Associations between genomic ancestry, genome size and capitula morphology in the invasive meadow knapweed hybrid complex (Centaurea × moncktonii) in eastern North America

Susanne Lachmuth1,2,3, Jane Molofsky1, Lindsey Milbrath4, Jan Suda5,6 and Stephen R. Keller1*

1University of Vermont, Department of Plant Biology, Burlington, VT 05405, USA, 2Martin Luther University Halle Wittenberg, Institute of Biology, Geobotany & Botanical Garden, 06108 Halle (Saale), Germany, 3German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany, 4United States Department of Agriculture, Agricultural Research Service (USDA-ARS), Ithaca, NY 14853, USA, 5Charles University Prague, Department of Botany, 128 01 Prague, Czech Republic, 6Czech Academy of Sciences, Institute of Botany, 252 43 Průhonice, Czech Republic

*Corresponding author's e-mail address: sskeller@uvm.edu

Associate Editor: Xavier Picó

Citation: Lachmuth S, Molofsky J, Milbrath L, Suda J, Keller SR. 2019. Associations between genomic ancestry, genome size and capitula morphology in the invasive meadow knapweed hybrid complex (Centaurea × moncktonii) in eastern North America. AoB PLANTS 11: plz055; doi: 10.1093/aobpla/plz055

Abstract

Plant invasions are prime opportunities for studying hybridization and the nature of species boundaries, but hybrids also complicate the taxonomic treatment and management of introduced taxa. In this study, we use population genomics to estimate the extent of genomic admixture and test for its association with morphology and genome size in a hybrid complex of knapweeds invasive to North America: meadow knapweed (Centaurea × moncktonii) and its parental species (C. jacea and C. nigra). We sampled 20 populations from New York and Vermont, USA, and used genotyping by sequencing to identify single nucleotide polymorphisms in order to estimate genome-wide ancestry and classify individuals into hybrid genotype classes. We then tested for association between degree of genomic introgression and variation in a subset of traits diagnostic for the parental taxa, namely capitula morphology and monoploid genome size. Genomic clustering revealed two clearly defined lineages, as well as many admixed individuals forming a continuous gradation of introgression. Individual assignments to hybrid genotype classes revealed many advanced generation intercrosses and backcrosses, suggesting introgression has been extensive and unimpeded by strong reproductive barriers between taxa. Variation in capitula traits between the two unadmixed, presumed parental, lineages exhibited continuous, and in some cases transgressive, segregation among introgressed hybrids. Genome size was also divergent between lineages, although advanced generation hybrids had smaller genomes relative to additive expectations. Our study demonstrates deep introgression between the porous genomes of a hybrid invasive species complex. In addition to strong associations among genomic ancestry, genome size and morphology, hybrids expressed more extreme phenotypic values for capitula traits and genome size, indicating transgressive segregation and genome size reduction interact to confer invasiveness. Future studies will apply these results to experimentally test how introgression, transgressive segregation and genome size reduction interact to confer invasiveness.

Keywords: Asteraceae; black knapweed; brown knapweed; Centaurea jacea; Centaurea nigra; genomic admixture; genotyping by sequencing; introgression; single nucleotide polymorphisms.
Introduction

The invasion of numerous exotic plant species has been preceded or accompanied by interspecific hybridization. Consequently, hybridization has been hypothesized to be a key evolutionary mechanism stimulating invasiveness (Ellstrand and Schierenbeck 2000; Schierenbeck and Ellstrand 2009; Hovick and Whitney 2014). Admixture between parental genomes may increase the fitness of hybrids through short-term hybrid vigour (F1 generation) and/or increased evolutionary potential in later recombinant generations (Lavergne and Molofsky 2007; Keller and Taylor 2010; Keller et al. 2014; Parepa et al. 2014; Li et al. 2018). However, many plant taxa capable of hybridization also show partial reproductive isolation as a result of genetic incompatibilities, meiotic irregularities or gene dosage effects due to differences in genome size or copy number (ploidy), leading to either F1 hybrid sterility or outbreeding depression in advanced recombinant generations (Lynch 1991; Harrison and Larson 2014; Goulet et al. 2017). Consequently, the ways in which hybridization may facilitate invasion depend on whether hybrid genotypes are restricted to early generations (or maintained through clonal propagation), or rather if hybrid genomes are porous to introgression through advanced generations of recombination (Bock et al. 2015). Despite widespread appreciation of the association between hybridization and invasion (Ellstrand and Schierenbeck 2000), few studies have actually tested whether hybridization during invasion reflects short-term vs. advanced generations of introgression (Dlugosch et al. 2015).

One potential constraint on hybridization and how it resolves is the difference in genome size between parental taxa. In plants, hybridization is often associated with genome size variation, either due to genome duplication (polyploidy) or smaller-scale gains and losses of chromosomes (aneuploidy) or chromosomal segments during homoplody hybridization. Population-level cytogenetic estimation of genome size has proven successful in resolving complex hybrid taxonomies, delimiting species boundaries and highlighting the origins of hybrids (Mahelka et al. 2005; Ekr et al. 2010). Additionally, frequent gains or losses of large genomic regions may have fitness consequences (Hovick and Whitney 2014), and there is a growing body of literature that supports the hypothesis that small holoploid (1C) genome size promotes invasion success (Pandit et al. 2014; Suda et al. 2014; Schmidt et al. 2017). Given the malleability of genome size to evolve differences between parental lineages and their hybrids (Baack and Rieseberg 2007), associating ancestry and genome size in an actively hybridizing invasive species complex could aid in distinguishing between parental types and hybrids, and may also yield insight into the reported associations of invasiveness with hybridization and genome size (Hovick and Whitney 2014; Suda et al. 2014).

An additional consequence of the increased phenotypic variance that often accompanies hybridization is that it may be difficult or impossible to determine the ancestry status of hybrids and their parental types using morphological or cytological data alone. This can complicate the study and management of invasive species as a consequence of taxonomic confusion over invading taxa, which are often recognized based on visually diagnostic morphological traits. Recent developments in population genomics provide vast opportunity for investigations into the genomic consequences of hybridization, even in non-model organisms lacking a reference genome (Goulet et al. 2017). The increased resolution of large, genome-wide sequencing data sets can prove advantageous to deciphering the pattern of ancestry and introgression, especially in morphologically difficult species complexes (Vilà et al. 2000). One such putatively hybrid species is meadow knapweed (Centaurea × moncktonii)—one of the invasive representatives of the extremely polymorphic Centaurea subg. jacea (Asteraceae), sometimes referred to as the C. jacea/nigra complex (e.g. Gardou 1972). Meadow knapweed is reported to be a fertile hybrid between C. jacea (brown knapweed) and C. nigra (black knapweed), with either brown or black knapweed as the female parent and the hybrid capable of freely backcrossing with either parental species (Roché and Coombs 2003). Hybrids of C. × moncktonii were likely introduced to North America directly, along with the parental species (Roché and Roché 1991), from the late 1800s onwards, but the occurrence of more recent post-introduction hybridization events is also possible. Based on morphological identification, pure parental forms do not appear to persist at least in western North American hybrid zones (Roché and Roché 1991). So far, invasive C. × moncktonii populations are reported to be tetraploid, although both tetraploids and diploids occur in Europe (Roché and Roché 1991). Despite their relevance as noxious weeds in the USA (Roché and Coombs 2003), members of the complex have only been studied in Europe where it is native and not in North America (Garcia-Jacas et al. 2000, 2006; Vanderhoeven et al. 2002), and no genetic studies have resolved the ancestry of parental and hybrid types within this complex on either continent. Furthermore, it is not clear whether North American populations represent a hybrid complex of multiple knapweed species or whether only C. × moncktonii persists as advanced generation hybrids following the original introduction. Consequently, the potential effects of genomic recombination and genome size variation on invasive spread remained entirely unexplored.

Here, we report on a study that combines a population genomic survey of single nucleotide polymorphisms (SNPs) with morphometric analyses on reproductive traits and assessments of genome size for samples of the C. jacea/nigra complex collected from 20 locations in the northeastern USA (New York and Vermont). Our goal is to integrate multiple sources of evidence to deconvolute this apparent hybrid swarm into its constituent sources of ancestry and to determine the level of introgression that exists among parental species and their hybrids. Unravelling the extent of genomic introgression and mixed ancestry present within hybrid zones of introduced species is prerequisite for addressing the broader goal of how admixture and introgression affect fitness and contribute to invasion success (Keller and Taylor 2008; Dlugosch et al. 2015). Our results reveal an abundance of advanced generation hybrids (F2 and backcrosses) and the persistence of a much smaller number of individuals with unadmixed genomic ancestry. The genomic differences among ancestry groups are corroborated by strong associations between ancestry, the capita traits measured and genome size. We discuss how combining multiple sources of evidence (genomics, cytometry, morphology) provide distinct advantages in deciphering the composition of plant hybrid zones, and can help unravel the complex systems of admixture that often arise during biological invasion.

Materials and Methods

Study species

Centaurea jacea and C. nigra (Asteraceae) are short-lived, perennial knapweeds native throughout Europe and introduced to North
According to Susanna and Roché (2011), *C. jacea* (Ochsmann 2006) by recognizing as two and *C. jacea* and *C. nigra* following the Flora of North America (Keil and Ochsmann 2006) in recognizing the nothospecies *C. pratensis* (Hardy et al. 2000). The hybrid *C. × moncktonii* prefers moister habitats than its parental species, such as meadows, irrigated pastures, riparian zones and moist forest openings (Roché and Roché 1991; Miller and Lucero 2014). The presence of the hybrid is undesirable because of its persistence in native plant communities and it also reduces the quality of forage produced in pastures and grass hayfields (Roché and Coombs 2003).

In their native Eurasian ranges, *C. jacea* has the widest native range with a rather continuous distribution spanning from Morocco to Norway and from Spain to Eastern Russia. The longitudinal distribution of *C. nigra* is more restricted, with easternmost records from Poland and Norway (Botanic Garden and Botanical Museum of Berlin-Dahlem 2006). The occurrence of *C. jacea × C. nigra* hybrids has been reported from several European countries, including the British Isles (Marsden-Jones and Turril 1954), Spain (Dostál 1975), Belgium (Vanderhoeven et al. 2002) and France (Orfeo Crosa and Bancheva 2006). In North America, occurrences of *C. jacea* and *C. nigra* were reported before records of *C. × moncktonii* occurred, while the increasing spread of *C. × moncktonii* went along with an apparent decline of the parental taxa since the 1950s (Roché and Roché 1991). The species were thought to have been introduced unintentionally via ship ballast as well as intentionally as a hay or forage crop and as a pollen source for honeybees (Roché and Coombs 2003). The first reports of *C. jacea* and *C. nigra* in North America are both from the Pacific Northwest. *Centaurea jacea* was found at Vancouver Island, 1887 and *C. nigra* in Pullman, WA, in 1895 (Roché and Roché 1991). The earliest records of hybrid meadow knapweed are from Oregon near Eugene, Lane County in 1918 (Roché and Roché 1991). Before 1960, *C. × moncktonii* was additionally reported from British Columbia, Washington and Montana according to (Roché and Roché 1991), who were able to show a substantial increase of the hybrid in these states by the time of their 1991 survey and added first records in Idaho. In recent decades, meadow knapweed has rapidly expanded its range and now occurs in ~25 of the US and four Canadian provinces (Miller and Lucero 2014), ranging from the Atlantic to the Pacific and South to North Carolina (USDA, NRCS 2013 in Miller and Lucero 2014). It has greatly increased in abundance in western coastal states (Miller and Lucero 2014) and is common in parts of New York State and New England (Eckel 2012). Other Eurasian *Centaurea* species introduced to North America are not known to be reproductively compatible with the *C. jacea* s.l. species complex, with the exception of *C. nigrescens* which appears closely related and suspected to hybridize (Keil and Ochsmann 2006), and also rare observations of a sterile hybrid between *C. × moncktonii* and yellow starthistle, *C. solstitialis* (Susanna and Roché 2011).

*Centaurea jacea*, *C. nigra* and *C. × moncktonii* hybrids share various morphological and life history traits. All taxa are short-lived perennials flowering typically from July through October. They are insect-pollinated and self-incompatible (Hardy et al. 2001), reproducing primarily by seed. Whereas the taxa are hardly distinguishable based on vegetative traits (Keil and Ochsmann 2006), various caputula traits have proven diagnostic for members of the species complex (Hardy et al. 2000; Vanderhoeven et al. 2002; Keil and Ochsmann 2006) and are summarized for the parental species in Supporting Information—Table S1.3. North American meadow knapweed hybrid populations are quite variable morphologically and approach either parental species in appearance. In North America, *C. nigra* and *C. jacea* are commonly found in fields, pastures, roadsides, disturbed areas, clearings and waste places, although observations of *C. nigra* encompass a lower altitudinal range (0–300 m) compared to *C. jacea* (50–1300 m) (Keil and Ochsmann 2006). The hybrid *C. × moncktonii* prefers moister habitats than its parental species, such as meadows, irrigated pastures, riparian zones and moist forest openings (Roché and Roché 1991; Miller and Lucero 2014). The presence of the hybrid is undesirable because of its persistence in native plant communities and it also reduces the quality of forage produced in pastures and grass hayfields (Roché and Coombs 2003).

Hardy (2000) reported that *C. jacea* and *C. nigra* (studied collectively as *C. jacea* s.l. in that paper) occur as two distinct cytotypes in Europe: a diploid (2n = 22) and a tetraploid (2n = 44) (Hardy et al. 2000). Hybridization within cytotypes is relatively common, but hybridization between different cytotypes is rare (Hardy et al. 2000; 2001; Kouteky 2008). Therefore, hybridization appears to be homoploid (i.e. 2n × 2n or 4n × 4n) and produces offspring with the same ploidy level as their parents. Segregation of allozyme loci in controlled crosses between tetraploid genotypes suggests inheritance patterns vary among loci from strict tetrasomic to intermediate between tetrasomic and disomic (Hardy et al. 2001; Stift et al. 2008). Thus, it is likely the 4n cytotype is in the process of diploidization of the homeologous parental chromosomes. The tetraploids also appear more phenotypically variable and have a wider ecological amplitude and geographical distribution than diploids (Gardou 1972).

**Sampling**

In 2015, 20 populations (see Fig. 1; Table 1) of the *C. jacea/nigra* complex were sampled from each of 10 locations in New York and Vermont. For each population, a whole young leaf (100 mg) and 1–2 mature capitula were collected from each of a minimum of 20 individuals. Leaf samples were transported back to the lab, and stored frozen at ~80 °C until DNA extraction. Capitula were stored dried in envelopes in the lab until morphological analysis.

Voucher specimens from each population were collected and have been submitted to the Liberty Hyde Bailey Hortorium at Cornell University (herbarium@cornell.edu) [see Supporting Information—Table S1.1].

**SNP genotyping**

We extracted total genomic DNA from leaf tissue using the DNeasy 96 Plant kit (Qiagen, Valencia, CA), quantified each sample fluorometrically using QUANT-IT (Invitrogen) and normalized concentrations to 10 ng μL⁻¹. Normalized DNA was used to prepare 96-plex genotype-by-sequencing (GBS) libraries following the Elshire protocol (Elshire et al. 2011). We individually digested 100 ng of sample DNA with the methylation-sensitive restriction enzyme ApeK1 (New England Bio Labs) for 16 h at 75 °C, followed by ligation with 1.2 ng barcoded adapter sequences (adapter sequences followed Elshire et al. 2011) for 1 h at 22 °C and then heat killed at 65 °C for 30 min. Following ligation, 5 μL of each barcoded sample were pooled together (96 samples total/library), cleaned using the Qiagen PCR cleanup kit
Table 1. Sampling locations of the *C. jacea/nigra* species complex in New York State and Vermont, USA. N-GBS, sample size for genomic analyses (after removal of samples with >70 % missing genotypes); N-GS, sample size genome size data (number of seed families); N-morpho, sample size morphometric data (number of individuals); Taxon, taxon to which the majority of the population was assigned in Admixture (Alexander et al. 2009) analysis for $K = 2$; Gen. size, mean holoploid genome size (1C value, pg).

| Population ID | Locality                                      | State | Latitude (°N) | Longitude (°W) | N-GBS | N-GS | N-morpho | Taxon       | Gen. size (1C, pg) |
|---------------|-----------------------------------------------|-------|---------------|----------------|-------|------|----------|-------------|-------------------|
| CB            | Central Bridge, Intersection of Route 7 and 30A | NY    | 42.71138      | 74.31467       | 17    | 6    | 17       | Hybrid      | 1.931             |
| CO            | Cobleskill, County Route 1                    | NY    | 42.67663      | 74.23293       | 20    | 10   | 10       | Hybrid      | 1.902             |
| FL            | Finger Lakes National Forest                  | NY    | 42.52917      | 76.7743        | 15    | 7    | 7        | Hybrid      | 1.902             |
| FP            | Fort Plain, Brockman’s Corners Road           | NY    | 42.90015      | 74.7652        | 8     | 5    | 5        | Hybrid      | 1.895             |
| JF            | Dryden, Gee Hill Road                         | NY    | 42.49607      | 76.24740       | 4     | 2    | 2        | C. cf. nigra | 2.025             |
| JV            | Jacksonville, Route 96                        | NY    | 42.50268      | 76.6073        | 17    | 8    | 10       | Hybrid      | 1.891             |
| LH            | Middleburgh, Lawton Hollow Road               | NY    | 42.58310      | 74.27177       | 10    | 6    | 9        | Hybrid      | 1.905             |
| MC            | Mclean Bog                                    | NY    | 42.54553      | 76.26913       | 16    | 6    | 10       | Hybrid      | 1.926             |
| MP            | Ithaca, Mt. Pleasant                          | NY    | 42.46377      | 76.37228       | 17    | 8    | 6        | Hybrid      | 1.986             |
| WV            | Willsieville, Willsieville Road               | NY    | 42.29737      | 76.38383       | 10    | 6    | 5        | C. cf. nigra | 1.952             |
| CH            | Colchester, Chimney Hill                      | VT    | 44.59865      | 73.17804       | 13    | 6    | 3        | C. cf. jacea | 1.893             |
| DF            | Shoreham, Doolittle Farm                      | VT    | 43.93722      | 73.30215       | 13    | 4    | 3        | C. cf. jacea | 1.903             |
| FC            | Addison, Farr Cross Road                     | VT    | 44.30018      | 73.28369       | 16    | 7    | 3        | C. cf. jacea | 1.889             |
| HV            | Milton, Happy Valley Road                     | VT    | 44.66431      | 73.19869       | 15    | 8    | 5        | C. cf. jacea | 1.891             |
| PR            | Jericho, Packard Road                         | VT    | 44.50658      | 72.97207       | 14    | 7    | 5        | C. cf. jacea | 1.896             |
| RM            | Richmond, Park and Ride                      | VT    | 44.43262      | 73.00874       | 13    | 6    | 3        | C. cf. jacea | 1.889             |
| SH            | South Hero, Ferry Road                       | VT    | 44.67213      | 73.33817       | 15    | 7    | 3        | C. cf. jacea | 1.890             |
| SP            | South Burlington, Spear Street               | VT    | 44.36105      | 73.19783       | 11    | 4    | 10       | Hybrid      | 1.882             |
| TH            | Marshfield, Thistle Hill Road                | VT    | 44.37105      | 72.30404       | 14    | 7    | 9        | C. cf. nigra | 1.955             |
| WR            | South Burlington, Williston Road             | VT    | 44.46829      | 73.16777       | 15    | 7    | 8        | Hybrid      | 1.885             |
and eluted into 30 μL EB buffer. We used 6 μL of this cleaned, pooled library in a 50-μL PCR reaction (1× NEB Taq Master Mix; 25 pmol each forward and reverse primers) for 18 cycles, following conditions reported in Elshire et al. (2011). The resulting PCR library was cleaned with a GeneJet NGS cleanup kit (Thermo Fisher) using the entire 50 μL and run on a Agilent 2100 Bioanalyzer to verify the expected size distribution and lack of contaminating adapted sequences. Libraries passing QC were then sequenced (one 96-plex library per lane) for 100 bp single-end reads on an Illumina HiSeq 1500 by the Vermont Integrative Genomics Resource (VIGR) at the University of Vermont. In total, we sequenced four 96-plex libraries, encompassing 380 samples.

We assessed the quality of the raw sequence reads using FASTQC v0.10.1 (http://hannonlab.cshl.edu/fastx_toolkit). We then used the genotyping pipeline GBS-SNP-CROP (Melo et al. 2016) to clean and demultiplex reads, align them to the reference, call SNPs and genotypes, and apply filters to retain high-confidence variants for downstream analysis. Briefly, this consisted of the following steps: We cleaned raw reads using Trimmomatic 0.36 (Bolger et al. 2014) to remove bases leading and trailing bases with Q-score < 30 and truncate reads when the average Q-score in a 4-bp window fell below 30, followed by demultiplexing based on individual barcodes. Since no reference genome is available for C. jacea s.l. we assembled a GBS-specific reduced-representation mock reference de novo using the Velvet assembler (Zerbino and Birney 2008). The assembly was based on sequences from five individuals (CB_18, LH_3, MC_10, CH_15 and LH_18) that comprised a large number of reads and possessed caprula traits representative of the morphological variation found across our sample. We optimized assembly parameters using the wrapper program VelvetOptimiser (https://github.com/tseemann/VelvetOptimiser), and searched a range of k-mer lengths between 19 and 31 (command line: VelvetOptimiser.pl -s 19 -e 31 -f 'short -fastq CB_18.R1.fastq LH_3.R1.fastq MC_10.R1.fastq CH_15.R1.fastq LH_18.R1.fastq'). The optimal settings had a hash length of 29, which produced an assembly of 854 446 contigs with an N50 of 100 bp, for a total assembled length of 94.2 Mb. While larger kmer values could have been searched and may have yielded longer contigs in some cases, larger kmer values are also computationally more demanding to search, and the highly fragmented nature of single-end GBS sequence ‘tags’ is not expected to assemble into large contiguous sequences (Andrews et al. 2016). We also explored de novo assembly of a mock reference by clustering sequences based on percent identity using the VSEARCH algorithm, which does not depend on kmer length. The average cluster length with VSEARCH (82 bp; range 32–123 bp) was very close but slightly smaller than obtained with Velvet, indicating that assembly of the GBS tags was not overly sensitive to assembly algorithm. We used the Velvet assembly as a reference sequence to map reads from all samples using the bwa-mem algorithm in BWA 0.7.12 with default parameters (Li 2013).

We used the SAMTools 1.1 (Li et al. 2009) mpileup to generate an alignment of potential SNPs, using only uniquely mapped reads and applying the –C50 flag to down-weight mapping quality in reads with excessive mismatches. The mpileup output was parsed, putative indels excluded and the resulting genotyping master matrix of all putative SNP positions across all individuals was used for SNP and genotype calling.

We applied a series of quality control filters to eliminate low-confidence SNP and genotype calls. To reduce PCR bias or sequencing and alignment artefacts, we filtered for: (i) minimum and maximum average depth per genotype of 3 and 200 reads, respectively, (ii) a minimum frequency of 70 % of individuals with genotype data to accept a SNP and (iii) a minimum of three individuals possessing the putative secondary allele at a minimum depth of three reads. We applied the additional constraints of: (iv) minimum depth of three reads required to call homozygotes when the secondary allele depth was 0, (v) minimum primary read depth of 3 required to call homozygotes when the secondary allele depth was 1 and (vi) minimum read depth of 3 required for each allele (primary and secondary) to call heterozygotes. To eliminate SNPs that were not strongly bi-allelic, we required (vii) a minimum of 90 % of non-primary allele reads to be the secondary allele. This filter was specifically applied to reduce multi-allelic SNPs and enrich for loci exhibiting diosomic inheritance. Lastly, we (viii) eliminated SNPs in contigs that likely represent microbial contamination by running nucleotide searches of the mock reference against the NCBI nr database with BLASTn (Altschul et al. 1990) and excluded SNPs located in contigs that matched non-plant genomes.

After filtering, the genomic data set consisted of 25 999 polymorphic SNPs for 380 samples. To reduce the impact of missingness on downstream analyses, we imputed missing genotypes using the LD-KNNi algorithm as implemented in LinkimputeR (Money et al. 2017). We performed test runs to optimize imputation accuracy using read depth thresholds between 3 and 12 and missingness (individual and site) thresholds between 0.4 and 0.7 in increments of 0.1. For these test runs, we masked 5000 genotypes with a site read depth >20 assumed to represent accurately called genotypes, and estimated imputation accuracy as the proportion of masked genotypes that were imputed correctly. For all test runs, we filtered for a maximum combined read depth of 100 and a minor allele frequency of 0.05. Based on the test runs, we then imputed genotypes that had a combined read depth >4 after first excluding SNPs and then individuals with >70 % missing values. These settings lead to a high imputation accuracy of 0.9 in test runs, while retaining a reasonable number of individuals (273) and SNPs (17 035) in the data set. We then used bcftools (http://samtools.github.io/bcftools/) to convert genotypes with an imputed probability <0.9 to missing, followed by removing SNPs with >50 % missing values. Lastly, to eliminate loci departing from disomic, Mendelian segregation due to formation of tetravalents or gene paralogy, we filtered SNPs with significant heterozygote excess (P < 5 e-4 based on Bonferroni correction). Positions with heterozygote deficit were not removed since this is an expected phenomenon in a hybrid zone due to a Wahlund effect. The final data set contained SNP genotype data for 273 individuals at 10 348 polymorphic loci.

**Population structure, genomic admixture and hybrid classification**

To identify signals of genetic ancestry and assess the degree of admixture in our sample, we performed maximum likelihood genotypic clustering using the program ADMIXTURE v1.3 (Alexander et al. 2009) to probabilistically assign individuals ancestry in one or more K genetic clusters. We tested different models of K from 2 to 20, and determined values of K most consistent with the data based on minimizing the 5-fold cross-validation error. To complement the ADMIXTURE analyses with a model-free clustering approach that does not make any genetic assumptions (i.e. Hardy–Weinberg, linkage equilibrium, disomic or tetrasomic inheritance) (Ortego et al. 2017), we performed a principal component analysis (PCA) using the R package adegenet (Jombart and Ahmed 2011).

Assignments of genotypic ancestry can provide evidence for admixture in a sample, but cannot distinguish between different
hybrid genotype classes (e.g. F1, F2, backcrosses, etc.). We used the program NewHydrïds (Anderson and Thompson 2002) to explore whether admixed individuals were first generation (F1) hybrids or instead represented more complex scenarios of introgression arising from advanced generation inter-mating (F2+) or backcrossing. We used the R package hybriddetective (Wringe et al. 2017b) to assemble and test diagnostic marker panels and create simulated data sets of parental taxa and four hybrid genotype classes. To this end, we assigned 27 samples to each of the two putative parental species based on ancestry (Q) values > 0.9 in the K = 2 ADMIXTURE run (Vähä and Primmer 2006). Individuals in these clusters morphologically resembled C. jacea and C. nigra to a large degree, and were thus labelled ad hoc as C. cf. jacea and C. cf. nigra, respectively (see also next section). We emphasize the use of ‘cf.’ here, from the Latin ‘confer’ or ‘compare to’, as an indication of the resemblance of our sample to the above referenced taxa without asserting a formal taxonomic assignment. These data were further subdivided to use half the samples per parental taxon for marker selection and hybrid simulation, respectively.

NewHydrïds assumes loci are independent (i.e. unlinked) and show strong allele frequency differences between parental taxa. To select an optimal subset of loci, we used the getTopLoc function to select diagnostic markers with low linkage disequilibrium within a parental taxon (assessed with FLINK; Purcell et al. 2007) and high F’ (assessed with hierfstat; Goudet 2005) between parental taxa. To account for choice of parental taxon when developing diagnostic marker panels, we performed marker panel selection separately for both the C. cf. jacea subsample and the C. cf. nigra subsample. In application, we found that these two diagnostic panels returned slightly different estimates for some hybrid classes (see Results), probably reflecting unknown differences in genomic structure and recombination in each lineage. Rather than combine the two sets of markers into a single panel, we instead report hybrid assignments from each panel separately so that sensitivity and agreement in the hybrid assignments, we measured genome size by flow cytometry. Genome size (1C values) reported in the literature for tetraploid C. jacea is generally ~2 pg, with reported variation among closely related species as well as within cytotypes of C. jacea (Bancheva and Greilhuber 2005; Dydak et al. 2009). Thus, correspondence between genetic ancestry and genome size provides additional confidence in parental and hybrid taxon assignments that may be less subjective and environmentally plastic compared to morphometric traits. We sampled fresh leaf tissue for genome size from three open-pollinated offspring per field-collected individual. Seeds were germinated, grown in a greenhouse at the University of Vermont, and leaf tissue sampled when plants were large rosettes. Tissue was kept cool but not frozen and shipped to the Laboratory of Flow Cytometry at the Institute of Botany, Czech Academy of Sciences. Genome size was determined following the protocol described in (Doležel et al. 2007), using Solanum lycopersicum ‘Stupicke polni rane’ (1C = 0.98 pg) as an internal reference standard loaded with each sample.

We analysed differences in mean genome size and morphometric traits as a function of genetic ancestry and hybrid class using linear mixed models, with population as a random effect. When the main effect of hybrid class was significant, we performed pair-wise comparisons of hybrid class using linear mixed models, with population as a random effect. When the main effect of hybrid class was significant, we performed Tukey’s post hoc tests for pair-wise comparisons of hybrid classes using the R package ‘multcomp’ (Hothorn et al. 2008). All analyses performed in R used R version 3.4.1 (R Core Team 2017).

Results

Population structure, genomic admixture and hybrid classification

The composition of genetic ancestry in our sample estimated by ADMIXTURE showed two relatively pure clusters at K = 2. These clusters separated individuals visually resembling C. nigra and C. jacea based on a subset of diagnostic capitula traits (treated tentatively here as C. cf. nigra and C. cf. jacea; see also section ‘Morphometric trait variation among the identified genetic clusters and hybrid classes’) as well as a large number of individuals with admixed genetic ancestry (Fig. 2A, upper panel, red for C. cf. nigra and blue for C. cf. jacea). Geographic
populations that were assigned by majority to one of the two unadmixed clusters usually also contained some individuals with mixed ancestry (i.e. maximum assignment probability < 0.9 in the K = 2 run).

At K = 3, the model best supported by cross-validation in ADMIXTURE, most admixed individuals formed a separate cluster (orange), suggesting the presence of a distinct hybrid lineage with intermediate ancestry between C. cf. nigra and C. cf. jacea (Fig. 2A, lower panel). Of the 102 individuals that showed a majority assignment to the third hybrid cluster, 99 were collected from seven populations in New York that had majority assignment to C. cf. nigra (NY hybrids), with one additional individual in NY population MP (Figs 1 and 2A) and two individuals from Vermont populations HV and SH. Two additional admixed populations were from Vermont (SP, WR; VT hybrids) harbouring 26 individuals whose ancestries included all three clusters, but with highest assignment probability for C. cf. nigra (Figs 1 and 2A lower panel; Table 1).

Inference of each individual’s hybrid genotype class by NEWHYBRIDS (Fig. 2B) largely reflected the ancestry scores from ADMIXTURE; however, the hybrid class assignments slightly differed depending on whether the marker panel was based on C. cf. jacea or C. cf. nigra (see also Supporting Information—Fig. S1.2), with the assignments of backcross hybrid classes showing a trend towards the parent taxon that served as the source of marker selection. In both cases, several individuals from
populations containing a majority of *C. cf. nigra* or *C. cf. jacea* were classified as hybrids, congruent with the *ADMITTURE* results. Most individuals assigned to the orange hybrid cluster in the *ADMITTURE* analysis were classified as F2 (or higher generation) hybrids based on the *C. cf. nigra* panel, or as second (or higher) generation backcrosses towards *C. cf. jacea* based on the *C. cf. jacea* panel (Fig. 2B). By contrast, the VT hybrids assigned to all three clusters in the *ADMITTURE* analysis were mostly classified as F2 (or higher) hybrids in both *NEWHYBRIDS* analyses. No F1 hybrids and few first generation backcrosses were identified, suggesting most hybrids in our sample represent advanced generations of introgression.

The PCA based on all 10,348 SNPs identified three genomic clusters that closely corresponded with the *ADMITTURE* and *NEWHYBRIDS* results based on the *C. cf. nigra* marker panel (Fig. 3A and B). The first two clusters consisted of parental *C. cf. jacea* and *C. cf. nigra* individuals, respectively, separated along principal component (PC) 1, while individuals with hybrid ancestry were intermediate (Fig. 3C). However, hybrid genotypes separated out into different positions along PC2—one group of hybrids composed primarily of F2 and BC1-jacea genotype classes showed high PC2 scores, corresponding to the orange *ADMITTURE* cluster. The remaining hybrid individuals from genotype classes BC2-jacea, BC1-nigra, BC2-nigra had lower PC2 scores and clustered more closely with their respective parental taxa (Fig. 3A, B and D).

**Morphometric trait variation among the identified genetic clusters and hybrid classes**

Due to better correspondence of hybrid genotype classes with genomic ancestry (*ADMITTURE*, PCA) and capitula morphology

---

**Figure 3.** Genetic structure of 273 individuals of the *C. jacea/nigra* species complex sampled in New York State and Vermont based on 10,348 SNPs. (A, B) Principal component analysis with individuals labelled by source population and assigned hybrid genotype class. (C) Relationship between the *ADMITTURE* assignment probability to the *C. cf. nigra* (red) cluster for *K* = 2 (Fig. 2A, upper panel) and the scores of genetic principal component 1. (D) Relationship between the *ADMITTURE* assignment probability to the *C. × moncktonii* hybrid cluster (grey) for *K* = 3 (Fig. 2A, lower panel) and the scores of genetic principal component 2. Hybrid classes were derived from the *NEWHYBRIDS* analysis using the *C. cf. nigra* marker panel (Fig. 2B, upper panel).
(see below), we hereafter focus on the hybrid class assignments made using the C. cf. nigra marker panel. The corresponding results based on the C. cf. jacea marker panel are presented in Supporting Information—Appendix S2. Morphometric capitula traits showed distinct variation between but also high variation within the two putative parental taxa inferred based on unadmixed genomic ancestry [see Supporting Information—Table S1.3; Fig. S1.1]. However, three traits differed significantly (and in accordance with expectations) between the two presumed parental taxa: the percentage of pectinate bract rows was significantly higher in C. cf. nigra than C. cf. jacea ($\chi^2_{(1)} = 30.7, P < 0.001$), whereas the ratio of bract appendage centre width to length ($\chi^2_{(1)} = 6.2, P < 0.05$) and the relative bract appendage centre width ($\chi^2_{(1)} = 8.7, P < 0.01$) were both significantly lower in C. cf. nigra than C. cf. jacea.

Consequently, the C. cf. jacea and C. cf. nigra individuals separated remarkably well in a PCA involving these three traits (Fig. 4). The hybrids, in particular the F2, exhibited transgressive segregation, exceeding parental trait values on the first two morphometric PC axes, which collectively explained 92% of the trait variance. When comparing morphometric PC scores of the hybrid classes (Fig. 5A and B), we found significant differentiation between C. cf. jacea and C. cf. nigra individuals along morphometric PC1 ($\chi^2_{(1)} = 13.1, P < 0.05$ for hybrid class main effect), whereas all classes of hybrids did not differ significantly from either of the presumed parental taxa in morphometric PC1 scores (Fig. 5A). For morphometric PC2, we did not find significant differences between the presumed parental taxa and hybrids except for significantly higher scores in C. cf. jacea than first generation backcrosses to C. cf. nigra ($\chi^2_{(1)} = 14.8, P < 0.05$ for hybrid class main effect), but this difference should be interpreted with caution since the sample size in the latter hybrid class was quite small (Fig. 5B). Morphometric differentiation between hybrid classes was less pronounced when they were assigned based on the C. cf. jacea marker panel [see Supporting Information—Fig. S2.3A and B].

The association between capitula morphology and genomic ancestry was further strengthened when regressing the morphometric PC scores against the genetic PC scores (Fig. 5C and D). Genomic ancestry (genetic PC1) significantly predicted variation in capitula morphology (morphometric PC1 $\chi^2_{(1)} = 10.6, P < 0.01$; PC2 $\chi^2_{(1)} = 5.5, P < 0.05$). These relationships were primarily driven by the clear morphological separation of genotypes classified as unadmixed parental taxa. By contrast, admixed hybrid individuals with intermediate values of genetic PC1 displayed highly variable capitula morphologies, but with a

![Figure 4](image-url)

Figure 4. Principal component analysis based on three morphometric capitula traits [see Supporting Information—Table S1.3; Fig. S1.1] derived from diagnostic characteristics from previous morphometric analyses on European samples of the C. jacea/nigra complex (Hardy et al. 2000; Vanderhoeven et al. 2002). The traits were measured for 133 individuals using standardized digital images of one or two field-collected capitula per individual. Coloured labels correspond to hybrid classes derived from the NewHybrids analysis using the C. cf. nigra marker panel (Fig. 2B, upper panel).
mean centred near the predicted value based on their ancestry (Fig. 5C and D). When categorizing hybrid classes based on the C. cf. jacea marker panel, the majority of this morphologically highly variable group of individuals was assigned to either C. cf. jacea or backcrosses to this parental taxon [see Supporting Information—Fig. S2.4A and C]. We did not find any significant relationships between genetic PC2 and either morphometric PC, which suggests that genomic differences that separate hybrid genotypes along genetic PC2 (Fig. 3B) are not reflected in capitula morphology [see Supporting Information—Figs S1.3A and B; S2.4B and D].

Genome size variation among the identified genetic clusters and hybrid classes

Holoploid (1C) genome size of our samples ranged between 1.83–2.28 pg, consistent with all sampled individuals being tetraploid (Bancheva and Greilhuber 2005; Dydak et al. 2009). Sample variation within individuals was low (CV_jaccea = 3.19 %; CV_black = 2.75 %), indicating we achieved good precision for estimating individual genome size. Despite the uniformity in DNA ploidy, we found considerable genome size differences that reflected genomic ancestry. Among NewHybrids genotype classes, C. cf. jacea as well as all types of backcrosses towards this parental taxon and F2 hybrids had significantly smaller genomes than C. cf. nigra (Fig. 6A; χ²(1) = 18.3, P < 0.01 for hybrid class main effect). The intermediate genome sizes of backcrosses to C. cf. nigra did not differ significantly from any other hybrid class. Interestingly, advanced hybrid generations (F2 and second generation backcrosses) consistently tended towards smaller genome sizes as compared to first generation backcrosses. Similar trends for hybrid class assignments were observed based on the C. cf. jacea marker panel [see Supporting Information—Fig. S2.5A].

Figure 5. Associations between genetic ancestry and reproductive morphology. Mean individual scores for principal component 1 (A, C) and 2 (B, D) from the morphometric PCA (Fig. 4) predicted by NewHybrids hybrid class (A, B) or individual scores for genetic PC1 (C) and PC2 (D). Hybrid classes that do not share a lower-case letter differ significantly (P < 0.05) according to Tukey’s post hoc tests.
In accordance with these results, we found a significant relationship between genetic PC1 and genome size (Fig. 6B; see Supporting Information—Figs S1.5B and S2.5C). However, this effect was no longer significant when an extreme outlier (C. cf. nigra with a 1C value > 2.15) was excluded. Again, this relationship was mostly driven by genome size differences between the two parental taxa, whereas in the middle range of genetic PC1, genome size was highly variable. There were no significant relationships between genetic PC2 and genome size (see Supporting Information—Figs S1.4 and S2.5C), or between morphometric PC scores and genome size.

**Discussion**

Interspecific hybridization is common in plant evolution and is an important source of novel trait architectures and adaptive variation (Rieseberg et al. 2003), yet most species are sufficiently reproductively isolated to prevent total dissolution of species boundaries. When weakly diverged taxa come into secondary contact, lack of strong reproductive barriers makes the parental genomes permeable to introgression, leading to extensive backcrossing and the formation of hybrid swarms (Harrison and Larson 2014). This process of genomic admixture and introgression may increase trait variance through the release of latent genetic variation, but also blurs the lines that keep species distinct. Such appears to be the case for the hybrid swarm formed between C. jacea, C. nigra and their hybrid, C. × moncktonii. This highly diverse, polymorphic species complex has been suggested to hybridize throughout its native European range, as well as in North America where it is introduced. However, much confusion exists about the separation of C. × moncktonii from its parental taxa based on morphology (Vonica and Cantor 2011), suggesting a population genomics approach could prove useful at shedding light on the composition of this hybrid swarm (Vanderhoeven et al. 2002).

Our results provide the first glimpse into the composition of genomic ancestry in the Centaurea × moncktonii hybrid species complex, focusing on introduced populations in the northeastern USA where hybrid meadow knapweed appears invasive. By associating genome-wide ancestry with variation in a subset of capitula traits and genome size, we obtained five key results: (i) both advanced generation hybrids as well as unadmixed parental types were found to occur in a geographically structured distribution across New York and Vermont; (ii) capitula morphology and genome size correspond significantly with genomic ancestry, reinforcing the phenotypic and genomic distinctiveness of the unadmixed lineages while also providing a detailed genetic view of the extent of their introgression; (iii) hybrid individuals showed great variation in floral traits and genome size, including transgressive segregation for capitula traits; (iv) many hybrids exhibited a C. cf. nigra-like capitula morphology although the majority of them had higher genomic ancestry from the C. cf. jacea parent; and (v) many advanced generation hybrids had relatively small genome sizes given their level of ancestry. We discuss these findings in detail below with an emphasis on how the combination of genomic, cytological and morphological data have provided our best view to date into the ancestry composition of this incredibly reticulate hybrid complex. The extensive introgression uncovered here offers opportunities for future research to shed light on how hybridization affects the ecological and evolutionary dynamics of invasive populations.

**Evidence of occurrence and geographic distribution of unadmixed parental-like taxa**

In contrast to regions of the Pacific Northwest where C. × moncktonii hybrids have reportedly largely replaced populations of the parental taxa (Roché and Roché 1991), our results for the northeastern USA suggest that unadmixed genotypes still remain in the vicinity of an extensive array of advanced generation admixed genotypes (Figs 1 and 2A). Whether these taxa correspond precisely to the recognized taxonomy described...
for European C. jacea and C. nigra is uncertain at present, and will require the inclusion of native European samples in future analyses. Our preliminary classification of unadmixed individuals as C. cf. jacea and C. cf. nigra is based on congruent evidence from genomic, cytological and morphometric analyses, which all agree on the existence of genetically distinct lineages that correspond in most (but not all; see below) aspects of reproductive morphology to C. jacea and C. nigra. A third related taxon, C. nigrescens is also introduced to North America and is sometimes treated as part of the C. jacea s.l. complex, and C. × moncktonii is noted to sometimes combine features of C. nigrescens, notably the presence of involucres with distinct green phyllaries not fully covered by black appendages (Keil and Ochsmann 2006). Individuals possessing these traits were present but uncommon in our study area. However, another C. nigrescens-like trait was not clearly evident—the involucre being longer than wide (Keil and Ochsmann 2006). Thus, we cannot exclude the possibility that C. nigrescens has contributed some ancestry to this already complex hybrid swarm, but we think the dominant sources of ancestry in the hybrids we sampled originate from C. jacea and C. nigra. To our knowledge, no other reproductively compatible Centaurea species occur in our region.

The most common morphological characteristics used for taxonomic classifications of the genus Centaurea including hybrids are floral bract traits (Vanderhoeven et al. 2002; Blair and Hufbauer 2010; Veronica and Cantor 2013). Despite the low intraspecific differentiation and high intraspecific polymorphism within this species complex (Hardy et al. 2009), capitula morphology differed significantly between individuals we genetically identified as putative parental taxa for three traits that characterized the colour and shape of the bract appendages [see Supporting Information—Fig. S1.1] as well as for the first principal component based on these traits. Moreover, holoploid genome size, which had previously been used as a taxonomically diagnostic trait to distinguish taxa with weak morphological differentiation (Mahelka et al. 2005; Ekrt et al. 2010), differed significantly between unadmixed C. cf. jacea and C. cf. nigra samples (Fig. 6A). However, we were not able to measure a key trait used previously to distinguish C. nigra from C. jacea—the absence of sterile ray florets (Vanderhoeven et al. 2002), and visual observations and vouchered specimens taken at the time of collection indicate that all populations in our New York samples possessed ray florets, suggesting they may in fact be products of advanced introgression rather than ‘pure’ C. nigra (L. Milbrath, pers. obs.). Therefore, the presence of ray-like florets at sites assigned to C. cf. nigra indicates that the parent species as described in various treatments is not present (or deviates from the current understanding), given that the lack of ray-like florets is considered one diagnostic trait for C. nigra (Vanderhoeven et al. 2002; Keil and Ochsmann 2006; Veronica and Cantor 2011).

Comparing individual assignments between the K = 2 and the most supported K = 3 ADMIXTURE runs (Fig. 2A), we found that two of the three identified clusters (red and blue clusters in Fig. 2A) comprised high numbers of relatively unadmixed individuals that were assigned to these clusters with >0.9 in both runs, whereas individuals assigned to the third cluster seemed to be of mixed ancestry (i.e. putative hybrids). These findings were supported by the genetic PCA analysis, which clearly separated the presumed parental taxa along PC1 with admixed individuals showing intermediate values (Fig. 3A–C). Most of the unadmixed individuals identified as genetically distinct in ADMIXTURE and PCA were also classified as parental types by NEWHYBRIDS (Fig. 2B), along with a preponderance of advanced generation (F2 and backcross) hybrids evident. This provides the best evidence to date for extensive introgression between C. jacea and C. nigra, but also suggests the persistence of relatively unadmixed genotypes that are morphologically similar to the parental taxa (at least C. jacea).

Unadmixed C. cf. jacea occurred exclusively in Vermont across seven populations, whereas populations with a predominantly unadmixed C. cf. nigra genetic background were more scarce and found in only one population in Vermont and three populations in New York (Figs 1 and 2). These results further reinforce the conclusion that the hybrid swarm of C. × moncktonii has retained some (though perhaps not all) genomic and morphological features of its presumed parental lineages, and that these can be confidently separated using a combination of genomic, morphological and cytological evidence.

Genomic ancestry and geographic distribution of Centaurea × moncktonii populations

Previous studies have observed the presence of apparent C. × moncktonii hybrids in the Niagara region of New York (Eckel 2012), but we are unaware of any reports in the floristic literature on the origins of invasive populations in eastern North America—i.e. if they were introduced directly from Europe or are the result of a secondary introduction from the Pacific Northwest. Our study shows the presence of two types of C. × moncktonii hybrid populations in our study region, which is supported by both ADMIXTURE and genetic PCA analyses. One type of hybrid that is frequent in our New York samples corresponds to the ADMIXTURE lineage that emerges at K = 3 (orange ancestry in Figs 1 and 2A) and is assigned F2 or BC-jacea status in the NewHybrids analysis (Fig. 2B). This hybrid type also shared a higher amount of genomic ancestry with C. cf. jacea than C. cf. nigra based on the K = 2 results (Fig. 2A, upper). Both the closer relationship with C. cf. jacea, and also clear differentiation from the two presumed parental taxa are supported by the genetic PCA analysis (Fig. 3A, C and D). By contrast, the less abundant Vermont hybrid type is geographically restricted to just two of our sampled populations (SP and WR), and individual typically had ancestry assignments to all three clusters in the ADMIXTURE K = 3 analysis, with higher assignment to C. cf. nigra than to C. cf. jacea for K = 2 (Fig. 2A).

Whereas most of the Vermont hybrids are classified as advanced generation hybrids by NEWHYBRIDS based on both marker panels, assignment of the New York hybrids depended on the marker panel used. Based on the C. cf. nigra marker panel, New York hybrids were classified as predominantly F2’s or backcrosses to C. cf. jacea (Fig. 2B, upper), which corresponds well with the capitula morphology as well as with the ADMIXTURE results (K = 2). In contrast, hybrid class assignment based on the C. cf. jacea marker panel assigned most individuals as first or second generation backcrosses to C. cf. jacea (Fig. 2B, lower). The sensitivity of hybrid class assignments to the choice of marker panel likely reflects differences in the structure of linkage disequilibrium within each parental species genome, and the degree to which different maker sets are affected by shared ancestral polymorphism between species. Because the C. cf. nigra marker panel gave hybrid class assignments that more closely reflected the ADMIXTURE ancestry proportions (K = 2), and also gave results for the capitula morphometrics that fit with the degree of separation of the parental taxa, we suggest the analyses based on the C. cf. nigra marker panel are more robust. However, regardless of marker panel, both results point to advanced hybrid generations (either F2 or backcrosses) being commonplace.

The fact that we found almost exclusively advanced generation hybrids and no F1’s suggests that hybrid populations
throughout our study region have persisted over multiple generations, and have diverged through either genetic drift or assortative mating. Our sampling in two regions of New York did not reveal any C. cf. jacea populations, although the hybrid individuals had higher genomic ancestry from this parental taxon than from C. cf. nigra. It is possible that C. cf. jacea has been completely introgressed in this region. Alternatively, New York hybrids may be descendants of a hybrid lineage formed in the more distant past (possibly prior to introduction) and now represents a stabilized hybrid gene pool. This latter scenario would seem to fit the genetic differentiation of this hybrid group along PCZ (Figs 2A and 3A, B and D), the very similar levels of admixed ancestry within and across these populations, and the trend towards smaller than expected genome sizes (Tamaki et al. 2017). Stable genomic composition in hybrids can be established within a few generations and thereafter be maintained over long periods, particularly if genomic rearrangements isolate the hybrid from mating with the parental taxa (Rieseberg 1997; Baack and Rieseberg 2007).

In contrast, all three identified hybrid populations in Vermont are located in the urban area of Burlington, which may have facilitated the more recent formation and persistence of hybrids through anthropogenic disturbance and/or gene flow through human infrastructure (Grabenstein and Taylor 2018). These populations may either have been formed through recent hybridization, as both parental taxa are still present in the region, or they may have been introduced via long distance dispersal, e.g. along traffic routes.

Morphological and genome size variation in relation to genomic ancestry in *Centaurea × moncktonii* hybrids

Broadly speaking, strong concordance between genomic ancestry and taxonomic characteristics of genome size and capitula morphology was apparent in the hybrids, as exemplified by the intermediate mean values of various hybrid classes (Figs 5A, B and 6A) and the significant relationships of genetic with morphometric principal components as well as with genome size (Figs 5C, D and 6B). However, the hybrids encompassed considerable variation in genome size (Fig. 5), and capitula morphology exhibited transgressive segregation, exceeding the parental trait values, with the greatest variation observed in F2 and higher generation hybrids (Figs 4 and 5). The occurrence of transgressive phenotypes is a common phenomenon in advanced generation hybrid taxa after recombination (Rieseberg et al. 2003; Bell and Travis 2005) and may contribute to their success in novel environments (Baack and Rieseberg 2007; Hovick and Whitney 2014). Unequal crossing over or illegitimate recombination and other genomic rearrangements may also generate novel genetic architectures that increase trait variance in hybrids (Baack and Rieseberg 2007; Soltis et al. 2015). We observed advanced generation hybrids, particularly the F2+ hybrids, exhibited variable but smaller genome sizes on average relative to mid-parent expectations, with some individuals exhibiting the smallest genome sizes in the sample (Fig. 6). This could indicate progressive genomic downsizing in these hybrids, but whether or not smaller holoploid genome size leads to a higher invasive potential in *Centaurea × moncktonii*, as in other taxa (e.g. Pandit et al. 2014), awaits future experimental validation.

Interestingly, most hybrids exhibited a C. cf. nigra-like capitula morphology (Figs 4 and 5) although the majority of them (i.e. the New York hybrids) had a higher genomic ancestry from C. cf. jacea (Fig. 2A and B lower panel, see also Supporting Information—Figs S2.1 and 2.2). Mismatches between ancestry and morphology are not uncommon (Tamaki et al. 2017). Within the genus *Centaurea*, capitula traits have been studied in hybrids of diffuse and spotted knapweed (Blair et al. 2012), and suggested to be controlled by a small number of genes. Further, dominance relationships among the alleles determining the floral traits may skew the phenotype more towards one parental type (Fishman et al. 2002). These results emphasize the limitations of morphology for hybrid identification, and reinforce the utility of pairing ancestry data with genome size when discriminating parental and hybrid taxa, as genome size is unaffected by dominance.

Conclusions and future perspectives

Invasive populations of *C. jacea/nigra* form a highly polymorphic species complex with previously unknown composition of genetic ancestry. By combining population genomics, genome size and morphometrics, our study revealed extensive introgression and a high abundance of advanced generation hybrids. However, in some locations relatively unadmixed populations of the putative parental taxa (C. cf. jacea and C. cf. nigra) persist. The potential for hybridization among co-invading reproductively compatible taxa is hypothesized to be a major mechanism for fuelling the establishment and spread of introduced plants (Hovick and Whitney 2014; Dlugosch et al. 2015). The extensive introgression we report here among these species of introduced *Centaurea* could increase their propensity for invasion if hybrids have higher short-term fitness (e.g. heterosis) or increased quantitative genetic variance (evolvability) compared to the parental types. This study provides a foundation for further work aimed at addressing the impact of admixture, introgression and changes in genome size on the fitness of invasive hybrids.

Increasing attention is being paid to the role of genomic and cytogenetic features, including genome size (Kubešová et al. 2010; Pandit et al. 2014; Suda et al. 2014; Schmidt et al. 2017) and genomic admixture following interspecific hybridization (Schierenbeck and Ellstrand 2009; Hovick and Whitney 2014) as potentially unifying traits promoting invasion success. The fact that species are transported globally with increasing frequency will lead to continued introduction of invasive hybrids to new regions as well as to the formation of novel hybrid taxa. Therefore, the hypothesized connection between hybridization, genome size variation and invasion success is likely to remain a critical issue for conservation of native species (Todesco et al. 2016) as well as fundamental to invasion and eco-evolutionary research (Hovick and Whitney 2014; Bock et al. 2015).

Data

All data and code necessary to reproduce these results are publically available from https://github.com/SusanneLachmuth/Centaurea-moncktonii-AoR-Plants.

Supporting Information

The following additional information is available in the online version of this article—

**Table S1.1.** Specimen voucher information for *Centaurea* samples analysed in this study. Vouchers have been archived at the Liberty Hyde Bailey Hortorium at Cornell University.

**Table S1.2.** Assignment criteria for *Centaurea* cf. jacea and C. cf. nigra and their hybrids using NEWHYBRIDS.

**Table S1.3.** List of measured and derived morphometric
capitula traits diagnostic for the Centaurea jacea/nigra complex.  
Figure S1.1. Boxplots of morphometric capitula traits (see also Supporting Information—Table S1.2) measured using standardized digital images of one or two capitula per individual collected in Centaurea jacea/nigra complex field populations in New York State and Vermont.

Figure S1.2. Comparison of the hybrid class assignments resulting from the NEWHYBRIDS analyses using each a set of 1000 selected single nucleotide polymorphism (SNP) loci that showed highest global Fst and no linkage disequilibrium among either 27 Centaurea cf. jacea individuals or among 27 C. cf. nigra individuals.

Figure S1.3. Relationships of genetic ancestry and capitula morphology.

Figure S1.4. Relationships of genetic ancestry and genome size.

Figure S2.1. Genetic structure of 273 Centaurea jacea/nigra species complex individuals sampled in New York State and Vermont based on 10 348 single nucleotide polymorphisms (SNPs).

Figure S2.2. Principal component analysis based on three morphometric capitula traits.

Figure S2.3. Boxplots for variation in capitula morphology of the different hybrid classes.

Figure S2.4. Relationships of genetic ancestry and capitula morphology.

Figure S2.5. Relationships of genetic ancestry and genome size.

Figure S2.6. Relationships of capitula morphology and genome size.

Sources of Funding

This work was supported by a United States Department of Agriculture Co-Operative agreement to J.M., S.R.K. and L.M. (S8-1907-4-032), and by the German Research Foundation (DFG) through a Research Fellowship grant to S.L. (grant LA 3434/3-1).

Contributions by the Authors

J.M., L.M., and S.R.K. planned and designed the study. J.M. and L.M. conducted the field sampling, J.S. performed the flow cytometry for genome size determination. J.M. and S.L. performed measurements of morphological traits. S.R.K. oversaw the GBS library construction and sequencing. S.L. performed the bioinformatics and statistical analyses under guidance from S.R.K. S.L. and S.R.K. interpreted the results, with contributions from J.M. and L.M. S.L. and S.R.K. drafted the manuscript, and S.L., S.R.K., J.M. and L.M. contributed to the submitted version and revisions.

Conflict of Interest

None declared.

Acknowledgements

We thank M. O’Neil, K. Palacio-Lopez and J. Biazzo for assistance in collecting field samples and preparing voucher specimens; E. Murray and S. Kogut for help with collection of morphology data; G. Seidler for preparation of Fig. 1; M. Hassett and C. Lynch for DNA extraction and building GBS libraries; J. Vitová, M. O’Neil, K. Palacio-Lopez, M. Sorrentino, B. Verrico, K. Fetter and V. Chhatre for assistance with sampling genome size; A. Melo for assistance with GBS-SNP-CROP; and D. Money for help with LinkImputeR.

Literature Cited

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655–1664.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403–410.

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

Andrews KR, Good JM, Miller MR, Luikart G, Hoehnlohe FA. 2016. Harnessing the power of RA Dseq for ecological and evolutionary genomics. Nature Reviews Genetics 17:81.

Baack EJ, Rieseberg LH. 2007. A genomic view of introgression and hybrid speciation. Current Opinion in Genetics & Development 17:513–518.

Bancheva S, Greilhuber J. 2005. Genome size in Bulgarian Centaurea s.l. (Asteraceae) Plant Systematics and Evolution 257:95–117.

Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48.

Bell MA, Travis MP. 2005. Hybridization, transgressive segregation, genetic covariation, and adaptive radiation. Trends in Ecology & Evolution 20:358–361.

Blair AC, Blumenthal D, Hufbauer RA. 2012. Hybridization and invasion: an experimental test with diffuse knapweed (Centaurea diffusa Lam.). Evolutionary Applications 5:17–28.

Blair AC, Hufbauer RA. 2010. Hybridization and invasion: one of North America’s most devastating invasive plants shows evidence for a history of interspecific hybridization. Evolutionary Applications 3:40–51.

Bock DG, Caseys C, Cousens RD, Hahn MA, Heredia SM, Hübner S, Turner KG, Whitney KD, Rieseberg LH. 2015. What we still don’t know about invasion genetics. Molecular Ecology 24:2277–2297.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Botanic Garden and Botanical Museum of Berlin-Dahlem. 2006. The Euro+Med Plantbase Project. http://ww2.bgbm.org/EuroPlusMed/PTaxonDetail.asp?NameCache=Centaurea+jacea (19 December 2017).

Burnat E, Briquet J-I, Cavillier FG. 1931. Flore des Alpes maritimes. Volume VII (Composites Cynaroides). Geneva, Switzerland: Biodiversity Heritage Library.

Dlugoš KM, Anderson SR, Braasch J, Cang FA, Gillette HD. 2015. The devil is in the details: genetic variation in introduced populations and its contributions to invasion. Molecular Ecology 24:2095–2111.

Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. Nature Protocols 2:2233–2244.

Dostal J. 1976. Centaureas L. In: Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA, eds. Flora Europaea. London, New York and Melbourne: Cambridge University Press, 254–301.

Dydał M, Kolano B, Nowak T, Siwińska D, Maluszynska J. 2009. Cytogenetic studies of three European species of Centaurea L. (Asteraceae). Hereditas 146:152–161.

Eckel PM. 2012. Meadow or protein knapweed in the Niagara frontier region. Claytonia 27:5–7.

Ekrt L, Holubová R, Trávnícek P, Suda J. 2010. Species boundaries and frequency of hybridization in the Dryopteris carthusiana (Dryopteridaceae) complex: a taxonomic puzzle resolved using genome size data. American Journal of Botany 97:1208–1219.

Ellstrand NC, Schierenbeck KA. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? Proceedings of the National Academy of Sciences 97:7043–7050.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379.

Fishman L, Kelly AJ, Willis JH. 2002. Minor quantitative trait loci underlie floral traits associated with mating system divergence in Mimulus. Evolution 56:2138–2155.

Garcia-Jacas N, Susanna A, Mozaffarian V, Iarslan R. 2000. The natural delimitation of Centaurea (Asteraceae: Cardueae): ITS sequence analysis of the Centaurea jacea group. Plant Systematics and Evolution 223:185–199.

Garcia-Jacas N, Uysal T, Romashchenko K, Suarez-Santiago VN, Erugrul K, Susanna A. 2006. Centaurea revisited: a molecular survey of the jacea group. Annals of Botany 98:741–753.

Gardou C. 1972. Recherches biosystématiques sur la Section Jacea Cass. et quelques sections voisines du genre Centaurea L. (Asteraceae) Plant Systematics and Evolution 6:311–472.

Goudet J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes 5:184–186.
Goulet BE, Roda F, Hopkins R. 2017. Hybridization in plants: old ideas, new techniques. *Plant Physiology* 173:65–78.

Grabenstein KC, Taylor SA. 2018. Breaking barriers: causes, consequences, and experimental utility of human-mediated hybridization. *Trends in Ecology & Evolution* 33:198–212.

Hardy OJ, de Loose M, Vekemans X, Meerts P. 2001. Allozyme segregation and inter-cytotype reproductive barriers in the polyploid complex *Centaurea jacea*. *Heredity* 87:136–145.

Hardy OJ, Vanderhoeven S, De Loose M, Meerts P. 2000. Ecological, morphological and allozyme differentiation between diploid and tetraploid knapweeds (*Centaurea jacea*) from a contact zone in the Belgian Ardennes. *Neu Phytologist* 146:281–290.

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

Hothorn T, Bretz F, Westfall P. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50:346–363.

Hovick SM, Whitney KD. 2014. Hybridization is associated with increased fecundity and size in invasive taxa: meta-analytic support for the hybridization-invasion hypothesis. *Ecology Letters* 17:1464–1477.

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

Keil DJ, Ochsmann J. 2006. *Centaurea*: In: *Flora of North America* Editorial Committee, eds. 1993+. *Flora of North America* North of Mexico. 19+ vols. New York and Oxford. Vol. 19, 20 and 21, pp. 52, 57, 58, 67, 83, 84, 96, 171, 172, 176, 177.

Keller SR, Fields PD, Berardi AE, Taylor DR. 2014. Recent admixture generates heterozygosity-fitness correlations during the range expansion of an invading species. *Journal of Evolutionary Biology* 27:616–627.

Keller SR, Taylor DR. 2008. History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecology Letters* 11:852–866.

Keller SR, Taylor DR. 2010. Genomic admixture increases fitness during a biological invasion. *Journal of Evolutionary Biology* 23:1720–31.

Koutecy, P. 2008. Hybridization between diploid and tetraploid taxa of *Centaurea* sect. Jace: morphological variation and role of unreduced gametes. *Oesterreichische Botanische Zeitschrift* 13:344–348.

Kubešová M, Moravcová L, Suda J, Jarolíšek V, Pyšek P. 2010. Naturalized plants have smaller genomes than their non-invading relatives: a flow cytometric analysis of the Czech alien flora. *Preslia* 82:81–96.

Lavergne S, Molofsky J. 2007. Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences of the United States of America* 104:3883–3888.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs guided genotype calling and imputation for non-model organisms. *Genome Research* 23:101–104.

Ortega J, Gugger PF, Sork VL. 2017. Impacts of human-induced environmental disturbances on hybridization between two ecologically differentiated Californian oak species. *New Phytologist* 213:942–955.

Pandit MK, White SM, Pooch MC. 2014. The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis. *New Phytologist* 203:697–703.

Parepa M, Fischer M, Krebs C, Bossdorf O. 2014. Hybridization increases invasive knotweed success. *Evolutionary Applications* 7:413–420.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PJ, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81:559–575.

R Core Team. 2017. R: a language and environment for statistical computing. Vienna, Austria: R Found. Stat. Comput.

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

Rieseberg LH, Widmer A, Arnitz AM, Burke JM. 2003. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 358:1141–1147.

Roché CT, Coombs E. 2003. Meadow knapweed (*Centaura x pratensis* Thunb.), Pacific Northwest Extension Publication PNW0566.

Roché CT, Roché BF, Jr. 1991. Meadow knapweed invasion in the Pacific Northwest, U.S.A. and British Columbia, Canada. *Northwest Science* 65:53–61.

Schierenbeck KA, Ellstrand NC. 2009. Hybridization and the evolution of invasiveness in plants and other organisms. *Biological Invasions* 11:1093.

Schmidt JP, Drake JM, Stephens P. 2017. Residence time, native range size, and genome size predict naturalization among angiosperms introduced to Australia. *Ecology and Evolution* 7:10289–10300.

Schneider CA, Rasband WS, Elcom SK. 2012. NIH Image to Image: 25 years of image analysis. *Nature Methods* 9:671.

Soltis PS, Marchant DB, de Peer Y, Soltis DE. 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics & Development* 35:119–125.

Stift M, Berenos C, Kuperus P, Van Tienderen PH. 2008. Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to *Rorippa* (yellow cress) microsatellite data. *Genetics* 179:2113–2123.

Suda J, Meyerson LA, Leitch JJ, Pyšek P. 2014. The hidden side of plant invasions: the role of genome size. *New Phytologist* 205:994–1007.

Susanna A, Roché C. 2011. Unlikely pairing of two alien species in Oregon produces unwanted offspring. *Kalmia* 18:24–29.

Tamaki I, Yoichi W, Matsuki Y, Suyama Y, Mizuno M. 2017. Inconsistency between morphological traits and ancestry of individuals in the hybrid zone between two *Rhododendron japonocophotamerum* varieties revealed by a genotyping-by-sequencing approach. *Tree Genetics & Genomes* 13:4.

Todesco M, Pascual MA, Owens GL, Ostevik KL, Meyers BT, Hübner S, Heredia SM, Hahn MA, Caseys C, Bock DG, Rieseberg LH. 2016. Hybridization and extinction. *Evolutionary Applications* 9:892–908.

Vahá J-P, Primiceri CR. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology* 15:63–72.

van Soest JL. 1947. *Centaurea* sect. *Jacea* en het voorkomen in Nederland. *Nederlands Kruiskundig Archief* 54:68–106.

Vanderhoeven S, Hardy O, Vekemans X, Lefebvre C, De Loose M, Lambinon J, Meerts P. 2002. A morphometric study of populations of the *Centaurea jacea* complex (*Asteraceae*) in Belgium. *Plant Biology* 4:403–412.

Vilà M, Weber E, Antonio CMD. 2000. Conservation implications of invasion by plant hybridization. *Biological Invasions* 2:207–217.

Vinata G, Cantor M. 2011. The polymorphism and hybridization of the *Centaurea* species. *Bulletin UASVM Horticulture* 68:444–450.

Wringe BF, Stanley RE, Jeffery NW, Anderson EC, Bradbury IR. 2017a. *parallelnewhybrid*: an R package for the parallelization of hybrid detection using newhybrids. *Molecular Ecology Resources* 17:91–95.

Wringe BF, Stanley RE, Jeffery NW, Anderson EC, Bradbury IR. 2017b. *hybridetective*: a workflow and package to facilitate the detection of hybridization using genomic data in *R*. *Molecular Ecology Resources* 17:275–284.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18:821–829.