Strengths and potential pitfalls of hay-transfer for ecological restoration revealed by RAD-seq analysis in floodplain *Arabis* species

Hannes Dittberner[1], Christian Becker[2], Wen-Biao Jiao[3], Korbinian Schneeberger[3], Norbert Hölzel[4], Aurélien Tellier[5], Juliette de Meaux*[1]

Affiliations: [1]Institute of Botany, University of Cologne, 50674 Cologne, Germany; Cologne Center for Genomics, University of Cologne, 50931 Cologne, Germany; [3]Max-Planck-Institute for Plant Breeding Research, 50829 Cologne, Germany; [4]Institute of Landscape Ecology, University of Münster, 48149 Münster, Germany;[5]Center of Life and Food Sciences Weihenstephan, Technical University of Munich, 85354, Freising, Germany

*Corresponding Author E-mail: jdemeaux@uni-koeln.de

Abstract

Achieving high intraspecific genetic diversity is a critical goal in ecological restoration as it increases the adaptive potential and long-term resilience of populations. Thus, we investigated genetic diversity within and between pristine sites in a fossil floodplain and compared it to sites restored by hay-transfer between 1997 and 2014. RAD-seq genotyping revealed that the stenoecious flood-plain species *Arabis nemorensis* is co-occurring with individuals that, based on ploidy, ITS-sequencing and morphology, probably belong to the close relative *Arabis sagittata*, which has a documented preference for dry calcareous grasslands but has not been reported in

floodplain meadows. We show that hay-transfer maintains genetic diversity for both species. Additionally, in *A. sagittata*, transfer from multiple genetically isolated pristine sites resulted in restored sites with increased diversity and admixed local genotypes. In *A. nemorensis,* transfer did not create novel admixture dynamics because genetic diversity between pristine sites was less differentiated. Thus, the effects of hay-transfer on genetic diversity also depend on the genetic makeup of the donor communities of each species, especially when material is mixed. Our results demonstrate the efficiency of hay-transfer for habitat restoration and emphasize the importance of pre-restoration characterization of micro-geographic patterns of intraspecific diversity of the community to guarantee that restoration practices reach their goal, i.e. maximize the adaptive potential of the entire restored plant community. Overseeing these patterns may differentially alter the adaptive potential of species present, thereby altering the balance between species in the community. Additionally, our comparison of summary statistics obtained from *de novo* and reference-based RAD-seq pipelines shows that the genomic impact of restoration can be reliably monitored in species lacking prior genomic knowledge.

Keywords

genetic diversity; RAD-seq; restoration genetics; hybridization; population structure; reference genome

Introduction

Habitat degradation is an ever growing problem in our modern world, causing unprecedented loss of biodiversity and essential ecosystem services (Baillie, Hilton-Taylor, & Stuart, 2004). Thus, there is a growing demand for ecological restoration, i.e. measures assisting the recovery of ecosystems that have been degraded, damaged or destroyed (Society for Ecological Restoration International Science & Policy Working Group, 2004). However, ecological restoration is a difficult process rarely leading to full ecosystem recovery (Benayas, Newton, Diaz, & Bullock, 2009). Thus, the young field of restoration ecology, which studies ecological processes in the light of restoration, is vital to improve restoration practices (Bullock, Aronson, Newton, Pywell, & Rey-Benayas, 2011; Roberts, Stone, & Sugden, 2009; Suding, 2011).

One of the main goals of ecological restoration is the recovery of biodiversity, including both species richness and intraspecific genetic diversity (henceforth called genetic diversity). Genetic diversity has generally positive effects on ecosystems (reviewed in Hughes, Inouye, Johnson, Underwood, & Vellend, 2008). For example, experimentally increasing the genetic diversity of *Solidago altissima* increased primary above-ground biomass productivity and arthropod diversity (Crutsinger et al., 2006). Genetic diversity also boosts resistance of populations to invasion and environmental fluctuations, presumably because it enhances the adaptive potential of populations (Reed & Frankham, 2003; Vrijenhoek, 1994). For example, high diversity experimental *Arabidopsis thaliana* populations showed higher resistance against invasion by *Senecio vulgaris* than low diversity populations (Scheepens, Rauschkolb, Ziegler, Schroth, & Bossdorf, 2017). Moreover, *Zoestra marina* populations with higher genetic diversity

showed increased biomass production, plant density and faunal abundance during an extremely warm period (Reusch, Ehlers, Hämmerli, & Worm, 2005). Concordantly, restored populations of *Z. marina* with increased genetic diversity showed longer plant survival, grew more rapidly and provided enhanced ecosystem services, measured by increased primary productivity, invertebrate density and nitrogen retention. This effect was stable in a range of environmental conditions along a water-depth gradient (Reynolds, McGlathery, & Waycott, 2012). These examples demonstrate the importance of genetic diversity for ecosystem function and stability and hence the need to consider population genetics in the design and evaluation of restoration practices, which is the focus of restoration genetics.

Restoration genetics can not only inform the planning of restoration efforts, e.g. by identifying suitable source populations, but also help evaluating the success of restoration projects, e.g. by monitoring genetic diversity in restored populations (Mijangos, Pacioni, Spencer, & Craig, 2015; Williams, Nevill, & Krauss, 2014). In fact, studies comparing the level of genetic diversity in pristine and restored populations frequently report limited success, with a reduction of genetic diversity in restored populations. This decline in genetic diversity may be caused by genetic bottlenecks in plant nurseries, biases introduced by seed harvesting strategies, founder effects during recolonization and/or unreliable commercial seeds (reviewed in Mijangos, Pacioni, Spencer, & Craig, 2015). By contrast, the transfer of seed-containing hay from pristine (donor) to restoration (donee) sites, termed hay-transfer, is expected to limit the loss of genetic diversity and maintain site-specific local adaption (Hufford & Mazer, 2003; Kiehl, Kirmer, & Shaw, 2014). In addition, this method has the unique feature that it can, theoretically, restore

81 an entire community without altering the genetic composition of populations and thus is the
82 best method available for restoring entire ecosystems (Hölzel & Otte, 2003; Kiehl, Kirmer,
83 Donath, Rasran, & Hölzel, 2010). So far, however, there is no empirical support for the efficiency
84 of this practice (Bucharova et al., 2017), especially since many species maintain seed banks in
85 the soil. Indeed, the genetic diversity specific to the seed bank will not be sampled with the hay,
86 although it is known that it can contribute significantly to the maintenance of diversity (Tellier,
87 Laurent, Lainer, Pavlidis, & Stephan, 2011).

88 The field of restoration genetics has witnessed a major technological shift over recent years.
89 Restoration genetics studies have initially relied on microsatellites and AFLP markers and thus
90 provided a limited overview on patterns of genetic variation within and between restored or
91 pristine populations (reviewed in Mijangos et al., 2015). Now, genotyping-by-sequencing (GBS)
92 methods are beginning to be more broadly adopted (Gruenthal et al., 2014; Massatti, Doherty,
93 & Wood, 2018; O'Leary, Hollenbeck, Vega, Gold, & Portnoy, 2018; Torres-Martinez & Emery,
94 2016). These methods drastically reduce sequencing costs through strategies to sequence a
95 reduced portion of the genome, e.g. Restriction site Associated DNA-sequencing (RAD-seq)
96 (Elshire et al., 2011; Etter, Bassham, Hohenlohe, Johnson, & Cresko, 2011; Peterson, Weber,
97 Kay, Fisher, & Hoekstra, 2012). In contrast to previous methods (AFLP, microsatellites), GBS
98 approaches sample proportions of the genome that are sufficiently large to allow resolving
99 patterns of genetic diversity and spatial structure even at very local scale where overall levels of
100 genetic diversity are low (Bradbury et al., 2015; Jeffries et al., 2016; Reitzel, Herrera, Layden,
101 Martindale, & Shank, 2013). In principle, GBS approaches have a third major advantage: they

102 are well suited to unravel genetic diversity in non-model species without prior genomic
103 information. Yet, the accuracy of genotyping in the absence of a reliable reference genome has
104 been questioned (Shafer et al., 2016). Since target species in restoration projects rarely coincide
105 with species or genera with advanced prior genomic knowledge, it is important to assess
106 whenever possible, whether conclusions from RAD-seq-based restoration genetics studies
107 depend on the availability of a reference genome.

108 Flood-plain meadows are species-rich ecosystems, accommodating many endangered and
109 stenoecious species adapted to the variable moisture regime. Due to increased agricultural
110 land-use and river regulation a large proportion of flood-plain ecosystems are degraded and
111 have become a target for ecological restoration. Here, we use RAD-seq to evaluate whether
112 genetic diversity is maintained in flood-plain meadows restored by hay-transfer in the Upper
113 Rhine valley in Germany (Hölzel & Otte, 2003, Donath, Bissels, Hölzel, & Otte, 2007). We
114 focused on *Arabis nemorensis*, a stenoecious species typically restricted to flood-plain
115 meadows, and thus strongly endangered in Central Europe (Schnittler & Günther, 1999). *A.*
116 *nemorensis* is a short-lived, mostly biennial hemicryptophyte that is known to maintain a long-
117 lived soil seed bank (Hölzel & Otte, 2004). It is part of the *Arabis hirsuta* species aggregate,
118 which comprises several morphologically similar but ecologically diverse species, which diverged
119 about 1.2 million years ago (Karl & Koch, 2014).

120 We performed RAD-seq for over 130 plants collected across pristine and restored sites. This
121 dataset allows us to ask the following questions: i) what is the level of genetic diversity and
122 structuration of the pristine sites that served as source populations for restoration? ii) do

restored sites show a lower level of genetic diversity than the pristine sites? iii) how did restoration affect the distribution of diversity within and among restored sites? iv) is the use of a reference genome necessary to reliably characterize the impact of restoration on genetic diversity? This work demonstrates that a thorough genetic analysis of source and restored sites reveals the complex dynamics at stake in the restoration process. Through genetic analysis, we reveal the unanticipated co-occurrence of *A. nemorensis* with *A. sagittata,* a morphologically similar and closely related species from the *A. hirsuta* aggregate, which normally exhibits a preference for drier habitats such as calcareous grasslands (Hand & Gregor, 2006). We further show that restoration by hay-transfer has maintained and can even enhance local diversity. Our analysis further shows that the use of a reference genome yields higher estimates of genetic diversity, but does not affect the resolution of patterns of genetic variation within and between sites, indicating that this approach can find broad applications in the field.

Methods

Plant material and DNA extraction

The sampling area comprises the fossil, dyke-protected flood plain of the River Rhine near Riedstadt in Hessen, Germany. The area is dominated by arable fields, but also contains remnants of pristine flood meadow communities in low lying depressions that are submerged by ascending groundwater during high floods of the River Rhine. Since ca. 20 years, new flood meadow communities have been restored on ex-arable land using the transfer of green hay

143 from the pristine sites as donors to overcome significant dispersal limitation (Hölzel & Otte,
144 2003). During this process, hay from different donor sites was placed in distinct yet adjacent
145 patches, making admixture possible. Since restoration is still ongoing, restored sites differ in age
146 (Table S1).

147 From previous monitoring of the sites we knew that populations that *Arabis* plants were not
148 present every year in all sites. Thus, we sampled in two consecutive years to maximize the
149 number of study sites. In the few sites sampled in both years, the genetic composition of the
150 samples was not markedly different across years, so samples from both years were bulked for
151 these sites (see Table S1, Figure S1). We harvested seeds from a total of 134 plants of *Arabis*
152 *nemorensis/sagittata* in nine sites named from A to I, in order of collection. The presence of *A.*
153 *sagitata* was not previously reported in these sites and was thus unexpected. Although some
154 individuals showed the reduced stem leaf density and shorter siliques typical of *A. sagittata*,
155 these phenotypic criteria were not always clearly distinguishable in the field, especially at the
156 end of the season when siliques matured, so that the presence of the two species has remained
157 overlooked in previous studies (Burmeier, Eckstein, Donath, & Otte, 2011). Species identity was
158 thus determined by post-hoc analysis, after the RAD-seq analysis revealed the presence of two
159 taxonomic units.

160 Four sites were sampled in pristine habitat (B, C, D and I) and five in restored habitat (A, E, F, G
161 and H, Figure 1). In site A two distinct stands of plants separated by about 100 m were sampled
162 and were treated as sub-sites A-1 and A-2 throughout this study. To produce material for DNA
163 extraction, we stratified seeds on wet filter paper for 6 days at 4°C in darkness. Afterwards we

sowed seeds in soil (33% VM, 33% ED-73, 33% Seramis (clay granules)). After 4 to 6 weeks of growth in the greenhouse, we harvested about 200mg of leaf material from one offspring of each wild parent (genotype). We homogenized freshly harvested leaf material using a Precellys Evolution homogenizer (Bertin technologies) for 2x20 seconds at 6800 rpm. We extracted DNA using the NucleoSpin Plant II Mini kit (Macherey-Nagel) following the manufacturer's instructions. We verified DNA quality using gel-electrophoresis with a 0.8% agarose gel. We measured DNA quantity using Qubit (broad-range kit) following manufacturer's instructions.

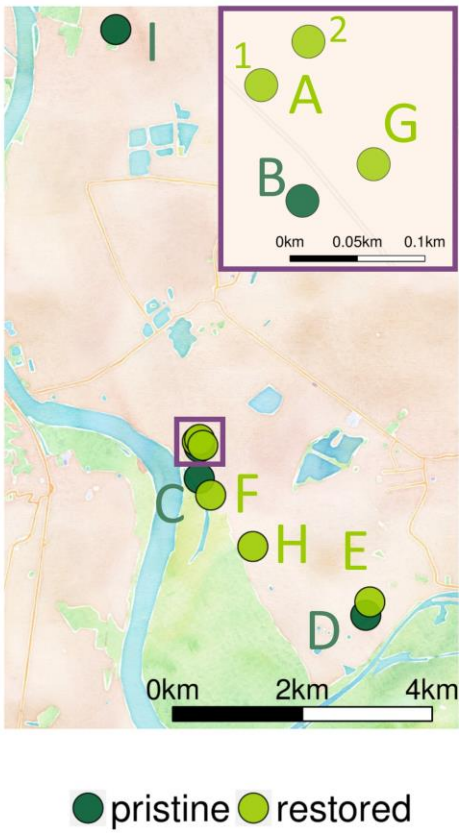


Figure 1: Overview map of studied sites
Each point on the map represents one site. The color of the points represents the type of the site, i.e. whether it is pristine or restored. The purple square represents the area shown in the zoomed inset. All sites, except site F are protected by a dyke.

176

177 ## Draft genome assembly

178 To facilitate genotype calling we assembled a draft genome for one *A. nemorensis* accession (ID
179 29). Library preparation and sequencing was done at the Cologne Center for Genomics. Three
180 libraries were created: one paired-end library with approximately 280 bp insert size creating an
181 overlap of 20 bp with 150 bp reads, and two mate-pair libraries with 3 kbp and 6 kbp inserts,
182 respectively. The paired-end library was created using the Illumina TruSeq® Nano DNA Library
183 Prep Kit, with 2µg of input DNA, without PCR. The mate-pair libraries were created using the
184 Illumina Nextera® Mate Pair Library Prep Kit, with 4 µg of input DNA and 10 cycles of PCR for the
185 3 kbp library and 15 cycles of PCR for the 6 kbp library. All libraries were sequenced together as
186 150bp paired-end reads as part of an Illumina HiSeq 4000 lane for a total of 66 Gbp.

187 We used *FastQC* (Andrews, 2010) to quality-check the resulting reads. We filtered the resulting
188 reads to remove reads shorter than 100bp and trimmed Illumina adapters using *Cutadapt*
189 (Martin, 2011). We assembled reads using the *ALLPATHS-LG* assembler (Gnerre et al., 2011)
190 with default settings, running it on the CHEOPS high-performance computing cluster of the
191 University of Cologne. The resulting contig assembly had a size of 199 Mbp and an N50 of 47
192 kbp. The scaffold assembly had a size of 206 Mbp and an N50 102 kbp. To further scaffold the
193 genome, we generated 2.9 Gbp of PacBio sequence data. Library preparation and sequencing
194 was done at the Max-Planck-Institute for Plant Breeding Research (Cologne, Germany). We
195 scaffolded the genome using *OPERA-LG* with default settings (Gao, Bertrand, Chia, & Nagarajan,
196 2016). This increased the N50 to 150kbp. To achieve chromosome-level assembly, we made use

197 of the available reference genome of *Arabis alpina* (Jiao et al., 2017; Willing et al., 2015). To
198 facilitate whole-genome alignment, repetitive regions in the genome were masked using
199 RepeatMasker (Smit, Hubley, & Green, 2013). We used 'brassicaceae' as the search term ('-
200 species' option) for the RepeatMasker repeat-database. We created the pseudo-chromosome
201 assembly using the *CoGe* website (Lyons & Freeling, 2008). We aligned our draft genome with
202 the *Arabis alpina* genome using *SynMap2* (Haug-Baltzell, Stephens, Davey, Scheidegger, &
203 Lyons, 2017) and afterwards performed synthenic path assembly (Lyons, Freeling, Kustu, &
204 Inwood, 2011) to assemble the chromosomes. The pseudo-chromosomes had a total size of 192
205 Mbp and were used for all following analyses.

206 The size of the genome was estimated by flow-cytometry, which was performed commercially
207 at Plant Cytometry Services. The estimated genome size was 274 Mbp. Thus, the assembly size
208 was 70% of the genome size.

209 ## Annotation

210 To detect and annotate Transposable elements (TE) specific to *A. nemorensis* (not in the
211 previously used database) we used the softwares RepeatModeler (Smit & Hubley, 2008) and
212 RepeatMasker (Smit et al., 2013). The consensus repeat library was first constructed by
213 RepeatModeler and then used by RepeatMasker to search TEs. Protein-coding genes were
214 annotated by integrating predictions of *ab initio* gene annotation tools and alignments of
215 homologous proteins. Three different tools including Augustus v3.2.3 (Stanke & Waack, 2003),
216 GlimmerHMM v3.0 (Majoros, Pertea, & Salzberg, 2004) and SNAP v2013 (Korf, 2004) were used
217 to predict the initial gene models. Protein sequences from *A. thaliana*, *A. lyrata* and *A. alpina*

218 (Arabidopsis Genome Initiative, 2000; Hu et al., 2011; Willing et al., 2015) were aligned to the
219 assembly by the tool Exonerate v2.2.0 (Slater & Birney, 2005). Then, the *ab initio* predictions
220 and protein alignment hits were further combined to build the consensus gene models by the
221 tool EVidenceModeler (EVM) v2012 (Haas et al., 2008). Finally, TE related genes in these models
222 were annotated by checking the TE annotation, blastp (Altschul, Gish, Miller, Myers, & Lipman,
223 1990) alignments with Plant TE related proteins and blastp alignments with *A. thaliana* proteins.
224 If a gene's protein sequence had blastp alignment identity and coverage both larger than 50%
225 with a TE related protein, or at least 30% of its exon regions overlapped with TEs but had no
226 good blastp hit (identity >50% and coverage >50%) from *A. thaliana* protein sequences, this
227 gene would be marked as a TE related gene.

228 ## RAD-sequencing and SNP calling

229 We genotyped 134 samples using the original RAD-sequencing (RADseq) protocol (Etter et al.,
230 2011), with the following modifications (Document S1): i) we used the enzyme KpnI-HF (New
231 England Biolabs) for DNA digestion, ii) we ligated digested DNA with complementary adapters
232 containing one of ten different barcodes and a stretch of five random nucleotides, used for
233 post-hoc removal of PCR duplicates (Table S2), iii) we created 14 pools of 10 barcoded samples
234 each in equal amounts, iv) we used indexed reverse primers for amplification, described in
235 (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012), to allow multiplexing of pools. Libraries were
236 sequenced on two Illumina HiSeq 4000 lanes with 2x150bp.

237 We used *FastQC* (Andrews, 2010) to quality-check the resulting reads. We trimmed adapters
238 and removed reads shorter than 100bp using *Cutadapt* (Martin, 2011). We removed PCR

239 duplicates based on a 5 bp stretch of random nucleotides at the end of the adapter, using the
240 *clone_filter* module of *Stacks* version 1.37 (Catchen, Hohenlohe, Bassham, Amores, & Cresko,
241 2013). We de-multiplexed samples using the *process_radtags* module from *Stacks*. We filtered
242 reads with ambiguous barcodes (allowed distance 2) and cut-sites, reads with uncalled bases
243 and low-quality reads (default threshold).

244 For reference-based genotyping, we mapped reads using *BWA* (Li & Durbin, 2009) with default
245 settings. We filtered mapped reads using *SAMtools* (Li et al., 2009) and custom python scripts
246 using the following criteria to remove reads: mapping quality < 30, number soft-clipped bases >
247 30; reads were unpaired; mates mapped on different chromosomes; mate mapping distance >
248 700. We called genotypes using *SAMtools mpileup* and *VarScan2* (Koboldt et al., 2012) with the
249 following options: base quality > 20; re-calculation of base quality on the fly (-E option); read
250 depth > 14; strand filter de-activated; SNP calling p-value < 0.01. We filtered genotyped loci
251 using *VCFtools* (Danecek et al., 2011) and custom python scripts removing loci with missing data
252 in more than 5% of individuals and loci in masked (repetitive) regions. We clustered contiguous
253 loci spaced less than 100 bp apart into RAD-regions. Regions with excessively low or high
254 coverage are likely results of allele dropout or paralogous mapping, respectively. Thus, we
255 removed regions fulfilling one of the following criteria: mean coverage of the region greater
256 than twice the overall mean coverage; mean coverage of the region smaller than a third of the
257 overall mean coverage; region maximum coverage greater than twice the mean maximal
258 coverage over all regions; region shorter than 250bp; region longer than 1300bp.

259 After depth filtering we still found loci with a high frequency of heterozygotes (up to 100%),
260 which is unlikely in highly selfing species. Notably, in over 90% of these loci only one of the two
261 homozygous genotypes was observed. Moreover, these highly heterozygous loci clustered in
262 high density on single RAD-fragments. Thus, these loci likely resulted from paralogous mapping,
263 which passed the depth filter due to sequencing depth variation among RAD-regions. Therefore,
264 we removed RAD-regions containing loci with a frequency of heterozygotes greater than 20%.
265 This threshold was picked to limit the impact of noise from mapping artifacts while still allowing
266 for reasonable levels of heterozygosity, since some low- level outcrossing likely occurs. From the
267 resulting genotype dataset, we extracted single nucleotide polymorphisms (SNPs) using
268 *VCFtools* (Danecek et al., 2011).

269 For *de-novo*-genotyping we used the *Stacks* 2.2 *denovo_map.pl* pipeline (Catchen et al., 2013).
270 The aim of the de-novo analysis was to test whether one would reach similar results and
271 conclusions without the use of a reference genome. Thus, we performed this analysis without
272 knowledge about the two species (since we gained this from the reference-based analysis) and
273 ran *Stacks* for all samples combined. As recommended by the authors of the tool (Rochette &
274 Catchen, 2017), we first used a subset of 15 representative genotypes to tune the parameters (-
275 M and -n) of the algorithm, which control the number of mismatches between stacks within (M)
276 and between (n) individuals. We varied M and n from 1 to 9. For each set of parameters, we
277 analysed the number of loci shared between 80% of the samples. This measure peaked at the
278 value six for M and n. Thus, we used this value for both parameters for the full analysis. We ran
279 the *denovo_map.pl* pipeline using 0.01 as the p-value threshold for calling genotypes and SNPs,

280 and otherwise default options. We used the *populations* program to create a VCF-file for further
281 analysis using the following filters: 5% maximum missing data per locus; 20% maximum
282 observed heterozygosity per locus; locus must be present in all sites.

283

284 Population genetics statistics

285 We did all statistical analysis using R version 3.4.4 (R Development Core Team, 2008) and
286 provide a supplemental R Markdown file (Document S2). The following packages were used for
287 plotting: *ggplot2* (Wickham, 2009), *ggmap* (Kahle & Wickham, 2013), *ggthemes* (Arnold et al.,
288 2017), *ggsn* (Baquero, 2017) and *heatmap3* (Zhao, Guo, Sheng, & Shyr, 2015). We performed all
289 analysis for the reference-based and the *de-novo*-based dataset and compared the results. We
290 used the *vcfR* package (Knaus & Grünwald, 2017) to load VCF-files into R and make the SNP data
291 available for processing with other libraries. Based on our annotation we determined whether
292 SNPs are in coding regions and whether they are synonymous or non-synonymous using the
293 *PopGenome* package (Pfeifer, Wittelsbuerger, Li, & Handsaker, 2018). We performed principal
294 component analysis (PCA) of SNP data for all samples using the *adegenet* package (Jombart et
295 al., 2016). Missing data was scaled to the mean for PCA.

296 Based on the first principal component most individuals (83%) could be assigned to two distinct
297 taxonomic groups. We used molecular methods to assign species labels to the two taxonomic
298 groups. We sequenced the internal transcribed spacer (ITS) sequence of nine individuals, three
299 from the first (left) and five from the second (right) cluster and one located between the

300 clusters (Table S3). Primers for amplification were taken from (Mummenhoff, Franzke, & Koch,
301 1997). We used the sequences as input to the taxonomy tool of the *Brassibase* website (Kiefer
302 et al., 2014). Interpretation of the output required ploidy information. Leaf samples were
303 therefore collected from twelve individuals (seven putative *A. sagittata*, two putative *A.*
304 *nemorensis*, three putative hybrids; Table S4) and their ploidy determined by Plant Cytometry
305 Services (Didam, Netherlands).

306 We conducted all population genetic analysis separately for the two species. Individuals, which
307 were not assigned to any species, were likely interspecific hybrids and excluded from further
308 analysis. We used the *pegas* library to calculate within-site genetic diversity (Nei's π; average
309 pairwise nucleotide differences) for each site (Paradis, Jombart, Schliep, Potts, & Winter, 2016),
310 excluding sites with less than two individuals per respective species. To scale the estimates of π,
311 we divided the average number of pairwise nucleotide differences among samples by the total
312 number of successfully genotyped bases, excluding all bases which failed any of our previously
313 described genotype filters (missing data, region heterozygosity, region depth and length). For
314 the de-novo pipeline we extracted the total number of genotyped bases from *Stacks* output. We
315 calculated correlation coefficients between reference-based and *de-novo*-based π estimates
316 using Pearson's method. We calculated pairwise F_{ST} (Nei, 1987) and genetic distance (Cavalli-
317 Sforza & Edwards, 1967) between all pairs of sites using the *hierfstat* package (Goudet &
318 Jombart, 2015). Negative F_{ST} values were set to zero. Differences of genetic distance and F_{ST}
319 among pristine and restored sites were tested using a Wilcoxon-rank-sum-test on pairwise

320 distance matrices. We tested for correlation between the distance matrices of the reference
321 and *de novo* datasets using a Mantel test with 10000 permutations.

322 Admixture analysis

323 For ADMIXTURE analysis (Alexander, Novembre, & Lange, 2009), we converted VCF-files to bed-
324 files using PLINK (Purcell, 2009; Purcell et al., 2007). First, we conducted ADMIXTURE analysis
325 for all samples combined for K=1 to K=10 (reference-pipeline only). Then for each of the species
326 and pipeline (reference/*de novo*), we ran ADMIXTURE analysis for K=2 to K=10, with 10
327 iterations of cross-validation each. Before plotting, we normalized clusters across runs using
328 CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). We created plots using
329 a custom R-script.

330 Results

331 RAD-sequencing uncovers two hybridizing species

332 Unless otherwise stated, all described results were obtained using the reference-based pipeline.
333 We genotyped 134 individuals from 10 sites – 4 pristine and 6 restored (Figure 1) – yielding 3.6
334 Mb of sequence of which 32,880 single nucleotide positions were polymorphic (SNPs). Only 20%
335 of SNPs were in coding regions, 40% and 56% of which were synonymous and non-synonymous,
336 respectively (4% unassigned). To visualize patterns of genetic diversity across sites, we
337 conducted principal component analysis (PCA) for all individuals (Figure 2, top). Almost 80% of
338 the total genetic variation was explained by the first principal component, which separated
339 most individuals into two clearly defined clusters, likely representing taxonomic units. Based on

340 phylogenetic analysis of the ITS region of 10 individuals (Table S3), we identified one taxonomic
341 unit as *Arabis nemorensis* (Figure S2). The ITS region of individuals from the other taxonomic
342 unit was highly similar to *A. sagittata* and *Arabis hirsuta* (Figure S3). Both are sibling species
343 from the *A. hirsuta* complex, but *A. hirsuta* is tetraploid and *A. sagittata* diploid. Since all twelve
344 tested samples were diploid (Table S4), we conclude that the second cluster most likely
345 corresponds to *A. sagittata*. Average genetic distance (d_{XY}) between *A. nemorensis* and *A.*
346 *sagittata* was 8.1e-03, i.e. 8 fixed SNP differences for 1000 bp.

347 Twenty-three individuals showed a positioning along the first PC that was intermediate between
348 the two clusters, suggesting they were interspecific hybrids. Most of these hybrids were closer
349 to *A. sagittata* on the first PC. Since sterile F1-hybrids should be located exactly in the middle
350 between the two species, this suggests that they are somewhat fertile and preferentially back-
351 cross with *A. sagittata.* Laboratory observations have shown that hybrids are indeed fertile
352 (Dittberner, personal observation). Additionally, in an ADMIXTURE analysis of all samples with
353 K=2, most hybrids showed ancestry over 50% from *A. sagittata*, as expected from preferential
354 back-crossing to this parent (Figure S4).

355 Overall our sample was composed of 31% *A. nemorensis*, 52% *A. sagittata* and 17% hybrids. The
356 species composition differed among sites (Figure 2, bottom). *A. nemorensis* was present in 7
357 sites (3 pristine, 4 restored), *A. sagittata* in 9 sites (3 pristine, 6 restored) and hybrids in 6 sites
358 (3 pristine, 3 restored). Notably, the pristine site D was dominated by hybrids with over 56%.

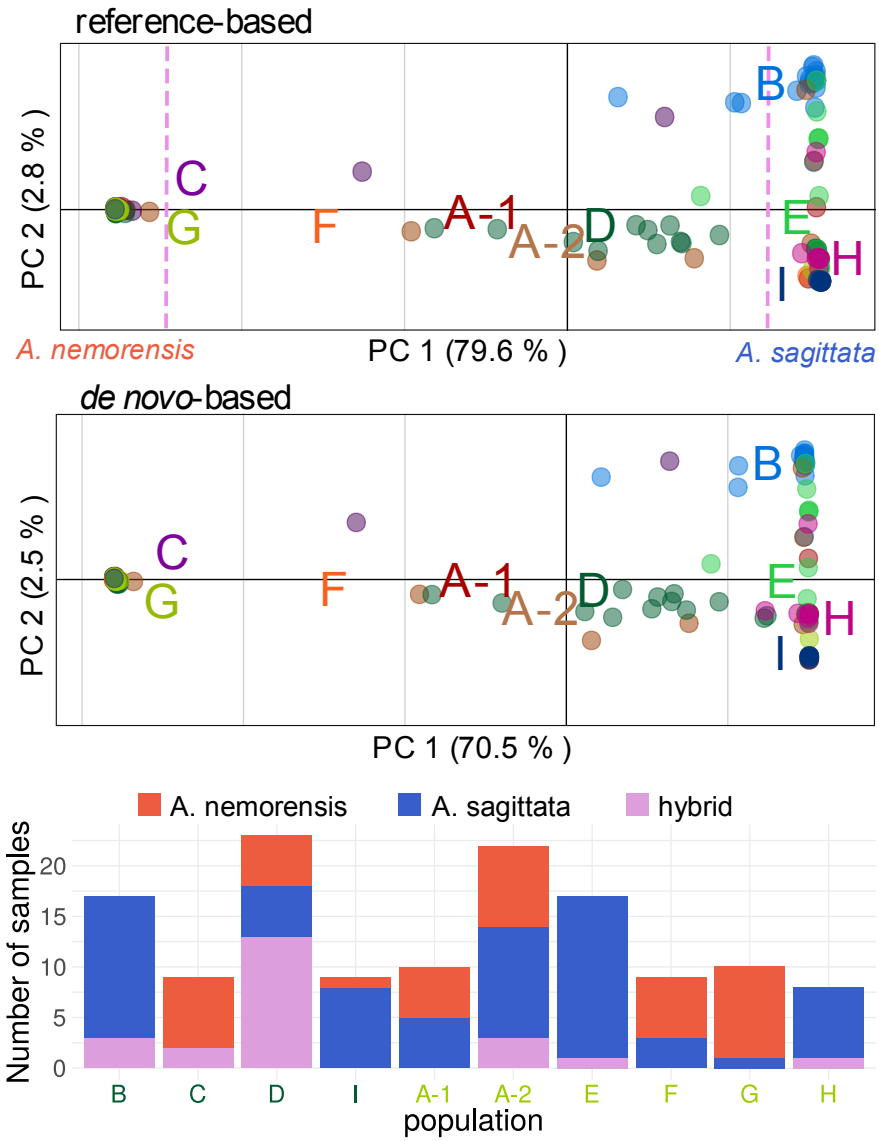


Figure 2: Presence of two hybridizing *Arabis* species in study sites.
Top) Plot of the first two principal components (PC) of genetic variation based on the reference-based pipeline. Text labels and colors distinguish the different sites. Text labels are approximately in the centroid of the respective site (with offset to avoid overlaps). Values in brackets in the axis labels show the amount of variance explained by each PC. Note that the first PC explains about 80% of the total variation and splits most individuals into two clearly defined clusters. The clusters on the left and on the right end of PC1 correspond to two distinct species, identified here as *Arabis nemorensis* and *Arabis sagittata*. Species identity is indicated by text labels below the plot. Points in between the two clusters are likely natural interspecific hybrids. The dashed lines show the thresholds used for distinction between species and hybrids throughout this study. Middle) Plot showing the same analysis as above but based on the de-novo-pipeline. Bottom) Stacked barplot showing the distribution of species among the

study sites. Population labels are colored by type: darkgreen = pristine, lightgreen = restored. Species are distributed heterogeneously among sites.

No reduction of genetic diversity in restored sites

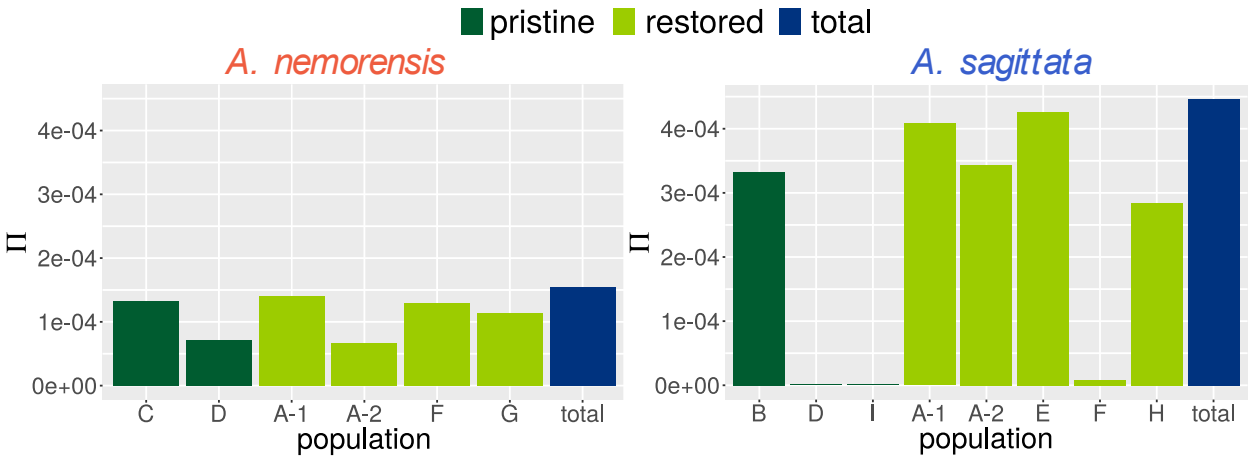
We computed species-specific estimates of genetic diversity within each site, excluding hybrid genotypes. The *A. nemorensis* dataset consisted of 2746 SNPs. Levels of genetic diversity (π) varied up to two-fold among sites, ranging from 6.6e-05 in A-2 to 1.4e-04 in A-1 (Figure 3**Fehler! Verweisquelle konnte nicht gefunden werden.**, left; Table S5). Total diversity was 1.5e-04. However, pristine and restored sites did not differ significantly in their level of diversity (mean difference= +10% in restored; W = 4, p = 1).

The *A. sagittata* dataset consisted of 6366 SNPs. Total genetic diversity in *A. sagittata* was about three times as high as in *A. nemorensis*. Yet, in contrast to *A. nemorensis*, genetic diversity differed strongly among sites, ranging from 1.03e-06 in site I to 4.26e-04 in site E (Figure 3**Fehler! Verweisquelle konnte nicht gefunden werden.**, right; Table S5). Notably, genetic diversity was low in two of three pristine sites. In contrast, we found high levels of diversity in all restored sites, except site F. However, the overall difference between pristine and restored sites was not significant (mean difference= +163% in restored sites; W=13, p=0.14).

Since hybridization potentially enables gene flow between the two species we also compared levels of genetic diversity for both species combined, including the hybrids. Overall genetic diversity increased by an order of magnitude and mixed sites were more diverse than mostly

391 pure sites, as would be expected (Figure S5). Again, restored sites did not show significantly

392 different levels of genetic diversity from pristine sites (mean difference= +22%; W=13, p=0.91).



393

394 **Figure 3: No reduction in genetic diversity in restored sites**
395 Barplot of average pairwise genetic diversity (π) within each site of each species. Bar-color indicates the
396 type of site.

397 ## Restoration reduces population structure and facilitates recombination

398 To quantify the degree of population structure, we estimated genetic distance and

399 differentiation (F_{ST}) among all pairs of sites (Table S6). Genetic distance among *A. nemorensis*

400 sites ranged from 0.03 to 0.31 and F_{ST} estimates from 0 to 0.5 (Figure 4, A+C). Population

401 structure was slightly more pronounced among pristine sites than restored sites: Mean genetic

402 distance was 0.26 among pristine sites and 0.13 among restored sites (W=17, p=0.047); mean

403 F_{ST} was 0.37 among pristine sites and 0.25 among restored sites (W=14, p=0.26).

404 In *A. sagittata*, genetic distance ranged from 0.32 to 0.34 and F_{ST} estimates from 0 to 0.91

405 (Figure 4, B+D). The difference in population structure between pristine and restored sites was

406 stronger than in *A. nemorensis*: mean genetic distance was 0.33 among pristine sites and 0.12

among restored sites (W=45, p=0.002); mean F_{ST} was 0.81 among pristine sites and 0.2 among restored sites (W=43, p=0.015).

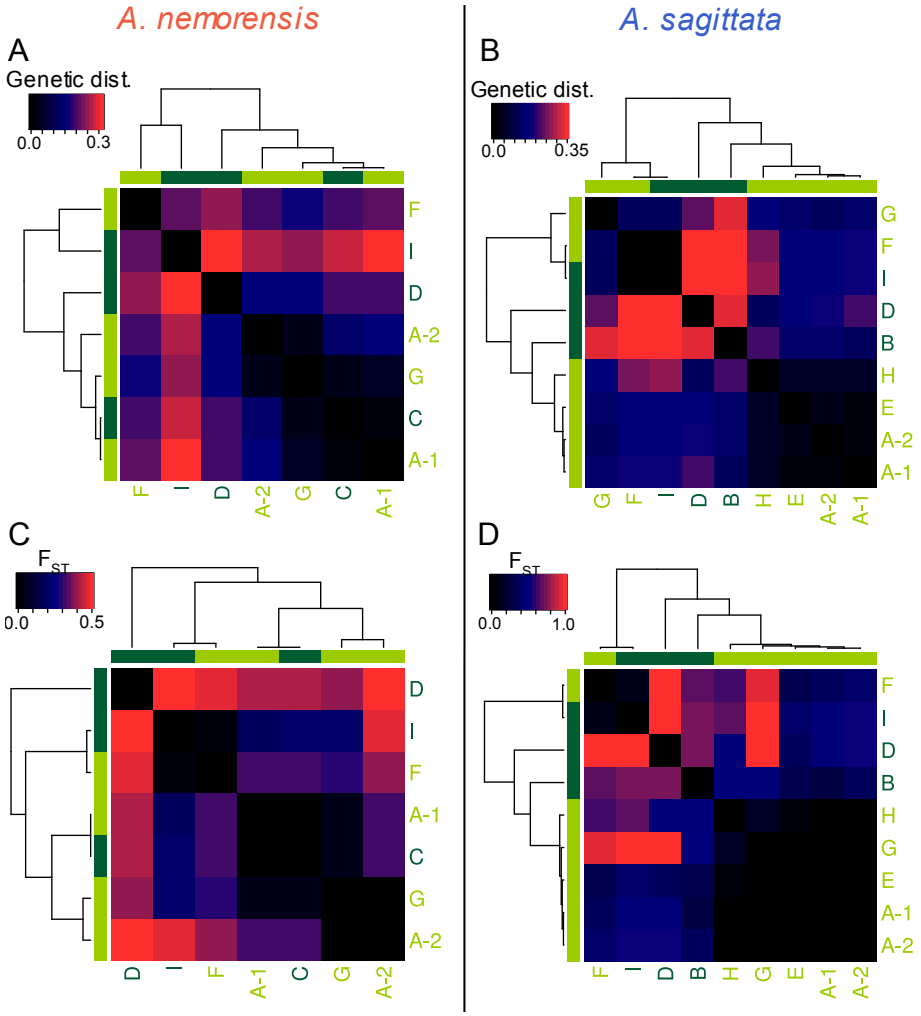


Figure 4: Overall strong population structure with signatures of admixture in restored sites
This panel shows different measurements of population structure of each species. A-D) Heatmaps showing pairwise comparisons between all sites of genetic distance (Cavalli-Sforza chord distance) and FST, respectively. Values of the respective variable are indicated by tile color, corresponding to the color scale in the top-left corner of each heatmap. Sites were clustered based on the respective variable, as indicated by the dendrograms. Colors of text labels and bars next to the grid indicate the type of the site (same color scheme as Figure 3).

418 Reduced differentiation among restored sites suggests that genetic material of pristine sites was
419 mixed in restored sites. To visualize this in more detail, we conducted ADMIXTURE analysis for
420 each species. This analysis assumes a given number (K) of genetic clusters (ancestral
421 populations) and assigns cluster ancestry proportions to each individual, allowing for mixed
422 ancestry. For *A. sagittata*, K=6 was estimated as the optimal value (Figure 5). Up to K=4, the
423 three pristine sites consisted of three distinct genetic clusters, which were mixed in all but one
424 of the restored sites. For higher values of K, pristine population B consisted in two clusters, with
425 pure and admixed individuals in equal proportions, in agreement with the higher genetic
426 diversity observed at this site. All individuals from other pristine sites had pure ancestry from a
427 single cluster per population.

428 For *A. nemorensis*, K=7 was estimated as the optimal value for clustering (Figure 5). In contrast
429 to *A. sagittata*, we found individuals with different or mixed ancestry within populations of *A.*
430 *nemorensis*, even at low values of K, indicating more genetic mixture than in *A. sagittata*. Since
431 genetic structure was less pronounced for *A. nemorensis*, the restoration procedure did not
432 impact the genetic distribution of this species, by contrast with *A. sagittata*.

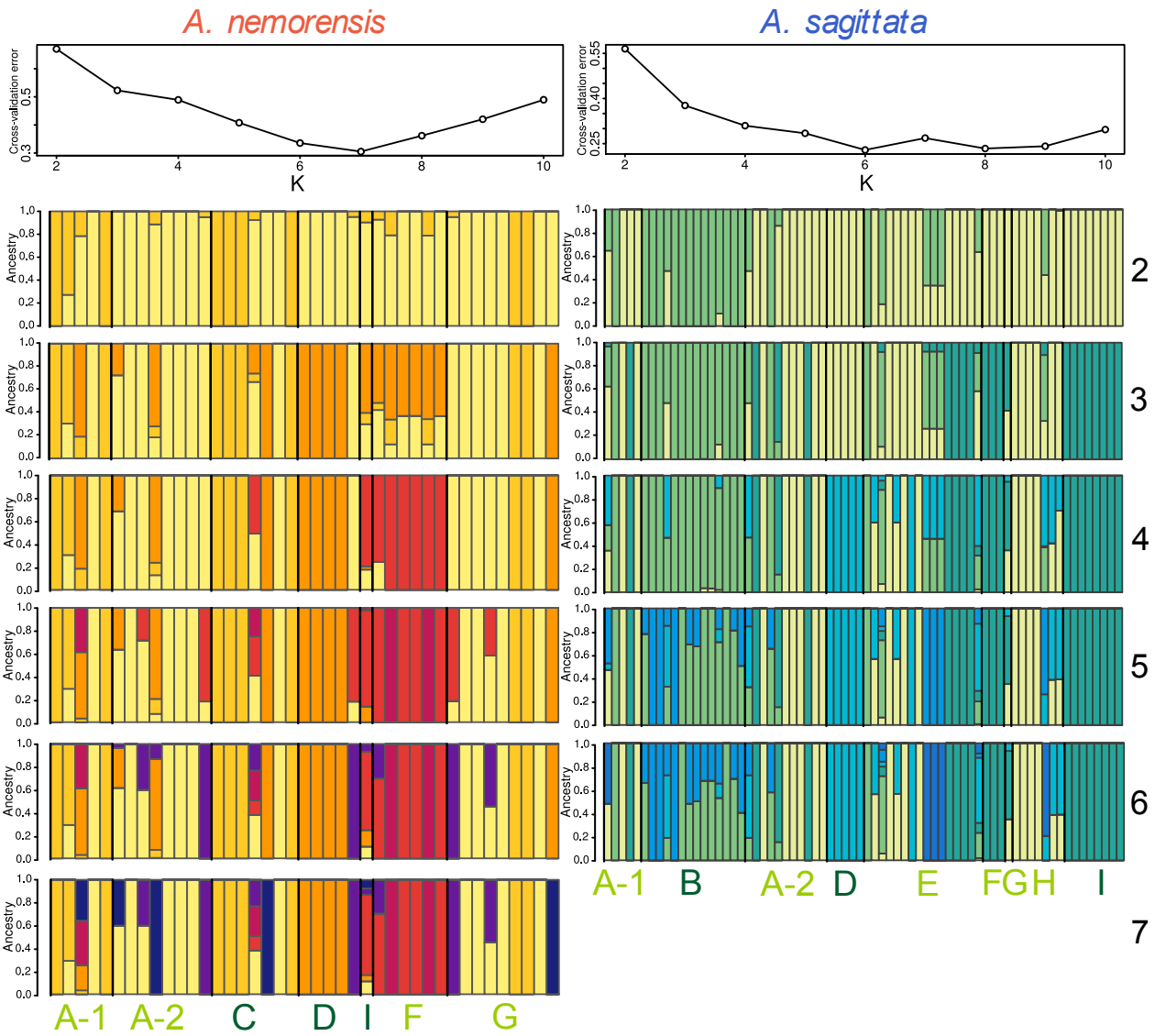


Figure 5: Increased impact of genetic admixture of pristine sites in *A. sagittata*
The figures shows a comparison of ADMIXTURE results for both species. The top panels show the cross-validation (CV) error for all tested values of K. The following panels show the ancestry proportions of each cluster for each individual for all values of K until the first minimum of the CV-error curve. The numbers on the right represent the value of K. Site labels are colored according to the type of the site: light-green = restored, dark-green = pristine.

441 ### *De-novo*- and reference-based summary statistics reach the same conclusion

442 Finally, we tested whether a reference genome is required to determine the impact of
443 restoration on genetic diversity by comparing results from a reference-based and a *de novo*
444 pipeline. We found that estimates of genetic diversity, genetic distance or F_{ST} yielded by the two
445 methods were highly correlated, with all correlation coefficients being greater than 0.95
446 (maximum $p < 0.001$, Figure 6). However, we observed that estimates of genetic diversity
447 generated without a reference genome were deflated, especially for high diversity sites (Figure
448 6), by a median factor of 0.83 for *A. nemorensis* and 0.36 for *A. sagittata*. Estimates of genetic
449 distance were underestimated by a median factor of 0.61 in both species. Interestingly,
450 however, both pipelines yielded almost identical estimates of F_{ST} (median factor of 0.93) and
451 revealed a very similar extent of admixture between sites (Figures S6-7). Moreover, the
452 presence of the two species and their hybrids was detected with both methods (Figure 2
453 middle). We concluded that, in this study, the use of a reference genome was not required to
454 determine the impact of restoration on genetic diversity. Since the two pipelines coincide, we
455 could further conclude that the distribution of variation in two species and across pristine and
456 restored sites reported above is not the result of possible mapping biases to the reference
457 genome.

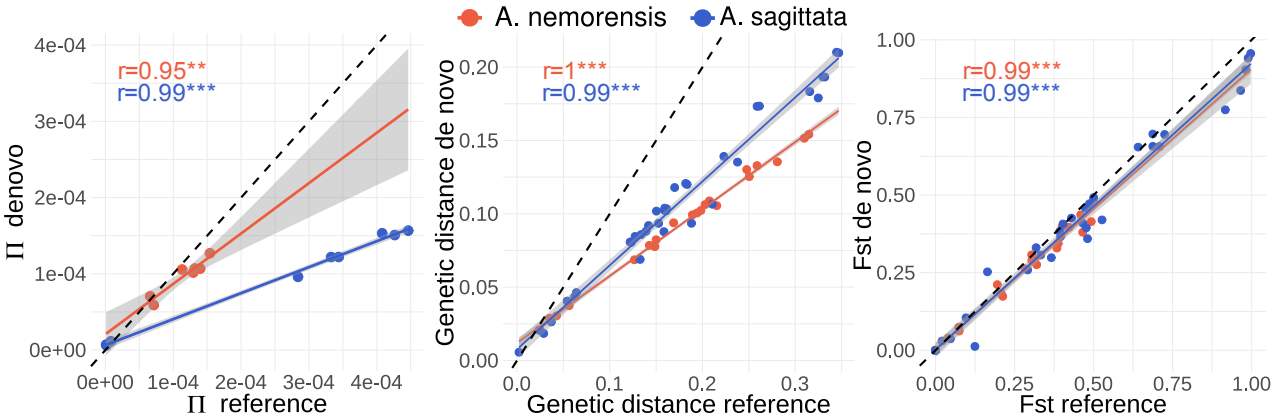


Figure 6: De-novo- and reference-based summary statistics are highly correlated.
Plots show the correlation of summary statistics based on the reference-based and the de-novo-based pipeline. Line and dot colors represent the species. Each dot represents one individual. Lines represent a linear fit through the dots and the gray shadows indicate the error of the fit. The dashed line represents a hypothetical 1:1 relationship of the variables. Correlation-coefficients (Pearson's product-moment correlation) are indicated as colored text corresponding to the species. Stars indicate the level of the coefficient: **p<0.01, ***p<0.001.

Discussion

A reference-genome is not required to characterize the impact of restoration

While RAD-seq and related methods are a cheap tool to acquire genotype information across the genome without need for a reference genome (Elshire et al., 2011; Etter et al., 2011; Peterson et al., 2012), the reliability of *de novo* assembly pipelines has been questioned (Shafer et al., 2016). The availability of a reference genome allowed us to ask whether reference-based read mapping pipelines yielded distinct conclusions from a pipeline based on *de novo* read assembly. Conversely, this also allowed verifying that the use of an *A. nemorensis* reference genome to map A. sagittata samples did not bias the outcome of our study. The results from both pipelines were highly correlated. Thus, comparative analysis of sites was reliable with both

477 pipelines. However, the *de-novo*-pipeline (*Stacks*, Catchen et al., 2013) underestimated the
478 amount of genetic diversity compared to the reference-based pipeline. This is in contrast to a
479 previous study comparing different RAD-seq pipelines (Shafer et al., 2016), where *Stacks*
480 produced slightly inflated estimates compared to reference-based pipelines. However, in the
481 same study, other de novo pipelines produced substantially lower estimates of genetic diversity
482 (Shafer et al., 2016). Thus, the magnitude of genetic diversity estimates might depend on the
483 study system and pipelines/parameters used, and caution is advised when comparing these
484 estimates across studies.

485 In contrast, both pipelines agreed for analyses comparing diversity between species or sites (e.g.
486 F_{ST}, ADMIXTURE). Thus, we conclude that RAD-seq is an efficient tool to characterize the
487 distribution of diversity, even in the absence of a reference genome. We therefore hope that
488 this study will pave the way for exploring how species, with diverse life-history and
489 uncharacterized genomes, will be maintained after hay transfer or how modalities of hay
490 transfer affects not only single species but the balance between multiple species in the
491 community.

492 ## First documented presence of *A. sagittata* in floodplain meadow habitat

493 We detected two independent genetic clusters. Based on ITS sequences, ecological preference
494 and morphological descriptors, the first one appears to represent *bona fide A. nemorensis*
495 individuals. Taxonomic boundaries in the *Arabis hirsuta* clade can be tenuous, yet ITS sequence
496 differences between the two clusters strongly suggests that they belong to distinct species (Karl
497 & Koch, 2014). ITS sequences, ploidy level and morphological evidence collectively indicated

498 that individuals in the second genetic cluster are likely to belong to *Arabis sagittata*. Species
499 barriers are often incomplete, and patterns of divergence between closely related species of the
500 same genus can be strongly intertwined. Population genomics approaches and wider sampling
501 will be needed to consolidate our understanding of the *A. sagittata* taxonomic group and its
502 historical relationship with the *A. nemorensis* taxon. Population genomics can indeed yield
503 detailed view on the process and history of speciation in plant genera (see e.g. Novikova et al.,
504 2016).

505 We did not anticipate the presence of the *A. sagittata* in flood-plain habitats because it is
506 predominantly found in warm and dry habitats (Hand & Gregor, 2006). However, agricultural
507 land-use and flood regulation have considerably modified the flood-plain ecosystem: flooding is
508 contained by a dyke and the ground water level has decreased (Hölzel & Otte, 2001). Man-made
509 modifications of the environment can change selection regimes and facilitate the establishment
510 of non-native species (Byers, 2002; Crooks, Chang, & Ruiz, 2011; Fukasawa, Miyashita,
511 Hashimoto, Tatara, & Abe, 2013; Tyrrell & Byers, 2007). Thus, it is possible that the migration of
512 *A. sagittata* into the flood-plain ecosystem was facilitated by the decreased frequency and
513 severity of flooding caused by human river regulation. We want to stress that the species co-
514 occurred both in pristine and restored sites. Thus, the contact between the species was not
515 caused by the restoration efforts. In fact, we observed that *A. sagittata* is more frequent than *A.*
516 *nemorensis* in our sample, which raises the concern that this species may be in the process of
517 displacing *A. nemorensis*. This trend could be enhanced by climate change, which may lead to
518 conditions that further favor the xero-thermophylic *A. sagittata*. Additionally, if *A. sagittata* is

519 often mistaken for *A. nemorensis* in flora reports on flood-plain environments, the remaining *A.*
520 *nemorensis* population in Central Europe might be even smaller than is currently assumed.

521

522 The parallel transfer of two sympatric relatives shows that hay transfer maintains genetic
523 diversity

524 Hay-transfer has proven to be a particularly successful method for establishing new populations
525 of target species in ecological restoration across a variety of herbaceous vegetation types
526 (Coiffait-Gombault, Buisson, & Dutoit, 2011; Hölzel & Otte, 2003; Kiehl et al., 2010; Kiehl &
527 Wagner, 2006; Török et al., 2012). Furthermore, hay-transfer is often seen as the gold standard
528 to preserve local levels of genetic diversity and adaptation (Vander Mijnsbrugge, Bischoff, &
529 Smith, 2010). The latter is probably the main reason, why it is increasingly used in in ecological
530 restoration (Kiehl et al., 2014). The aim of this study was to characterize the level of diversity in
531 the pristine source sites and document the impact of hay-transfer on the genetic diversity in
532 restored sites. Although our initial plan was to focus on *A. nemorensis*, a typical representative
533 of species-rich floodplain meadows, the unanticipated presence of *A. sagittata* in our sample
534 allowed us to compare the genetic effects of restoration by hay-transfer on the two species.

535 Several years after restoration, we did not find a significant difference in genetic diversity
536 between pristine and restored sites for either of the two species. Thus, the hay-transfer method
537 can restore populations with levels of diversity indistinguishable from the source populations in
538 the long-term. Our findings are in agreement with studies on population life-stage structure and

539 dynamics comparing pristine and restored sites of *A. nemorensis/A. sagittata* in the same region
540 (Burmeier et al., 2011). This outcome is particularly remarkable given that *A. nemorensis* has a
541 long-term seed bank, with up to 25,000 germinable seeds*m^{-2}, which was not transferred to
542 restored sites (Burmeier et al., 2011). Although species for which the genetic diversity present in
543 the seed bank tends to differ more strongly from the above ground diversity may fare
544 differently after hay-transfer, we note that populations restored with alternative methods, e.g.
545 spontaneous recolonization (Vandepitte et al., 2012) or propagated seed mixtures (Espeland et
546 al., 2017; Fant, Holmstrom, Sirkin, Etterson, & Masi, 2008), both excluded the seed bank and
547 revealed a reduction of diversity (Mijangos et al., 2015). Thus, hay-transfer might be superior to
548 other restoration methods not only in restoration success (Hölzel & Otte, 2003; Kiehl et al.,
549 2010) but also in transferring genetic diversity.

550

551 ## The modalities of habitat restoration can modify the relative adaptive potential of species
552 in the ecosystem

553 Genetic diversity was low in *A. nemorensis* and *A. sagittata,* compared to other outcrossing or
554 selfing Brassicaceae species (Mattila, Tyrmi, Pyhäjärvi, & Savolainen, 2017; Onge, Källman,
555 Slotte, Lascoux, & Palmé, 2011). In fact, genetic distance between species was only on the order
556 of magnitude as diversity found within populations of *Arabidopsis lyrata* (Mattila et al., 2017).
557 Genetic diversity within our species was similar to that reported for *Arabis alpina* populations in
558 Scandinavia (Laenen et al., 2018). As in our study, these populations are selfing and located on
559 the margin of the species' range (Jalas & Suominen, 1994), two factors often coinciding with

560 lower levels of genetic diversity. Since populations with low genetic diversity may suffer from
561 increased genetic load and decreased adaptive potential, the transfer of non-local material is
562 often envisaged to preserve endangered species (Breed, Stead, Ottewell, Gardner, & Lowe,
563 2013; Weeks et al., 2011). This approach is however controversial. Strategies that introduce
564 non-local seeds can decrease population fitness either by introducing maladapted genotypes
565 (Crémieux, Bischoff, Müller-Schärer, & Steinger, 2010; McKay, Christian, Harrison, & Rice, 2005)
566 or by causing outbreeding depression (Frankham et al., 2011). As a compromise, a strategy of
567 mixing regional seeds was recently proposed (Bucharova et al., 2018). Our results show that this
568 strategy was incidentally implemented in the examined restoration effort: ADMXITURE analysis
569 showed that pristine *A. sagittata* sites are dominated by one or two ancestral groups, with very
570 low genetic diversity within each group. Yet, in some of the restored sites, admixture of low
571 diversity *A. sagittata* groups took place, leading to a strong increase in genetic diversity and
572 decrease in population structure. ADMIXTURE analysis also revealed that this led to genetic
573 recombination in *A. sagittata,* possibly increasing the adaptive potential of restored sites in this
574 species. Increased genetic connectivity between populations has indeed been shown to help
575 maintain or even increase genetic diversity in the long term (DiLeo, Rico, Boehmer, & Wagner,
576 2017). Fitness assays of recombined *A. sagittata* genotypes are needed to verify whether this
577 local admixture has reinforced the establishment of this species in restored floodplain
578 meadows.

579

580 Interestingly, we find less pronounced population structure in pristine sites of *A. nemorensis*
581 than in *A. sagittata*, suggesting that a higher level of gene flow helps maintain diversity within
582 pristine sites. The restoration of this species will therefore not directly benefit from post-
583 transfer admixture. More so, our results suggest that it is possible that *A. nemorensis* is at
584 increased disadvantage in restored habitats, if admixture and resulting recombination were to
585 favor the emergence of more competitive genotypes were to evolve in *A. sagittata* thanks to
586 admixture.

587 The coexistence of these species in the Rhine floodplain ecosystem will be further impacted by
588 the ongoing hybridization dynamic that our analysis uncovers for the first time in this system
589 and which occurred independently of the restoration effort. *A. nemorensis* could receive alleles
590 conferring drought-adaptation from the xero-thermophilic *A. sagittata* that may enhance its
591 ability to cope with increased drought exposure in its habitat. The genomic composition of the
592 hybrids, however, also indicates that hybrids back-cross preferentially with *A. sagittata*. Gene-
593 flow from *A. nemorensis* could facilitate adaptation of *A. sagittata* to the flood-plain
594 environment. Such adaptive introgressions could also potentially accelerate the extinction of *A.*
595 *nemorensis*. While hybridization is common in plants (Mallet, Besansky, & Hahn, 2016), well
596 documented cases of adaptive introgression are rare and require elaborate experiments
597 (Goulet, Roda, & Hopkins, 2017; Suarez-Gonzalez, Lexer, & Cronk, 2018). Such experiments are
598 now warranted to determine the impact of hybridization in this sympatric species complex.

599

600 ## Conclusions

601 Clearly, genetic analysis helps with species identification as sibling species can be difficult to
602 distinguish morphologically even for specialists. A unique feature of hay-transfer restoration
603 approaches is that the whole plant community can be transplanted. It is therefore particularly
604 important to determine the composition of the source populations to limit the spread of non-
605 target species (Bickford et al., 2007). Our study demonstrates that hay-transfer has maintained
606 genetic diversity in restored populations. However, we also note that it may have inadvertently
607 contributed to increase the genetic diversity and adaptive potential of only one of the two
608 species, due to differences in the genetic makeup of the donor populations. This might lead to a
609 competitive advantage of one over the other species in the long term, potentially disturbing the
610 balance in the community. On the one hand this shows that restoration by hay-transfer can
611 indeed enhance the adaptive potential, which is especially important in the face of a rapid
612 climate change. On the other hand, this also highlights that understanding the underlying
613 genetics of the community to be transferred is a pre-requisite for the design of restoration
614 strategies that promote the maintenance of both endangered ecosystems and endangered
615 species.

616 # Acknowledgements

617 We thank Matthias Harnisch for providing information about the populations, Markus Koch for
618 insightful discussions about the *A. hirsut*a tribe, Eric Schranz for his assistance with the pseudo-
619 chromosome assembly and Gregor Schmitz for helpful feedback. Further, we thank Janine

Altmüller and the team of the Cologne Center for Genomics for their assistance in RAD-seq optimization, library preparation and sequencing. This work was partly funded by the German Research Foundation 'Deutsche Forschungsgemeinschaft DFG [DFG priority program 1529 'ADAPTOMICS'].

References

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. https://doi.org/10.1101/gr.094052.109

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data (Version 0.11.3). Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, *408*(6814), 796–815. https://doi.org/10.1038/35048692

Arnold, J. B., Daroczi, G., Werth, B., Weitzner, B., Kunst, J., Auguie, B., … London, J. (2017). ggthemes: Extra Themes, Scales and Geoms for "ggplot2" (Version 3.4.0). Retrieved from https://cran.r-project.org/web/packages/ggthemes/index.html

Baillie, J., Hilton-Taylor, C., & Stuart, S. N. (2004). *2004 IUCN Red List of threatened species : a global species assessment*. Retrieved from https://portals.iucn.org/library/node/9830

Baquero, O. S. (2017). ggsn: North Symbols and Scale Bars for Maps Created with "ggplot2" or "ggmap" (Version 0.4.0). Retrieved from https://CRAN.R-project.org/package=ggsn

Benayas, J. M. R., Newton, A. C., Diaz, A., & Bullock, J. M. (2009). Enhancement of Biodiversity and Ecosystem Services by Ecological Restoration: A Meta-Analysis. *Science*, *325*(5944), 1121–1124. https://doi.org/10.1126/science.1172460

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., … Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*(3), 148–155. https://doi.org/10.1016/j.tree.2006.11.004

Bradbury, I. R., Hamilton, L. C., Dempson, B., Robertson, M. J., Bourret, V., Bernatchez, L., & Verspoor, E. (2015). Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. *Molecular Ecology*, *24*(20), 5130–5144. https://doi.org/10.1111/mec.13395

Breed, M. F., Stead, M. G., Ottewell, K. M., Gardner, M. G., & Lowe, A. J. (2013). Which provenance and where? Seed sourcing strategies for revegetation in a changing

environment. *Conservation Genetics*, *14*(1), 1–10. https://doi.org/10.1007/s10592-012-0425-z

Bucharova, A., Bossdorf, O., Hölzel, N., Kollmann, J., Prasse, R., & Durka, W. (2018). Mix and match: regional admixture provenancing strikes a balance among different seed-sourcing strategies for ecological restoration. *Conservation Genetics*. https://doi.org/10.1007/s10592-018-1067-6

Bucharova, A., Michalski, S., Hermann, J.-M., Heveling, K., Durka, W., Hölzel, N., … Bossdorf, O. (2017). Genetic differentiation and regional adaptation among seed origins used for grassland restoration: lessons from a multispecies transplant experiment. *Journal of Applied Ecology*, *54*(1), 127–136. https://doi.org/10.1111/1365-2664.12645

Bullock, J. M., Aronson, J., Newton, A. C., Pywell, R. F., & Rey-Benayas, J. M. (2011). Restoration of ecosystem services and biodiversity: conflicts and opportunities. *Trends in Ecology & Evolution*, *26*(10), 541–549. https://doi.org/10.1016/j.tree.2011.06.011

Burmeier, S., Eckstein, R. L., Donath, T. W., & Otte, A. (2011). Plant Pattern Development during Early Post-Restoration Succession in Grasslands—A Case Study of Arabis nemorensis. *Restoration Ecology*, *19*(5), 648–659. https://doi.org/10.1111/j.1526-100X.2010.00668.x

Byers, J. E. (2002). Impact of non-indigenous species on natives enhanced by anthropogenic alteration of selection regimes. *Oikos*, *97*(3), 449–458. https://doi.org/10.1034/j.1600-0706.2002.970316.x

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

Cavalli-Sforza, L. L., & Edwards, A. W. F. (1967). Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, *19*(3 Pt 1), 233–257.

Coiffait-Gombault, C., Buisson, E., & Dutoit, T. (2011). Hay Transfer Promotes Establishment of Mediterranean Steppe Vegetation on Soil Disturbed by Pipeline Construction. *Restoration Ecology*, *19*(201), 214–222. https://doi.org/10.1111/j.1526-100X.2010.00706.x

Crémieux, L., Bischoff, A., Müller-Schärer, H., & Steinger, T. (2010). Gene flow from foreign provenances into local plant populations: Fitness consequences and implications for biodiversity restoration. *American Journal of Botany*, *97*(1), 94–100. https://doi.org/10.3732/ajb.0900103

Crooks, J. A., Chang, A. L., & Ruiz, G. M. (2011). Aquatic pollution increases the relative success of invasive species. *Biological Invasions*, *13*(1), 165–176. https://doi.org/10.1007/s10530-010-9799-3

Crutsinger, G. M., Collins, M. D., Fordyce, J. A., Gompert, Z., Nice, C. C., & Sanders, N. J. (2006). Plant Genotypic Diversity Predicts Community Structure and Governs an Ecosystem Process. *Science*, *313*(5789), 966–968. https://doi.org/10.1126/science.1128326

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DiLeo, M. F., Rico, Y., Boehmer, H. J., & Wagner, H. H. (2017). An ecological connectivity network maintains genetic diversity of a flagship wildflower, Pulsatilla vulgaris. *Biological Conservation*, *212*, 12–21. https://doi.org/10.1016/j.biocon.2017.05.026

Donath, T. W., Bissels, S., Hölzel, N., & Otte, A. (2007). Large scale application of diaspore transfer with plant material in restoration practice – Impact of seed and microsite limitation. *Biological Conservation*, *138*(1), 224–234. https://doi.org/10.1016/j.biocon.2007.04.020

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Espeland, E. K., Emery, N. C., Mercer, K. L., Woolbright, S. A., Kettenring, K. M., Gepts, P., & Etterson, J. R. (2017). Evolution of plant materials for ecological restoration: insights from the applied and basic literature. *Journal of Applied Ecology*, *54*(1), 102–115. https://doi.org/10.1111/1365-2664.12739

Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology (Clifton, N.J.)*, *772*, 157–178. https://doi.org/10.1007/978-1-61779-228-1_9

Fant, J. B., Holmstrom, R. M., Sirkin, E., Etterson, J. R., & Masi, S. (2008). Genetic Structure of Threatened Native Populations and Propagules Used for Restoration in a Clonal Species, American Beachgrass (Ammophila breviligulata Fern.). *Restoration Ecology*, *16*(4), 594–603. https://doi.org/10.1111/j.1526-100X.2007.00348.x

Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R., & Fenster, C. B. (2011). Predicting the probability of outbreeding depression. *Conservation Biology: The Journal of the Society for Conservation Biology*, *25*(3), 465–475. https://doi.org/10.1111/j.1523-1739.2011.01662.x

Fukasawa, K., Miyashita, T., Hashimoto, T., Tatara, M., & Abe, S. (2013). Differential population responses of native and alien rodents to an invasive predator, habitat alteration and plant masting. *Proceedings of the Royal Society of London B: Biological Sciences*, *280*(1773), 20132075. https://doi.org/10.1098/rspb.2013.2075

Gao, S., Bertrand, D., Chia, B. K. H., & Nagarajan, N. (2016). OPERA-LG: efficient and exact scaffolding of large, repeat-rich eukaryotic genomes with performance guarantees. *Genome Biology*, *17*, 102. https://doi.org/10.1186/s13059-016-0951-y

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., … Jaffe, D. B. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences*, *108*(4), 1513–1518. https://doi.org/10.1073/pnas.1017351108

Goudet, J., & Jombart, T. (2015). hierfstat: Estimation and Tests of Hierarchical F-Statistics (Version 0.04-22). Retrieved from https://cran.r-project.org/web/packages/hierfstat/index.html

Goulet, B. E., Roda, F., & Hopkins, R. (2017). Hybridization in Plants: Old Ideas, New Techniques. *Plant Physiology*, *173*(1), 65–78. https://doi.org/10.1104/pp.16.01340

Gruenthal, K. M., Witting, D. A., Ford, T., Neuman, M. J., Williams, J. P., Pondella, D. J., … Larson, W. A. (2014). Development and application of genomic tools to the restoration of green abalone in southern California. *Conservation Genetics*, *15*(1), 109–121. https://doi.org/10.1007/s10592-013-0524-5

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., … Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, *9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Hand, R., & Gregor, T. (2006). Die Verbreitung von Arabis sagittata In Deutschland. Ergebnisse einer Herbarstudie. *Kochia*, *1*, 21–31.

Haug-Baltzell, A., Stephens, S. A., Davey, S., Scheidegger, C. E., & Lyons, E. (2017). SynMap2 and SynMap3D: web-based whole-genome synteny browsers. *Bioinformatics*, *33*(14), 2197–2198. https://doi.org/10.1093/bioinformatics/btx144

Hölzel, N., & Otte, A. (2001). The impact of flooding regime on the soil seed bank of flood-meadows. *Journal of Vegetation Science*, *12*(2), 209–218. https://doi.org/10.2307/3236605

Hölzel, N., & Otte, A. (2003). Restoration of a species-rich flood meadow by topsoil removal and diaspore transfer with plant material. *Applied Vegetation Science*, *6*(2), 131–140. https://doi.org/10.1111/j.1654-109X.2003.tb00573.x

Hölzel, N., & Otte, A. (2004). Assessing soil seed bank persistence in flood-meadows: The search for reliable traits. *Journal of Vegetation Science*, *15*(1), 93–100. https://doi.org/10.1111/j.1654-1103.2004.tb02241.x

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., … Guo, Y.-L. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics*, *43*(5), 476–481. https://doi.org/10.1038/ng.807

Hufford, K. M., & Mazer, S. J. (2003). Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends in Ecology & Evolution*, *18*(3), 147–155. https://doi.org/10.1016/S0169-5347(03)00002-8

Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*, *11*(6), 609–623. https://doi.org/10.1111/j.1461-0248.2008.01179.x

Jalas, J., & Suominen, J. (1994). *Atlas Florae Europaeae. Distribution of Vascular Plants in Europe 10: Cruciferae (Sisymbrium to Aubrieta)* (Vol. 10). The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki.

Jeffries, D. L., Copp, G. H., Handley, L. L., Olsén, K. H., Sayer, C. D., & Hänfling, B. (2016). Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, Carassius carassius, L. *Molecular Ecology*, *25*(13), 2997–3018. https://doi.org/10.1111/mec.13613

Jiao, W.-B., Accinelli, G. G., Hartwig, B., Kiefer, C., Baker, D., Severing, E., … Schneeberger, K. (2017). Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. *Genome Research*, gr.213652.116. https://doi.org/10.1101/gr.213652.116

Jombart, T., Kamvar, Z. N., Lustrik, R., Collins, C., Beugin, M.-P., Knaus, B., … Calboli, F. (2016). adegenet: Exploratory Analysis of Genetic and Genomic Data (Version 2.0.1). Retrieved from https://cran.r-project.org/web/packages/adegenet/index.html

Kahle, D., & Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, *5*(1), 144–161.

Karl, R., & Koch, M. A. (2014). Phylogenetic signatures of adaptation: The Arabis hirsuta species aggregate (Brassicaceae) revisited. *Perspectives in Plant Ecology, Evolution and Systematics*, *16*(5), 247–264. https://doi.org/10.1016/j.ppees.2014.06.001

Kiefer, M., Schmickl, R., German, D. A., Mandáková, T., Lysak, M. A., Al-Shehbaz, I. A., … Koch, M. A. (2014). BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. *Plant & Cell Physiology*, *55*(1), e3. https://doi.org/10.1093/pcp/pct158

Kiehl, K., Kirmer, A., Donath, T. W., Rasran, L., & Hölzel, N. (2010). Species introduction in restoration projects - evaluation of different techniques for the establishment of semi-natural grasslands in Central and Northwestern Europe. *Basic and Applied Ecology*, *11*(4), 285–299.

Kiehl, K., Kirmer, A., & Shaw, N. (2014). *Guidelines for Native Seed Production and Grassland Restoration*. Cambridge Scholars Publishing.

Kiehl, K., & Wagner, C. (2006). Effect of Hay Transfer on Long-Term Establishment of Vegetation and Grasshoppers on Former Arable Fields. *Restoration Ecology*, *14*(1), 157–166. https://doi.org/10.1111/j.1526-100X.2006.00116.x

Knaus, B. J., & Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., … Wilson, R. K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, *22*(3), 568–576. https://doi.org/10.1101/gr.129684.111

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 59. https://doi.org/10.1186/1471-2105-5-59

Laenen, B., Tedder, A., Nowak, M. D., Toräng, P., Wunder, J., Wötzel, S., … Slotte, T. (2018). Demography and mating system shape the genome-wide impact of purifying selection in Arabis alpina. *Proceedings of the National Academy of Sciences*, 201707492. https://doi.org/10.1073/pnas.1707492115

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

820 Lyons, E., & Freeling, M. (2008). How to usefully compare homologous plant genes and
821 chromosomes as DNA sequences. *The Plant Journal*, *53*(4), 661–673.
822 https://doi.org/10.1111/j.1365-313X.2007.03326.x
823 Lyons, E., Freeling, M., Kustu, S., & Inwood, W. (2011). Using Genomic Sequencing for Classical
824 Genetics in E. coli K12. *PLOS ONE*, *6*(2), e16717.
825 https://doi.org/10.1371/journal.pone.0016717
826 Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open
827 source ab initio eukaryotic gene-finders. *Bioinformatics (Oxford, England)*, *20*(16), 2878–
828 2879. https://doi.org/10.1093/bioinformatics/bth315
829 Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays*, *38*(2),
830 140–149. https://doi.org/10.1002/bies.201500149
831 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing
832 reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200
833 Massatti, R., Doherty, K. D., & Wood, T. E. (2018). Resolving neutral and deterministic
834 contributions to genomic structure in <Emphasis Type="Italic">Syntrichia
835 ruralis</Emphasis> (Bryophyta, Pottiaceae) informs propagule sourcing for dryland
836 restoration. *Conservation Genetics*, *19*(1), 85–97. https://doi.org/10.1007/s10592-017-
837 1026-7
838 Mattila, T. M., Tyrmi, J., Pyhäjärvi, T., & Savolainen, O. (2017). Genome-Wide Analysis of
839 Colonization History and Concomitant Selection in Arabidopsis lyrata. *Molecular Biology*
840 *and Evolution*, *34*(10), 2665–2677. https://doi.org/10.1093/molbev/msx193
841 McKay, J. K., Christian, C. E., Harrison, S., & Rice, K. J. (2005). "How Local Is Local?"—A Review of
842 Practical and Conceptual Issues in the Genetics of Restoration. *Restoration Ecology*,
843 *13*(3), 432–440. https://doi.org/10.1111/j.1526-100X.2005.00058.x
844 Mijangos, J. L., Pacioni, C., Spencer, P. B. S., & Craig, M. D. (2015). Contribution of genetics to
845 ecological restoration. *Molecular Ecology*, *24*(1), 22–37.
846 https://doi.org/10.1111/mec.12995
847 Mummenhoff, K., Franzke, A., & Koch, M. (1997). Molecular phylogenetics of Thlaspi s.l.
848 (Brassicaceae) based on chloroplast DNA restriction site variation and sequences of the
849 internal transcribed spacers of nuclear ribosomal DNA. *Canadian Journal of Botany*,
850 *75*(3), 469–482. https://doi.org/10.1139/b97-051
851 Nei, M. (1987). *Molecular evolutionary genetics*. Columbia university press.
852 Novikova, P. Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., … Nordborg, M.
853 (2016). Sequencing of the genus *Arabidopsis* identifies a complex history of
854 nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics*,
855 *48*(9), 1077–1082. https://doi.org/10.1038/ng.3617
856 O'Leary, S. J., Hollenbeck, C. M., Vega, R. R., Gold, J. R., & Portnoy, D. S. (2018). Genetic mapping
857 and comparative genomics to inform restoration enhancement and culture of southern
858 flounder, Paralichthys lethostigma. *BMC Genomics*, *19*(1), 163.
859 https://doi.org/10.1186/s12864-018-4541-0
860 Onge, K. R. S., Källman, T., Slotte, T., Lascoux, M., & Palmé, A. E. (2011). Contrasting
861 demographic history and population structure in Capsella rubella and Capsella

grandiflora, two closely related species with different mating systems. *Molecular Ecology*, *20*(16), 3306–3320. https://doi.org/10.1111/j.1365-294X.2011.05189.x

Paradis, E., Jombart, T., Schliep, K., Potts, A., & Winter, D. (2016). pegas: Population and Evolutionary Genetics Analysis System (Version 0.9). Retrieved from https://cran.r-project.org/web/packages/pegas/index.html

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Pfeifer, B., Wittelsbuerger, U., Li, H., & Handsaker, B. (2018). PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses (Version 2.6.1). Retrieved from https://CRAN.R-project.org/package=PopGenome

Purcell, S. (2009). PLINK (Version 1.07).

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reed, D. H., & Frankham, R. (2003). Correlation between Fitness and Genetic Diversity. *Conservation Biology*, *17*(1), 230–237. https://doi.org/10.1046/j.1523-1739.2003.01236.x

Reitzel, A. M., Herrera, S., Layden, M. J., Martindale, M. Q., & Shank, T. M. (2013). Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology*, *22*(11), 2953–2970. https://doi.org/10.1111/mec.12228

Reusch, T. B. H., Ehlers, A., Hämmerli, A., & Worm, B. (2005). Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(8), 2826–2831. https://doi.org/10.1073/pnas.0500008102

Reynolds, L. K., McGlathery, K. J., & Waycott, M. (2012). Genetic Diversity Enhances Restoration Success by Augmenting Ecosystem Services. *PLOS ONE*, *7*(6), e38397. https://doi.org/10.1371/journal.pone.0038397

Roberts, L., Stone, R., & Sugden, A. (2009). The Rise of Restoration Ecology. *Science*, *325*(5940), 555–555. https://doi.org/10.1126/science.325_555

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, *12*(12), 2640–2659. https://doi.org/10.1038/nprot.2017.123

Scheepens, J. F., Rauschkolb, R., Ziegler, R., Schroth, V., & Bossdorf, O. (2017). Genotypic diversity and environmental variability affect the invasibility of experimental plant populations. *Oikos*, *127*(4), 570–578. https://doi.org/10.1111/oik.04818

Schnittler, M., & Günther, K.-F. (1999). Central European vascular plants requiring priority conservation measures – an analysis from national Red Lists and distribution maps. *Biodiversity & Conservation*, *8*(7), 891–925. https://doi.org/10.1023/A:1008828704456

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2016). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*(8), 907–917. https://doi.org/10.1111/2041-210X.12700

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31. https://doi.org/10.1186/1471-2105-6-31

Smit, A., & Hubley, R. (2008). RepeatModeler Open-1.0 (Version 1.0). Retrieved from http://www.repeatmasker.org

Smit, A., Hubley, R., & Green, P. (2013). *RepeatMasker Open-4.0*. Retrieved from http://www.repeatmasker.org

Society for Ecological Restoration International Science & Policy Working Group. (2004). *The SER International Primer on Ecological Restoration*. www.ser.org & Tucson: Society for Ecological Restoration Internationa.

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics (Oxford, England)*, *19 Suppl 2*, ii215-225.

Suarez-Gonzalez, A., Lexer, C., & Cronk, Q. C. B. (2018). Adaptive introgression: a plant perspective. *Biology Letters*, *14*(3), 20170688. https://doi.org/10.1098/rsbl.2017.0688

Suding, K. N. (2011). Toward an Era of Restoration in Ecology: Successes, Failures, and Opportunities Ahead. *Annual Review of Ecology, Evolution, and Systematics*, *42*(1), 465–487. https://doi.org/10.1146/annurev-ecolsys-102710-145115

Tellier, A., Laurent, S. J. Y., Lainer, H., Pavlidis, P., & Stephan, W. (2011). Inference of seed bank parameters in two wild tomato species using ecological and genetic data. *Proceedings of the National Academy of Sciences*, *108*(41), 17052–17057. https://doi.org/10.1073/pnas.1111266108

Török, P., Miglécz, T., Valkó, O., Kelemen, A., Tóth, K., Lengyel, S., & Tóthmérész, B. (2012). Fast restoration of grassland vegetation by a combination of seed mixture sowing and low-diversity hay transfer. *Ecological Engineering*, *44*, 133–138. https://doi.org/10.1016/j.ecoleng.2012.03.010

Torres-Martinez, L., & Emery, N. C. (2016). Genome-wide SNP discovery in the annual herb, Lasthenia fremontii (Asteraceae): genetic resources for the conservation and restoration of a California vernal pool endemic. *Conservation Genetics Resources*, *8*(2), 145–158. https://doi.org/10.1007/s12686-016-0524-0

Tyrrell, M. C., & Byers, J. E. (2007). Do artificial substrates favor nonindigenous fouling species over native species? *Journal of Experimental Marine Biology and Ecology*, *342*(1), 54–60. https://doi.org/10.1016/j.jembe.2006.10.014

Vandepitte, K., Gristina, A. S., Hert, K. D., Meekers, T., Roldán-Ruiz, I., & Honnay, O. (2012). Recolonization after habitat restoration leads to decreased genetic variation in populations of a terrestrial orchid. *Molecular Ecology*, *21*(17), 4206–4215. https://doi.org/10.1111/j.1365-294X.2012.05698.x

Vander Mijnsbrugge, K., Bischoff, A., & Smith, B. (2010). A question of origin: Where and how to collect seed for ecological restoration. *Basic and Applied Ecology*, *11*(4), 300–311. https://doi.org/10.1016/j.baae.2009.09.002

Vrijenhoek, R. C. (1994). Genetic diversity and fitness in small populations. In V. Loeschcke, S. K. Jain, & J. Tomiuk (Eds.), *Conservation Genetics* (pp. 37–53). Basel: Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-8510-2_5

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., … Hoffmann, A. A. (2011). Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications*, *4*(6), 709–725. https://doi.org/10.1111/j.1752-4571.2011.00192.x

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Williams, A. V., Nevill, P. G., & Krauss, S. L. (2014). Next generation restoration genetics: applications and opportunities. *Trends in Plant Science*, *19*(8), 529–537. https://doi.org/10.1016/j.tplants.2014.03.011

Willing, E.-M., Rawat, V., Mandáková, T., Maumus, F., James, G. V., Nordström, K. J. V., … Schneeberger, K. (2015). Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. *Nature Plants*, *1*(2), 14023. https://doi.org/10.1038/nplants.2014.23

Zhao, S., Guo, Y., Sheng, Q., & Shyr, Y. (2015). heatmap3: An Improved Heatmap Package (Version 1.1.1). Retrieved from https://cran.r-project.org/web/packages/heatmap3/index.html

Data availability

The genome assembly, annotation, raw reads and RAD-seq samples will be uploaded to EMBL ENA upon acceptance (available upon request for review). VCF files and custom script will be uploaded to Dryad repository upon acceptance (available upon request for review). An R Markdown file describing the population genetic analysis is provided in the supplemental material.

Author contributions

JdM, HD, AT and NH conceived the study. HD and NH collected plant material; HD and CB prepared material for sequencing; WBJ, KS and HD were responsible for bioinformatic

processing. JdM, HD and NH analyzed data and wrote the manuscript with significant contributions from CB, WBJ, KS and AT.