

RESEARCH ARTICLE

Open Access



RAD-seq reveals genetic structure of the F₂-generation of natural willow hybrids (*Salix* L.) and a great potential for interspecific introgression

Susanne Gramlich , Natascha Dorothea Wagner and Elvira Hörandl

Abstract

Background: Hybridization of species with porous genomes can eventually lead to introgression via repeated backcrossing. The potential for introgression between species is reflected by the extent of segregation distortion in later generation hybrids. Here we studied a population of hybrids between *Salix purpurea* and *S. helvetica* that has emerged within the last 30 years on a glacier forefield in the European Alps due to secondary contact of the parental species. We used 5758 biallelic SNPs produced by RAD sequencing with the aim to ascertain the predominance of backcrosses (F₁ hybrid × parent) or F₂ hybrids (F₁ hybrid × F₁ hybrid) among hybrid offspring. Further, the SNPs were used to study segregation distortion in the second hybrid generation.

Results: The analyses in STRUCTURE and NewHybrids revealed that the population consisted of parents and F₁ hybrids, whereas hybrid offspring consisted mainly of backcrosses to either parental species, but also some F₂ hybrids. Although there was a clear genetic differentiation between *S. purpurea* and *S. helvetica* (F_{ST} = 0.24), there was no significant segregation distortion in the backcrosses or the F₂ hybrids. Plant height of the backcrosses resembled the respective parental species, whereas F₂ hybrids were more similar to the subalpine *S. helvetica*.

Conclusions: The co-occurrence of the parental species and the hybrids on the glacier forefield, the high frequency of backcrossing, and the low resistance to gene flow via backcrossing make a scenario of introgression in this young hybrid population highly likely, potentially leading to the transfer of adaptive traits. We further suggest that this willow hybrid population may serve as a model for the evolutionary processes initiated by recent global warming.

Keywords: Population genomics, Hybrid evolution, Population structure, Sex chromosomes, Climate change

Background

Natural hybridization due to secondary contact has been observed in many plant and animal species. Especially in North America and Northern and Central Europe, a major driving force for secondary contact is the ongoing recolonization after the retreat of glaciers [1]. This process is amplified by human-induced global warming that also causes rapid range shifts of species [2–5], especially in mountain regions, leading to secondary contact

of previously allopatric species [6]. The absence of strong pre- or postzygotic reproductive barriers may then lead to hybridization.

Many studies have investigated the evolutionary relevance of hybridization. Although there are some documented cases of homoploid hybrid speciation [7–9], speciation seems to be a rather rare outcome compared to the plenty of reported incidents of hybridization [10]. Important requirements for homoploid hybrid speciation seem to be strong ecological or geographical barriers that restrict gene flow between hybrids and the parental species [7, 11–13]. Thus, hybrid speciation is closely connected to the availability of novel or extreme habitats

* Correspondence: susanne.gramlich@biologie.uni-goettingen.de
Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), University Goettingen, Untere Karspüle 2, 37073 Goettingen, Germany



[14, 15]. Chromosomal rearrangements can also rapidly establish crossing barriers between parents and hybrids [9]. Generally, interspecific hybridization seems more likely to result in introgression than in speciation [7, 14]. In extreme cases, introgression can lead to genetic swamping threatening species integrity and posing a severe problem especially in small populations or rare species [16]. On the other hand, adaptive introgression can lead to the transfer of favourable alleles [10, 13, 17–19]. Introgression of favourable traits can increase the species' genetic and phenotypic diversity, and hence the potential to adapt to novel environments [9, 20]. The outcome of an incipient hybridization event is not easy to predict because it depends on many factors like the fitness of the hybrids and their offspring [21], the impact of endogenous and exogenous selection, interactions of certain genotypes with the environment [22], and habitat availability [23].

To assess the evolutionary impact of a hybridization event, it is crucial to know the extent to which a genome is susceptible to the introgression of heterospecific alleles. Segregation distortion, the deviation from expected Mendelian segregation ratios, can be used as a measure of the resistance of the hybridizing species' genomes to introgression [24, 25]. Further, it can be assumed that distorted loci are linked to genes that affect the viability or fitness of hybrids or their gametes [25, 26]. Thus, segregation distortion is also connected to reproductive barriers and the suppression of interspecific gene flow [27–29]. Segregation distortion can also arise from low recombination rates on sex chromosomes or in sex-determining regions [30, 31]. In dioecious plants, female-biased sex ratios are connected to segregation distortion at distorter loci [32]. The search for loci or regions under segregation distortion has therefore been applied, even in nonmodel species, as a basis to draw conclusions about the potential underlying causes of reproductive barriers between species [33–37], or to identify loci responsible for environmental adaptation [29, 38].

In this study, we investigate hybridization in a zone of secondary contact between two willow species, *Salix purpurea* L. and *S. helvetica* Vill., which are situated on the forefield of the Rhône Glacier in central Switzerland. *Salix helvetica* is a shrub that occurs naturally in the subalpine to alpine zone. *Salix purpurea*, on the other hand, is a widespread lowland species that was recently able to colonize higher altitudes due to global warming and subsequent glacier retreat [39]. Secondary contact and hybridization of these species takes place on glacier forefields that have recently become ice free. These glacier forefields are covered with sparse vegetation and offer plenty of space and different niches for the settlement of pioneer species like willows [39–41]. Hybridization even across sections and between distantly related species is a

common phenomenon in *Salix* [42, 43]. Although *S. purpurea* and *S. helvetica* belong to different sections of the genus [44], they form natural hybrid zones in the European Alps [39]. The composition of such hybrid zones would provide important clues for an assessment of the evolutionary consequences of these hybridization events. A predominance of backcrosses would render introgression of genes between the parental species more likely, whereas the domination of F_2 hybrids (i.e. F_1 hybrid \times F_1 hybrid) might be an indication of the potential for hybrid swarm formation and further hybrid evolution. In an earlier study on the willow population at the Rhône Glacier, an attempt has been made to determine the exact class of the hybrids (F_1 , F_2 , backcrosses) on the glacier forefield based on genotyping with microsatellite markers [39]. These markers clearly separated the two parental species and confirmed the hybrid origin of phenotypically intermediate individuals, but their resolution was not sufficient for an unequivocal assignment of all individuals to a certain hybrid class [39]. Thus, the precise composition of the hybrid zone is still uncertain. However, we found that the hybrids between *S. purpurea* and *S. helvetica* are fertile and produce viable seeds in the natural population, and thus confirmed that hybridization can proceed beyond the F_1 hybrid generation [45]. The offspring raised from these naturally formed seeds offered the opportunity to study not only progeny classes of second generation hybrids, but also putative segregation distortion and phenotypic traits.

In order to overcome the limitations caused by a low number of markers, we used restriction-site associated DNA sequencing (RAD-seq) to generate a genome-wide set of thousands of single-nucleotide polymorphisms in this nonmodel species. High-quality biallelic SNPs were used to (i) determine the class of the hybrids on the glacier forefield and of offspring produced by F_1 hybrids in order to predict the consequences of this hybridization event. Further, (ii) we were looking for deviations from expected segregation patterns at individual loci in F_2 hybrids and backcrosses to determine if alleles of one parental species were favoured over the other. Population genomic analyses were accompanied by morphometric measurements to (iii) get insights into the variation of a selected, potentially adaptive phenotypic character (plant height) in the respective second generation hybrid classes.

Results

RAD-seq and SNP calling

RAD-seq of *S. purpurea*, *S. helvetica* and their hybrids yielded an average of 7.5×10^6 reads per individual (SD 2.4×10^6). The average per base sequence quality was very high with a Phred score of 40 for all positions in the reads in all samples. The average depth of read coverage was 59x (SD 18). The STACKS-pipeline initially

generated 49,081 loci. After the application of all filters, 5758 nuclear loci remained. After mapping the loci to the plastid genome no match was observed. However, of the filtered loci 933 reads aligned to coding regions in the genome of *P. trichocarpa*. The results of population genetic and progeny analyses did not change when the SNPs lying in coding regions were excluded, and thus we performed all analyses with all 5758 SNPs.

Population genetic structure

The STRUCTURE analysis confirmed that the most likely number of populations (K) in the sample was two (Additional file 1: Figure S1). It can thus be assumed that all hybrids were crosses between *S. purpurea* and *S. helvetica* with no third species involved (Fig. 1a). The F_{ST} value between *S. purpurea* and *S. helvetica* was 0.53 for the filtered loci, but only 0.24 when all 49,081 unfiltered loci were considered, indicating a strong dependence of F_{ST} values on locus selection. The NewHybrids analysis confirmed the parental classes and revealed that all hybrid individuals sampled on the glacier forefield of the Rhône Glacier were F_1 hybrids. Although the dataset had to be restricted to 300 loci because NewHybrids cannot handle a larger number of loci, the results for the assignment to the different classes had 100% posterior probability support for all individuals. Accordingly, *S. purpurea* and *S. helvetica* were clearly separated along the first axis in the PCoA (Fig. 2) with the hybrid individuals sampled at the Rhône Glacier clustering exactly in the middle along the second axis.

The offspring of the F_1 hybrids consisted of backcrosses to *S. purpurea*, backcrosses to *S. helvetica*, and of F_2 hybrids (crosses between F_1 hybrids). F_2 hybrids were only observed among the offspring of one of the five mother-plants (Fig. 1b). All other plants produced backcrosses in both directions. Overall, significantly more backcrosses to *S. purpurea* ($n=40$) than to *S. helvetica* ($n=16$) were detected (binomial test, 2-sided, $p=0.001$, $n=57$). The different hybrid classes also formed well-separated clusters in the PCoA analysis (Fig. 2). The backcrosses clustered in the respective parental half without overlapping with the purebred individuals. The F_2 hybrids formed a cluster of their own, clearly separated from the other individuals along the second axis. The F_1 individual that clusters with the F_2 hybrids is the mother of the F_2 hybrids.

Segregation distortion in the second generation hybrids

The filtering of all loci for $F_{ST}=1$ between *S. purpurea* and *S. helvetica* retrieved 396 species-specific SNPs, which included SNPs in coding regions of the genome. Of these, 334 loci could be aligned to the *S. purpurea* genome. The alignment showed that they were represented on all 19 chromosomes of the *Salix* genome (Table 1).

There were no significant deviations from the expected segregation patterns neither in the backcrosses nor in the F_2 hybrids. Although the distribution of homozygous and heterozygous genotypes was skewed at some loci in

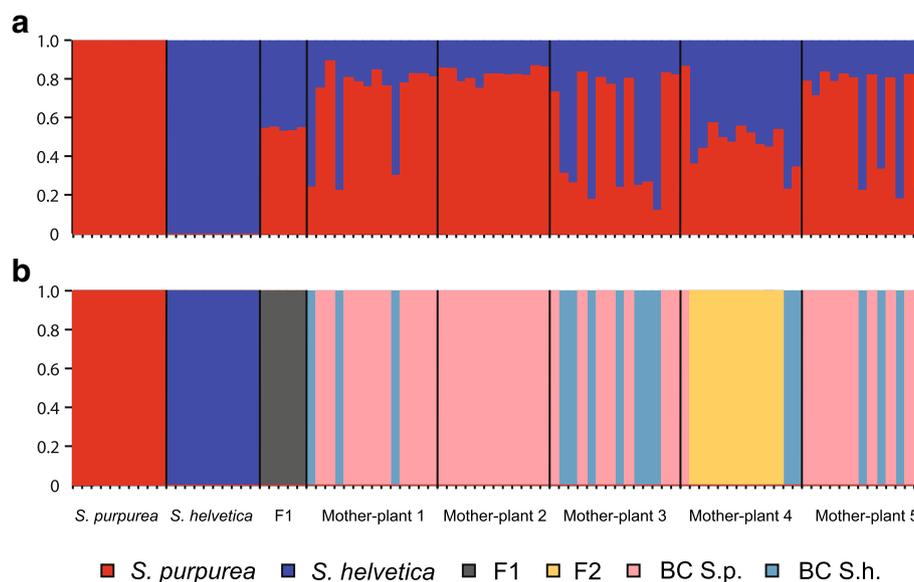
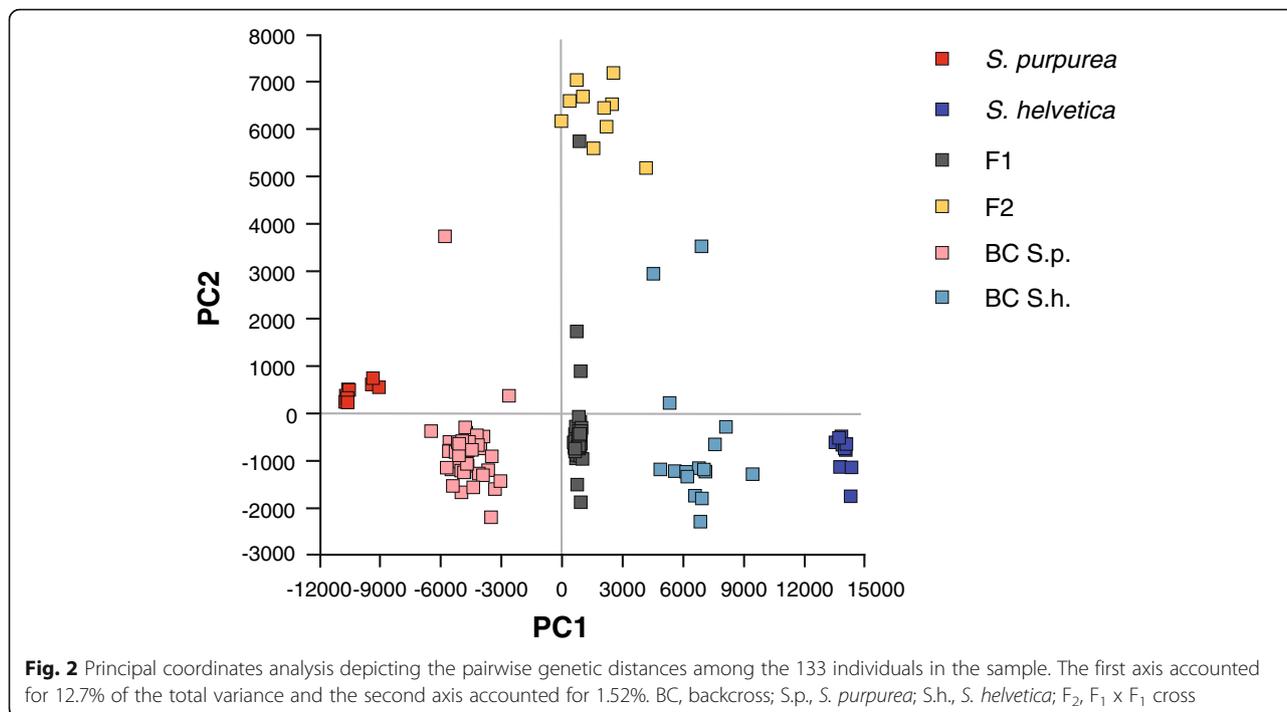


Fig. 1 Genetic structure of parental species and their hybrids. Admixture proportions (q_i) based on two clusters ($K=2$) estimated by STRUCTURE (a) and assignment of hybrid class in NewHybrids (b). The plots include the five F_1 mother-plants representing all 45 F_1 hybrid individuals in the sample (for admixture proportions and assignment of hybrid class of the remaining 40 hybrids see Additional file 1: Table S1 and S2). The offspring of the mother-plants is sorted by mother-plant. The order of the individuals within the different groups is the same in both plots



the F₂ hybrids (Additional file 2: Table S3), the number of individuals seemed to be too low to support significant deviations.

Table 1 Distribution of 334 species specific SNPs detected in *S. purpurea* and *S. helvetica* on the 19 chromosomes of *S. purpurea*

Chromosome	Nr of loci
I	26
II	37
III	24
IV	9
V	26
VI	35
VII	16
VIII	21
IX	25
X	35
XI	6
XII	5
XIII	19
XIV	18
XV	2
XVI	20
XVII	3
XVIII	5
XIX	2

Plant height in the second generation hybrids

The one-way ANOVA revealed significant differences in plant height between the three groups ($F_{2,67} = 15.064$, $MS = 219.71$, $p < 0.001$, Fig. 3). Backcrosses to *S. purpurea* ranged from 15 to 70 cm ($M = 41.8$ cm, $SD = 14.2$) and were significantly taller than backcrosses to *S. helvetica* (12.5–38 cm, $M = 24$ cm, $SD = 6.3$) or F₂ hybrids (16–46.5 cm, $M = 29.4$ cm, $SD = 10.2$).

Discussion

Composition of the natural hybrid population

Our study confirms the existence of a natural secondary contact hybrid zone of *S. helvetica* and *S. purpurea* on the Rhône glacier forefield. RAD-seq data enabled a much better resolution of the hybrid classes than the microsatellite markers applied in our previous study [39]. While the results based on genotyping with microsatellites had suggested that two-thirds of the hybrids were probably later generation hybrids (F₂ hybrids or backcrosses), the analysis of the same hybrids with RAD-seq data showed with maximum statistical support that all hybrid individuals sampled on the forefield of the Rhône Glacier were F₁ hybrids. This discrepancy in the hybrid classification is probably due to the low number of microsatellite loci used in the previous study that had to be restricted to primers amplifying in both

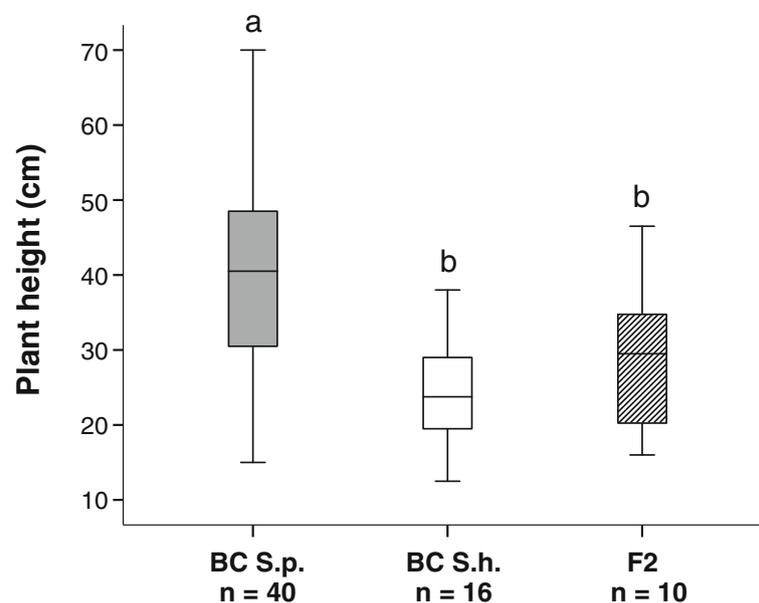


Fig. 3 Boxplots of the plant height of backcrosses to *S. purpurea* (BC S.p.), backcrosses to *S. helvetica* (BC S.h.), and F₂ hybrids (F₁ × F₁). Significant differences between groups are indicated by letters. The median is indicated by the black horizontal line. The bottom and top of the boxes show the 25 and 75 percentiles, respectively. The whiskers extend to the highest and lowest value that is not an outlier

species. Further, the lack of species-specific alleles at the microsatellite loci made it difficult to determine from which species an allele was inherited in the admixed genotypes. Similar discrepancies in the classification of hybrids were also observed in studies on *Populus* hybrid zones. While no hybrids were classified as F₁ hybrids based on genotyping with microsatellites [46], subsequent genotyping-by-sequencing revealed that most hybrids belonged to the F₁ generation [47]. The authors explained the different results with the shortcomings of microsatellites, like allele dropout [47]. We think that the results based on RAD-seq data are more reliable due to the large number of loci that produce a much higher resolution than the DNA fingerprinting techniques used so far [48]. Further, all hybrids were assigned to the respective hybrid class with 100% posterior probability in the RAD-seq analysis, whereas the SSR analysis gained less than 95% posterior probability for many individuals. Altogether RAD-seq clearly performs better for studies on interspecific hybridization than microsatellites.

We expected to find some later generation hybrids on the glacier forefield because the analysis of offspring grown from seeds that had been collected from naturally pollinated F₁ hybrids at the forefield of the Rhône Glacier [45] suggested that backcrosses and F₂ offspring can be regularly formed in this population. However, the results revealed only F₁ generation hybrids. We assume that this lack of second generation hybrids is due to the sampling strategy. We included only material from adult individuals because the leaves, flowers and fruits of

juvenile plants and seedlings did not yet show the typical phenotypic characteristics of adult willows. In a previous study, we already concluded that the hybrid population on this recently emerged glacier forefield is probably not more than 20–30 years old [39]. We thus believe that second generation hybrids on the glacier forefield were not yet present or very rare in the adult generation during our sampling. In order to clarify whether backcrosses and F₂ hybrids have meanwhile grown up in the natural population, an extended sampling strategy that includes juvenile plants and a broad representation of phenotypes would have to be applied.

We have no reason to assume that backcrosses to either parent as well as F₂ hybrids cannot establish on the glacier forefield. In a study on hybrid fertility, we found that the seed output of hybrids was reduced compared to the parents, but that seeds showed high germination ability, and seedlings developed well [45]. Thus it seems unlikely that there are no backcrosses or F₂ hybrids on the glacier forefield, although their numbers may probably be still lower compared to F₁ hybrids. Alternatively, habitat mediated selection may act against the establishment of later generation hybrids on the parental sites, as it has been observed in *Rhododendron* hybrid zones [49]. At the time the first F₁ hybrids were formed, the glacier forefield was still in an early state of succession with less vegetation cover so that the conditions were more favourable for the establishment of willows, which are pioneer species. Later in time, when the backcrosses and F₂ hybrids were produced, the vegetation may have been

denser so that the conditions for the establishment may have become more difficult. However, it should also be kept in mind that glacier retreat is ongoing and that open pioneer sites for colonization will be continuously available. Although we did not find later generation hybrids in our present sampling, it may still be concluded that the hybrid population on the glacier forefield is able to develop beyond the F_1 generation so that hybridization may have further consequences as discussed in the next section.

Classes of hybrid offspring

In contrast to the findings made in the natural population on the glacier forefield, the offspring raised from seeds formed in the wild consisted of F_2 hybrids (F_1 crossed with F_1) and backcrosses to both parental species. Interestingly, F_2 hybrids were only found among the offspring of one of the five mother-plants. On the glacier forefield, this female plant stands less than three metres away from a male F_1 hybrid (S. Gramlich, unpublished observation). It thus seems that F_2 hybrids are only produced in high numbers, when male and female F_1 hybrids stand close together. This arrangement is quite rare on the glacier forefield so that female F_1 hybrids are closer to male individuals of *S. purpurea* or *S. helvetica* in most cases (S. Gramlich, unpublished observation). The parental species and the hybrids occur evenly dispersed over the whole area so that there is no spatial structure like a clumped distribution or a cline from one parental species to the other [39]. Therefore it is more likely that female F_1 hybrids are pollinated by one of the purebred species so that they will produce backcrosses. Accordingly, our results showed that the offspring of the sampled F_1 hybrids consisted mainly of backcrosses to the parental species and only few F_2 hybrids. Similar findings were made in poplars where purebred female plants produced exceptionally large amounts of backcross seedlings when they were surrounded by F_1 hybrids [50]. Overall, there were more backcrosses to *S. purpurea* than to *S. helvetica* in the sample, yet the reason for this result is unclear. Possible causes could be a greater overlap of flowering time between *S. purpurea* and the hybrids, postzygotic selection against backcrosses to *S. helvetica*, stochastic factors like a closer proximity between male *S. purpurea* and female F_1 hybrids than between *S. helvetica* males and F_1 hybrids, or sampling bias due to the choice of mother plants in the analysis, the limited number of progenies, or conditions for pollination in the year of sampling. Another possible interpretation could be that pollen limitation is stronger for pollen of *S. helvetica* than for pollen of *S. purpurea*. Accordingly, a reduced seed set was found in purebred *S. helvetica* compared to purebred *S. purpurea* [45]. The seed set in willows is pollen

limited [51], and thus the pollen of *S. helvetica* could also be transported less efficiently. Pollen-pistil incongruences also represent a strong prezygotic crossing barrier in willows [52]. Pollen tube growth could act differentially between *S. purpurea* and *S. helvetica*, as the former species has a much shorter style and a capsule without a beak, while the latter has long styles and beaked capsules. However, the determination of the exact causes of the observed pattern requires further research. Irrespective of the direction of backcrossing it can be concluded that, at least in this early stage of secondary contact, a higher production of backcrosses than of F_2 hybrids renders a future trajectory of introgression more likely than hybrid speciation.

Second generation hybrids show no signs of segregation distortion

This is one of the first studies on nonmodel plant species where segregation distortion was analyzed with RAD-seq data. The power of this marker system for detecting segregation distortion and linkage groups was demonstrated e.g. on hybrid fish [29] and on white cypress pine [38]. The fact that there are only two alleles per locus makes it difficult to determine the species of origin of an allele in a hybrid individual, especially when the alleles are evenly distributed in the parental species. Therefore we restricted our analyses to species-specific loci. We believe that this subsampling is representative of genome-wide patterns of hybridization because the markers are located on all 19 chromosomes of the *Salix* genome.

We did not detect significant deviations from the expected Mendelian segregation ratios in the F_2 hybrids or backcrosses. It is expected that the magnitude of distorted loci correlates with the level of divergence between the parental species [53]. The divergence between *S. purpurea* and *S. helvetica* turned out to be quite low with a F_{ST} value of 0.3 in our study based on microsatellite makers [39]. Divergence based on RAD-seq loci is also quite low with an F_{ST} value of 0.24 for completely unfiltered loci, but moderate for the filtered loci ($F_{ST} = 0.53$). This increase of F_{ST} is thought to be due to the removal of loci that do not discriminate the parental species. Thus we think that the low F_{ST} value based on the unfiltered loci gives a more realistic estimate of the population divergence. Another hint for the low divergence between *S. purpurea* and *S. helvetica* is that they hybridize easily although they belong to different, unrelated sections or clades of the genus *Salix* [44, 54]. However, phylogenetic studies showed in general a low genetic divergence between species and sections in the genus *Salix*, especially in the shrub species [55–57]. Recently, the phylogeny of the whole subgenus comprising the shrub willows could be resolved using RAD sequencing while more conservative

markers had failed [54]. Thus, a low genetic divergence between the parental species seems to be a likely explanation for the absence of segregation distortion in the hybrids.

Genetic incompatibilities that cause hybrid sterility or inviability and thus act as postzygotic reproductive barriers accumulate with evolutionary divergence of the parental species [58, 59]. We observed that F_1 hybrids produced less seeds than *S. purpurea* or *S. helvetica* but that the seeds they produced were viable and developed equally well as seedlings from the purebred species [45]. Due to the shallow genetic divergence of the parental species there seems to be a certain degree of postzygotic (i.e. intrinsic) selection before seed maturation during meiosis, pollination, fertilization or seed development, but the absence of segregation distortion suggests that heterospecific alleles are not selectively purged. Because large parts of the genome seem to be unaffected by segregation distortion, it can be assumed that the genome is susceptible to introgression, as it was also concluded for backcrosses in *Iris* [25].

Another interesting finding is that only two species-specific loci are located on chromosome XV that carries the sex determination locus in *Salix* [60]. Other studies found that sex chromosomes were highly divergent due to suppressed recombination and the accumulation of species-specific differences [30]. However, in *Salix*, as well as in *Populus*, no heteromorphic sex chromosomes have been discovered yet. Stölting et al. [61] did also not identify fixed SNPs between *Populus* species on chromosome XIX that carries the sex determining locus in *Populus* [62–64]. They concluded that, against the predictions, the incipient sex chromosome of *Populus* is not resistant to gene flow and introgression. Accordingly, Macaya-Sanz et al. [65] also detected gene flow on chromosome XIX in *Populus*. This finding seems to be reflected in *Salix* due to the low number of species-specific SNPs on chromosome XV detected in this study.

In contrast to the genomic data, segregation became obvious in phenotypic traits in one-year old juvenile plants. *Salix helvetica* is a shrub that reaches a height of ca. 50–80 cm [43]. *Salix purpurea* can reach up to 6 m in the lowland [66] but on the glacier forefield the shrubs were ca. 160–180 cm high (S. Gramlich, unpublished observation). With respect to plant height, the backcrosses seem to keep the traits of the recurrent parent, as expected, whereas the F_2 hybrids adopted the lower height of *S. helvetica*, even in the absence of external selection under equal garden conditions. This is striking because a typical feature for alpine shrubs is the reduction of plant height (typically < 50 cm) as the plants are better protected by snow cover during freezing periods [54, 67]. Growth height thus appears to be a

promising candidate for studying an adaptive trait in this hybrid system.

Conclusions

Range shifts initiated by climate change will increase the likelihood of secondary contact hybridization in some species [68]. Comparisons of the outcome of diverse hybridization events induced by climate change are important in order to assess the effects of such events on biodiversity so that conservation measures can be initiated if necessary [69]. Which effect will the hybridization event have on the genetic diversity of the hybridizing willow species? We found that introgression is highly likely because intrinsic barriers against hybridization and gene flow between *S. purpurea* and *S. helvetica* are low. Further, hybrids and the purebred species occur in a mixed stand on the glacier forefield leading to a continuing formation of F_1 hybrids and backcrosses. Introgression might be asymmetric because there were more backcrosses to *S. purpurea* in the sampling. Due to the isolated location, introgression might be highly localized affecting mainly the gene pools of *S. purpurea* and *S. helvetica* individuals on the glacier forefield and the surrounding slopes. On the other hand, we already discovered another population of hybrids between *S. purpurea* and *S. helvetica* at a higher altitude on the Morteratsch glacier [39], and thus it can be assumed that further hybrid populations will emerge at other locations in the European Alps due to the ongoing retreat of glaciers. Many localized, independent hybridization events would make a wider distribution of introgressed alleles more likely.

In general, hybridization appears to increase genetic and phenotypic variability in the offspring population. Interspecific exchange of genes via introgression is considered to be an important evolutionary force because it may lead to the transfer of adaptations [70]. Adaptation is viewed as the most important process that promotes divergence during speciation [13, 71]. In this way, introgression of adaptive traits could lead to the formation of ecotypes or even new species [70]. However, the long generation turnover of shrubs and the time needed to establish populations makes it difficult to predict the adaptive value of traits. Long term monitoring of such hybrid populations is essential to draw final conclusions.

Willow hybrids may also serve as models for the evolutionary processes initiated by global warming. Due to their properties as pioneer species, range shift and establishment of willows in novel habitats may be more rapid than in other species. The observations made in this model system may thus help to anticipate evolutionary processes that might affect species with lower dispersal rates much later in time.

Methods

Sampling

Leaf samples were collected at the forefield of the Rhône Glacier in central Switzerland (46°34′03.0″N, 08°22′12.3″E) from a mixed stand of *S. purpurea*, *S. helvetica*, and their hybrids. All plant samples were collected with the permission of the Canton du Valais, Service des forêts et du paysage. All individuals have already been genotyped at nine microsatellite loci in a previous study, which also included some reference populations sampled outside the glacier forefield [39]. For the present study, we sampled leaves from six individuals of *Salix purpurea* and nine individuals of *S. helvetica* from the glacier forefield. To extend the data of the purebred species, we also included four individuals of *S. purpurea* from three additional locations in Germany (51°18′53.0″N, 11°54′19.8″E; 51°44′32.2″N, 10°43′31.8″E; 49°21′13.0″N, 8°14′15.0″E) and one *S. helvetica* individual from Austria (46°49′21.6″N, 10°59′25.0″E).

We included a comprehensive sampling of 45 hybrids from the glacier forefield. These plants were classified as hybrids by both an intermediate phenotype between the parental species and genetic analysis that had been conducted in a previous study [39]. The identification of all specimens was done by S. Gramlich. Herbarium vouchers of each purebred and hybrid sample were deposited in the herbarium of the University of Göttingen (GOET).

Among these 45 hybrids, we selected five hybrids that had a >95% probability of being a F₁ hybrid in the NewHybrids analysis of our previous study [39]. To investigate the second hybrid generation, seeds that had been collected from these five naturally pollinated F₁ hybrids at the forefield of the Rhône Glacier were germinated under controlled conditions (for details see [45]). Seedlings were grown for 1 year under equal conditions in climate growth chambers (see below). Out of hundreds of juvenile plants, a subset of 20–30 progenies per mother plant was selected that represented the phenotypic diversity among the offspring. From each of five F₁ mother plants 13–14 progeny (overall $n = 68$) were finally sampled for RAD-seq analysis. The five mother plants from the natural population at the Rhône Glacier were also included in the sampling. The final dataset for RAD-seq analysis comprised 133 individuals.

DNA extraction, RAD-seq

DNA was extracted from silica-dried leaves using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The DNA concentration was assessed with a Qubit 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and the samples were normalized to a concentration of 30 ng/μl.

Aliquots of 3 μg DNA were then submitted to Florigenex Inc. (Eugene, OR, USA) for library preparation and single-end RAD sequencing (following the protocol of [72]). The total genomic DNA was digested with the restriction enzyme *Pst*I. The size selection of 300 bp – 500 bp with a Pippin Prep (Sage Science, Beverly, MA, USA) was followed by the ligation of sequencing adaptors and unique 10 bp barcodes for each sample. The samples were sequenced on an Illumina HiSeq 2500 Instrument (Illumina Inc., San Diego, CA, USA) and raw reads were delivered in FASTQ format trimmed to 100 bp.

Bioinformatic analysis of raw data and SNP calling

The software STACKS v. 1.44 [73, 74] was used for demultiplexing, SNP discovery, and genotyping. First, we demultiplexed the raw reads and removed low-quality reads using the *process_radtags* program implemented in STACKS with the default parameters. In this step, the 10 bp barcodes were removed from the reads so that the final length of the reads was shortened to 90 bp. Afterwards, the quality of each sample was assessed with FastQC v. 0.11.4 [75]. Loci were assembled de novo using the *denovo_map* pipeline that merges RAD-tags into loci in each sample (*ustacks*), creates a catalog containing the merged loci from multiple samples (*cstacks*), and finally matches the loci from each sample against the catalog (*sstacks*). The first step within the pipeline is the creation of so called stacks (matching reads) out of the raw reads of each individual. The stacks provide the basis for building loci [73]. The minimum number of matching raw reads (minimum depth of coverage) required to create a stack (*m*) was set to 10. Thus, calling of heterozygotes requires at least 10 reads of each allele. The maximum number of nucleotide mismatches allowed between two stacks was set to 5 both for processing loci within individuals (*M*) and between individuals (*n*) when building the catalog. Appropriate values for *M* and *n* were determined in preliminary test runs. In these runs, *M* and *n* had the same value set between 2 and 7, and the total number of loci as well as the number of polymorphic loci was recorded. Following the recommendations of Viricel et al. [76] we chose the set of parameters where the total number of loci as well as the number of polymorphic loci reached an asymptote. Finally we used the *populations* program in STACKS to extract loci that were present in all three groups (*S. purpurea*, *S. helvetica*, hybrids) in at least 70% of the individuals in each group (*r*). Data analysis was restricted to the first SNP at each locus in order to obtain unlinked SNPs required for population genetic analysis. PLINK v 1.0.7 [77] was used to filter out SNPs with a minor allele frequency < 0.05, and SNPs that were out of the Hardy–Weinberg equilibrium at $p < 0.05$ in one or both parental

populations. Further, individuals with a genotyping rate < 90% were also excluded from the analysis leading to the exclusion of two individuals of the hybrid offspring. We also applied further filters to remove potentially paralogous loci resulting from the recent ‘salicoid’ duplication event [78]. Collapsed paralogous copies at such loci should be characterized by an excess of heterozygosity and an increased coverage depth [79]. Thus, we removed loci with an observed heterozygosity (H_o) ≥ 0.6 , $F_{IS} < 0$, or a coverage depth that was greater than twice the standard deviation. Further, we also removed loci with a F_{ST} -value of 0 between *S. purpurea* and *S. helvetica* in order to restrict the analyses to loci with variation between the parental species. H_o , F_{IS} , and F_{ST} -values for each locus were generated by the *populations* program. The filtered loci were aligned to the *Populus trichocarpa* genome [78] to detect SNPs that were in coding, putative highly conservative regions, and to the plastome of *Salix suchowensis* [80] (GenBank: KM983390.1) to remove maternally inherited plastid markers. However, since no match to the plastid genome was observed, all loci appear to belong to the nuclear genome. Alignments were performed in Geneious R 10.0.9 [81]. After the filtering, 5758 loci remained for population genomic analysis.

Genetic structure of the hybrid zone and the progenies

Pairwise genetic distances between individuals were calculated using the genetic distance measure of Smouse and Peakall [82] implemented in the R package *PopGenReport* [83]. Based on these genetic distances we performed a PCoA implemented in the R package *ade4* [84]. We used the program STRUCTURE v. 2.3.4 [85] to confirm that all hybrids originated from crosses between *S. purpurea* and *S. helvetica* without the involvement of a third species. All 5758 loci were included in the STRUCTURE analysis. We tested K -values ranging from 1 to 7 without prior population information under the admixture model assuming independent allele frequencies. Five runs were performed per tested K -value and the most likely K -value was determined using the method of Evanno et al. [86]. Finally, we determined the parental plants and the hybrid categories (F_1 , F_2 , backcrosses) of each hybrid individual with the program NewHybrids [87]. We designated only the hybrid \times hybrid class as F_2 hybrids, while the progeny as a whole was called second generation hybrids. Due to the young age of the hybrid zone on the forefield (approximately 20–30 years) it can be assumed that all individuals still belong to the early hybrid generations so that an assignment to exact hybrid classes is possible. NewHybrids cannot handle large datasets and thus we restricted the dataset to the first 300 loci of the whole dataset. We also ran NewHybrids with 300 loci that were selected randomly to ensure that the patterns were consistent across different

subsets of the genome. The results were the same but we chose to use the first 300 loci to ensure the reproducibility of the results. STRUCTURE and NewHybrids were run using a burn-in period of 10,000 followed by 50,000 MCMC iterations. Longer run times were tested using reduced data sets, but did not change the results substantially.

Analysis of segregation distortion in the second generation hybrids

We selected loci showing fixed differences between *S. purpurea* and *S. helvetica* for the analysis of segregation distortion so that the origin of an allele in a hybrid individual could be unequivocally determined. We checked whether all 45 F_1 hybrids were heterozygous at these loci. Overall, 396 loci met both criteria. The R package *introgress* [88] was used to count the number of alleles derived from each of the parental populations for each hybrid individual at each locus. These counts were used as the basis to detect deviations from the expected segregation patterns. In the F_2 hybrids, we tested for the deviation from the expected 1:2:1 distribution of homozygous and heterozygous genotypes found at each locus. In the backcrosses, we tested for the deviation from the expected 1:1 distribution of homozygous and heterozygous genotypes. χ^2 goodness-of-fit tests were performed in R [89]. In the analysis of the F_2 hybrids, p -values were computed by Monte Carlo simulation due to the low number of individuals. We corrected for multiple testing using the false discovery rate (FDR) method of [90] with $\alpha = 0.10$.

We performed a BLAST search of all RAD-loci containing species specific SNPs against the *S. purpurea* genome using Phytozome 12 [91] to determine on which chromosomes the loci were located. The alignment was accepted when the reads showed > 98% identity over the whole read length of 90 bp.

Evaluation of plant height

The seedlings grown from five naturally pollinated F_1 hybrids at the forefield of the Rhône Glacier [45] were raised in climate chambers for about 1 year at 18 °C with a 16-h light period (c. 250 $\mu\text{mol m}^{-2} \text{s}^{-1}$) under equal soil and watering conditions. At this age, plants had leaves typical for adults (for some examples see Additional file 1: Figure S2), but did not yet produce flowers. Before these juvenile plants were transferred to pots for outdoor cultivation, the length of the longest shoot was measured to the nearest 0.5 cm. A one-way ANOVA was performed to test for differences between groups (according to the NewHybrids analysis: F_2 , backcross to *S. purpurea*, backcross to *S. helvetica*), followed by the Games–Howell test as post hoc test. The type I error rate was $\alpha = 0.05$. The ANOVA was performed using SPSS version 24 (IBM Corp., Armonk, NY).

Additional files

Additional file 1: Supporting information for the Methods and Results section. Contains **Figure S1**, plot of the ΔK -values for the range of K -values tested in STRUCTURE, **Figure S2**, examples of the leaf shape of *S. purpurea*, *S. helvetica* and their hybrids, **Table S1**, admixture proportions of 40 F₁ hybrids, **Table S2**, assignment of hybrid class of 40 F₁ hybrids. (PDF 288 kb)

Additional file 2: Table S3. Distribution of homozygous and heterozygous genotypes found at each locus in the F₂ hybrids and distribution of parental alleles in the backcrosses. (XLSX 29 kb)

Acknowledgements

We thank Jennifer Krüger for extracting the DNA and Silvia Friedrichs for nursing the seedlings. The referee's comments were of great value for improving the manuscript.

Funding

The study was funded by the German Research Fund (DFG project Ho 5462 7-1) to E.H. The DFG was not included in study design or any other operational part of this project.

Availability of data and materials

All demultiplexed read data were submitted to the NCBI Sequence Read Archive: accession number SRP133640, BioProject ID PRJNA429746. The dataset generated for the population genetic analysis (STRUCTURE input file) is available in the Dryad Digital Repository, <https://doi.org/10.5061/dryad.4k3v0kg>.

Authors' contributions

EH and SG designed the research, SG and NDW analyzed the data, SG wrote the paper with assistance of NDW and EH. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 8 June 2018 Accepted: 21 November 2018

Published online: 03 December 2018

References

- Hewitt GM. Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Mol Ecol*. 2001;10:537-49.
- Kelly AE, Goulden ML. Rapid shifts in plant distribution with recent climate change. *Proc Natl Acad Sci U S A*. 2008;105:11823-6.
- Lenoir J, Gégout JC, Marquet PA, de Ruffray P, Brisse H. A significant upward shift in plant species optimum elevation during the 20th century. *Science*. 2008;320:1768-71.
- Harsch MA, Hulme PE, McGlone MS, Duncan RP. Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecol Lett*. 2009;12:1040-9.
- Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD. Rapid range shifts of species associated with high levels of climate warming. *Science*. 2011;333:1024-6.
- Soltis PS, Soltis DE. The role of hybridization in plant speciation. *Annu Rev Plant Biol*. 2009;60:561-88.
- Rieseberg LH. Hybrid origins of plant species. *Annu Rev Ecol Syst*. 1997;28:359-89.
- Gross BL, Rieseberg LH. The ecological genetics of homoploid hybrid speciation. *J Hered*. 2005;96:241-52.
- Yakimowski SB, Rieseberg LH. The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. *Am J Bot*. 2014;101:1247-58.
- Mallet J. Hybridization as an invasion of the genome. *Trends Ecol Evol*. 2005;20:229-37.
- Grant V. *Plant speciation*. 2nd ed. New York: Columbia University Press; 1981.
- Mallet J. Hybrid speciation. *Nature*. 2007;446:279-83.
- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, et al. Hybridization and speciation. *J Evol Biol*. 2013;26:229-46.
- Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH. The likelihood of homoploid hybrid speciation. *Heredity*. 2000;84:441-51.
- Gompert Z, Fordyce JA, Forister ML, Shapiro AM, Nice CC. Homoploid hybrid speciation in an extreme habitat. *Science*. 2006;314:1923-5.
- Rhymer JM, Simberloff D. Extinction by hybridization and introgression. *Annu Rev Ecol Syst*. 1996;27:83-109.
- Martin NH, Bouck AC, Arnold ML. Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics*. 2006;172:2481-9.
- Whitney KD, Randell RA, Rieseberg LH. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytol*. 2010;187:230-9.
- Lexer C, Widmer A. The genic view of plant speciation: recent progress and emerging questions. *Philos Trans R Soc B Biol Sci*. 2008;363:3023-36.
- Kim M, Cui M-L, Cubas P, Gillies A, Lee K, Chapman MA, et al. Regulatory genes control a key morphological and ecological trait transferred between species. *Science*. 2008;322:1116-9.
- Rieseberg LH, Carney SE. Plant hybridization. *New Phytol*. 1998;140:599-624.
- Campbell DR, Waser NM. Genotype-by-environment interaction and the fitness of plant hybrids in the wild. *Evolution*. 2001;55:669-76.
- Arnold ML. Natural hybridization as an evolutionary process. *Annu Rev Ecol Syst*. 1992;23:237-61.
- Arnold ML, Ballerini ES, Brothers AN. Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana irises. *Heredity*. 2012;108:159-66.
- Bouck A, Peeler R, Arnold ML, Wessler SR. Genetic mapping of species boundaries in Louisiana irises using IRRE retrotransposon display markers. *Genetics*. 2005;171:1289-303.
- Pritchard VL, Dimond L, Harrison JS, Velázquez CCS, Zieba JT, Burton RS, et al. Interpopulation hybridization results in widespread viability selection across the genome in *Tigriopus californicus*. *BMC Genet*. 2011;12:54.
- Rieseberg LH, Baird SJ, Gardner KA. Hybridization, introgression, and linkage evolution. *Plant Mol Biol*. 2000;42:205-24.
- Harushima Y, Nakagahra M, Yano M, Sasaki T, Kurata N. A genome-wide survey of reproductive barriers in an intraspecific hybrid. *Genetics*. 2001;159:883-92.
- Recknagel H, Elmer KR, Meyer A. A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3 genes, genomes, Genet*. 2013;3:65-74.
- Qvarnström A, Bailey RL. Speciation through evolution of sex-linked genes. *Heredity*. 2009;102:4-15.
- De Carvalho D, Ingvarsson PK, Joseph J, Suter L, Sedivy C, Macaya-Sanz D, et al. Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Mol Ecol*. 2010;19:1638-50.
- Pucholt P, Hallingbäck HR, Berlin S. Allelic incompatibility can explain female biased sex ratios in dioecious plants. *BMC Genomics*. 2017;18:251.
- Myburg AA, Vogl C, Griffin AR, Sederoff RR, Whetten RW. Genetics of postzygotic isolation in Eucalyptus: whole-genome analysis of barriers to introgression in a wide interspecific cross of *Eucalyptus grandis* and *E. globulus*. *Genetics*. 2004;166:1405-18.
- Bodénès C, Chancerel E, Ehrenmann F, Kremer A, Plomion C. High-density linkage mapping and distribution of segregation distortion regions in the oak genome. *DNA Res*. 2016;23:115-24.
- Bradshaw HD, Stettler RF. Molecular genetics of growth and development in *Populus*. II. Segregation distortion due to genetic load. *Theor Appl Genet*. 1994;89:551-8.
- Fishman L, Kelly AJ, Morgan E, Willis JH. A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics*. 2001;159:1701-16.

37. Fishman L, Willis JH. A novel meiotic drive locus almost completely distorts segregation in *Mimulus* (monkeyflower) hybrids. *Genetics*. 2005;169:347–53.
38. Sakaguchi S, Sugino T, Tsumura Y, Ito M, Crisp MD, Bowman DMJS, et al. High-throughput linkage mapping of Australian white cypress pine (*Callitris glaucophylla*) and map transferability to related species. *Tree Genet Genomes*. 2015;11:121.
39. Gramlich S, Sagmeister P, Dullinger S, Hadacek F, Hörandl E. Evolution *in situ*: hybrid origin and establishment of willows (*Salix* L.) on alpine glacier forefields. *Heredity*. 2016;116:531–41.
40. Burga C. Vegetation development on the glacier forefield Morteratsch (Switzerland). *Appl Veg Sci*. 1999;2:17–24.
41. Stöcklin J, Bäumler E. Seed rain, seedling establishment and clonal growth strategies on a glacier foreland. *J Veg Sci*. 1996;7:45–56.
42. Argus GW. An experimental study of hybridization and pollination in *Salix* (willow). *Can J Bot*. 1974;52:1613–9.
43. Hörandl E, Florineth F, Hadacek F. Weiden in Österreich und angrenzenden Gebieten. 2nd ed. Wien: Arbeitsbereich Ingenieurbiologie und Landschaftsbau, Inst. für Landschaftsplanung und Ingenieurbiologie, Univ. für Bodenkultur; 2012.
44. Skvortsov AK. Willows of Russia and adjacent countries. Joensuu: University of Joensuu; 1999.
45. Gramlich S, Hörandl E. Fitness of natural willow hybrids in a pioneer mosaic hybrid zone. *Ecol Evol*. 2016;6:7645–55.
46. Lindtke D, Buerkle CA, Barbará T, Heinze B, Castiglione S, Bartha D, et al. Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. *Mol Ecol*. 2012;21:5042–58.
47. Lindtke D, Gompert Z, Lexer C, Buerkle CA. Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Mol Ecol*. 2014;23:4316–30.
48. Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogdan R, Limborg MT, et al. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Mol Ecol Resour*. 2011;11:123–36.
49. Milne RI, Terzioglu S, Abbott RJ. A hybrid zone dominated by fertile F1s: maintenance of species barriers in *Rhododendron*. *Mol Ecol*. 2003;12:2719–29.
50. Bialozyt R, Rathmacher G, Niggemann M, Ziegenhagen B. Reconstructing explicit mating schemes in poplar hybrids - a case study in the *Populus nigra* L. - *Populus x canadensis* Moench complex. *Silvae Genet*. 2012;61:157–67.
51. Totland O, Sottocornola M. Pollen limitation of reproductive success in two sympatric alpine willows (Salicaceae) with contrasting pollination strategies. *Am J Bot*. 2001;88:1011–5.
52. Mosseler A. Interspecific pollen-pistil incongruity in *Salix*. *Can J For Res*. 1989;19:1161–8.
53. Jenczewski E, Gherardi M, Bonnin I, Prospero JM, Olivieri I, Huguet T. Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). *Theor Appl Genet*. 1997;94:682–91.
54. Wagner ND, Gramlich S, Hörandl E. RAD sequencing resolved phylogenetic relationships in European shrub willows (*Salix* L. subg. *Chamaetia* and subg. *Vetrix*) and revealed multiple evolution of dwarf shrubs. *Ecol Evol*. 2018;8(16):8243–55.
55. Chen JH, Sun H, Wen J, Yang YP. Molecular phylogeny of *Salix* L. (Salicaceae) inferred from three chloroplast datasets and its systematic implications. *Taxon*. 2010;59:29–37.
56. Lauron-Moreau A, Pitre FE, Argus GW, Labrecque M, Brouillet L. Phylogenetic relationships of American willows (*Salix* L., Salicaceae). *PLoS One*. 2015;10:e0121965.
57. Wu J, Nyman T, Wang D-C, Argus GW, Yang Y-P, Chen J-H. Phylogeny of *Salix* subgenus *Salix* s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. *BMC Evol Biol*. 2015;15:31.
58. Matute DR, Butler IA, Turissini DA, Coyne JA. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science*. 2010;329:1518–21.
59. Moyle LC, Nakazato T. Hybrid incompatibility “snowballs” between *Solanum* species. *Science*. 2010;329:1521–3.
60. Pucholt P, Rönnerberg-Wästjung A-C, Berlin S. Single locus sex determination and female heterogamety in the basket willow (*Salix viminalis* L.). *Heredity*. 2015;114:575–83.
61. Stölting KN, Nipper R, Lindtke D, Caseys C, Waeber S, Castiglione S, et al. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Mol Ecol*. 2013;22:842–55.
62. Yin T, Difazio SP, Gunter LE, Zhang X, Sewell MM, Woolbright SA, et al. Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Res*. 2008;18:422–30.
63. Pakull B, Groppe K, Meyer M, Markussen T, Fladung M. Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). *Tree Genet Genomes*. 2009;5:505–15.
64. Paolucci I, Gaudet M, Jorge V, Beritognolo I, Terzoli S, Kuzminsky E, et al. Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. *Tree Genet Genomes*. 2010;6:863–75.
65. Macaya-Sanz D, Suter L, Joseph J, Barbará T, Alba N, González-Martínez SC, et al. Genetic analysis of post-mating reproductive barriers in hybridizing European *Populus* species. *Heredity*. 2011;107:478–86.
66. Schiechl HM. Weiden in der Praxis: Die Weiden Mitteleuropas, ihre Verwendung und ihre Bestimmung. Berlin-Hannover: Patzer Verlag; 1992.
67. Körner C. Alpine plant life: functional plant ecology of high mountain ecosystems. 2nd ed. Berlin: Springer; 2003.
68. Chunco AJ. Hybridization in a warmer world. *Ecol Evol*. 2014;4:2019–31.
69. Allendorf FW, Leary RF, Spruell P, Wenburg JK. The problems with hybrids: setting conservation guidelines. *Trends Ecol Evol*. 2001;16:613–22.
70. Rieseberg LH, Wendel JF. Gene flow and its consequences in plants. In: Harrison RG, editor. Hybrid zones and the evolutionary process. Oxford: Oxford University Press; 1993. p. 70–109.
71. Coyne JA, Orr HA. Speciation. Sinauer: Sunderland, MA; 2004.
72. Baird NA, Etter PD, Atwood TS, Currey KM, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. 2008;3:e33376.
73. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH, De Koning D-J. Stacks: building and genotyping loci de novo from short-read sequences. *G3 Genes, Genomes, Genetics*. 2011;1:171–82.
74. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 2013;22:3124–40.
75. Brahm Bioinformatics. FastQC 0.11.4. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
76. Viricel A, Pante E, Dabin W, Simon-Bouhet B. Applicability of RAD-tag genotyping for interfamilial comparisons: empirical data from two cetaceans. *Mol Ecol Resour*. 2014;14:597–605.
77. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
78. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 2006;313:1596–604.
79. Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol Ecol Resour*. 2011;11:117–22.
80. Wu Z. The whole chloroplast genome of shrub willows (*Salix suchowensis*). *Mitochondrial DNA part A*. 2016;27:2153–4.
81. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9.
82. Smouse PE, Peakall R. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*. 1999;82:561–73.
83. Adamack AT, Gruber B. PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol Evol*. 2014;5:384–7.
84. Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw*. 2007;22:1–20.
85. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
86. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;14:2611–20.
87. Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*. 2002;160:1217–29.
88. Gompert Z, Alex Buerkle C. INTROGRESS: a software package for mapping components of isolation in hybrids. *Mol Ecol Resour*. 2010;10:378–84.
89. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016.

90. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 1995;57: 289–300.
91. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 2012;40:1178–86.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

