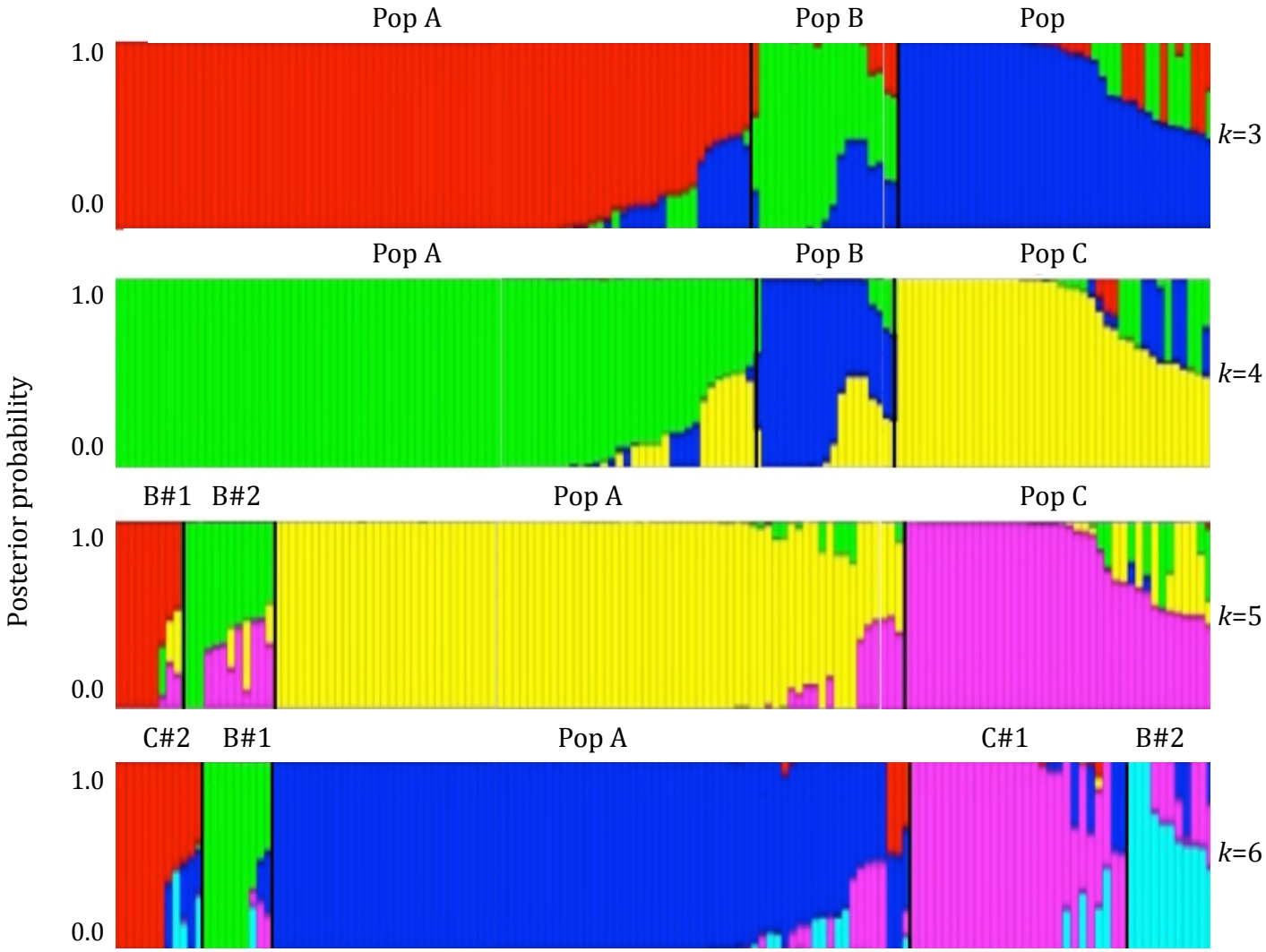

Figure F.3. Population structure estimated by *Structure* v. 2.3.4 (Pritchard *et al.*, 2000) of all 143 sampled individuals of *C. jemtlandica* and *C. lepidocarpa* based on 493 SNPs. Bar heights represents proportion of ancestry for each individual. Each cluster is separated by vertical lines and is represented in different colour. Length of each coloured segments represent the estimated proportion of individuals belonging to each cluster. Pop A consists of *C. jemtlandica*, and both Pop B (#1 and #2) and Pop C (#1 and #2) consist of *C. lepidocarpa*.

References

Earl, D.A. & vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*(2): 359-361. DOI: 10.1007/s12686-011-9548-7

Pritchard, J.K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *GENETICS*, 155(2): 945-959. Available for download at: http://pritchardlab.stanford.edu/structure.html