

RESEARCH

Open Access



How well do genetic markers inform about responses to intraspecific admixture? A comparative analysis of microsatellites and RADseq

Yeşerin Yıldırım , Anders Forsman  and Johanna Sunde* 

Abstract

Background: Fitness consequences of intraspecific genetic admixture can vary from positive to negative depending on the genetic composition of the populations and environmental conditions. Because admixture has potential to influence the success of management and conservation efforts, genetic similarity has been suggested to be used as a proxy to predict the outcome. Studies utilizing microsatellites (a neutral marker) to investigate associations between genetic distance and admixture effects show conflicting results. Marker types that yield information on genome-wide and/or adaptive variation might be more useful for predicting responses to inter-population hybridization. In this study we utilized published data for three populations of pike (*Esox lucius*) to investigate associations between offspring performance (hatching success) and parental genetic similarity in experimentally purebred and admixed families, based on neutral (microsatellites), genome-wide neutral (RADseq SNPs), and adaptive (SNPs under selection) markers.

Results: Estimated similarity varied among the markers, likely reflecting differences in their inherent properties, but was consistently higher in purebred than admixed families. A significant interaction between marker type and admixture treatment reflected that neutral SNPs yielded higher estimates than adaptive SNPs for admixed families whereas no difference was found for purebred families, which indicates that neutral similarity was not reflective of adaptive similarity. When all samples were pooled, no association between similarity and performance was found for any marker. For microsatellites, similarity was positively correlated with hatching success in purebred families, whereas no association was found in admixed families; however, the direction of the effect differed between the population combinations.

Conclusions: The results strengthen the notion that, as of today, there is no proxy that can reliably predicted the outcome of admixture. This emphasizes the need of further studies to advance knowledge that can shed light on how to safeguard against negative consequences of admixture, and thereby inform management and promote conservation of biological diversity.

Keywords: *Esox lucius*, Hybridization, Interbreeding, Microsatellites, Offspring performance, Outbreeding, Pike, RADseq, SNP

* Correspondence: johanna.sunde@lnu.se

Centre for Ecology and Evolution in Microbial Model Systems, EEMIS,
Department of Biology and Environmental Science, Linnaeus University,
SE-392 31 Kalmar, Sweden



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Intraspecific genetic admixture (henceforth ‘admixture’) occurs when separated populations start interbreeding [1]. It occurs naturally in many terrestrial and aquatic species in the wild as a consequence of dispersal [2], and can also occur as a result of anthropogenic activities, for example management actions to support populations (e.g. supplementations and translocations) [3, 4], removal of dispersal barriers [5], and escapes of farmed individuals [6].

Admixture affects the genetic composition of the involved populations and tends to increase the genetic variation in the receiving population. Increased genetic variation is generally considered to positively influence aspects of population performance. For example, it has been shown that genetically and phenotypically more diverse populations are better able to cope with environmental change, and to colonize novel environments [2, 7–10]. Admixture can also have positive fitness effects by allowing creation of novel gene combinations and dampening inbreeding depression by masking detrimental effects of deleterious recessive alleles (heterosis) [1, 3]. However, the influx of new genetic material can also result in negative fitness effects. If the involved populations are highly differentiated, genomic incompatibilities might exist (e.g. as a result of chromosomal rearrangements in the parental populations) [11, 12], and admixture may thus lead to outbreeding depression. In addition, admixture between populations that have adapted to different environmental conditions might dilute favorable alleles [1], give rise to offspring with intermediate phenotypes that are not optimal in either of the parental environments [4, 13], and has potential to impair fitness in subsequent generations by breaking up co-adapted gene complexes or by underdominance (i.e. heterozygote disadvantage) [1, 4, 13–15]. The net outcome (with regards to both magnitude and direction) of the response to admixture will be determined by an interplay between these mechanisms, and will thus depend on the genetic composition of the parental populations, local adaptations and environmental conditions [13].

The motivation for management and conservation efforts resulting in admixture is to increase productivity, viability, and adaptability of populations [3, 11, 14]. To avoid undesirable outcomes of such efforts, it would be valuable to have a reliable proxy that could be used to predict the response to admixture. Genetic similarity has been put forward as such a candidate proxy [11, 14]. However, empirical studies show conflicting results [3, 11, 16–19], and further studies are therefore required to increase the understanding about potential associations between genetic similarity and the response to admixture, and to understand the reasons for the observed

inconsistencies. Previous studies have commonly estimated genetic similarity based on a modest number (4–32) of microsatellite markers [20–23]. Microsatellites are mainly neutral markers that only occasionally reflect functional evolution, e.g. by residing within coding or regulatory regions or by being linked to functional loci [24, 25]. Microsatellites have high mutation rates, and loci with relatively high allelic variation are commonly selected during the marker development [26, 27]. As a consequence, genetic diversity estimates based on small numbers of microsatellites generally do not reflect the genome-wide diversity [26, 28]. It is therefore possible that the inconsistencies in associations between genetic similarity and admixture effects in previous studies [3, 11, 14, 16, 20, 29] can be partly attributed to the use of non-representative estimates of genetic diversity, and estimates of genome-wide diversity might be better able to predict the outcome of admixture. In addition, estimates of parental genetic similarity might be more informative than population-based similarity, as the former accounts for the inter-individual variation [23].

The development of next generation sequencing techniques, such as Restriction-site associated DNA sequencing (RADseq) has enabled to genotype many markers at a low cost, even for non-model species [30]. RADseq commonly yields thousands of single nucleotide polymorphisms (SNPs), which provides better estimates of genome-wide diversity. It has also been shown that even relatively low numbers of SNPs (≥ 50) generally has the same, if not more, statistical power compared to 20 microsatellites in relatedness studies [31], and that RADseq SNPs generally tend to outperform microsatellites in population genetic studies [32], but see [33]. In addition, it is functional – not neutral – genetic variation that is key for the adaptive potential of populations [34] with potential to influence the outcome of admixture. It can therefore be hypothesized that estimates of genetic similarity based on adaptive genetic variation (or divergence) offer a better predictor of the response to admixture [35]. RADseq SNPs thus offer a viable alternative as they can provide information on both neutral and adaptive genetic variation and differentiation.

The overall aim of the present study was to evaluate the potential of neutral (microsatellites), genome-wide neutral (RADseq SNPs), and adaptive (outlier SNPs under selection) parental genetic similarity to predict admixture effects on offspring performance (hatching success) resulting from experimental matings of individuals representing genetically separated and locally adapted natural populations. To this end, we used previously published genetic and phenotypic data for three populations (Harfjärden, Lerviksbacken, and Okneback; henceforth Harfjärden, Lervik and Okne) of anadromous Baltic Sea pike (*Esox lucius*) [32, 36]. Pike is a long-lived

fish, which is both ecologically and socio-economically important. Unfortunately, it has suffered decreases during the last decades, and has therefore been the target of extensive management efforts [37–40]. All three populations included in this study are neutrally genetically differentiated based on both microsatellites [32, 39, 41] and RADseq SNPs [32], but the degree of differentiation differ among the population pairs. Differentiation between the two adjacent streams (Lervik and Okne) is low ($F_{ST} = 0.044–0.071$), and some gene flow may occur [32, 39, 41]. Comparisons of the geographically more separated populations (Harfjärden compared to both other populations), showed no signs of gene flow and higher differentiation ($F_{ST} = 0.136–0.226$) [32, 39, 41]. Studies also indicate that the populations are adaptively differentiated and display different local adaptations [36, 42–45]. More specifically, we investigated whether: *i*) variation in parental similarity estimates among pairs of experimentally mated males and females were consistent or differed among the markers; *ii*) parental similarity was associated with offspring hatching success; and *iii*) the associations between parental similarity and offspring performance were consistent or depended on whether parental similarity was estimated based on neutral (microsatellites), genome-wide neutral (RADseq SNPs) or adaptive (RADseq SNPs under selection) genetic markers.

Results

Comparison of parental similarity based on microsatellites and RADseq SNP data

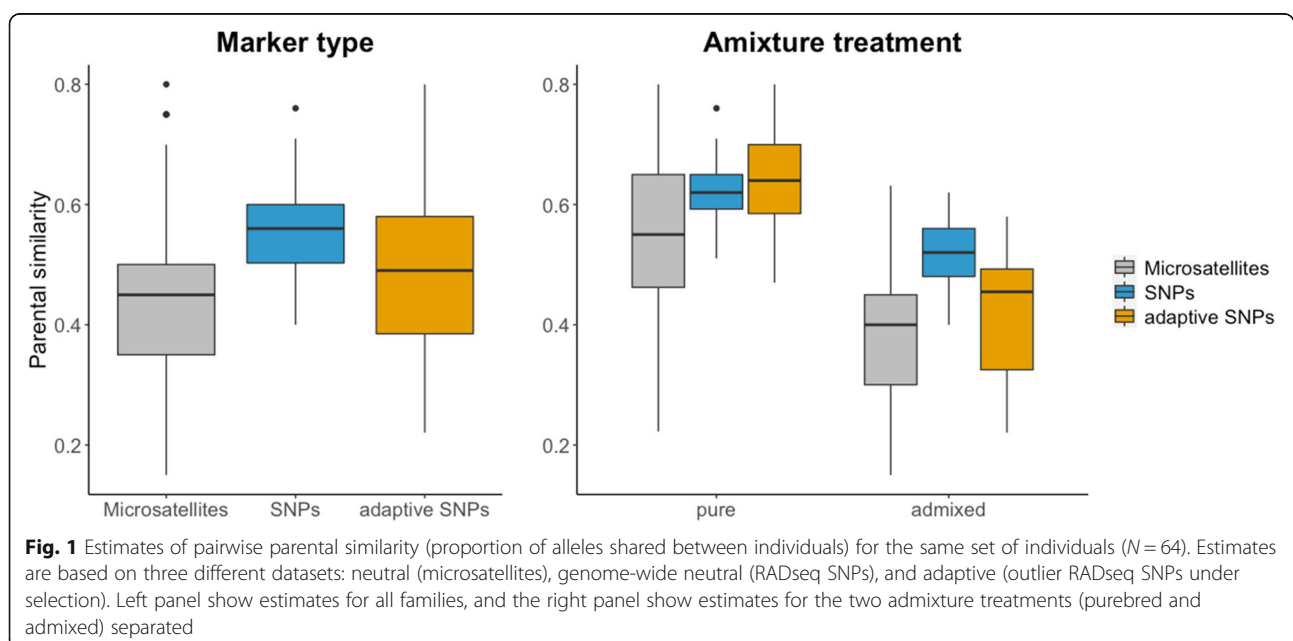
Estimates of pairwise parental similarity based on the three different datasets (microsatellites, neutral SNPs and adaptive SNPs) differed ($F_{2,178} = 36.38$, $P < 0.001$,

Fig. 1), and differences were evident between all pairwise comparisons (Tukey's test: microsatellites – adaptive SNPs: $P = 0.014$; microsatellites – neutral SNPs: $P < 0.001$; adaptive SNPs – neutral SNPs: $P = 0.003$). Estimated parental similarity was highest for neutral SNPs (mean \pm SD: 0.55 ± 0.07), intermediate for adaptive SNPs (mean \pm SD: 0.49 ± 0.15), and lowest for microsatellites (mean \pm SD: 0.43 ± 0.15). In addition, the range of similarity estimates was smaller for neutral SNPs than for both microsatellites and adaptive SNPs (Fig. 1).

Parental similarity estimates were higher for purebred families than for admixed families for all three datasets (Student's *t*-test: microsatellites: $t = -5.88$, $df = 49.39$, $P < 0.001$; neutral SNPs: $t = -9.63$, $df = 66.36$, $P < 0.001$; adaptive SNPs: $t = -11.61$, $df = 72.61$, $P < 0.001$) (Fig. 1). In addition, the analysis revealed a significant interaction effect between marker type and admixture treatment (purebred or admixed) ($F_{2,176} = 10.16$, $P < 0.001$), which reflected that neutral SNPs yielded higher estimates for the admixed families than did microsatellites and adaptive SNPs (Tukey's test: microsatellites – adaptive SNPs: $P = 0.15$; microsatellites – neutral SNPs: $P < 0.001$; adaptive SNPs – neutral SNPs: $P < 0.001$), whilst there was no difference between neutral SNPs and adaptive SNPs for purebred families (Tukey's test: microsatellites – adaptive SNPs: $P < 0.001$; microsatellites – neutral SNPs: $P = 0.008$; adaptive SNPs – neutral SNPs: $P = 0.70$) (Fig. 1).

No association between parental similarity and hatching success

When all the samples were pooled, a large variation in hatching success was evident throughout the range of



parental similarity for all three datasets (Fig. 2). Parental similarity was not associated with hatching success for any of the marker types (microsatellites: $F_{1,177} = 1.23$, $P = 0.27$, neutral SNPs: $F_{1,177} = 0.04$, $P = 0.83$, adaptive SNPs: $F_{1,177} = 0.27$, $P = 0.61$).

When the samples were classified according to admixture treatment (purebred or admixed), to test whether associations between parental similarity and hatching success differed between purebred and admixed families, no significant interaction effect was found for either neutral SNPs (effect of treatment: $F_{1,176} = 0.24$, $P = 0.62$; effect of parental similarity: $F_{1,176} = 0.045$, $P = 0.85$; effect of interaction between parental similarity and admixture treatment: $F_{1,175} = 0.67$, $P = 0.41$) or adaptive SNPs (effect of treatment: $F_{1,176} = 1.83$, $P = 0.18$; effect of parental similarity: $F_{1,176} = 0.27$, $P = 0.18$; interaction between parental similarity and admixture treatment: $F_{1,175} = 3.07$, $P = 0.08$) (Fig. 3). However, for the microsatellite data, there was a significant effect of the interaction between admixture treatment and parental similarity ($F_{1,175} = 4.31$, $P = 0.04$), reflecting that there was a positive relationship between similarity and hatching success for purebred families (a positive slope), whereas parental similarity was not associated with hatching success in the admixed families (Fig. 3). Although not statistically significant, similar trends (a positive slope for purebred families and no association for admixed families) were observable also for the two other marker types (neutral SNPs and adaptive SNPs) (Fig. 3).

When the data was further separated into specific population combinations (3 purebred and 3 admixed groups, based on the source population of each of the parental individuals) (Fig. 4), Student's T-tests showed that the association between parental similarity and hatching success differed among the purebred populations for the microsatellite dataset ($F_{1,2} = 7.47$, $P < 0.001$), but not for the neutral SNP dataset ($F_{1,2} = 0.72$, $P = 0.49$)

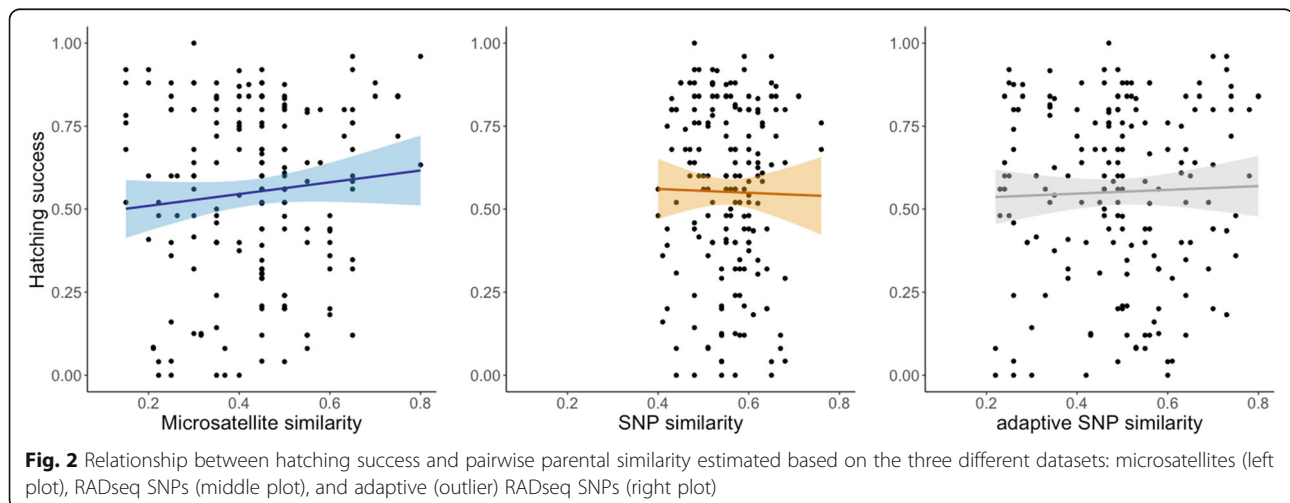
whereas adaptive SNPs approached marginal significance ($F_{1,2} = 2.81$, $P = 0.061$). No significant interaction effects between parental similarity and population combination were found for the admixed population combinations for any of the marker types (microsatellites: $F_{1,2} = 0.95$, $P = 0.3880$; neutral SNPs: $F_{1,2} = 0.81$, $P = 0.446$; adaptive SNPs: $F_{1,2} = 1.73$, $P = 0.177$).

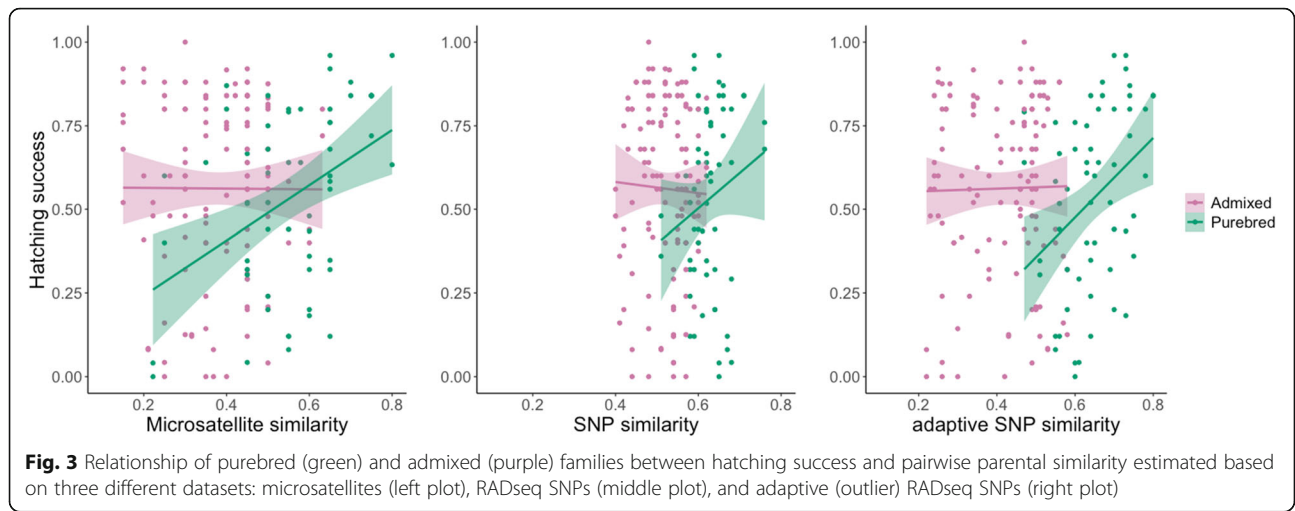
Discussion

Increased knowledge about the effects of admixture can further the understanding about evolution of genetic structure and what shapes patterns of diversity, and also help avoid undesirable effects associated with conservation measures and management actions. In the present study, we used data for pike to evaluate whether estimates of parental genetic similarity based on three different marker types (microsatellites, neutral SNPs and adaptive SNPs) could predict the outcome of admixture. To our knowledge, this is the first attempt to systematically evaluate and compare the utility of different markers as proxies of parental compatibility and predictors of offspring performance. The main findings were that: *i*) estimated parental similarity differed between the marker types, and were consistently higher for purebred families compared to admixed families; *ii*) parental similarity was not consistently or clearly associated with hatching success for any of the marker types; and *iii*) the association between hatching success and parental similarity as estimated based on microsatellites was different for purebred and admixed groups, and also differed between population combinations.

Parental similarity differed between marker types

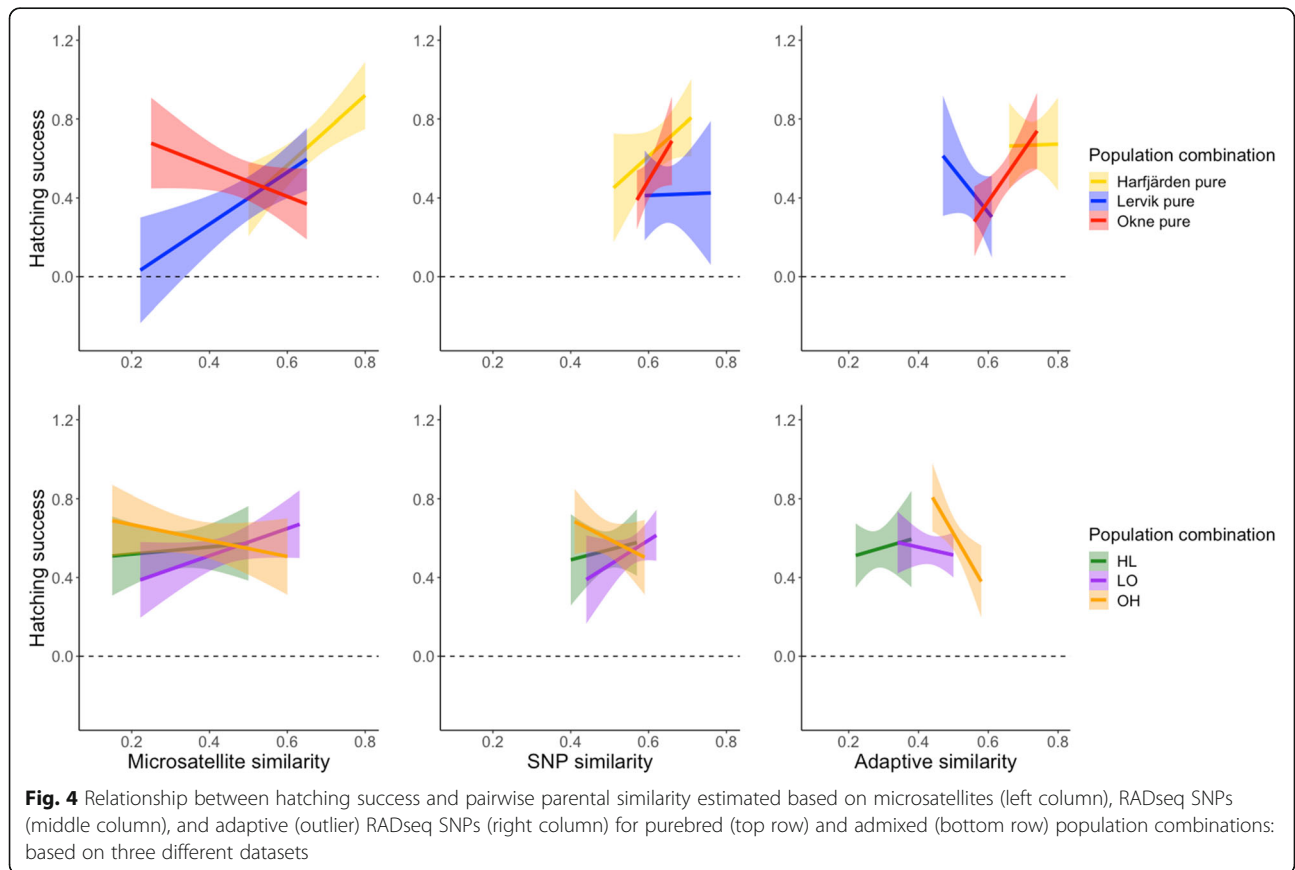
The findings that both values and ranges of estimated parental similarity differed between the marker types (Fig. 1) were expected, and likely reflect differences in the inherent properties of the marker types and datasets.





The higher similarity estimates obtained for both of the SNP datasets (genome-wide neutral and adaptive) compared to microsatellites (neutral) is likely explained by the higher number of alleles per locus for microsatellites [26, 27], which might lead to overestimation of differentiation between individuals [26]. Moreover, the denser RADseq SNP data (~ 1500 SNPs) is more likely to represent genome-wide diversity than the moderate number

of microsatellites [28]. The wider range of similarity for both microsatellites and adaptive SNPs compared to neutral SNPs likely in part reflect the relatively low number of loci used for the former two, as each allele will have a large effect on the estimated similarity. In addition, the adaptive dataset consists of loci that are associated with environmental variables (temperature and salinity). It is therefore likely that the wide range for



adaptive SNPs also in part reflects the range of environmental differences among the populations. Taken together, this calls for caution when comparing results of studies that have used different markers, and comparisons should be based on ranking rather than absolute values.

The finding that purebred families had higher parental similarity than admixed families for all datasets (Fig. 1) was also expected as the three study populations are both neutrally and adaptively differentiated [32, 41, 46]. That similarity estimates were higher for neutral SNPs than for adaptive SNPs for admixed families but did not differ for purebred families, likely reflects differences in neutral and adaptive evolution. Neutral loci are mainly affected by neutral and stochastic processes, whereas functional loci are also affected by deterministic processes such as selection [30, 47]. For admixed families, the loci in the adaptive dataset are probably under diversifying selection (as the algorithm used in the study identifies the outlier loci associated with environmental variables [48]), which would explain the higher degree of differentiation observed for adaptive SNPs.

No association between parental similarity and hatching success

While there was no overall association between hatching success and parental similarity for any of the marker types (Fig. 2), the results indicated that the effect of parental similarity might differ between purebred and admixed families, and that the association was positive only for purebred families (Fig. 3). The large variation in hatching success for both admixed and purebred families throughout the similarity ranges for all marker types (Fig. 3) and the lack of any consistent association across marker types between parental similarity and offspring hatching success argues against the utility of parental genetic similarity estimates as a reliable predictive proxy for admixture effects in this system. It is possible that the range of genetic differentiation between the study populations was not wide enough to get a complete picture, that effects of admixture manifest more strongly in natural environments likely reflecting differences in the selective regimes [49–51], and that some admixture effects are not expressed until the F2 generation [10, 14, 16]. That our analyses, like some previous investigations of other species [17–20, 22, 23, 52], failed to detect any association between parental similarity and offspring performance cannot be taken as evidence that the genetic resemblance between parents is of no importance. However, that such an association apparently is difficult to detect is both disappointing and problematic, particularly from an applied conservation perspective.

Sadly, the conclusion that parental similarity is a poor predictor of the response to admixture extends to other

candidate proxies, such as geographic distance, neutral genetic differentiation, genetic diversity or environmental similarity. Empirical studies using these different proxies [3, 11, 14, 16–18, 29, 53, 54] show conflicting results. This inconsistency is likely reflective of the complex interactions between environmental factors and inherent properties of the parental populations [15, 54].

Conclusions

The present study showed that parental similarity was not consistently or clearly associated with hatching success for any of the marker types (neutral, genome-wide neutral, and adaptive). Our present study thus strengthens the conclusion that, as of today, there is no proxy that can reliably predict the outcome of admixture. There is therefore a clear need for further studies and different approaches to advance knowledge that can shed light on how to safeguard against negative consequences of admixture, and thereby inform management and promote successful conservation of biological diversity.

Methods

Study species

Pike is a long-lived fish that inhabits both freshwater and brackish water systems [55]. As one of the most common large predatory fishes in the Baltic Sea, it fills an important function in many systems where it regulates the abundance of species in lower trophic levels through top-down trophic cascades [56, 57]. As a valued species in both commercial and recreational fishing, pike is also socio-economically important [37, 58]. It has also emerged as a model species for studies of ecology and evolution [59]. Unfortunately, pike populations in the Baltic Sea have been experiencing declines during the last decades [37, 38, 57, 60]. Several different factors, such as eutrophication, habitat loss, overfishing, and altered species interactions, have been proposed to have contributed to the decrease [37, 38, 57, 61, 62]. Due to the importance of pike, management actions, e.g., restoration of spawning locations (wetlands), and large-scale stocking programs have therefore been carried out to support and revitalize the populations [39, 40, 63, 64].

Study populations

The three populations of anadromous pike included in this study reproduce in different spawning habitats in the Kalmar Sound region [36]. Two of the localities (Lervik and Okne) are closely located (approximately 20 km shortest waterway distance) in the southeast of the Swedish mainland (Lervik: N57° 04.414'; E16°31.246', Okne: N57° 01.200'; E16° 26.700'), and the third locality (Harfjärden) is located on the east coast of the island of Öland (N56° 49.063'; E16° 48.673'; approximately 120

and 135 km from Lervik and Okne, respectively) [32]. All three populations are significantly genetically differentiated from each other as indicated by results from analyses based on both microsatellites [32, 39, 41] and RADseq SNPs [32]. Among the three populations, Harfjärden forms the most distinct genetic cluster with high genetic differentiation compared to both other populations ($F_{ST} = 0.136\text{--}0.226$, $P < 0.01$) [32, 39, 41] and no evidence of gene flow to the Swedish mainland populations [32]. Despite evidence of low levels of gene flow, the two closely located streams (Lervik and Okne) also form genetically distinct populations but with low differentiation ($F_{ST} = 0.044\text{--}0.071$, $P < 0.01$) [32, 39, 41], and the distinctiveness of the clusters becomes more evident with increasing numbers of samples/loci included in the analyses [32, 41]. In addition, common garden and translocation experiments indicate that the study populations exhibit local adaptations for several traits including early life history traits and reproductive investment [42], salinity [36] and temperature tolerance [45], growth rate and adult body size [43] and vertebrae count [44]. The local adaptations have been attributed to environmental differences among the three spawning grounds, such as differences in the salinity and temperature regimes [45, 65], and in the amount of suspended materials [42]. Moreover, the study populations (exactly the same set of individuals as in the present study) exhibited genetic signatures of selection associated with salinity and temperature [32].

Estimates of hatching success

We obtained estimates of hatching success from the study by Sunde, et al. [36] that investigated effects of admixture on F1 offspring performance [46]. More specifically, the study investigated whether and how admixture affected offspring quality in different population combinations, and whether the effects were population-specific [36]. In that study, gametes were collected from a total of 66 individuals from the three populations (Lervik, Okne and Harfjärden). To include the exact same set of individuals in both the hatching success dataset and genotyping dataset in the present study, we decided to omit two of the samples (that did not pass the quality control in the RADseq pipeline) before proceeding with the analyses (see the subsection Estimates of parental similarity below), thus resulting in a total of 64 parental individuals (Harfjärden: $N_{\text{males}} = 12$, $N_{\text{females}} = 10$, Lervik: $N_{\text{males}} = 12$, $N_{\text{females}} = 9$, Okne: $N_{\text{males}} = 10$, $N_{\text{females}} = 11$). In short, separate batches of eggs from each female were artificially fertilized with milt from one male from each population, thus creating one purebred (male from the same population) and two admixed (male from one of the other populations) treatments per female. Each combination was done in duplicates, and this resulted in a

total of 180 units from 90 female/male pairs (30 purebred and 60 admixed). Eggs, and subsequently hatched fry were reared in a common garden environment, and three offspring performance measures: 1) hatching success (proportion of eggs that hatched), 2) fry survival (15 days following hatching), and 3) fry body length (at 15 days post hatch) were estimated (for details see [36]). Because Sunde, et al. [36] found that hatching success was affected by the treatment we chose not to include the other offspring performance measures in the present study, to avoid using potentially biased estimates and low statistical power resulting from differences in hatching success, and the associated unequal and small sample sizes.

Estimates of parental similarity

We obtained the genotype data used for estimation of pairwise parental genetic similarity from the study by Sunde, et al. [32] where the relative performance of the two marker types (microsatellites and RADseq SNPs) to detect genetic structure was evaluated. The microsatellite data was retrieved from the Dryad Digital Repository [66], and RADseq SNPs from NCBI Sequence Read Archive (BioProject accession code PRJNA586770) [67]. In the study by Sunde, et al. [32], individuals were genotyped for ten microsatellite loci and 1580 SNPs. All ten microsatellites were found to be neutral, and the full RADseq dataset also reflected neutral evolution. In the present study we therefore used these datasets to represent partial neutral (microsatellites) and genome-wide neutral variation (the dataset referred to as 'neutral SNPs'). Sunde, et al. [32] further searched for adaptive SNPs with multiple outlier analyses, including tests of locus-specific effects and gene-environment associations (GEAs). Because of the low number of populations included in their study, the tests of locus-specific effects suffered from low statistical power and were not able to detect any signals of selection despite clear indications from previous common garden and translocation studies that the populations are adaptively differentiated [36, 42–45]. In the GEA analysis with latent factor mixed model (LFMM), on the other hand, Sunde, et al. [32] detected loci associated with two environmental variables of importance for pike (salinity in the spawning ground during spawning and temperature at initiation of spawning) were identified. In the present study we therefore used the loci that they identified as outliers in the LFMM analyses to represent adaptive variation (the dataset referred to as 'adaptive SNPs'). However, in the present study we chose to use a somewhat more liberal approach for classifying SNPs as outliers (adaptive) than used in the original study by Sunde, et al. [32] (in the present study we used a P -value cut-off of 0.01 instead of applying FDR correction for multiple testing) to not

exclude potential outliers, and to increase the number of included outlier loci. This resulted in a total of 17 loci identified as putatively under selection (for details see [Table S1](#) in Additional file 1). Based on the three datasets (microsatellites, neutral SNPs and adaptive SNPs), we then estimated pairwise parental similarity as the proportion of alleles shared between the two individuals in each family. This was calculated separately for each of the three datasets using R Studio v1.1.383 [68] with R v.3.2.2 [69].

Statistical analysis

General or generalized linear mixed models (depending on the distribution of response variables) were used to test whether and how parental similarity varied among the three marker types (microsatellites, neutral SNPs, and adaptive SNPs), whether hatching success was associated with parental similarity, and to explore whether and how associations between hatching success and parental similarity were affected by admixture treatment (differed between purebred and admixed families). For this, we used the lme4 package v1.1–15 in RStudio with R. The different models were chosen based on the response distribution of the data: general linear mixed models for parental similarity (with a normal response distribution) and generalized linear mixed models with a logit-link function for hatching success (with a binomial response distribution).

For all tests, we treated marker type as a fixed categorical factor, parental similarity as a fixed continuous factor, and family (female/male pair) as a random factor. Statistical significance was assessed with Type III partitioning and an α -level of 0.05, and the Satterthwaite's method was used to approximate degrees of freedom. For tests of interactions between factors, we excluded the interaction term and rerun the analysis in case of no significant interaction effect. For tests with significant terms, we further analyzed the data with Student's t-test with non-pooled SD or Tukey's test to determine the nature of the interactions and evaluate pairwise differences. *P*-values were adjusted using the FDR method [70] to account for multiple comparisons.

Abbreviations

FDR: False discovery rate; LFMM: Latent factor mixed model; PCA: Principal component analysis; RADseq: Restriction-site associated DNA sequencing; SD: Standard deviation; SNP: Single-nucleotide polymorphism

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12863-021-00974-3>.

Additional file 1: Table S1. Information on the 17 outlier loci that were used as adaptive dataset in the present study.

Acknowledgements

We would like to thank our colleagues in the evolutionary ecology and fish ecology research groups for stimulating discussions and valuable input on this study.

Authors' contributions

J.S. and A.F. conceived the study. J.S. and Y.Y. analyzed the data. J.S., A.F. and Y.Y. contributed to interpreting the results. J.S. and Y.Y. drafted the first manuscript. The author(s) read and approved the final manuscript.

Funding

Funding was provided by the Linnaeus University (to A.F.), and The Swedish Research Council FORMAS (to A.F. grant Dnr. 2017–00346). Open Access funding provided by Linnaeus University.

Availability of data and materials

The datasets supporting the conclusions of this article were obtained from previously published studies and are publicly available in Dryad Digital repository (Microsatellite genotype data: <https://doi.org/10.5061/dryad.31zcrjdgvg> [66]; data on hatching success: file "offspring.xlsx", <https://doi.org/10.5061/dryad.dd64hf3> [46]), and the NCBI Sequence Read Archive (RADseq genotype data: BioProject PRJNA579326, <https://www.ncbi.nlm.nih.gov/sra/PRJNA586770> [67]).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 17 February 2021 Accepted: 20 May 2021

Published online: 28 June 2021

References

- Lynch M. The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution*. 1991;45(3):622–9. <https://doi.org/10.1111/j.1558-5646.1991.tb04333.x>.
- Rius M, Darling JA. How important is intraspecific genetic admixture to the success of colonising populations? *Trends Ecol Evol*. 2014;29(4):233–42. <https://doi.org/10.1016/j.tree.2014.02.003>.
- Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ, Miller KA, et al. Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evol Appl*. 2011;4(6):709–25. <https://doi.org/10.1111/j.1752-4571.2011.00192.x>.
- Hughes J, Goudkamp K, Hurwood D, Hancock M, Bunn S. Translocation causes extinction of a local population of the freshwater shrimp *Paratya australiensis*. *Conserv Biol*. 2003;17(4):1007–12. <https://doi.org/10.1046/j.1523-1739.2003.01636.x>.
- Verhoeven KJ, Macel M, Wolfe LM, Biere A. Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proc Biol Sci*. 2011;278:2–8.
- Glover KA, Pertoldi C, Besnier F, Wennevik V, Kent M, Skaala Ø. Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. *BMC Genet*. 2013;14(1):74. <https://doi.org/10.1186/1471-2156-14-74>.
- Forsman A, Wennersten L. Inter-individual variation promotes ecological success of populations and species: evidence from experimental and comparative studies. *Ecography*. 2016;39(7):630–48. <https://doi.org/10.1111/ecog.01357>.
- Forsman A. Effects of genotypic and phenotypic variation on establishment are important for conservation, invasion and infection biology. *Proc Biol Sci*. 2014;111:302–7.
- Hamilton JA, Miller JM. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv Biol*. 2016;30(1):33–41. <https://doi.org/10.1111/cobi.12574>.

10. Frankham R. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Mol Ecol.* 2015;24(11):2610–8. <https://doi.org/10.1111/mec.13139>.
11. Frankham R, Ballou JD, Eldridge MD, Lacy RC, Ralls K, Dudash MR, et al. Predicting the probability of outbreeding depression. *Conserv Biol.* 2011; 25(3):465–75. <https://doi.org/10.1111/j.1523-1739.2011.01662.x>.
12. Edmands S, Timmerman CC. Modeling factors affecting the severity of outbreeding depression. *Conserv Biol.* 2003;17(3):883–92. <https://doi.org/10.1046/j.1523-1739.2003.02026.x>.
13. Whitlock R, Stewart GB, Goodman SJ, Piortney SB, Butlin RK, Pullin AS, et al. A systematic review of phenotypic responses to between-population outbreeding. *Environ Evid.* 2013;2(1):13. <https://doi.org/10.1186/2047-2382-2-13>.
14. Edmands S. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Mol Ecol.* 2007;16(3):463–75. <https://doi.org/10.1111/j.1365-294X.2006.03148.x>.
15. Todesco M, Pascual MA, Owens GL, Ostevik KL, Moyers BT, Hübner S, et al. Hybridization and extinction. *Evol Appl.* 2016;9(7):892–908. <https://doi.org/10.1111/eva.12367>.
16. Edmands S. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution.* 1999;53(6):1757–68. <https://doi.org/10.1111/j.1558-5646.1999.tb04560.x>.
17. Edmands S. Does parental divergence predict reproductive compatibility? *Trends Ecol Evol.* 2002;17(11):520–7. [https://doi.org/10.1016/S0169-5347\(02\)02585-5](https://doi.org/10.1016/S0169-5347(02)02585-5).
18. Raabová J, Münzbergová Z, Fischer M. Ecological rather than geographic or genetic distance affects local adaptation of the rare perennial herb, *Aster amellus*. *Biol Conserv.* 2007;139(3-4):348–57. <https://doi.org/10.1016/j.biocon.2007.07.007>.
19. Willi Y, van Buskirk J. Genomic compatibility occurs over a wide range of parental genetic similarity in an outcrossing plant. *Proc Biol Sci.* 2005;272:1333–8.
20. Amos W, Wilmer JW, Fullard K, Burg TM, Croxall JP, Bloch D, et al. The influence of parental relatedness on reproductive success. *Proc Biol Sci.* 2001;268(1480):2021–7. <https://doi.org/10.1098/rspb.2001.1751>.
21. Bean K, Amos W, Pomeroy PP, Twiss SD, Coulson TN, Boyd IL. Patterns of parental relatedness and pup survival in the grey seal (*Halichoerus grypus*). *Mol Ecol.* 2004;13(8):2365–70. <https://doi.org/10.1111/j.1365-294X.2004.02199.x>.
22. Jolivet C, Rogge M, Degen B. Molecular and quantitative signatures of biparental inbreeding depression in the self-incompatible tree species *Prunus avium*. *Heredity.* 2013;110(5):439–48. <https://doi.org/10.1038/hdy.2012.103>.
23. Phillips KP, Jorgensen TH, Jolliffe KG, Richardson DS. Evidence of opposing fitness effects of parental heterozygosity and relatedness in a critically endangered marine turtle? *J Evol Biol.* 2017;30(11):1953–65. <https://doi.org/10.1111/jeb.13152>.
24. Li YC, Korol AB, Fahima T, Nevo E. Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol.* 2004;21(6):991–1007. <https://doi.org/10.1093/molbev/msh073>.
25. Srivastava S, Avvaru AK, Sowpati DT, Mishra RK. Patterns of microsatellite distribution across eukaryotic genomes. *BMC Genomics.* 2019;20(1):153. <https://doi.org/10.1186/s12864-019-5516-5>.
26. Väli Ü, Einarsson A, Waits L, Ellegren H. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Mol Ecol.* 2008;17(17):3808–17. <https://doi.org/10.1111/j.1365-294X.2008.03876.x>.
27. Hodel RGJ, Chen S, Payton AC, McDaniel SF, Soltis P, Soltis DE. Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. *Sci Rep.* 2017;7(1):17598. <https://doi.org/10.1038/s41598-017-16810-7>.
28. Lemopoulos A, Prokkola JM, Uusi-Heikkilä S, Vasemägi A, Huusko A, Hyvärinen P, et al. Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — implications for brown trout conservation. *Ecol Evol.* 2019; 9(4):2106–20. <https://doi.org/10.1002/ece3.4905>.
29. Fenster CB, Galloway LF. Population differentiation in an annual legume: genetic architecture. *Evolution.* 2000;54(4):1157–72. <https://doi.org/10.1111/j.0014-3820.2000.tb00551.x>.
30. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet.* 2016;17(2):81–92. <https://doi.org/10.1038/nrg.2015.28>.
31. Santure AW, Stapley J, Ball AD, Birkhead TR, Burke T, Slate J. On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Mol Ecol.* 2010;19(7):1439–51. <https://doi.org/10.1111/j.1365-294X.2010.04554.x>.
32. Sunde J, Yıldırım Y, Tibblin P, Forsman A. Comparing the performance of microsatellites and RADseq in population genetic studies: analysis of data for pike (*Esox lucius*) and a synthesis of previous studies. *Front Genet.* 2020; 11. <https://doi.org/10.3389/fgene.2020.00218>.
33. Forstmeier W, Schielzeth H, Mueller JC, Ellegren H, Kempnaers B. Heterozygosity–fitness correlations in zebra finches: microsatellite markers can be better than their reputation. *Mol Ecol.* 2012;21(13):3237–49. <https://doi.org/10.1111/j.1365-294X.2012.05593.x>.
34. Charlesworth B, Charlesworth D. Population genetics from 1966 to 2016. *Heredity.* 2017;118(1):2–9. <https://doi.org/10.1038/hdy.2016.55>.
35. McClelland EK, Naish KA. What is the fitness outcome of crossing unrelated fish populations? A meta-analysis and an evaluation of future research directions. *Conserv Genet.* 2007;8(2):397–416. <https://doi.org/10.1007/s10592-006-9178-x>.
36. Sunde J, Tibblin P, Larsson P, Forsman A. Sex-specific effects of outbreeding on offspring quality in pike (*Esox lucius*). *Ecol Evol.* 2018;8(21):10448–59. <https://doi.org/10.1002/ece3.4510>.
37. Lehtonen H, Leskinen E, Selen R, Reinikainen M. Potential reasons for the changes in the abundance of pike, *Esox lucius*, in the western gulf of Finland, 1939–2007. *Fish Manage Ecol.* 2009;16(6):484–91. <https://doi.org/10.1111/j.1365-2400.2009.00701.x>.
38. Ljunggren L, Sandstrom A, Bergstrom U, Mattila J, Lappalainen A, Johansson G, et al. Recruitment failure of coastal predatory fish in the Baltic Sea coincident with an offshore ecosystem regime shift. *ICES J Mar Sci.* 2010; 67(8):1587–95. <https://doi.org/10.1093/icesjms/fsq109>.
39. Larsson P, Tibblin P, Koch-Schmidt P, Engstedt O, Nilsson J, Nordahl O, et al. Ecology, evolution, and management strategies of northern pike populations in the Baltic Sea. *Ambio.* 2015;44(5):451–61. <https://doi.org/10.1007/s13280-015-0664-6>.
40. Skov C, Nilsson PA. Evaluating stocking of YOY pike *Esox lucius* as a tool in the restoration of shallow lakes. *Freshwat Biol.* 2007;52(9):1834–45. <https://doi.org/10.1111/j.1365-2427.2007.01795.x>.
41. Nordahl O, Koch-Schmidt P, Sunde J, Yıldırım Y, Tibblin P, Forsman A, et al. Genetic differentiation between and within ecotypes of pike (*Esox lucius*) in the Baltic Sea. *Aquat Conserv Mar Freshwat Ecosyst.* 2019;29(11):1923–35. <https://doi.org/10.1002/aqc.3196>.
42. Berggren H, Nordahl O, Tibblin P, Larsson P, Forsman A. Testing for local adaptation to spawning habitat in sympatric subpopulations of pike by reciprocal translocation of embryos. *PLoS One.* 2016;11(5):e0154488. <https://doi.org/10.1371/journal.pone.0154488>.
43. Tibblin P, Forsman A, Koch-Schmidt P, Nordahl O, Johannessen P, Nilsson J, et al. Evolutionary divergence of adult body size and juvenile growth in sympatric subpopulations of a top predator in aquatic ecosystems. *Am Nat.* 2015;186(1):98–110. <https://doi.org/10.1086/681597>.
44. Tibblin P, Berggren H, Nordahl O, Larsson P, Forsman A. Causes and consequences of intra-specific variation in vertebral number. *Sci Rep.* 2016; 6(1):26372. <https://doi.org/10.1038/srep26372>.
45. Sunde J, Larsson P, Forsman A. Adaptations of early development to local spawning temperature in anadromous populations of pike (*Esox lucius*). *BMC Evol Biol.* 2019;19(1):148. <https://doi.org/10.1186/s12862-019-1475-3>.
46. Sunde J, Tibblin P, Larsson P, Forsman A. Data from: sex-specific effects of outbreeding on offspring quality in pike (*Esox lucius*). Dryad Digital Repository. 2018. <https://doi.org/10.5061/dryad.dd64hf3>.
47. de Villemereuil P, Gaggiotti OE. A new FST-based method to uncover local adaptation using environmental variables. *Methods Ecol Evol.* 2015;6(11):1248–58. <https://doi.org/10.1111/2041-210X.12418>.
48. Gain C, François O. LEA 3: factor models in population genetics and ecological genomics with R. *Mol Ecol Resour.* 2020. <https://doi.org/10.1111/1755-0998.13366>. <https://doi.org/10.1111/1755-0998.13366.2020.2010.2006.327783>.
49. Harshman LG, Hoffmann AA. Laboratory selection experiments using *Drosophila*: what do they really tell us? *Trends Ecol Evol.* 2000;15(1):32–6. [https://doi.org/10.1016/S0169-5347\(99\)01756-5](https://doi.org/10.1016/S0169-5347(99)01756-5).
50. Ellegren H, Sheldon BC. Genetic basis of fitness differences in natural populations. *Nature.* 2008;452(7184):169–75. <https://doi.org/10.1038/nature06737>.
51. Bicskei B, Taggart JB, Bron JE, Glover KA. Transcriptomic comparison of communally reared wild, domesticated and hybrid Atlantic salmon fry under stress and control conditions. *BMC Genet.* 2020;21(1):57. <https://doi.org/10.1186/s12863-020-00858-y>.

52. Lutz ML, Tonkin Z, Yen JDL, Johnson G, Ingram BA, Sharley J, et al. Using multiple sources during reintroduction of a locally extinct population benefits survival and reproduction of an endangered freshwater fish. *Evol Appl*. 2020;00:1–15.
53. Waser NM, Price MV, Shaw RG. Outbreeding depression varies among cohorts of *Ipomopsis aggregata* planted in nature. *Evolution*. 2000;54(2):485–91. <https://doi.org/10.1111/j.0014-3820.2000.tb00051.x>.
54. Rollinson N, Keith DM, Houde ALS, Debes PV, McBride MC, Hutchings JA. Risk assessment of inbreeding and outbreeding depression in a captive-breeding program. *Conserv Biol*. 2014;28(2):529–40. <https://doi.org/10.1111/cobi.12188>.
55. Craig JF. A short review of pike ecology. *Hydrobiologia*. 2008;601(1):5–16. <https://doi.org/10.1007/s10750-007-9262-3>.
56. Donadi S, Austin ÅN, Bergström U, Eriksson BK, Hansen JP, Jacobson P, et al. A cross-scale trophic cascade from large predatory fish to algae in coastal ecosystems. *Proc R Soc B-Biol Sci*. 2017;284(1859):20170045. <https://doi.org/10.1098/rspb.2017.0045>.
57. Nilsson J, Flink H, Tibblin P. Predator-prey role reversal may impair the recovery of declining pike populations. *J Anim Ecol*. 2019;88(6):927–39. <https://doi.org/10.1111/1365-2656.12981>.
58. Pierce RB, Tomcko CM, Schupp DH. Exploitation of northern pike in seven small north-Central Minnesota Lakes. *N Am J Fish Manage*. 1995;15(3):601–9. [https://doi.org/10.1577/1548-8675\(1995\)015<0601:EONPIS>2.3.CO;2](https://doi.org/10.1577/1548-8675(1995)015<0601:EONPIS>2.3.CO;2).
59. Forsman A, Tibblin P, Berggren H, Nordahl O, Koch-Schmidt P, Larsson P. Pike *Esox lucius* as an emerging model organism for studies in ecology and evolutionary biology: a review. *J Fish Biol*. 2015;87(2):472–9. <https://doi.org/10.1111/jfb.12712>.
60. Olsson J. Past and current trends of coastal predatory fish in the Baltic Sea with a focus on perch, pike, and pikeperch. *Fishes*. 2019;4(1):7. <https://doi.org/10.3390/fishes4010007>.
61. Laikre L, Miller LM, Palme A, Palm S, Kapuscinski AR, Thoreson G, et al. Spatial genetic structure of northern pike (*Esox lucius*) in the Baltic Sea. *Mol Ecol*. 2005;14(7):1955–64. <https://doi.org/10.1111/j.1365-294X.2005.02570.x>.
62. Nilsson J. Predation of northern pike (*Esox lucius* L.) eggs: a possible cause of regionally poor recruitment in the Baltic Sea. *Hydrobiologia*. 2006;553(1):161–9. <https://doi.org/10.1007/s10750-005-1949-8>.
63. Craig JF. Pike - biology and exploitation. London: Chapman & Hall; 1996.
64. Larsen PF, Hansen MM, Nielsen EE, Jensen LF, Loeschcke V. Stocking impact and temporal stability of genetic composition in a brackish northern pike population (*Esox lucius* L.), assessed using microsatellite DNA analysis of historical and contemporary samples. *Heredity*. 2005;95:136–43.
65. Sunde J, Tamario C, Tibblin P, Larsson P, Forsman A. Variation in salinity tolerance between and within anadromous subpopulations of pike (*Esox lucius*). *Sci Rep*. 2018;8(1):22. <https://doi.org/10.1038/s41598-017-18413-8>.
66. Sunde J, Yildirim Y, Tibblin P, Forsman A. Data from: comparing the performance of microsatellites and RADseq in population genetic studies: analysis of data for pike (*Esox lucius*) and a synthesis of previous studies. Dryad Digit Repository. 2020. <https://doi.org/10.5061/dryad.dd64hf3>.
67. Sunde J, Yildirim Y, Tibblin P, Forsman A. Data from: Comparing the performance of microsatellites and RADseq in population genetic studies: analysis of data for pike (*Esox lucius*) and a synthesis of previous studies. NCBI Seq Read Arch. 2020; BioProject PRJNA586770, submission ID: SUB6466329. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA586770>.
68. RStudio Team. RStudio: integrated development for R. Boston: MA; 2017.
69. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
70. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289–300.

Publisher's Note

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

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

