Hodel Richard G.J. (Orcid ID: 0000-0002-2896-4907)

Hybrid enrichment of adaptive variation revealed by genotype-environment associations in montane sedges

Richard G. J. Hodel[1,2,4]

Rob Massatti[3]

L. Lacey Knowles[1]

Affiliations

[1]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA

[2]Department of Botany, National Museum of Natural History, MRC 166, Smithsonian Institution, Washington, DC, 20013, USA

[3]U.S. Geological Survey, Southwest Biological Science Center, Flagstaff, AZ, 86001, USA

[4]Author for correspondence: richiehodel@gmail.com

Author Manuscript

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.16502

This article is protected by copyright. All rights reserved.

Author Manuscript

Abstract

The role of hybridization in diversification is complex and may result in many possible outcomes. Not only can hybridization produce new lineages, but those lineages may contain unique combinations of adaptive genetic variation derived from parental taxa that allow hybrid-origin lineages to occupy unique environmental space relative to one (or both) parents. We document such a case of hybridization between two sedge species, *Carex nova* and *Carex nelsonii* (Cyperaceae), that occupy partially overlapping environmental space in the southern Rocky Mountains, USA. In the region hypothesized to be the origin of the hybrid lineage, one parental taxon (*C. nelsonii*) is at the edge of its environmental tolerance. Hybrid-origin individuals display mixed ancestry between the parental taxa – of nearly 7,000 unlinked loci sampled, almost 30% showed evidence of excess ancestry from one parental lineage – approximately half displayed a genomic background skewed towards one parent, and half skewed towards the other. To test whether excess ancestry loci may have conferred an adaptive advantage to the hybrid-origin lineage, we conducted genotype-environment association analyses on different combinations of loci – with and without excess ancestry – and with multiple contrasts between the hybrids and parental taxa. Loci with skewed ancestry showed significant environmental associations distinguishing the hybrid lineage from one parent (*C. nelsonii*), whereas loci with relatively equal representation of parental ancestries showed no such environmental associations. Moreover, the overwhelming majority of candidate adaptive loci with respect to environmental gradients also had excess ancestry from a parental lineage, implying these loci have facilitated the persistence of the hybrid lineage in an environment unsuitable to at least one parent.

Keywords: Candidate Adaptive Loci, Genomic Cline, Genotype-Environment Association, Hybridization, RAD-Seq, Redundancy Analysis

Author Manuscript

Introduction

Physical barriers to genetic exchange often arise between interbreeding populations, which may result in reproductive isolation that leads to increased genetic structure between populations; ultimately, such processes may culminate in speciation (Rieseberg & Willis 2007). However, the instability of biological communities due to changing climates often results in secondary contact of previously isolated lineages. For example, during the Pleistocene, environmental change occurred repeatedly between glacial and interglacial extremes and on a time scale short enough to prevent reproductive isolation by genetic drift alone (Knowles & Richards 2005, Kou et al. 2020). When secondary contact occurs, there are a number of potential outcomes for diverging lineages ranging from the breakdown to the strengthening of reproductive barriers (Dion-Côté & Barbash 2017). The genomic age has confirmed that hybridization between diverging lineages is common (Rieseberg 1997, Rieseberg & Willis 2007); an estimated 25% of plant species hybridize with at least one other species (Mallet 2005). While much of this genomic evidence may indicate transient hybridization events with the potential to impact the evolutionary trajectories of parental species, most events do not result in stable hybrid lineages. However, instances in which hybrids persist illustrate how this process can be a key mechanism in diversification (Seehausen 2004) and for generating evolutionary novelty (Abbott et al. 2016). Furthermore, studying hybridization may provide valuable insights into the speciation process, including the identification of genomic regions that likely contribute to reproductive isolation (Taylor & Larson 2019, Wang et al. 2021).

A key insight resulting from genomic analyses of hybridization is that the genomes of hybridizing taxa are not monolithic – regions of the genome may be differentially exchanged between both hybrids and their parental species (Gompert & Buerkle 2011). Although the role of

Author Manuscript

hybridization in species' histories varies, genomic exchange facilitated by hybridization may contribute to adaptative evolution (Tywford & Ennos 2012) and have functional consequences (Abbott et al. 2016). When genetic variation introduced by hybridization events includes adaptive variation, this source of variation may in fact be more effective than new mutation (Grant & Grant 1994, Kim & Rieseberg 1999, Arnold & Martin 2009, Whitney et al. 2010, Kunte et al. 2011, Abbott et al. 2013). The genomic age has advanced our understanding of species boundaries and adaptive introgression, although it remains challenging to link hypothesized adaptive introgression to phenotypic variation and fitness (Suarez-Gonzalez et al. 2018). Nevertheless, powerful insights can be identified in the genomes of extant taxa, which show the effects of hybridization due to histories of ancient hybridization, adaptive introgression, and/or hybrid speciation (Taylor & Larson 2019). For example, genomic data illuminating rapid radiations in African cichlids reveal that hybridization can provide the genetic novelty that sparks adaptive radiations (Meier et al. 2017). Furthermore, hybridization has enabled fungi to colonize novel environments unsuitable for the parental taxa (Leducq et al. 2016). In some cases, hybrid lineages may readily occupy areas vacated by distributional changes as one or both parents track their niches, often as a response to climatic changes (Kadereit 2015). In effect, increasingly abundant genomic data confirm the role of hybridization in generating novel gene combinations upon which selection may act (Goulet et al. 2017).

Here, we investigate the consequences of a hybridization event that generated a distinct evolutionary lineage that has spread to occupy a substantial range across the southern Rocky Mountains of western North America. Our focus is on two species in *Carex* section *Racemosae* G.Don, *Carex nova* L.H.Bailey and *Carex nelsonii* Mack., and an undescribed, putatively hybrid-origin evolutionary lineage identified during the course of previous research (Massatti &

Author Manuscript

Knowles 2016). Putative hybrids, while first assumed to represent morphological variation within *C. nelsonii*, are now hypothesized to represent the southernmost extent of the *C. nelsonii* range where this parent species is largely absent; as such, these hybrids may occupy environmental space that is different from both putative parental species. Section *Racemosae* contains at least ten endemic species across this region, many of which are restricted to montane ecosystems or adjacent, topographically complex landscapes. This diversity likely arose since the start of the Pleistocene (Massatti & Knowles 2016), during which glacial cycling provided abundant opportunities for complex demographic processes among species' populations (Massatti & Knowles 2014, 2016, Hodel et al. 2021). While diversification within section *Racemosae* is consistent with models of speciation and differentiation in montane systems, the mechanistic basis of diversification remains unknown; understanding such processes may illuminate the origins of diversity in the speciose *Carex* genus, which occurs in many habitats and regions throughout the world.

To test the hypothesis that a hybridization event between *C. nelsonii* and *C. nova* produced an evolutionarily distinct lineage that persists in a unique environment compared to the parents, we first establish the phylogenetic history of the parental taxa and the hybrid lineage. Next, we identify the subsets of the genome (i.e., specific loci) that display either equal ancestry from both of the parental species or excess ancestry from one. In turn, we use the loci displaying excess ancestry to test the hypothesis that hybridization facilitated an adaptive shift that allowed the hybrid lineage to occupy a unique environmental distribution compared to the parental species. Specifically, we use loci with excess ancestry against the background of the parental genomes in a gene-by-environment analytical framework to test for significant relationships between specific loci and a set of independent environmental gradients. If excess ancestry as a

result of hybridization allowed the occupation of unique environmental space by hybrids, we predict that unique combinations of putative adaptive loci representative of both parental species will be identified in the hybrid taxon. Our analytical framework clearly supports an independent evolutionary lineage that has spread to occupy a substantial range throughout the southern Rocky Mountains. Moreover, this taxon persists in environments where the parental species that it most closely resembles (morphologically and ecologically; *C. nelsonii*) does not occur. These patterns suggest that the hybrid taxon may leverage adaptive genetic variation from both parental species that allows it to occupy unique environmental space and supports hybridization as a potential means of diversification in the hyperdiverse *Carex* genus.

Materials & Methods

Field sampling and genomic characterization

Samples were field-collected from eight *C. nelsonii* localities, 15 *C. nova* localities, and five putative hybrid sites (Fig. 1). Between two and ten individuals (mean = 8.5) separated by at least five meters were collected at each sampling locality (Supplemental Table S1, Supplemental Table S2). Putative hybrid-origin individuals collected and keyed out to *C. nelsonii* were only identified as hybrid-origin during the course of population genetic analyses. A total of 237 individuals – 54 *C. nelsonii*, 136 *C. nova*, and 47 putative hybrids – were used for genomic analyses (Fig. 1, Supplemental Table S1, Supplemental Table S2).

We performed double-digest RAD-Seq (Peterson et al. 2012) to obtain anonymous genomic loci. We used two restriction enzymes, *Eco*RI and *Mse*I, to digest genomic DNA, and ligated Illumina adaptor sequences and unique 10 bp barcodes to restriction sites. The ligation products for each library were pooled and PCR-amplified for 12 cycles, and 400-500 bp

Author Manuscript

fragments were size selected using a Pippin Prep (Sage Science, Beverly, MA). All libraries were sequenced on single-end Illumina HiSeq 2500 runs at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, Ontario). Raw sequencing reads were processed using ipyrad v9.42 (Eaton 2014, Eaton & Overcast 2020) in a single assembly that combined both species and the putative hybrids (Supplemental Table S2). The combined assembly approach facilitated the tracking of allele frequencies across all individuals and populations in downstream analyses. The ipyrad assembly was filtered using vcftools (Danacek et al. 2011) to allow 40% missing data per site (--max-missing = 0.6) and require a minor allele count of two or greater (--mac = 2). The resulting data matrix consisted of 6,946 unlinked SNPs (i.e., one randomly selected SNP per RAD-Seq locus). Because we were interested in the effects of putative adaptive loci associated with hybridization and not loci associated with selection within one of the parental species, we used PCAdapt (Luu et al. 2017) to remove F_{ST} outliers identified within either *C. nelsonii* or *C. nova*; in total, we removed 206 loci, which left 6,740 unlinked SNPs that were used in most subsequent analyses, except where noted. To verify that removing F_{ST} outliers did not bias our downstream results, we ran several analyses both with and without F_{ST} outliers, including analyses of excess ancestry, all genetic diversity and differentiation statistics, and redundancy analyses, as noted below.

A separate ipyrad assembly was performed to add an outgroup (*C. chalciolepis*) suitable for rooting the phylogeny of *C. nelsonii* + *C. nova* + hybrids and for use in downstream analyses (see hybridization analyses below). This assembly included the 237 individuals referenced above plus 152 individuals of *C. chalciolepis* and leveraged the same filtering strategy as the first assembly. The resulting SNP data matrix consisted of 6,069 unlinked SNPs, which were used in the SVDQuartets and TREEMIX analyses. The authors of HyDe recommend including invariant

Author Manuscript

sites in the genetic dataset, so we used all 69,373 sites from the assembly with the outgroup in the HyDe analysis (see below).

Environmental characterization

We characterized the environmental niche of each species – *C. nelsonii*, *C. nova*, and putative hybrids – by performing a Principal Components Analysis (PCA) on environmental variables across the study region. The environmental PCA was run using the 'dudi.pca' function with centering and scaling in the R package *ade4* (Dray & Dufour 2007). Specifically, we included 17 of 19 'bioclim' layers (Hijmans et al. 2005; http://worldclim.org) with a spatial resolution of 30-arc seconds; BIO8 and BIO9 were excluded because of bimodal distributions in the study region. The R packages *dismo* (Hijmans et al. 2017) and *raster* (Hijmans 2019) were used to extract environmental data from GPS points representing individuals at each of the sampling locations (Supplemental Table S3). The first three principal components, which explain 89.9% of the variation, were used to define the environmental space occupied by each taxon.

Genetic relationships between individuals and populations

The 'populations' program in STACKS v2.41 (Catchen et al. 2013) was used to estimate per-locus F_{ST}, pairwise F_{ST}, nucleotide diversity (π), observed heterozygosity (H_O), and % polymorphic loci. Herein, we used these statistics to compare taxa and sets of loci, but we caution against comparing their absolute magnitudes to values in other studies because they are sensitive to filtering parameters and the markers used. We estimated pairwise F_{ST} between all sampling localities within taxa as well as between *C. nelsonii*, *C. nova*, and putative hybrids. After discovering differences in occupied environmental space and in summary statistics

Author Manuscript

between hybrids from different populations and/or geographic regions (see Fig. 1), we reran analyses considering subsets of hybrids as separate groups. The above population genetic statistics were also calculated separately for sets of loci defined by excess ancestry due to hybridization, or a lack thereof (see 'Identifying loci with excess ancestry' below).

Furthermore, we implemented the Bayesian clustering program STRUCTURE (Pritchard et al. 2000) to infer population assignment and identify signals of hybridization. Specifically, we used an admixture model with correlated allele frequencies for $K = 1$ through $K = 6$ and ran 1,000,000 MCMC generations with a burnin of 200,000 generations. We also investigated genetic similarity among individuals using a PCA implemented in the R packages *SNPRelate* (Zheng et al. 2012) and *vegan* (Oksanen et al. 2017). SVDQuartets (Chifman & Kubatko 2014) was used to infer phylogenetic relationships between all *C. nelsonii*, *C. nova*, and hybrid individuals, as well as a rooted phylogeny using *C. nelsonii*, *C. nova*, hybrids, and an outgroup, *C. chalciolepis*. For consistency with previous analyses, STRUCTURE, PCAs, and unrooted SVDQuartets analyses used the SNP dataset with the 206 loci considered F_{ST} outliers in one parent removed.

Hybridization among taxa was assessed using HyDe (Hybrid Detector; Blischak et al. 2018) and TREEMIX (Pickrell & Pritchard 2012). HyDe uses phylogenetic invariants arising under a coalescent model that includes hybridization to detect and assign probability of hybridization of the three ingroup taxa (i.e., *C. nelsonii*, *C. nova*, and the putative hybrids). The parameter γ is the probability of gene trees having a hybrid population sister to parent X arising under the parental population trees while $1\text{-}\gamma$ is the probability that the hybrid population is sister to parent Y. We ran HyDe analyses exhaustively on all possible combinations of taxa, with *C. chalciolepis* set as the outgroup. To further investigate signals of hybridization, HyDe analyses

Author Manuscript

Author Manuscript

were complemented with TREEMIX, an approach that incorporates reticulation in a phylogeny. TREEMIX uses allele frequencies and a Gaussian approximation of genetic drift to maximize the likelihood of a population covariance matrix while allowing both population splits and admixture via a migration parameter. The change in the likelihood score as the number of migration events (m) ranged from zero to five was assessed in line with ipyrad analysis tools (https://ipyrad.readthedocs.io/en/latest/API-analysis/cookbook-treemix.html; Eaton & Overcast 2020) and the optimal m-value was inferred using OptM (Fitak 2021). We investigated a range of cutoffs (5% to 50%) for the minimum percent of samples required to be present in each species and used 10 random starting seeds for each run for a particular m-value. If one or more migration events are inferred, a reticulate topology explains the data better than a purely bifurcating one and is therefore consistent with a hybridization history. Based on STRUCTURE and SVDQuartets analyses, we used several strategies for grouping hybrid populations in the TREEMIX analysis.

Identifying loci with excess ancestry

A genomic cline approach, implemented in the software bgc (Gompert & Buerkle 2012), was used to identify loci with alleles displaying excessive amounts of ancestry from one parent relative to the average genomic background. The bgc analysis defined parental populations as those with a $K = 3$ STRUCTURE Q-score greater than 0.9, as recommended by the developers, resulting in the exclusion of three *C. nova* populations (Lizard Head, Pike's Peak, Red Lakes; see Fig. 2). In our use of this model, a positive genomic cline parameter α describes an increase in the probability that a hybrid has an excess of *C. nova* ancestry relative to the expectation based on the hybrid index (h), whereas a negative α represents a deficit of *C. nova* ancestry. The

genomic cline parameter β specifies an increase (positive β) or decrease (negative β) in the rate of transition from low to high probability of *C. nova* ancestry as a function of hybrid index (*h*). After convergence among 10 independent MCMC runs, each having 25,000 burnin generations followed by 50,000 generations, we assessed the α and β parameter of each locus. We identified loci displaying significant departures from null expectations (i.e., equal ancestry of alleles from each parent) by assessing whether 95% confidence intervals around estimates of α or β overlapped with zero. Such loci with excess ancestry from either parent as defined by exceptional values of α are termed 'outlier loci'. For brevity, outlier loci with excess *C. nova* ancestry are referred to as '*C. nova*-outliers' and outlier loci with excess *C. nelsonii* ancestry are referred to as '*C. nelsonii*-outliers'. Loci with no detectable excess ancestry are hereafter referred to as 'inlier loci'.

Environmental associations of loci

We used redundancy analysis (RDA), a multivariate ordination method that can simultaneously compare multiple loci and environmental variables, to determine the effect of environmental gradients on genetic variation in different sets of loci (Forester et al. 2018). RDA is capable of identifying signatures of weak and/or multi-locus selection and adaptation (Rellstab et al. 2015, Forester et al. 2018), and this approach has a documented low rate of false positives and high rates of true positives across a range of demographic histories (Capblanq et al. 2018). Here, we used RDA to test if loci displaying excess ancestry from one parent disproportionately influence the climatic niche of the hybrid taxon. In our use of RDA, genetic and environmental data are analyzed using multivariate linear regression resulting in a matrix of fitted values (Legendre & Legendre 2012). Next, input predictors (i.e., environmental variables) are

Author Manuscript

Author Manuscript

constrained to test if individual input axes explain more of the variation in a response variable (in this case genotype) compared to a PCA of the input axes (e.g., Forester et al. 2018, Massatti & Knowles 2020, Vanhove et al. 2021). To avoid pseudoreplication, we used genetic data based on populations instead of individuals; population-level allele frequencies were calculated using the 'makefreq' function in the R package *adegenet* (Jombart & Ahmed 2011). The environmental input variables used were the PC axes from an environmental PCA constructed in a similar fashion as described above in 'Environmental characterization,' except this time a single GPS coordinate was used per population to ensure one-to-one correspondence between genetic data (i.e., population-level allele frequencies) and environmental data in the RDA.

If hybridization allows the expansion of the occupied environmental space by the hybrid compared to one or both parents, we expect significant RDA results when investigating outlier loci between the hybrid and a parent. At the same time, we would not expect a significant environmental association when using inlier loci from the hybrid plus a parent. Accordingly, genotype-environment associations were tested using RDA for each set of outlier loci (i.e., *C. nelsonii*-outliers, *C. nova*-outliers) first considering the hybrid plus each parent. To ensure that any detected significant results using outlier loci in the hybrids are meaningful, we also used RDA to analyze all other possible combinations of loci sets (i.e., inlier loci, all loci) and taxa (i.e., the hybrid separately, each parent separately, and between the parents) as controls. Each overall RDA model and the individual RDA axes were tested for significance using permutation tests. Significant RDA axes are considered evidence that the environmental input variables explain a substantial amount of genetic variation.

For any RDAs that were significant, we identified putative adaptive loci across environmental gradients by following the methods of Forester et al. (2018). Candidate adaptive

loci may load significantly on any of the three environmental PC axes, and we assessed the loadings of candidate adaptive loci to infer environmental conditions that may be associated with adaptation. We used the RDAs including inlier loci as the background against which to test for candidate adaptive loci due to hybridization. In order to test the hypothesis that excess ancestry loci within the hybrid taxon allow it to occupy unique environmental space compared to the parental taxon that it most closely resembles (*C. nelsonii*), we visualized sampling sites of *C. nelsonii* and the hybrid taxon in RDA space using only *C. nelsonii*-outliers and *C. nova*-outliers. Candidate adaptive loci were defined as those outside of a two standard deviation cutoff relative to the background of inlier loci as determined by the RDA of *C. nelsonii* + the hybrid using only inlier loci. To clarify the distinction between outlier loci with respect to ancestry (i.e., α outliers detected using bgc) and loci that are statistical outliers in RDAs, we hereafter refer to RDA-outliers as 'candidate adaptive loci.'

Results

Genetic characterization of lineages and hybrid diagnosis

The HyDe analyses confirmed the hypothesized hybrid origin of the five putative hybrid populations and rejected hybrid status for all populations identified as *C. nova* and *C. nelsonii* (Table 1). Hybrid indices were skewed towards *C. nova* ancestry (average hybrid index = 0.6, where 0 = pure *C. nelsonii* and 1 = pure *C. nova*; Supplemental Table S4). STRUCTURE results using a $K = 2$ model support the hypothesis that the total genetic diversity across taxa represents the gene pools of two species; hybrid individuals largely clustered with *C. nelsonii* (Fig. 2). The $K = 3$ STRUCTURE solution clearly separates the five hybrid populations, with hybrid Q-scores corroborating the HyDe results (Fig. 2). In the genetic PCA, variation on PC1 corresponds to the

Author Manuscript

Author Manuscript

K = 2 STRUCTURE plot, with *C. nova* distinguished from *C. nelsonii* + hybrids (Supplemental Fig. S1). Notably, individuals from the Mt. Oso, Red Lakes, and Lizard Head hybrid populations are more diffuse in PC space than individuals from the Medicine Bows and Flat Tops hybrid populations (Supplemental Fig. S1). PC2 differentiates between the hybrid and *C. nova* + *C. nelsonii*, whereas PC3 shows the hybrid individuals intermediate between *C. nova* and *C. nelsonii*, and in some cases overlapping with the genetic space occupied by *C. nova* (Supplemental Fig. S1).

The SVDQuartets species tree recovered a strongly supported clade (bootstrap score = 100%) containing all *C. nelsonii* individuals as sister to a strongly supported clade containing all hybrids from Medicine Bows and Flat Tops, and most individuals from Lizard Head and Red Lakes and one Mt. Oso individual (Fig. 3). All hybrid individuals from Medicine Bows and Flat Tops were in a strongly supported monophyletic group (bootstrap score = 94%) nested within the hybrid clade (Fig. 3); hereafter this lineage is referred to as the 'northern hybrids' to promote discussion of evolutionary hypotheses. Nine out of ten Lizard Head hybrid individuals plus eight out of ten Red Lakes hybrids formed a clade with high bootstrap support (93%) sister to the northern hybrids; for brevity, this lineage is hereafter referred to as 'core southern hybrids' (Fig. 3). Southern hybrid individuals outside of the core southern hybrid clade—including all Mt. Oso hybrids, as well as one Lizard Head individual and two Red Lakes individuals—are described as 'Mt. Oso-like hybrids' hereafter and may fall outside the hybrid clade due to more recent backcrossing events with *C. nova* (e.g., Fig. 2); regardless, these patterns indicate a single origin for the hybrid-origin taxon.

The clade comprised of *C. nelsonii* plus most hybrids is sister to a clade containing all *C. nova* plus two hybrid individuals from Mt. Oso (Fig. 3). These results, including low support at

nodes and the various placement of hybrid individuals, are readily explained by the various patterns of ancestry coefficients in some hybrid and southern *C. nova* individuals displayed in Fig. 2. Phylogenetic relationships inferred using the *C. chalciolepis*-rooted dataset support the SVDQuartets inference (Supplemental Fig. S2), although there are some minor differences, which is common at such shallow-scale divergences (e.g., Wang et al. 2017). The TREEMIX phylogeny inferred the same branching topology as the SVDQuartets tree (Fig. 3, Supplemental Figs. S2 & S3) and consistently inferred *C. nelsonii* as sister to the northern and all southern hybrids with one migration event from *C. nova* to the southern hybrids (Fig. 3). Excluding the Mt. Oso-like hybrids or treating them as a distinct entity did not have an impact on the branching topology (Supplemental Fig. S3). When the northern, core southern, and Mt. Oso-like hybrids were considered as separate groups, the Mt. Oso-like hybrids were sister to *C. nova*, but *C. nelsonii* remained sister to the northern plus core southern hybrids, with one migration event from *C. nova* to the core southern hybrids (Supplemental Fig. S3).

Nucleotide diversity (π) across all 6,740 loci ranged from 0.021 (*C. nova*-Uintahs) to 0.186 (hybrid-Mt. Oso). At the taxon level, genetic diversity of *C. nova* ($\pi = 0.102$) is nearly twice that of *C. nelsonii* ($\pi = 0.055$), while the highest diversity is represented by all hybrids ($\pi = 0.161$) (Supplemental Table S5). The number of polymorphic sites in *C. nova* (4,025) is nearly four times greater than *C. nelsonii* (1,119), while the hybrids have an intermediate value (3,507) closer to that of *C. nova* than *C. nelsonii*. F_{ST} was lowest between the *C. nova* and hybrid populations (0.167), higher between *C. nova* and *C. nelsonii* (0.270), and highest between *C. nelsonii* and all hybrids (0.311) (Supplemental Table S6). Genetic diversity and differentiation statistics for each set of loci, and using different sets of hybrid individuals, are shown in Supplemental Tables S5 and S6.

Author Manuscript

Author Manuscript

Environmental characterization

The environmental space occupied by *C. nelsonii* is substantially smaller than that of *C. nova* (Fig. 4, Supplemental Fig. S4). Certain bioclimatic variables are strongly (i.e., $r^2 > 0.45$) and significantly ($p < 0.001$) associated with the PC axes (Supplemental Table S7, Fig. 4). Despite substantial overlap on PC1, there is a tendency for *C. nelsonii* to occupy more positive values on this axis, which is associated with lower temperatures and greater precipitation, while *C. nova* extended further into negative PC1 values indicative of higher temperatures and less precipitation (Fig. 4). *Carex nelsonii*'s distribution across PC2 is largely contained within the variation represented by *C. nova*, which also occupies a substantial and unique distribution across negative values of PC2 (corresponding to warmer habitats with higher precipitation seasonality). The hybrid lineage occupies a similarly sized, but largely distinct, environmental space compared to *C. nelsonii* (Fig. 4), overlapping with both parental species in a small region with intermediate values of both PC1 and PC2. This region corresponds to the far northern portion of *C. nova*'s range in northern Colorado and southern Wyoming (Fig. 1). In other words, the northern hybrids, which are the most differentiated from both parental species (Supplemental Table S6), overlap with the environmental space of each parental taxon; the Lizard Head, Red Lakes, and Mt. Oso hybrids from the southern part of the study range overlap only with *C. nova* in environmental space not occupied by *C. nelsonii* (Fig. 4).

Loci with signatures of excess ancestry from a parental lineage

Genomic cline analyses identified a substantial number of loci (29.6%) with significantly positive and negative α-values indicating excess ancestry from either *C. nova* or *C. nelsonii*,

Author Manuscript

respectively; no extreme β-values were detected. Out of 6,740 loci, 1,027 had excess *C. nova* ancestry and 971 had excess ancestry *C. nelsonii,* leaving 4,742 loci with showing equal ancestry from both parents ('inlier loci'). Outlier loci with excess *C. nova* ancestry are hereafter referred to as '*C. nova*-outliers' and outlier loci with excess *C. nelsonii* ancestry are referred to as '*C. nelsonii*-outliers', while loci with no detectable excess ancestry are hereafter referred to as 'inlier loci'. Using all loci, genetic differentiation between the parental species, as measured by pairwise F_{ST}, is positively correlated with the magnitude of both α (r^2 = 0.030, p < 2.2e-16) and β (r^2 = 0.080, p < 2.2e-16) estimates on a per locus basis. These regression coefficients are small but similar to results from previous studies (e.g., Parchman et al. 2013). When the 206 loci identified as F_{ST} outliers in either parent were included, the distribution of loci was qualitatively similar: out of 6,946 loci, 1,061 were *C. nova*-outliers, 998 were *C. nelsonii*-outliers, and 4,887 were classified as inlier loci.

Associations of loci with environmental conditions

The allelic makeup of the hybrid-lineage sampling sites versus *C. nelsonii* sampling sites clearly distinguishes populations of *C. nelsonii* from hybrids (Fig. 5A), with both sets of outlier loci showing significant associations with environmental conditions. The overall RDA model and the RDA1 canonical axis were significant for both (Table 2). Simultaneously, no environmental associations of inlier loci distinguished *C. nelsonii* from the hybrid lineage. Hybrid genotypes at outlier loci are generally associated with higher temperatures and precipitation than loci in *C. nelsonii*; the hybrid lineages trend towards positive PC1 (positively correlated with many precipitation variables: BIO12, BIO13, BIO16, BIO17, BIO18, BIO19; hereafter referred to as 'precipitation axis') and PC2 (positively correlated only with temperature

variables: BIO5, BIO10; hereafter referred to as 'temperature axis') space (Fig. 5A). There was also an association between the alleles of the two parental species and environmental gradients when using all 6,740 loci, and either outlier set of loci. An RDA using inlier loci was not significant, so the significance of the RDA using all loci may have been driven by outlier loci. However, when F_{ST} outliers within either parent were included (i.e., the 6,946 loci dataset), the RDA comparing both parents using inlier loci was significant (Supplemental Table S8), which may be attributed to the influence of loci that were F_{ST} outliers with respect to differentiation. No other comparisons were significant for any other combinations of taxa, and all RDAs using inlier loci were non-significant (Table 2).

We investigated SNPs from the significant RDAs using outlier loci with hybrids and *C. nelsonii* in more detail, revealing that candidate adaptive loci were overwhelmingly represented by the two sets of outlier loci as opposed to inlier loci (Fig. 5B). Against the background of inlier loci, we applied a two standard deviation cutoff, which is appropriate to identify loci experiencing different strengths of selection (one of the core strengths of RDA); 356 *C. nelsonii*-outlier loci and 251 *C. nova*-outlier loci were identified as candidate adaptive loci (607 out of 1,998 outlier loci; Fig. 5B, Table 3). In contrast, only 10 out of 4,742 inliers surpassed the two standard deviation cutoff. As such, >98% (607 out of 617) of candidate adaptive loci belong to the two sets of outlier, as opposed to inlier, loci.

Candidate adaptive loci loaded on both the temperature and precipitation axes (Fig. 5B, Table 3). Because results from environmental PCAs using multiple GPS coordinates per sampling location informed the description of environmental space (Fig. 4, Supplemental Fig. S4) whereas the identification of candidate adaptive loci in RDAs used one GPS point per sampling location (Fig. 5), we consider how candidate adaptive loci identified by RDAs

Author Manuscript

Author Manuscript

correspond to the overall environmental space of the three taxa. Candidate adaptive loci loading positively on the temperature axis (PC2 in the RDA; Fig. 5) correspond to the lower left quadrant of the environmental PCA (Fig. 4), while candidate adaptive loci loading negatively on the temperature axis correspond to the upper right quadrant of the environmental PCA. Meanwhile, candidate adaptive loci loading positively on the precipitation axis (PC1 in the RDA; Fig. 5) correspond to the lower right quadrant of the environmental PCA, while candidate adaptive loci loading negatively on the precipitation axis correspond to the upper left quadrant of the environmental PCA. Crucially, all candidate adaptive loci that load on the temperature or precipitation RDA axes – except the ones negatively correlated with the temperature axis – align with space in the environmental PCA (Fig. 4) that does not correspond to an environment occupied by *C. nelsonii*; in total, 499 of 607 candidate adaptive loci correspond to environmental space unsuitable for *C. nelsonii* (Fig. 4, Fig. 5B, Table 3).

Discussion

Our detailed examination of anonymous genomic loci within a hybrid and its two parents not only solidifies the role of hybridization in the formation of a new lineage of montane sedges, but it also shows that hybridization has facilitated an adaptive shift allowing hybrid individuals to occupy unique environmental conditions compared to the parent that it most closely resembles in terms of morphology and habitat preference. As such, our work directly addresses the role of hybridization in promoting diversification. Our results suggest that rather than simply representing a breakdown of reproductive isolation (and the potential for the loss of diversity through the merging of divergent species' lineages), hybridization studied herein represents a case in which alleles contributing to adaptation have been uniquely combined to promote the

Author Manuscript

occupation of a new climatic niche. Specifically, the hybrids are recognized genomically as an evolutionary lineage distinguishable from the parental species *C. nelsonii* and *C. nova*. Moreover, outlier loci with respect to ancestry show significant environmental associations distinguishing the hybrid lineage from *C. nelsonii*, whereas other loci do not. This suggests that hybridization has played a role in adaptive divergence. By mapping the environmental associations to axes of environmental variation, we also show the potential role of candidate adaptive loci in facilitating a potentially adaptive environmental shift as well as the nature of this environmental shift (e.g., contributing to a shift along a precipitation versus a temperature gradient) (Figs. 4 & 5, Table 3, Supplemental Table S7). Below, we discuss how this evidence contributes to our understanding of hybridization as a mechanism for diversification in montane sedges, and more generally how the dissection of the history of hybridization, and specifically the heterogeneous contribution of parental genomes to hybrid taxa, can be quantified analytically to identify candidate adaptive loci that have spread between species.

Candidate adaptive loci acquired through hybridization

The genomes of many species show evidence of ancient and/or recent hybridization (Taylor & Larson 2019). However, such hybridization is not always adaptive and could simply reflect the permeability of species boundaries (Twyford & Ennos 2012, Gompert et al. 2017). As such, additional evidence is required to evaluate whether there is an adaptive component to the gene flow (Suarez-Gonzalez et al. 2018). Moreover, adaptive gene flow via hybridization can simultaneously introduce novel alleles at many unlinked loci, potentially promoting rapid evolution by adaptation even when selected traits are polygenic (Abbott et al. 2013, Hedrick 2013, Arnold & Kunte 2017). A pool of alleles with potential adaptive value has been argued to

be key to the rapid diversification and/or shifts of lineages into a broad range of niches (e.g., in Darwin's finches, Lamichhaney et al. 2015; *Heliconius* butterflies, Kozak et al. 2015; cichlids, Albertson et al. 1999), and in many cases, adaptive alleles were often moved across species boundaries via hybridization and introgression. Heterozygosity of candidate genes has been implicated in promoting prezygotic reproductive isolation of a hybrid *Ostryopsis* species with its parents (Wang et al. 2021). For more recent hybridization events, studies demonstrate how candidate adaptive loci move between parental species and/or hybrids, potentially facilitating diversification, as in canids (Galaverni et al. 2017), *Populus* tree species (Chhatre et al. 2018), and *Lycaeides* butterflies (Gompert et al. 2010). Confirming adaptive introgression by explicitly linking putatively adaptive loci to fitness consequences is challenging (Suarez-Gonzalez et al. 2018). However, analyses of genomic clines offer an alternative approach to identifying introgressed alleles that were likely targets of selection (Gompert & Buerkle 2012).

In our study, there are multiple lines of evidence that suggest hybridization between two montane sedges involves potentially adaptive alleles at multiple loci. An analytical approach (i.e., the bgc method; Gompert & Buerkle 2012) identified outlier loci with an excess ancestral contribution of one of the two parental species (Supplemental Table S4). By itself, this is not sufficient to draw conclusions about the role of these loci, although they require explanation, and differential introgression due to selection is one possibility (among many, including constraints associated with genomic architecture, such as chromosomal re-arrangements; Rieseberg et al. 2003). However, with the complement of the genotype-environment method (RDA) in which we evaluate the potential adaptive role of outlier loci, we confirm that the outlier loci (i.e., loci with an excess ancestry of one of the two parental species) show specific environmental associations while inlier loci (i.e., loci whose ancestry traces to both parental species similarly) do not (Fig

Author Manuscript

5B). Taken together, the results make a compelling case that candidate adaptive loci are overwhelmingly associated with hybridization.

Other studies involving recent hybridization have also observed that candidate adaptive loci often have been disproportionately contributed by one of the two parental lineages. For example, the vast majority of loci with excess ancestry and putative adaptive effects on photoperiod and phenology came from one parent in poplar hybrids (see Chhatre et al. 2018); a similar pattern was found in pines (see Menon et al. 2021), where excess ancestry loci facilitated adaptation associated with freeze tolerance. However, in some species, and even when it is predicted that one parental species will contribute disproportionately to an adaptive shift in the hybrid based on environment data, an excess of ancestry solely from that parental species is not found (e.g., hybridization in sparrows between a salt marsh specialist and a recent salt marsh colonist concluded that introgression of loci associated with salt marsh tolerance surprisingly came from both parents; Walsh et al. 2018). In this regard, it is notable that we found a similar number of outlier loci with excess ancestry from each parent, as have other studies (e.g., Parchman et al. 2013; but see Chhatre et al. 2018). Additionally, the set of candidate adaptive loci traces that ancestry to both parents, in which environmental shifts are associated with precipitation and temperature gradients (Figs. 4 & 5). Nevertheless, and regardless of which parent contributed to excess ancestry at candidate adaptive loci, the majority of these loci align with regions outside of the environmental space preferred by *C. nelsonii* and therefore may be responsible for facilitating an environmental shift in the hybrid (Figs. 4 & 5, Table 3, Supplemental Table S7).

The role of candidate adaptive loci in facilitating environmental shifts in hybrids

Author Manuscript

Author Manuscript

Hybrids may occur where parental taxa overlap, or they may be defined by the limit of environmental tolerance of one of the parent species (Kadereit et al. 2015). Adaptive loci acquired via hybridization represent a potentially important source of genetic variation for adaptation to novel environments (Taylor et al. 2015). For example, hybridization in the sunflower genus *Helianthus* has been linked to the colonization of environments that are distinct from the parental species, including deserts, salt marshes, and sand dunes (Rieseberg et al. 2006). Likewise, candidate adaptive loci can be traced to hybridization in poplar trees, where the introgression is thought to facilitate adaptation at range limits (Chhatre et al. 2018). Moreover, candidate loci may promote the reproductive isolation of a hybrid from its parents in different ways; for example, hybrid speciation in *Ostryopsis* resulted in a stable hybrid that became reproductively isolated from one parent because of adaptation to iron-rich soils and from the other parent via loci associated with flowering time (Wang et al. 2021). Based on our results, hybridization in montane sedges serves as the source of potential adaptive loci (Fig. 5B), and these loci are associated with an environmental shift in which hybrids occupy an environment largely unsuitable for one of the parental species (i.e., *C. nelsonii*; see Fig. 4, Supplemental Table S7).

Our study also adds to growing evidence within the genus *Carex* that localized hybridization may be common between species (Drury 1956, Cayouette & Catling 1992, Waterway 1994, Więcław & Wilhelm 2014, Gizaw et al. 2016). Interestingly, our results highlight how hybridization can provide a source for adaptive divergence even when the hybrids occupy different environmental space (Fig. 4) and occur at geographic extremes separated by hundreds of kilometers (Fig. 1), as with the comparison of hybrids from northern Colorado/southern Wyoming (i.e., Flat Tops and Medicine Bows) with those from southern

Author Manuscript

Colorado (i.e., Lizard Head, Red Lakes, Mt. Oso). This geographic disjunction and differences in environmental space suggests the possibility of independent formation of hybrids (i.e., two separate origins for the hybrid taxon). However, this hypothesis is not supported by our data. For example, hybrids cluster together (with a few exceptions) in phylogenetic analyses (Fig. 3, Supplemental Fig. S2), and although the northern hybrids form a clade, the remainder of hybrid individuals do not form distinct clades based on geography (Fig. 3, Supplemental Fig. S2). The nesting of a clade of northern hybrids within the grade of hybrids from Lizard Head, Red Lakes, and Mt. Oso suggests that a founder event in the colonization of the north is a more likely explanation for observed genetic patterns (Fig. 3, Supplemental Fig. S2).

Under the hypothesis that hybrids all descend from a single hybridization event, whether the origin occurred in the north versus the south has important implications for interpreting the genetic and environmental differences between the two hybrid groups. The environmental conditions in the northern hybrids are within the environmental tolerances of *C. nelsonii,* whereas the environmental conditions experienced by all southern hybrids are well outside of the environmental space occupied by *C. nelsonii* (Fig. 4). Positing a northern origin of hybridization would suggest that the hybridization was a result of overlap of parental ranges, and hence the initial hybridization event reflected a breakdown of reproductive isolation. In contrast, a southern origin would suggest that an adaptive shift facilitated the formation of the hybrid taxon. Both hypotheses are consistent with candidate adaptive loci involvement in an environmental shift in the hybrid, but the difference is whether the environmental shift was secondary to the formation of the hybrid taxon (i.e., associated with range expansion) versus the origin of the hybrid taxon. The evidence suggests the latter hypothesis applies here. Specifically, candidate adaptive loci distinguish the hybrid taxon from *C. nelsonii*, but not *C. nova* (Fig. 5A). This suggests a southern

Author Manuscript

origin for the hybrid taxon that was driven by a shift in the environmental space hybrids occupy away from one parental species – in this case, *C. nelsonii* (Fig. 4). Moreover, when mapped onto the environmental PCA for both parental species and the hybrid, over 80% of candidate adaptive loci align with regions outside of the environmental space preferred by *C. nelsonii* (i.e., negative values of PC1 and PC2 in the environmental PCA; see Fig. 4). We interpret this to mean that although we do not know the precise environmental conditions promoting adaptation in specific candidate adaptive loci, many of the candidate adaptive loci may have increased the tolerance of hybrids to environments unsuitable for *C. nelsonii*. We note that only a couple 'pure' *C. nelsonii* individuals have been collected in the southern hybrid region (i.e., in the San Juan mountains) and this is the southern distribution limit of *C. nelsonii*, whereas *C. nova* is common. The hybridization event may have allowed the *C. nelsonii* genotype to persist in this region, albeit as a part of a new, reproductively isolated evolutionary lineage. The San Juans are also known to contain an abundance of endemic montane species (Fowler et al. 2014; Weber 2003), potentially because the rate of post-glacial climate change was high in the region (Briles et al. 2012). Taxa adapted to colder climates more typical of the more northern Rocky Mountains, such as *C. nelsonii*, either had to adapt to this new environmental space, or they were extirpated.

Caveats with identifying candidate adaptive loci in species with a complex history of hybridization

Hybridization between closely related species can occur in complex spatial and temporal contexts that make it challenging to infer the history of hybridization and interspecific gene flow, even with genomic data. Several analyses point to a complicated history in the focal taxa. For example, the hybrid index and population genetic statistics show the hybrid taxon's ancestry

skews towards *C. nova*, but other analyses such as STRUCTURE and SVDQuartets show a skew towards *C. nelsonii* (Figs. 2 & 3, Supplemental Fig. S2). Backcrossing between some hybrid individuals and a parent may explain some of the observed patterns, especially in the Mt. Oso-like hybrids, which show evidence of backcrossing with *C. nova* in the STRUCTURE and SVDQuartets analyses (Figs. 2 & 3, Supplemental Fig. S2). Furthermore, we cannot identify the location of outlier loci in the genome due to a lack of a reference genome for these taxa. Consequently, it is possible that genetic drift contributed to the pool of outlier loci, not selection (Gompert & Buerkle 2011). However, we consider the drift explanation unlikely given that candidate adaptive loci (Table 3) were overwhelmingly outlier loci (Fig. 5B). We expected alleles with excess ancestry from one or the other parental species to show an adaptive association with the environment space occupied by the hybrid. However, neither set of outlier loci showed a significant association with the environment when the hybrids were analyzed separately without either parent. This could mean a lack of adaptive importance in the outlier loci, or it could reflect limited statistical power due to the small sample sizes – there were only five hybrid populations. Furthermore, because we used RDAs with population allele frequencies as opposed to individual frequencies to avoid pseudoreplication, a large environmental effect on the loci was needed for detecting an association between candidate adaptive genes and local environmental conditions of the hybrid populations.

Processes other than selection may contribute to candidate adaptive loci detected among the outlier loci. The genus *Carex* has relatively high diversification rates, which are often attributed to chromosomal incompatibilities via holocentric chromosomes (Hipp 2007). Chromosomal incompatibilities can cause nearly instantaneous speciation, with offspring reproductively incompatible with their parents, but this is an unlikely explanation here because at

Author Manuscript

Author Manuscript

least one parent (*C. nova*) was reproductively compatible with the hybrid for some time (Fig. 3, Supplemental Fig. S2). Meiotic aberrations are also quite common in *Carex* and can cause these taxa to behave unlike most other angiosperms, especially with regard to hybridization (see Hipp 2007). There are not chromosome count data available for either of the parental species; counts in the genus range from $n = 6$ to $n = 62$ (Więcław et al. 2020), limiting our ability to consider the viability of this alternative explanation for the outlier loci. However, it seems peculiar that we observed a similar number of loci with extreme ancestry for each parent; this is consistent with a drift hypothesis, but the environmental association of outlier loci and their position in environmental space makes this unlikely. It may be possible that tight linkage (as with chromosomal rearrangements; see Rieseberg et al. 2003) could underlie the parity of the two parental species contribution to outlier loci. For example, if the outlier loci all occur on a single chromosome, some outlier loci may have nothing to do with an adaptive environmental shift in the hybrids. However, not all the outlier loci mapped to an environmental space distinct from *C. nelsonii* (Figs. 4 & 5B, Table 3). Although we consider the prospect unlikely, we cannot rule out the possibility that the candidate adaptive loci might reside on a single chromosome that is inherited as a linkage block, which would suggest that the hundreds of outlier loci we observe may represent many fewer independent loci with excess ancestry from one parent. Because of the documented chromosomal irregularities in *Carex,* this possibility needs to be considered, even if it statistically improbable (e.g., if the parental species have only six chromosomes, then > 1,000 loci would be expected to reside on a single chromosome, and with 607 candidate adaptive loci detected from the outlier loci, it is possible a single inversion could capture all candidate adaptive loci, but that would be a very large linkage block to remain in linkage equilibrium).

Author Manuscript

Conclusions

Genomic data considered in light of the geographic distribution of a hybrid taxon, in addition to the environmental space it occupies relative to the parental species, provide complementary evidence that supports the role of hybridization in contributing to an adaptive environmental shift in montane sedges. Our findings corroborate a hybrid taxon morphologically similar to *C. nelsonii,* but which occupies an environmental space inaccessible to *C. nelsonii*. Furthermore, given the distinctiveness of the hybrid's environmental distribution compared to the more northerly distributed *C. nelsonii*, but not the more southerly distributed *C. nova*, we suggest that the hybridization event, and the concomitant environmental shift, likely originated in the south (perhaps in the San Juan mountains in southern Colorado, where non-admixed *C. nelsonii* are rare, but *C. nova* is common), and hybrids subsequently colonized northern Colorado and southern Wyoming. Several lines of evidence support this historical scenario and suggest that candidate adaptive loci facilitated an environmental shift in the hybrid relative to the parental species *C. nelsonii*. Our results illustrate how a complementary set of analyses, in addition to independent datasets based on geography and the environment, can provide insight into the potential adaptive role of hybridization even when genomic resources (e.g., whole genome sequences, or a chromosomal scaffold) is lacking and in the absence of quantitative phenotypic data. In other words, the detailed examination of loci within hybrids relative to the genetic backgrounds of parental lineages not only identifies the role of hybridization in the formation of a new lineage of montane sedges, but also shows that hybridization has facilitated an adaptive shift in which hybrid individuals occupy unique environmental conditions.

Author Manuscript

Literature Cited

1. Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.
2. Abbott RJ, Barton NH, Good JM (2016) Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, **25**, 2325–2332.
3. Albertson RC, Markert JA, Danley PD, Kocher TD (1999) Phytogeny of a rapidly evolving clade: The cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 5107–5110.
4. Arnold ML, Martin NH (2009) Adaptation by introgression. *Journal of Biology*, **8**, 82.
5. Arnold ML, Kunte K (2017) Adaptive Genetic Exchange: A Tangled History of Admixture and Evolutionary Innovation. *Trends in Ecology and Evolution*, **32**, 601–611.
6. Blischak PD, Chifman J, Wolfe AD, Kubatko LS (2018) HyDe: A Python Package for Genome-Scale Hybridization Detection (D Posada, Ed,). *Systematic Biology*, **67**, 821–829.
7. Briles CE, Whitlock C, Meltzer DJ (2012) Last glacial–interglacial environments in the southern Rocky Mountains, USA and implications for Younger Dryas-age human occupation. *Quaternary Research*, **77**, 96-103.
8. Capblancq T, Luu K, Blum MGB, Bazin E (2018) Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, **18**, 1223–1233.
9. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–40.
10. Cayouette J, Catling PM (1992) Hybridization in the Genus *Carex* with Special Reference to North America. *The Botanical Review*, **58**, 351–438.
11. Chhatre VE, Evans LM, DiFazio SP, Keller SR (2018) Adaptive introgression and maintenance of a trispecies hybrid complex in range-edge populations of *Populus*. *Molecular Ecology*, **27**, 4820–4838.
12. Chifman J, Kubatko L (2014) Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics*, **30**, 3317–3324.
13. Danacek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
14. Dion-Côté AM, Barbash DA (2017) Beyond speciation genes: an overview of genome stability in evolution and speciation. *Current Opinion in Genetics and Development*, **47**, 17–23.
15. Dray S, Dufour AB (2007) The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.
16. Drury WH (1956) The ecology of the natural origin of a species of *Carex* by hybridization. *Rhodora*, **58**, 51–72.
17. Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.
18. Eaton DAR, Overcast I (2020) Ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, **36**, 2592–2594.

Author Manuscript

19. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
20. Fitak RR (2021) OptM: Estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols*, **6**.
21. Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, **27**, 2215–2233.
22. Fowler JF, Nelson BE, Hartman RL (2014) Vascular plant flora of the alpine zone in the Southern Rocky Mountains, USA. *Journal of the Botanical Research Institute of Texas*, **8**, 611-636.
23. Galaverni M, Caniglia R, Pagani L *et al.* (2017) Disentangling Timing of Admixture, Patterns of Introgression, and Phenotypic Indicators in a Hybridizing Wolf Population. *Molecular Biology and Evolution*, **34**, 2324–2339.
24. Gizaw A, Wondimu T, Mugizi TF *et al.* (2016) Vicariance, dispersal, and hybridization in a naturally fragmented system: the afro-alpine endemics Carex monostachya and C. runssoroensis (Cyperaceae). *Alpine Botany*, **126**, 59–71.
25. Gompert Z, Lucas LK, Fordyce JA, Forister ML, Nice CC (2010) Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive admixture and a patchy hybrid zone. *Molecular Ecology*, **19**, 3171–3192.
26. Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.
27. Gompert Z, Buerkle CA (2012) bgc : Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, **12**, 1168–1176.
28. Gompert Z, Mandeville EG, Buerkle CA (2017) Analysis of Population Genomic Data from Hybrid Zones. *Annual Review of Ecology, Evolution, and Systematics*, **48**, 207–229.
29. Goulet BE, Roda F, Hopkins R (2017) Hybridization in plants: Old ideas, new techniques. *Plant Physiology*, **173**, 65–78.
30. Grant PR, Grant BR (1994) Phenotypic and genetic effects of hybridization in Darwin's finches. *Evolution*, **48**, 297–316.
31. Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**, 4606–4618.
32. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
33. Hijmans RJ (2019) raster: Geographic Data Analysis and Modeling. R package version 3.0-2. https://CRAN.R-project.org/package=raster
34. Hijmans RJ (2017) dismo: Species Distribution Modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo
35. Hipp AL (2007) Nonuniform processes of chromosome evolution in sedges (*Carex*: Cyperaceae). *Evolution*, **61**, 2175–2194.
36. Hodel RGJ, Massatti R, Bishop SGD, Knowles LL (2021) Testing which axes of species differentiation underlie covariance of phylogeographic similarity among montane sedge species. *Evolution*, **75**, 349–364.

Author Manuscript

37. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
38. Kadereit JW (2015) The geography of hybrid speciation in plants. *Taxon*, **64**, 673–687.
39. Kim S-C, Rieseberg LH (1999) Genetic architecture of species differences in annual sunflowers: implications for adaptive trait introgression. *Genetics*, **153**, 965–977.
40. Knowles LL, Richards CL (2005) Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. *Molecular Ecology*, **14**, 4023–4032.
41. Kou Y, Zhang L, Fan D *et al.* (2020) Evolutionary history of a relict conifer, *Pseudotaxus chienii* (Taxaceae), in southeast China during the late Neogene: Old lineage, young populations. *Annals of Botany*, **125**, 105–117.
42. Kozak KM, Wahlberg N, Neild AFE *et al.* (2015) Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. *Systematic Biology*, **64**, 505–524.
43. Kunte K, Shea C, Aardema ML *et al.* (2011) Sex Chromosome Mosaicism and Hybrid Speciation among Tiger Swallowtail Butterflies (NA Moran, Ed,). *PLoS Genetics*, **7**, e1002274.
44. Lamichhaney S, Berglund J, Almén MS *et al.* (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371–375.
45. Legendre P, Legendre L (2012) *Numerical Ecology*. Amsterdam, The Netherlands: Elsevier.
46. Luu K, Bazin E, Blum MGB (2017) *pcadapt*: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, **17**, 67–77.
47. Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**, 229–237.
48. Massatti R, Knowles LL (2014) Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution*, **68**, 2833–2846.
49. Massatti R, Knowles LL (2016) Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Molecular Ecology*, **25**, 3974–3986.
50. Massatti R, Knowles LL (2020) The historical context of contemporary climatic adaptation: a case study in the climatically dynamic and environmentally complex southwestern United States. *Ecography*, **43**, 735–746.
51. Meier JI, Marques DA, Mwaiko S *et al.* (2017) Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, **8**, 14363.
52. Menon M, Bagley JC, Page GFM *et al.* (2021) Adaptive evolution in a conifer hybrid zone is driven by a mosaic of recently introgressed and background genetic variants. *Communications Biology*, **4**, 1–14.
53. Oksanen, J. F., Blanchet, G., & Friendly, M. et al. (2017). *vegan: Community ecology package*. R package version 2.4-4.
54. Parchman TL, Gompert Z, Braun MJ *et al.* (2013) The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, **22**, 3304–3317.

Author Manuscript

55. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, **7**.
56. Pickrell, J.K., Pritchard, J.K., 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet. 8, 1002967. https://doi.org/10.1371/journal.pgen.1002967
57. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
58. Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–4370.
59. Rieseberg LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, **28**, 359–389.
60. Rieseberg LH, Raymond O, Rosenthal DM *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211–1216.
61. Rieseberg LH (2006) Hybrid speciation in wild sunflowers. *Annals of the Missouri Botanical Garden*, **93**, 34–48.
62. Rieseberg LH, Willis JH (2007) Plant speciation. *Science*, **317**, 910–914.
63. Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology and Evolution*, **19**, 198–207.
64. Suarez-Gonzalez A, Lexer C, Cronk QCB (2018) Adaptive introgression: a plant perspective. *Biology Letters*, **14**, 20170688.
65. Taylor SA, Larson EL, Harrison RG (2015) Hybrid zones: Windows on climate change. *Trends in Ecology and Evolution*, **30**, 398–406.
66. Taylor SA, Larson EL (2019) Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution*, **3**, 170–177.
67. Twyford AD, Ennos RA (2012) Next-generation hybridization and introgression. *Heredity*, **108**, 179–189.
68. Vanhove M, Pina-Martins F, Coelho AC *et al.* (2021) Using gradient forest to predict climate response and adaptation in Cork oak. *Journal of Evolutionary Biology*, jeb.13765.
69. Walsh J, Kovach AI, Olsen BJ, Shriver WG, Lovette IJ (2018) Bidirectional adaptive introgression between two ecologically divergent sparrow species. *Evolution*, **72**, 2076–2089.
70. Wang X, Ye X, Zhao L *et al.* (2017) Genome-wide RAD sequencing data provide unprecedented resolution of the phylogeny of temperate bamboos (Poaceae: Bambusoideae). *Scientific Reports*, **7**, 1–11.
71. Wang Z, Jiang Y, Bi H *et al.* (2021) Hybrid speciation via inheritance of alternate alleles of parental isolating genes. *Molecular Plant*, **14**, 208–222.
72. Waterway MJ (1994) Evidence for the hybrid origin of *Carex knieskernii* with comments on hybridization in the genus *Carex* (Cyperaceae). *Canadian Journal of Botany*, **72**, 860–871.
73. Weber WA (2003) The Middle Asian element in the southern Rocky Mountain flora of the western United States: a critical biogeographical review. *Journal of Biogeography*, **30**, 649-685.

74. Whitney KD, Randell RA, Rieseberg LH (2010) Adaptive introgression of abiotic tolerance traits in the sunflower Helianthus annuus. *New Phytologist*, **187**, 230–239.
75. Więcław H, Wilhelm M (2014) Natural Hybridization within the *Carex flava* Complex (Cyperaceae) in Poland: Morphometric Studies. *Annales Botanici Fennici*, **51**, 129–147.
76. Więcław H, Kalinka A, Koopman J (2020) Chromosome numbers of *Carex* (Cyperaceae) and their taxonomic implications. *PLOS ONE*, **15**, e0228353.
77. Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high- performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326– 3328. https://doi.org/10.1093/bioinformatics/bts606

Data Accessibility

Supporting Information files, including DNA matrices, sampling locations, and GPS coordinates from vetted herbarium specimens are deposited in the Dryad Digital Repository (doi: https://doi.org/10.5061/dryad.sbcc2fr8f). Scripts for conducting redundancy analyses are available in a GitHub Repository (https://github.com/richiehodel/Carex_hybridization).

Acknowledgements

We thank Giorgia Auteri, Arnaud Becheler, Sasha Bishop, and Cecilia Fiorini for valuable discussions that improved this manuscript. We thank Marcella Baiz for assistance in running bgc software and Zach Gompert for providing helpful comments on a draft of this manuscript. Finally, we thank the associated editor and three anonymous reviewers for many valuable comments and suggestions during the review process.

Table 1. Results from Hybrid Detector (HyDe) for the three lineages to test hybrid status using phylogenetic invariants arising under a coalescent model with hybridization. The test statistics show the results of testing the null hypothesis that $\gamma = 0$, where the γ parameter describes the

Author Manuscript

probability of hybrid ancestry from one parent while 1-γ represents the probability of hybrid ancestry from the other parent. Only the case in which the hybrids were tested as the hybrid taxon rejected the null hypothesis.

P1	Hybrid	P2	*Z*-score	*P*-value	γ
C. nelsonii	hybrid	*C. nova*	3.119	0.001	0.822
C. nelsonii	*C. nova*	hybrid	-9.725	1.000	0.562
hybrid	*C. nelsonii*	*C. nova*	-2.559	0.995	1.376

Table 2. Regression coefficients from the overall RDAs are displayed for each combination of taxa and loci sets. The r^2 value in RDA needs to be adjusted to account for multiple predictor variables; accordingly adjusted r^2 is shown. The *P*-values for the overall RDA model are shown, as are any significant *P*-values for individual RDA axes – no individual axes other than RDA1 were significant. Asterisks indicate significance with an α level of 0.05.

	Loci set	Adjusted r^2	Overall RDA *P*-value	RDA1 *P*-value
C. nelsonii	inliers	-0.128	0.743	
	C. nova-outliers	-0.073	0.907	
	C. nelsonii-outliers	-0.157	0.929	
	all	-0.131	0.770	
C. nova	inlier	-0.004	0.544	
	C. nova-outliers	0.001	0.466	
	C. nelsonii-outliers	0.012	0.261	
	all	-0.002	0.524	
hybrid	inlier	-0.737	0.908	
	C. nova-outliers	-0.746	0.933	
	C. nelsonii-outliers	-0.948	0.917	
	all	-0.747	0.908	
C. nelsonii + hybrid	inlier	0.119	0.125	
	C. nova-outliers	0.242	0.010*	0.019*
	C. nelsonii-outliers	0.228	0.018*	0.029*
	all	0.148	0.085	
C. nova + hybrid	inlier	0.044	0.071	
	C. nova-outliers	0.033	0.090	
	C. nelsonii-outliers	0.029	0.118	
	all	0.040	0.076	
C. nelsonii + *C. nova*	inlier	0.014	0.245	
	C. nova-outliers	0.077	0.013*	0.016*

Author Manuscript

	C. nelsonii-outliers	0.080	0.047*	0.065
	all	0.081	0.008*	0.022*

Table 3. The candidate adaptive loci identified under three different cutoffs (i.e., 2, 2.5, 3 standard deviations from the mean of the inlier loci RDA). Candidate adaptive loci are shown separately for each set of outlier loci (*Carex nova*-outliers, top; *Carex nelsonii*-outliers, bottom). The loadings of the candidate adaptive loci onto the precipitation (PC1) and temperature (PC2) PCs from the RDA (Fig. 5) are shown in the four rightmost columns with values highlighted in gray indicating that they correspond to an environment outside of the suitable environment characterized for *C. nelsonii* (see Fig. 4).

C. nova-outliers

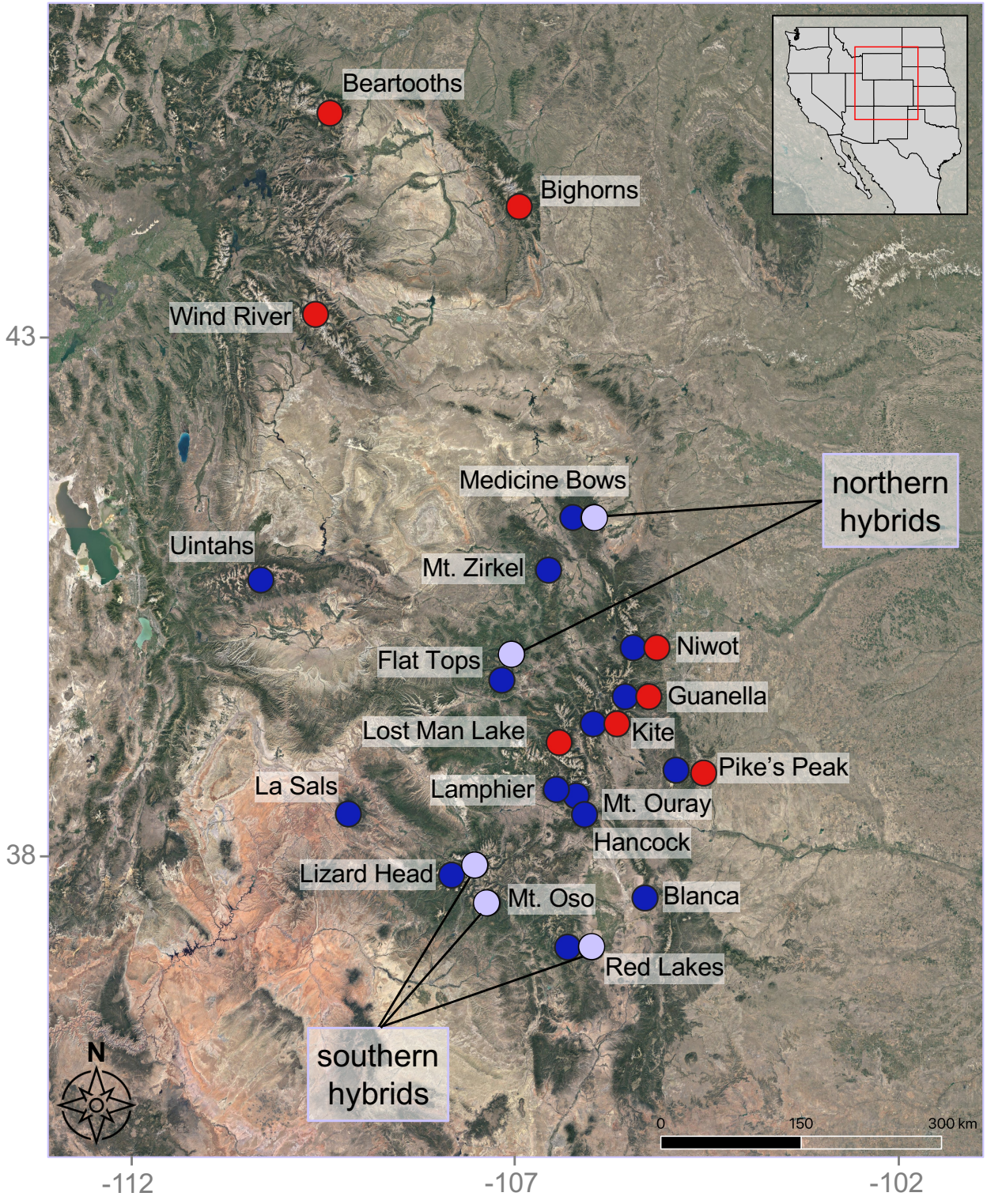
Standard deviations	Candidate loci	Positive correlation with PC1	Negative correlation with PC1	Positive correlation with PC2	Negative correlation with PC2
2	251	84	67	51	47
2.5	124	33	24	32	33
3	9	2	2	2	3

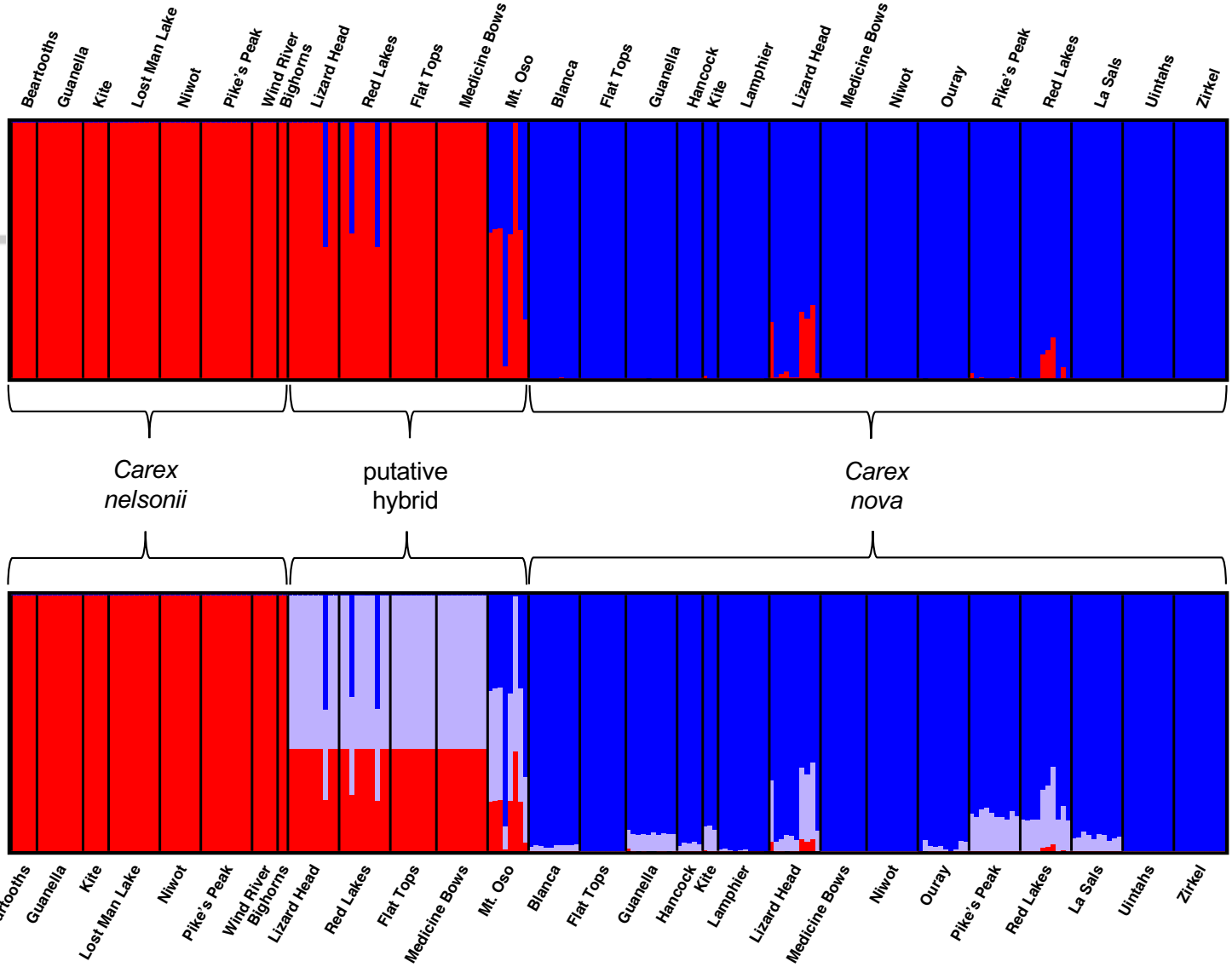
C. nelsonii-outliers

Standard deviations	Candidate loci	Positive correlation with PC1	Negative correlation with PC1	Positive correlation with PC2	Negative correlation with PC2
2	356	115	119	63	58
2.5	108	20	18	37	33
3	2	2	0	0	0

Author Manuscript

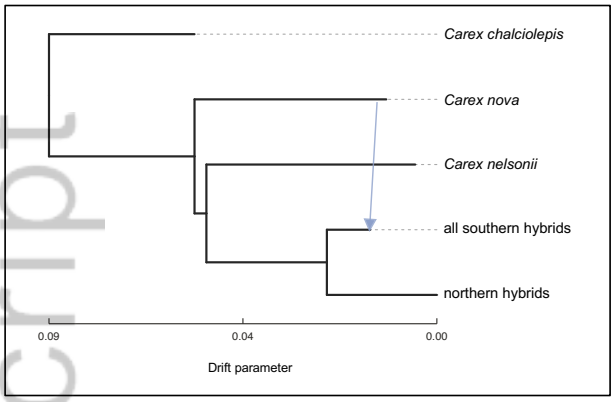
Author Manuscript

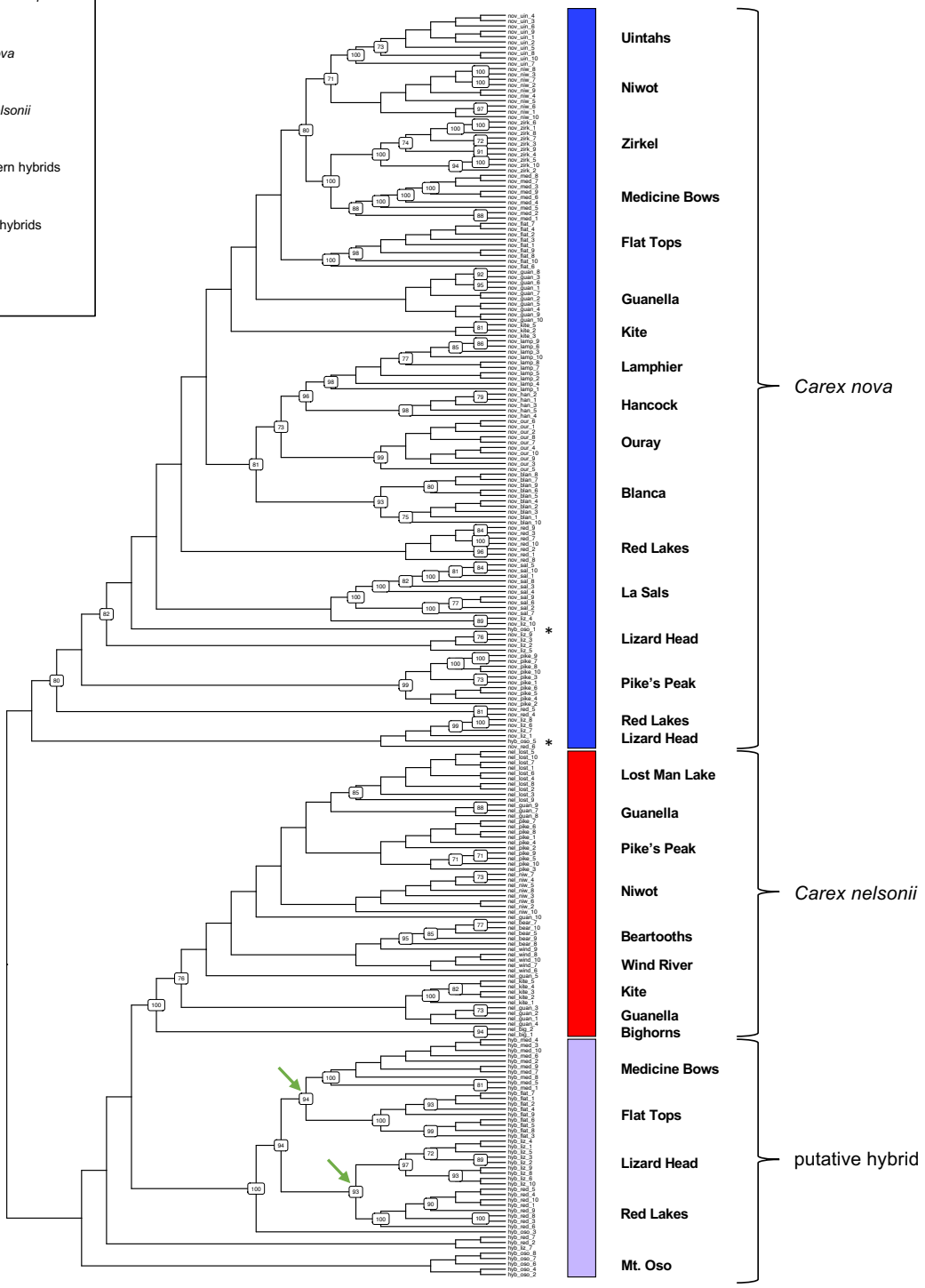




Author Manuscript

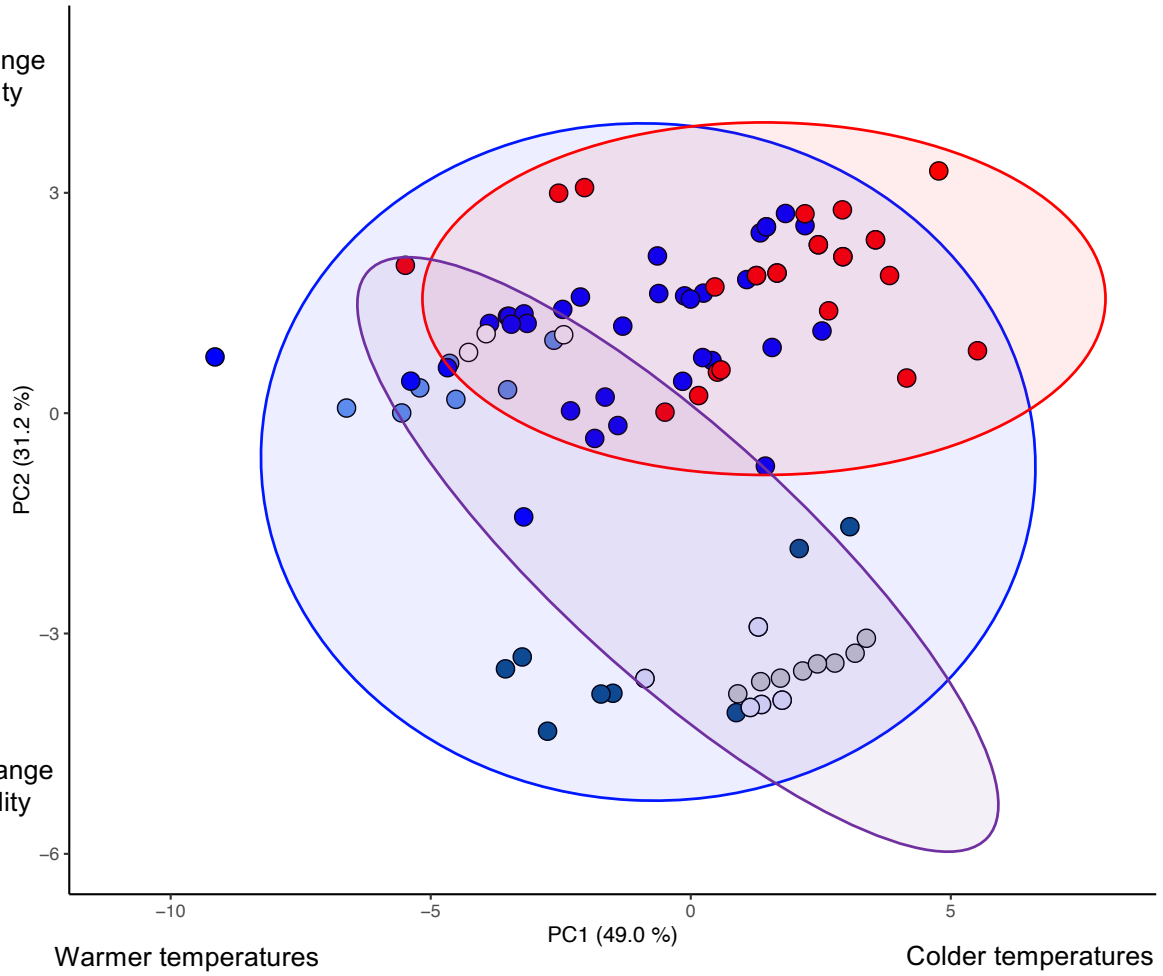
Author Manuscript





Author Manuscript

Lower diurnal temperature range
Lower precipitation seasonality



Higher diurnal temperature range
Higher precipitation seasonality

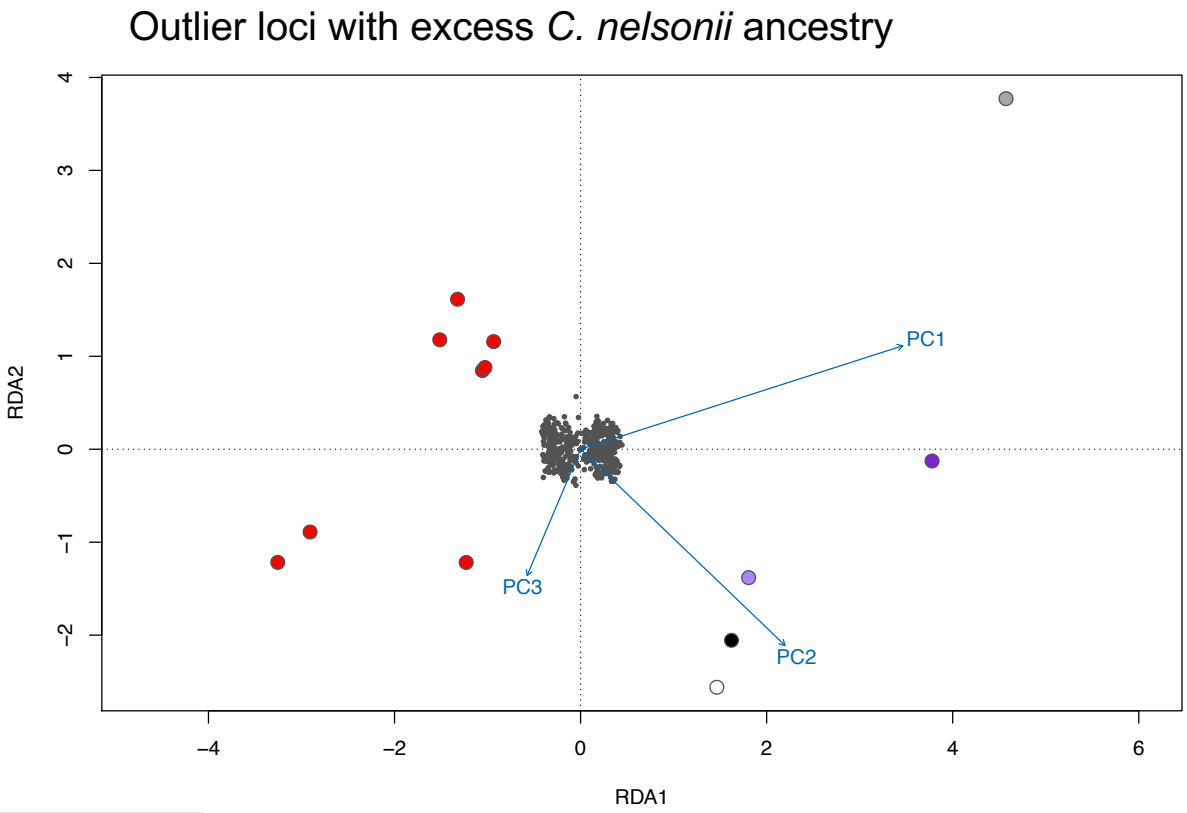
Warmer temperatures
Lower precipitation

Colder temperatures
Higher precipitation

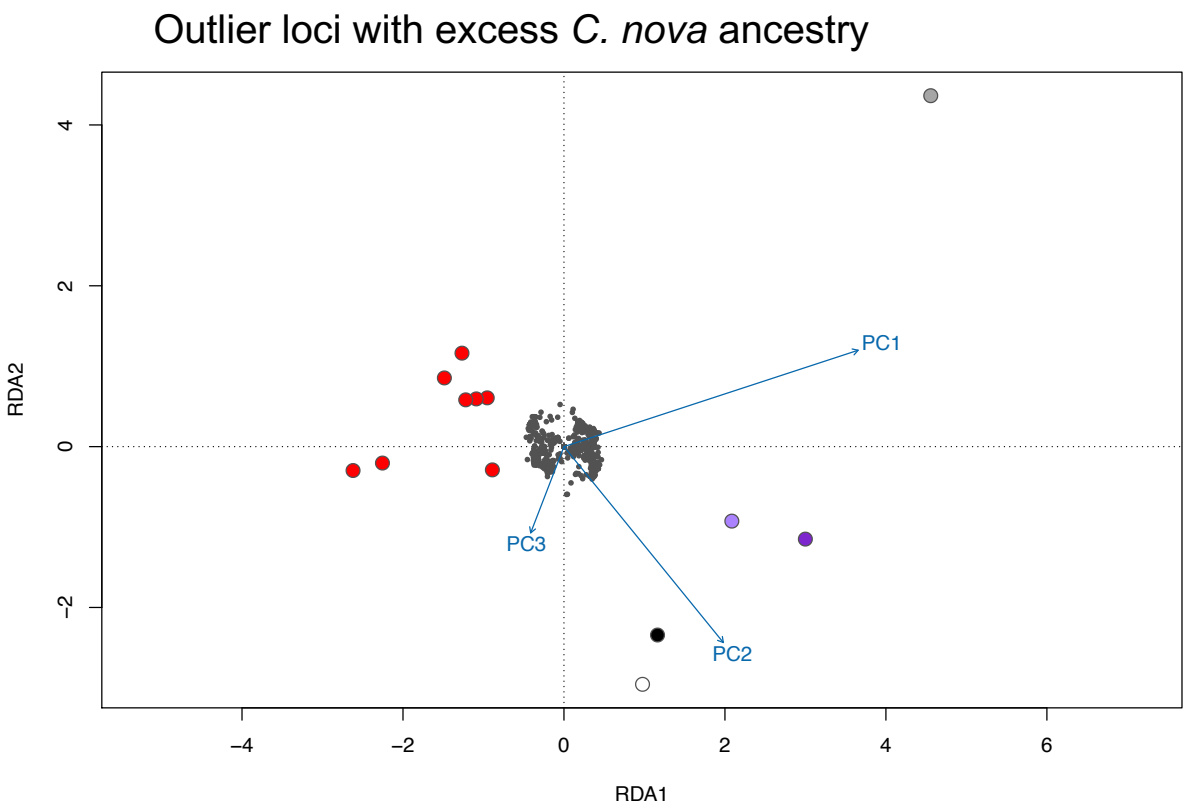
Species / Population

- C. nelsonii–Beartooths
- C. nelsonii–Bighorns
- C. nelsonii–Guanella
- C. nelsonii–Kite
- C. nelsonii–Lost Man Lake
- C. nelsonii–Niwot
- C. nelsonii–Pike's Peak
- C. nelsonii–Wind River
- Hybrid–Flat Tops
- Hybrid–Medicine Bows
- Hybrid–Lizard Head
- Hybrid–Red Lakes
- Hybrid–Mt. Oso
- C. nova–Flat Tops
- C. nova–Medicine Bows
- C. nova–Lizard Head
- C. nova–Red Lakes
- C. nova–Blanca
- C. nova–Guanella
- C. nova–Hancock
- C. nova–Kite
- C. nova–La Sals
- C. nova–Lamphier
- C. nova–Niwot
- C. nova–Ouray
- C. nova–Pike's Peak
- C. nova–Uintahs
- C. nova–Zirkel

A.

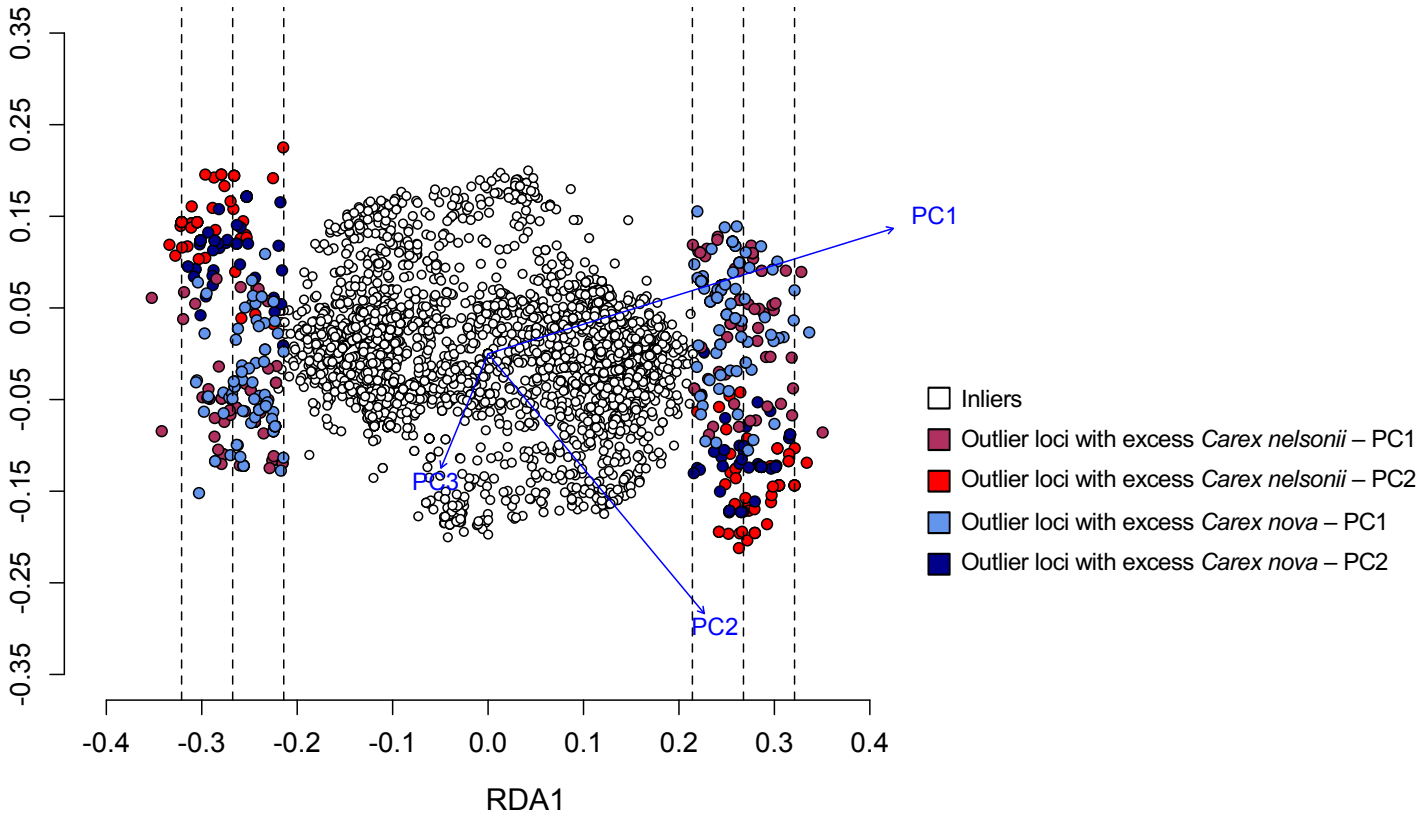


- Hybrid-Flat Tops
- Hybrid-Medicine Bows
- Hybrid-Lizard Head
- Hybrid-Red Lakes
- Hybrid-Mt. Oso
- *C. nelsonii*-all populations



Author Manuscript

B.



Author Manuscript

Author Manuscript

Figure 1. The geographic distribution of sampling sites for *Carex nelsonii* (red dots), *Carex nova* (blue dots), and the hybrid lineage (purple dots). Sampling locations have been jittered to reduce overlap.

Figure 2. Ancestry estimates for *K*=2 (top) and K=3 (bottom) based on STRUCTURE analyses; colors represent ancestry coefficients for individuals, which are demarcated by thin white lines. At *K* = 3, the hybrid origin of five populations is apparent. Individuals are labeled according to their sampling location (see Fig.1).

Figure 3. The SVDQuartets species tree inferred for all *Carex nelsonii*, *Carex nova*, and hybrid individuals using all 6,740 unlinked SNPs. The tree was rooted in a separate species tree analysis (see Supplemental Fig. S2) based on a subset of loci that included ten individuals from the *Carex chalciolepis*. Two Mt. Oso hybrid individuals marked by asterisks were nested within *C. nova*. Nodes with >70% bootstrap values are labeled. The two green arrows indicate the nodes defining monophyletic groups of hybrids—northern hybrids (Medicine Bows and Flat Tops individuals)—and core southern hybrids (most individuals from Lizard Head and Red Lakes. Inset in the upper left is the maximum likelihood tree inferred with migration using TREEMIX with separate lineages defined by geography for the northern hybrids (i.e., Flat Tops and Medicine Bows populations; see Fig. 1), the southern hybrids (i.e., Lizard Head, Mt. Oso, and Red Lakes populations; see Fig. 1), and rooted with *Carex chalciolepis*. The blue arrow depicts the direction of gene flow for the model with the optimal number of migration events (i.e., *m* = 1).

Author Manuscript

Figure 4. Environmental PCA inferred from 17 bioclimatic variables showing sampling locations for *Carex nelsonii*, *Carex nova*, and the hybrid lineage. PC1 and PC2 are strongly ($r^2 > 0.45$) and significantly ($p<0.05$) correlated with certain bioclimatic variables (Supplemental Table S7); general patterns are noted on the PC axes. Note the southern hybrids (Lizard Head, Mt. Oso, Red Lakes) occur outside of the environmental space occupied by *C. nelsonii*.

Figure 5.

A) RDA plots for *Carex nelsonii* and the hybrid lineage using the two sets of loci identified as outlier loci with excess ancestry from one of the parent species; *Carex nova*-outliers (1,027 loci) are shown in the top plot, and *C. nelsonii*-outliers (971 loci) in the bottom plot. Each plot shows RDA axis 1 versus RDA axis 2, with SNPs depicted by small gray dots. Populations are shown by color-coded dots. In both plots, the overall RDA and axis 1 were significant. The positions of the populations on RDA1, but not RDA2, indicate significant differences in the environmental association of the alleles in a population.

B) Detailed relationships among SNPs from the RDA plots (A). Inlier loci with no excess ancestry from either parent are shown as white dots, while candidate adaptive loci that showed a significant association with either the precipitation axis (PC1) or the temperature axis (PC2) are color coded dots. Maroon- and periwinkle-colored dots mark SNPs correlated with PC1, and red and navy dots mark SNPs correlated with PC2. Candidate adaptive loci with an excess of *Carex nelsonii* ancestry are shown in maroon and periwinkle, whereas candidate adaptive loci with an excess of *Carex nova* ancestry are shown in red and navy. The dashed vertical lines represent the threshold for SNPs that are 2, 2.5, or 3 standard deviations from the mean of the set of inlier loci.