

Incomplete bioinformatic filtering and inadequate age and growth analysis lead to an incorrect inference of harvested-induced changes

Wesley A. Larson¹  | Daniel A. Isermann²  | Zachary S. Feiner³

¹National Oceanographic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Auke Bay Laboratories, Juneau, AK, USA

²U.S. Geological Survey, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, Stevens Point, WI, USA

³Wisconsin Department of Natural Resources, Office of Applied Science, Science Operations Center, Madison, WI, USA

Correspondence

Wesley A. Larson, National Oceanographic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Auke Bay Laboratories, 17109 Point Lena Loop Road, Juneau, Alaska 99801, USA.

Email: wes.larson@noaa.gov

Abstract

Understanding the evolutionary impacts of harvest on fish populations is important for informing fisheries management and conservation and has become a growing research topic over the last decade. However, the dynamics of fish populations are highly complex, and phenotypes can be influenced by many biotic and abiotic factors. Therefore, it is vital to collect robust data and explore multiple alternative hypotheses before concluding that fish populations are influenced by harvest. In their recently published manuscript, Bowles et al, *Evolutionary Applications*, 13(6):1128 conducted age/growth and genomic analysis of walleye (*Sander vitreus*) populations sampled 13–15 years (1–2.5 generations) apart and hypothesized that observed phenotypic and genomic changes in this time period were likely due to harvest. Specifically, Bowles et al. (2020) documented differential declines in size-at-age in three exploited walleye populations compared to a separate, but presumably less-exploited, reference population. Additionally, they documented population genetic differentiation in one population pair, homogenization in another, and outlier loci putatively under selection across time points. Based on their phenotypic and genetic results, they hypothesized that selective harvest had led to fisheries-induced evolution (referred to as nascent changes) in the exploited populations in as little as 1–2.5 generations. We re-analyzed their data and found that (a) sizes declined across both exploited and reference populations during the time period studied and (b) observed genomic differentiation in their study was the result of inadequate data filtering, including retaining individuals with high amounts of missing data and retaining potentially undersplit and oversplit loci that created false signals of differentiation between time points. This re-analysis did not provide evidence for phenotypic or genetic changes attributable to harvest in any of the study populations, contrasting the hypotheses presented by Bowles et al. (2020). Our comment highlights the potential pitfalls associated with conducting age/growth analyses with low sample sizes and inadequately filtering genomic datasets.

This draft manuscript is distributed solely for purposes of scientific peer review. Its content is deliberative and predecisional, so it must not be disclosed or released by reviewers. Because the manuscript has not yet been approved for publication by the U.S. Geological Survey (USGS), it does not represent any official USGS finding or policy.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd. This article has been contributed to by US Government employees and their work is in the public domain in the USA.

KEYWORDS

age and growth analysis, bioinformatic filtering, fisheries harvest, population genomics, RADseq, walleye

1 | INTRODUCTION

Understanding whether harvest impacts the long-term viability of fish populations is vital for informing conservation and fisheries management (Dunlop, Eikeset, & Stenseth, 2015). Fortunately, this important topic is becoming more well-studied as fisheries managers are beginning to understand that harvest can threaten the long-term viability of populations even if population abundances are not decreasing (Hutchings & Kuparinen, 2020). Bowles, Marin, Mogensen, MacLeod, and Fraser (2020) add to the growing body of literature on this topic by investigating potential phenotypic and genetic changes resulting from harvest in populations of walleye (*Sander vitreus*) from Mistassini Lake, a large lake in Quebec, Canada. The paper documents phenotypic and genomic changes in these walleye populations over 1–2.5 generations that the authors hypothesize are likely due to harvest, a striking result given that most other studies do not find impacts of harvest on this short of a timescale. However, we believe that these conclusions are supported by inadequate age/growth data and genomic evidence that disappears when data are filtered correctly. We outline some specific issues with the age/growth and genomic data below and argue that if the data were analyzed correctly, the conclusions presented by the authors would not have been reached.

2 | BODY SIZE ANALYSIS

We agree that the Indigenous Knowledge (IK) obtained by the authors supports the assertion that harvest has affected size structure of walleye stocks examined in the study. However, whether those reductions can be directly attributed to selective removal of specific genotypes or phenotypes from a wild population, and that harvest differentially affected southern and northern walleye populations, requires robust data and sampling that occurs at least somewhat consistently over time (Dieckmann & Heino, 2007; Sinclair, Swain, & Hanson, 2002). Size distributions and length-at-age in individual walleye populations vary extensively on a temporal scale (Hansen & Nate, 2014; Isermann, 2007; Wuellner, Willis, Blackwell, & Lott, 2011) and can vary among populations or stocks at relatively small spatial scales (Pedersen et al., 2017; Sass, Hewett, Beard, Fayram, & Kitchell, 2004; Wang et al., 2012). Differences in recruitment, density, and environmental conditions are just a few of the many factors that have been shown to influence walleye size distributions or length-at-age (Lester, Shuter, Kushneriuk, & Marshall, 2000; Pedersen et al., 2017; Quist, Guy, Schultz, & Stephen, 2003; Sass & Kitchell, 2005). Changes in harvest may or may not affect size structure in walleye populations (Haglund, Isermann, & Sass, 2016; Hansen & Nate, 2014; Sass & Shaw, 2018, 2020; Sullivan, 2003); at

a fundamental level, the effects of harvest on fish populations will vary in relation to abundance and exploitation rate (i.e., percentage of population harvested on annual basis).

In an experimental sense, Bowles et al. (2020) must show that growth trends differed between the southern systems and northern reference system and that the only factor that changed between the two was walleye harvest. This can be done in whole lake experiments with explicit control of harvest (e.g., Haglund et al., 2016; Sass & Shaw, 2018), but is much more difficult in wild populations. Mistassini Lake is a large system, and the presumably low exploitation reference tributary (Takwa River) is located more than 150 km from the high exploitation tributaries in the south; this latitudinal separation alone is sufficiently large to suggest that the reference population might encounter different environmental conditions and exhibit different population dynamics than walleye farther south (as discussed in Dupont, Bourret, & Bernatchez, 2007). Indeed, Bowles et al. (2020) report that walleye spawning in southern tributaries “stay close” to these locations after spawning, represent the majority of fish harvested during the mixed fishery, and exhibit different growth patterns than northern fish (Dupont et al., 2007). Therefore, it is plausible that these populations were experiencing different environmental conditions throughout the study period that could influence life history trait expression.

Because of these complexities, the reliance of Bowles et al. (2020) on two observations of walleye length-at-age 12 years apart (2003 to 2015) provides little context for the variability or co-variability of growth among study populations. The authors provide no quantifiable information regarding abundance and exploitation rates of the walleye populations in question, or the size selectivity of fishers, making it impossible to discern the effects of harvest. Moreover, examination of the otoliths (conducted by Isermann's research group at University of Wisconsin-Stevens Point) revealed that the age structure of walleye from these tributaries was dominated by fish older than age 10 with a relatively high proportion of fish age 15 or older. In a species with an approximate maximum age of 20 years (including these populations, Figure 1; Venturelli, Lester, Marshall, & Shuter, 2010), this suggests that overall mortality for the study populations was relatively low and growth, maturation, and susceptibility to harvest was delayed (Lavigne, Lucotte, & Paquet, 2010). The lack of demographic and ecological context for these populations at a sufficiently fine temporal scale leads to a number of unanswered questions that are essential to addressing the authors' hypotheses. Did age composition change over time? How extensive was exploitation in each study population? Does exploitation occur at levels that might be expected to cause selective pressures, and is size selectivity sufficiently intense? What if some stocks were historically more abundant than others? What if exploitation rates vary significantly among populations and among years within a population? Did size

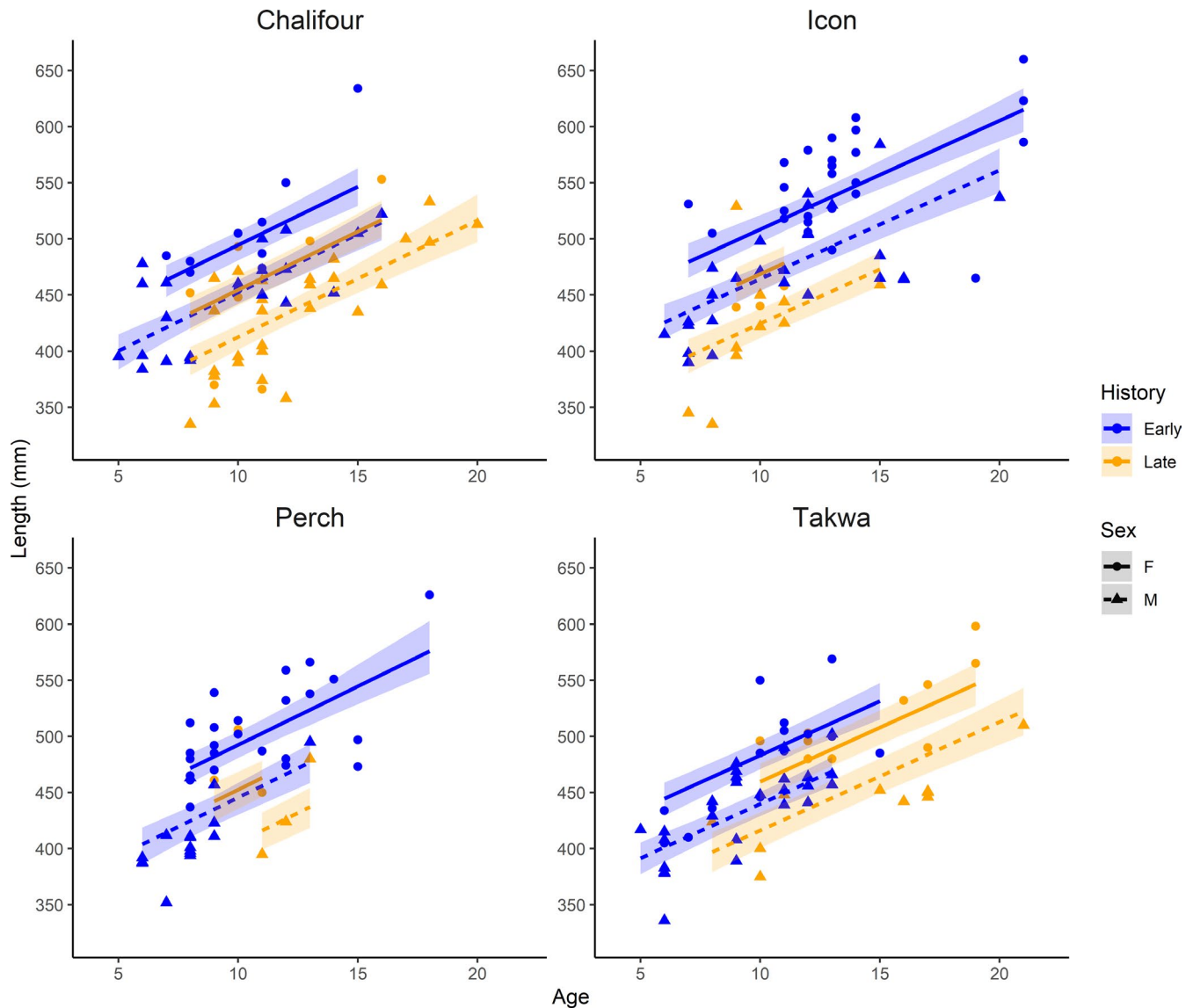


FIGURE 1 Length-at-age observations (points) and predicted model fits (lines with shaded 95% credible intervals) for walleye captured from the four tributaries documented in Bowles et al. (2020). Observations are color-coded by time period (historical, 2002 = blue; contemporary, 2015/2017 = orange) and sex (female = circles and solid lines, males = triangles and dashed lines). Length-at-age declined from the historical to contemporary period in all four populations

distributions and length-at-age of walleye from southern tributaries fluctuate over this period or were there detectable trends in these metrics that would at least suggest the possibility that harvest might be exerting an effect?

To expand on one of these questions as an example, declines in body size over time could be rooted in harvest, but could also reflect loss of older, larger fish from strong year classes that may have been produced more than a decade before samples were collected based on the age distribution of the walleye examined. In their reporting of harvest, Bowles et al. (2020) note that harvest was 54% higher in 2003 but varied little in other years from 1997, 2011, and 2015 (Table S2), despite increases in the local indigenous population. The production of a large year class in the years preceding 2003 could have led to increased harvest observed in 2003, reduced growth through density-dependent mechanisms of subsequent cohorts,

and a perceived reduction in old/large fish in the study's contemporary sampling as those cohorts would have comprised the older fish sampled for ages in 2015–2017. Therefore, based on the ages of fish collected by the authors, walleye abundance and size structure in the contemporary period were undoubtedly affected by recruitment events that occurred during the historic period. The authors never mention this important fact. Additionally, correlations among growth, mortality, and maturation complicate the interpretation of growth differences because the fish included in the study were all sampled while spawning. Walleye maturation is highly plastic (Feiner & Höök, 2015), and increased growth, condition, or energy reserves can increase the probability that small walleyes will mature and spawn (Feiner, Shaw, & Sass, 2019; Henderson & Morgan, 2002). If fishing or other factors (e.g., temperature, food availability) affected growth or condition of walleye in the lake, individuals experiencing

more favorable environmental conditions may become mature at smaller sizes for a given age, leading to the phenomenon of catching smaller fish on the spawning grounds as observed by Bowles et al. (2020) in the absence of any selective or genetic impacts of fishing itself. The authors did not account for any of these other factors in their analyses, and yet they suggest the differences they observed in their size and growth metrics may be the result of harvest-induced selection occurring within 1–2.5 generations. At best, harvest-induced evolution would be one of many factors that may explain the trends the authors observed, and there is no way to reliably identify the mechanisms driving size trends given the authors' study design.

3 | LENGTH-AT-AGE ANALYSIS

Because of the issues with the body size analysis we outlined above, evidence for phenotypic effects due to harvest in this study hinges on the length-at-age analysis. Our issues with the length-at-age analysis are twofold: First, we argue that the data are insufficient to examine patterns in growth, and second, we argue their length-at-age model does not show differential growth patterns between southern and northern walleye populations. Length-at-age analysis typically involves a stratified-random sampling approach where calcified structures are removed for age estimation from a specific number of fish of each sex within defined length categories (i.e., structures removed from 10 fish of each sex per 10-mm length group) and ages from these fish are extrapolated back to the entire sample (Coggins, Gwinn, & Allen, 2013). However, in Bowles et al. (2020), there appears to be no specific sampling target defined by total number of fish or percentage of overall samples, and sample sizes are generally small. The total number of walleye selected for age estimation in a year from a specific tributary was less than 25 in five of the ten annual samples that were used in analyses and was greater than or equal to 50 in only two instances, indicating that tributary-specific sample sizes for length-at-age were likely less than five for many ages. In the most extreme case, the contemporary age sample for the Perch River is reported as six fish total across all ages. For female walleye, total sample sizes used to characterize length-at-age over the three-year "contemporary" sampling period never exceeded 11 fish for any tributary. Attempting to characterize female length-at-age for a population using 11 fish or fewer in a single year is not reasonable, let alone over a three-year period. Therefore, sample sizes were insufficient for a legitimate analysis of female walleye body size or length-at-age.

The authors attempted to account for these issues, in addition to potential biases due to sampling gear and aging structures, by using a Bayesian hierarchical modeling approach. However, Bayesian models, while better able to handle difficult data distributions or hierarchical designs, are not a cure for flawed or insufficient sampling—the data are the data (Lecoutre, 1999; O'Hagan, 2006). For instance, error terms for sampling gear or aging structure are not included and the model appears to assume that the effects of age and sex

are independent, which is likely violated (Feiner et al., 2017). Most importantly, the model provides no direct comparison of length-at-age among populations, even though this is the phenotypic shift that represents the crux of the paper's argument. Rather, the length-at-age model assesses whether variation in length of walleye (not length-at-age) can be explained by age, sampling period (historic versus contemporary), location, or sex, essentially a duplication of their previous analyses regarding length of fish but using only the fish that were selected for age estimation. Here, assuming all other ecological factors remained constant, the specific questions of interest are 1) are walleye from southern tributaries now smaller at a given age (e.g., age 11) than they were in the historic period?, and 2) was this trend also apparent for fish from the reference tributary? At a minimum, an age*location*history interaction term would be needed to test whether the effect of age (i.e., growth) differed between locations and histories. However, the results of the model are unclear. In the Results, the model is described as suggesting "fish were 29.4 mm larger contemporaneously than historically" and "fish in southern rivers were 13.7 mm smaller relative to fish in the north in contemporary relative to historical samples." If the first statement is accurate, this refutes the hypothesis that harvest negatively influenced growth over time. Moreover, the 13.7 mm difference in temporal changes in size between locations was marginal, with a confidence interval that overlapped zero with 13% of the posterior less than zero. This suggests, as we do, that the phenotypic shifts in body size were likely caused by ecological mechanisms.

We re-analyzed the length-at-age data in an attempt to duplicate Bowles et al.'s (2020) results and test for contrasts in the predicted size of age-11 fish across populations and time periods to address the questions we posed above. Data were retrieved from the Dryad repository <https://doi.org/10.5061/dryad.5tb2rbp1z> and the file "Bowles_et_al-Evol_App-2020-Age_infoR.xlsx." We used a hierarchical Bayesian model parameterized following the one described in Bowles et al. (2020); however, our results did not match the reported results in the paper, we suspect because of some differences in model parameterization based on the relatively brief description of the model in the paper. In our model, we included random intercepts and slopes for the effect of age by river, random intercepts for the effect of sex by river, and fixed effects of the categorical variables for location (north or south) and history (historical or contemporary):

$$\mu \sim (\beta_0 + u_{0j} + \beta_{2j} \times \text{sex}) + (\beta_1 + u_{1j}) \times \text{age} + \beta_{3i} + \beta_{4k} + \beta_{5i,k}$$

$$\text{Length} \sim N(\mu, \sigma)$$

where β_0 is the grand mean, u_0 and u_1 are the random intercepts and slopes for age by river j , β_2 are the random effects of sex by river j , distributed around a mean effect $\beta_{2j} \sim N(\mu_{\beta_2}, \sigma_{\beta_2})$, β_3 is the effect of location k , β_4 is the effect of history k , and β_5 is the interaction between location i and history k . We used a vague Cholesky prior for the random intercepts and slopes of age, half-Cauchy priors for other parameter variances, and vague normal priors for parameter means. We calculated the predicted length of an age-11 fish for each time period and location and compared the change between periods between

north and south locations. We ran the model for 10,000 steps with a 5,000 warmup, thinning to every 10th step, and assured convergence using traceplots, effective sample size >1,000, and Gelman-Rubin statistics approaching 1. The model was fit in Stan in R via package “rstan” (Carpenter et al., 2017; R & Development Core Team, 2019; Stan & Development Team, 2020). Model code is available in supplemental file 1.

Our model predicted that all populations, including the reference (“unfished”) population, were reduced in size-at-age from historical to contemporary periods (Figure 1; see Table S1 for all parameter estimates). Age-11 fish in the north (i.e., Takwa) declined by approximately 24 mm (CI: 5.4–41.2 mm) in total length, while fish in the south (i.e., Chalifour, Perch, and Icon) declined in length by 40 mm (CI: 28.5–51.5 mm). The 95% credible intervals for this contrast in changes in length (South–North) overlapped zero (mean = 16 mm, CI: –4.2–36.1 mm) suggesting that differences in changes in size-at-age between regions may be marginal and reducing the strength of Bowles et al.’s argument that growth (measured by size-at-age) changed differentially among populations over time. Importantly, our re-analysis of these data illustrated that size-at-age was potentially reduced in both the northern (low exploitation) and southern (high exploitation) areas of the lake, providing evidence that this trend may be the result of similar effects of harvest in both populations, and/or environmental factors.

4 | Genomic analysis

The authors make three primary conclusions regarding the genetic characteristics of the walleye populations in their study: (a) genetic homogenization has occurred between two southern (i.e., heavily exploited) rivers over the study period (PER and ICO), (b) the genetic structure in at least one southern river (CHA) has increased and become significant, while no changes were observed in the northern (i.e., less exploited) river, and (c) there is evidence of selection between historic and contemporary samples in southern rivers. We re-analyzed the data presented in Bowles et al. (2020) and found that all three of these conclusions are likely the result of incorrect bioinformatic processing and/or inadequate data filtering. Specifically, we found that the apparent genetic homogenization between populations was likely the result of high levels of missing data (>50% for all individuals) in the PER03 sample and that signatures of population differentiation and selection were largely driven by low F_{IS} /high F_{ST} loci that are likely bioinformatic artifacts.

Understanding how different aspects of genomics datasets can influence downstream analyses is vital for conducting robust genomics research. This is especially true for datasets with low genetic structure, where small data inconsistencies such as differences in missing data or library effects can create patterns that appear to represent true population structure. Recently, O’Leary, Puritz, Willis, Hollenbeck, and Portnoy (2018) published an excellent overview of RAD data filtering, where they highlighted multiple potential issues with RAD datasets, how these issues might influence downstream

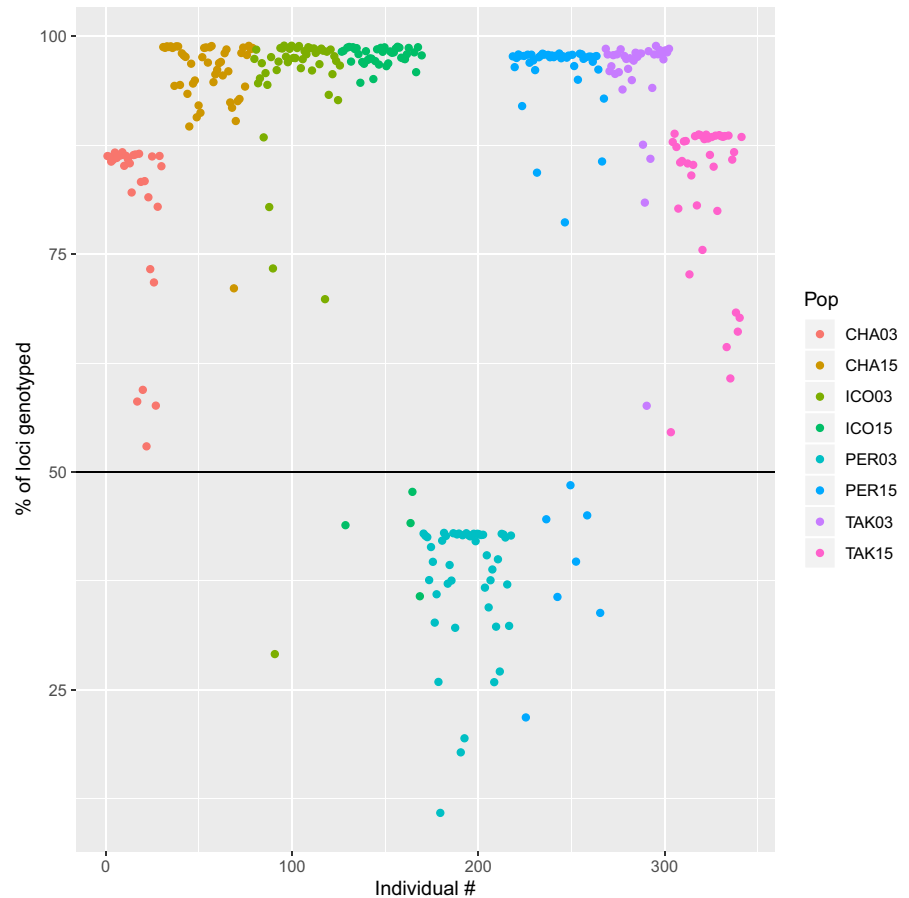
analyses, and how to identify and mitigate these issues. Two particular issues raised in O’Leary et al. (2018), inconsistent sequencing of loci (i.e., missing data) and clustering errors (i.e., oversplitting or undersplitting loci), are likely the cause of the spurious results reported in Bowles et al. (2020).

One of the first steps for data filtering recommended by O’Leary et al. (2018) is a visualization of missingness across loci and individuals. We conducted this analysis for the Bowles dataset (Figure 2, Figure S1) using the Bowles_et_al-EvolApp-2020-r12dWithOutOddBallSamps_includ-outlier_loci.genepop file from the DRYAD repository for this paper <https://doi.org/10.5061/dryad.5tb2rbp1z>. This dataset was used as the starting point for all analyses in our comment, all R code used to analyze the data are in supplementary files, and population abbreviations used throughout the comment are as follows: Chalifour (CHA), Icon (ICO), Perch (PER), and Takwa (TAK) with year appended (i.e., PER03 are walleye from the 2003 samples of the Perch River). Unfortunately, the raw sequence data and sex information were not available in the supplementary materials, so we were unable to replicate all of the authors analyses (i.e., sex-specific analysis, STACKs analysis). Genotyping rates were relatively consistent within populations but variable across populations (Figure 2). Most notably, all individuals in the PER03 population had >50% missing data, with some individuals displaying nearly 90% missing data, while the other populations sequenced relatively well (<20% missing data on average). The authors did not impose any missing data filter to remove individuals that were sequenced poorly.

Regarding locus-specific filtering, Bowles et al. (2020) started with a typical and what we believe appropriate method of retaining loci that were genotyped in at least 80% of individuals within each population (i.e., $-r$ 0.8 in the populations module of STACKs; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). However, they added an additional filtering component that resulted in the inclusion of many loci with 100% missing data in certain populations. Specifically, they retained loci if they met the 80% genotype rate criteria in 6/8 or 4/6 populations depending on which dataset they were analyzing (i.e., a locus could be missing 100% of data in 2 populations if there was <20% missing data in the other populations). This methodology resulted in a missing data pattern where the maximum percentage of missing data for a given locus was only ~40%, but certain populations contained large numbers of loci with 100% missing data (Figure S1). Most notably, over half of the loci were missing 100% of data in the PER03 population, and ~10% of loci were missing 100% of data in the CHA03 and TAK15 populations.

To test the influence of these patterns of missing data on population structure, we conducted individual-based PCAs in adegenet (Jombart, 2008) with the unfiltered dataset (i.e., the dataset presented in Bowles et al., 2020) and a dataset with only the loci that were genotyped in >50% of individuals in the PER03 population (Figure 3.). In the unfiltered dataset, the PER03 population is clearly separated from the PER15/ICO03/ICO15 cluster of points (Figure 3a). However, when loci missing >50% of data are removed, the PER03 individuals overlap completely with the PER15/ICO03/ICO15 cloud (Figure 3b). These results illustrate that the claim

FIGURE 2 Percentage of missing data for all individuals in the dataset. Each individual is represented by a dot, and populations are color-coded. See text for population abbreviations



that the PER population was historically differentiated from ICO and genetically homogenized with the ICO population between 2003 and 2015 is false. PER03 is highly similar to ICO03/ICO15/PER15 when data are filtered properly, and the apparent structure between PER03 and these other three populations was caused by missing data. In fact, analysis of the same samples with microsatellites (Dupont et al., 2007) also supports our conclusion that PER03 and ICO03 are genetically homogenous, providing further evidence that structure between PER03 and ICO03 is a bioinformatic artifact caused by missing data.

Missing data did not explain the apparent genetic differences between the CHA03 and CHA15 samples, as a similar PCA analysis with loci missing >50% in CHA03 removed still revealed strong structure between the two (Figure 4b). Rather, we believe population structure observed between the CHA03 and CHA15 samples was caused by inadequate bioinformatic filtering resulting in the retention of oversplit or undersplit loci. Adequate filtering is especially important when detecting loci under selection, as bioinformatic artifacts can often manifest as outlier loci with high F_{ST} values (O'Leary et al., 2018). The first step in vetting outlier loci used in our laboratory is visually examining their F_{IS} and F_{ST} values, as F_{IS} values that deviate substantially from zero can reveal either laboratory or bioinformatic errors (e.g., over or undersplitting) or interesting biological processes that create departures from Hardy-Weinberg equilibrium. When we were first exploring the Bowles et al. (2020) dataset, we sorted loci by F_{ST} and realized that many of the loci

with the highest F_{ST} s had highly negative F_{IS} values. We then isolated comparisons to population pairs sampled 12 years apart and graphed F_{ST} and F_{IS} values (WC 1984) calculated in Genepop V4 (Rousset, 2008; Figure 5). This analysis revealed that a large number of the high F_{ST} loci in the CHA comparison also displayed very negative F_{IS} values (Figure 5b), sharply contrasting results from other sample pairs (Figure 5). Notably, the CHA population pair has 70 loci with $F_{ST} > 0.2$ and $F_{IS} < -0.2$, whereas all other population pairs have only 1 or 0 loci that fall into this category. These highly negative F_{IS} values are the result of extremely high locus-specific heterozygosities in the CHA03 population (>90% in some cases) and extremely low locus-specific heterozygosities in all other populations (<5% in some cases; Figure 6). In other words, some of these loci are nearly fixed heterozygotes in the CHA03 population while being nearly invariant in the other populations in the study including in the same population 12 years (i.e., <3 generations) later (CHA15). Put simply, this result is not biologically possible and is most likely a bioinformatic artifact.

Over and undersplit loci are difficult to detect and remove, as noted by O'Leary et al. (2018). However, O'Leary et al. (2018) recommend plotting heterozygosity and allele depth as a potential way to identify splitting errors. These metrics are efficiently visualized in HDplot (McKinney, Waples, Seeb, & Seeb, 2017), which is integrated into our typical filtering workflow. HDplot analysis revealed that about half of the high F_{ST} -low F_{IS} loci that we identified deviated from typical patterns (Figures S2, S3). Most notably, a

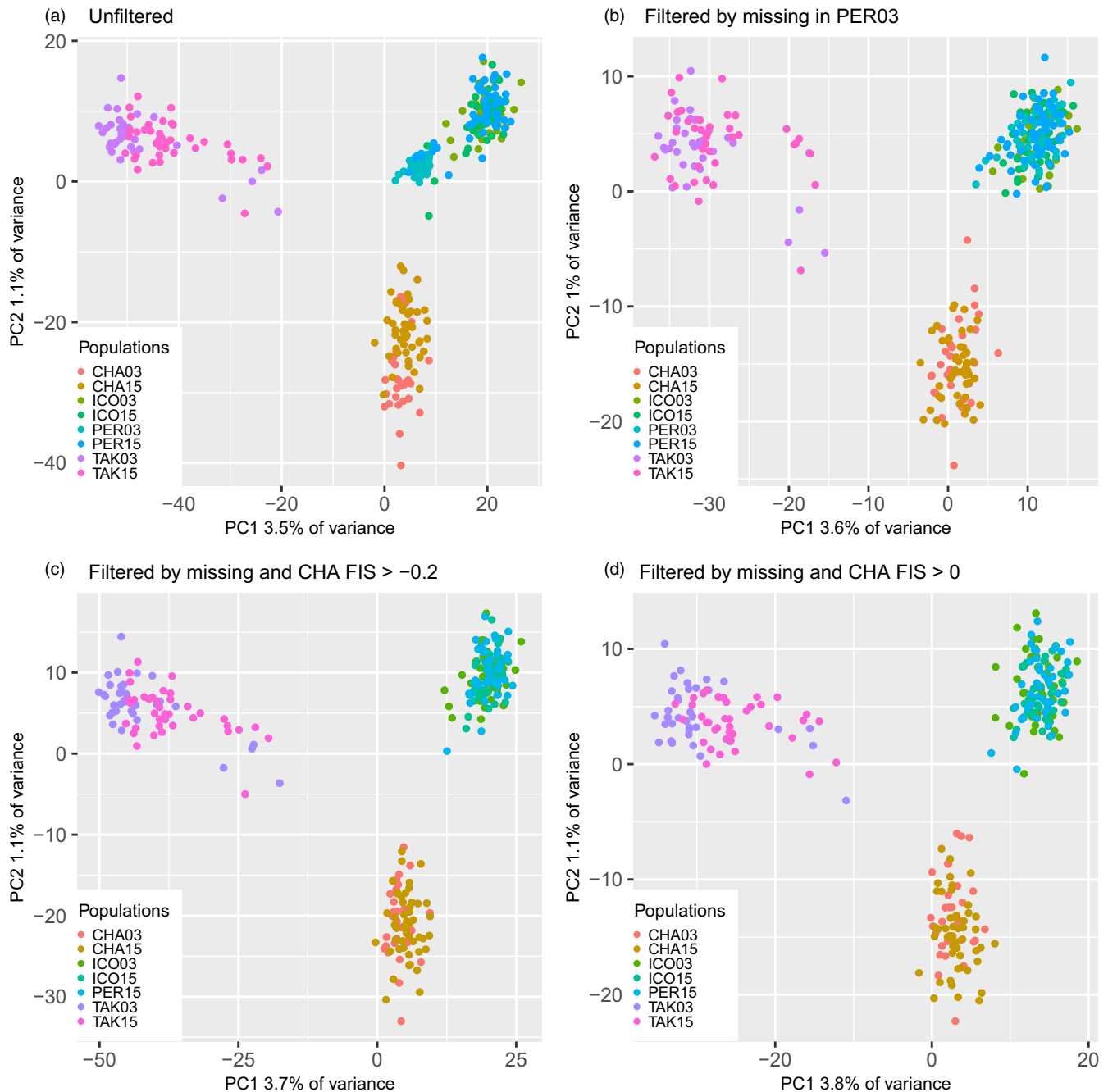


FIGURE 3 Individual-based PCAs for all populations with four different datasets. Datasets are (a) unfiltered data downloaded from DRYAD, (b) dataset including only loci with >50% genotyping rate in PER03, (c) dataset including only individuals genotyped at >50% of loci and loci with $F_{IS} > -0.2$ in the CHA samples, and (d) dataset including only individuals genotyped at >50% of loci and loci with $F_{IS} > 0$ in the CHA samples. Each dot represents an individual, and individuals are color-coded by population. See text for population abbreviations

large number of the loci had allele ratios near 0.75, which could indicate over merging of loci. However, many of these loci fell within normal ranges and would have been retained by our typical filters of deviation between -5 and 5 (Figures S2, S3). These results support the conclusion by O'Leary et al. (2018) that over and under-split loci are difficult to detect and remove. We believe that the most effective way to address this issue would be to re-analyze the data in STACKs; however, the raw data for this study are not available. Therefore, we attempted to filter the data in different

ways to determine the impact of these high F_{ST} low F_{IS} loci on patterns of population structure.

We analyzed four datasets to investigate the influence of filtering on genetic differentiation with an emphasis on the CHA population pair. These datasets were (1) the unfiltered data downloaded from DRYAD (8,728 loci), (2) a dataset with all SNPs retained but with individuals missing >50% of data removed (this removes all individuals from the PER03 population and up to 7 individuals from other populations), (3) a dataset with individuals

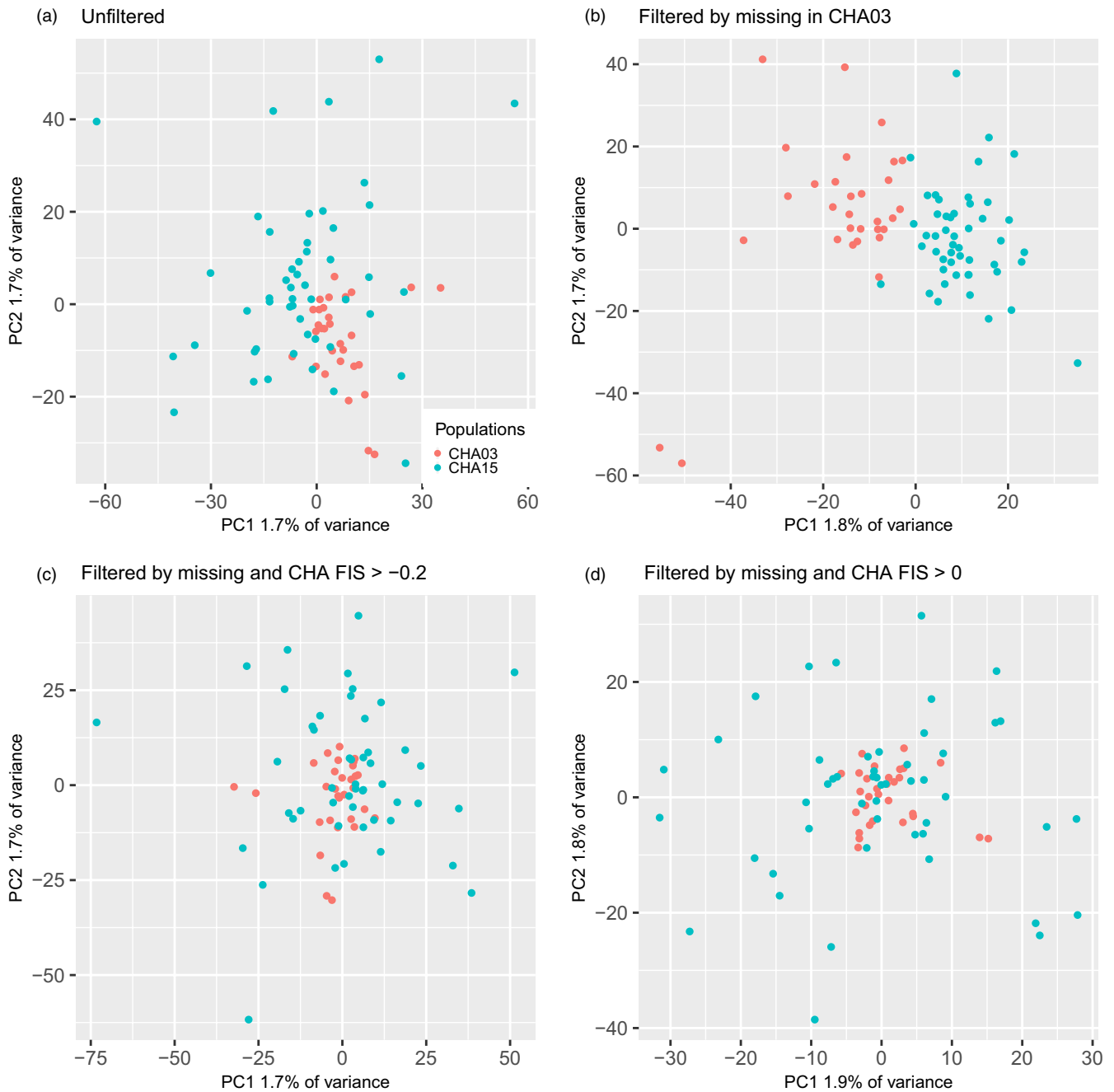


FIGURE 4 Individual-based PCAs for the CHA populations with four different datasets. Datasets are (a) unfiltered data downloaded from DRYAD, (b) dataset including only loci with >50% genotyping rate in CHA03, (c) dataset including only individuals genotyped at >50% of loci and loci with $F_{IS} > -0.2$ in the CHA samples, and (d) dataset including only individuals genotyped at >50% of loci and loci with $F_{IS} > 0$ in the CHA samples. Each dot represents an individual, and individuals are color-coded by population. See text for population abbreviations

missing >50% of data removed and with SNPs with $F_{IS} < -0.2$ in the CHA samples removed (8,560 loci), and (4) a dataset with individuals missing >50% of data removed and with SNPs with $F_{IS} < 0$ in the CHA samples removed (3,900 loci). First, we conducted individual-based PCA as described above and found that the apparent genetic structure between CHA03 and CHA15 disappears when either F_{IS} filter is applied (Figures 3, 4). This is most observable in the PCAs including only CHA comparisons (Figure 4), where the strong structure observed between CHA03 and CHA15 in panels

a and b disappears when loci are filtered by F_{IS} in both panels c and d.

We also evaluated genetic structure for these four datasets with estimates of pairwise F_{ST} (Weir & Cockerham, 1984) and exact tests of genetic differentiation in Genepop. F_{ST} estimates among most sample pairs were similar across datasets as expected (Table 1). However, pairwise F_{ST} values for the CHA03/CHA15 comparison decreased from 0.0036 for the unfiltered dataset to 0.0017 for the dataset excluding loci with $F_{IS} < -0.2$ to 0.0003 for the dataset

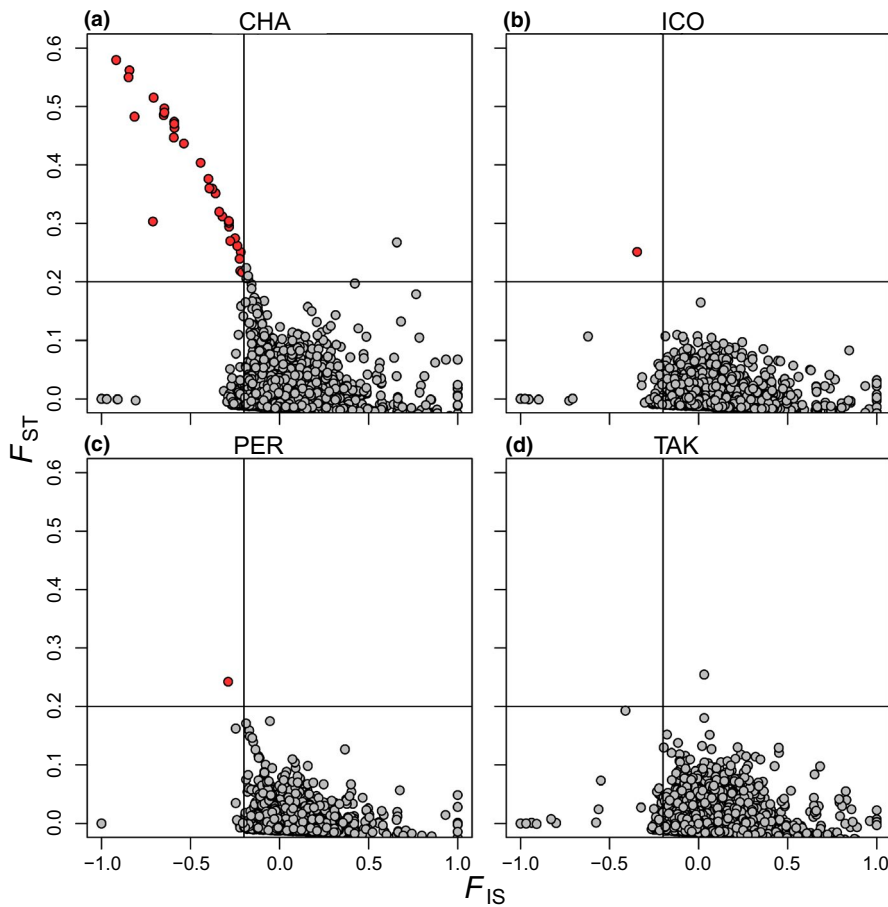


FIGURE 5 Distribution of F_{IS} and F_{ST} values for each population pair. F_{IS} values were calculated by pooling each sample from a given population pair. Loci with $F_{IS} < -0.2$ and $F_{ST} > 0.2$ (an arbitrary cutoff) are highlighted in red. Population abbreviations are found in the text

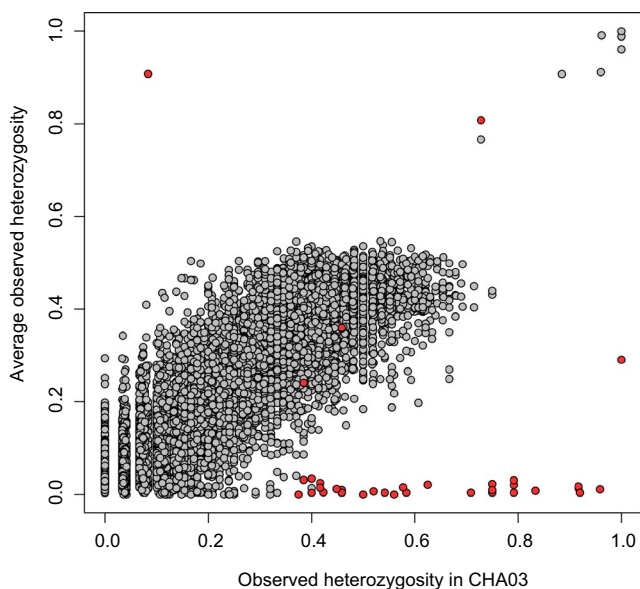


FIGURE 6 Observed heterozygosity in CHA03 (x-axis) and all other populations (y-axis). Loci with $F_{IS} < -0.2$ and $F_{ST} > 0.2$ in CHA samples are denoted in red

excluding loci with $F_{IS} < 0$. Notably, when the data were filtered properly, the TAK reference population actually has a higher F_{ST} value between time points than any of the temporal replicates in the southern rivers. F_{ST} estimates were significant for the CHA03/

CHA15 comparison with the unfiltered dataset and the dataset with low quality individuals removed, but were not significant with either of the datasets that were filtered based on F_{IS} . In summary, filtering the dataset based on F_{IS} values produced consistent results for all sample pairs except for the CHA03/CHA15 pair, where genetic structure decreased substantially and was not significant when loci with low F_{IS} values were removed. These results demonstrate that the bias in this dataset extends below the artificial threshold of $F_{IS} < -0.2$ and illustrates the difficulty of filtering out bioinformatic artifacts such as over or undersplit loci.

Although the analyses described above illustrate that there is no significant genetic differentiation between temporally replicated sample pairs across all loci, it is still possible that some outlier loci differentiating the populations could exist. To investigate this hypothesis, we explored the distributions of F_{ST} values for each population pair with dataset four described above (Table 2, Figure S4). Distributions were quantified and visualized using all samples and with populations randomly subsampling to the minimum sample size in the study ($N = 30$). F_{ST} distributions for all three population pairs were highly similar (Figure S4). For example, when population sizes are normalized, the TAK (reference) and CHA (high exploitation) populations have similar numbers of SNPs with $F_{ST} > 0.1$, 0.15, and 0.2 (Figure S4). These results illustrate that there are no significant qualitative or quantitative differences in the F_{ST} distributions among the three population pairs.

TABLE 1 Pairwise F_{ST} estimates for four datasets

Pop	CHA03	CHA15	ICO03	ICO15	PER03	PER15	TAK03
Unfiltered							
CHA15	0.0036						
ICO03	0.0218	0.0183					
ICO15	0.0214	0.0170	0.0001				
PER03	0.0210	0.0204	0.0020	0.0012			
PER15	0.0227	0.0186	0.0006	-0.0004	0.001		
TAK03	0.0568	0.0541	0.0777	0.0758	0.0803	0.0785	
TAK15	0.0623	0.0589	0.0834	0.0819	0.0838	0.0842	0.0009
Missingness filter							
CHA15	0.0036						
ICO03	0.0218	0.0183					
ICO15	0.0214	0.0168	0.0002				
PER03	NA	NA	NA	NA			
PER15	0.0225	0.0182	0.0008	-0.0003	NA		
TAK03	0.0568	0.0541	0.0775	0.0752	NA	0.0774	
TAK15	0.0623	0.0589	0.0832	0.0814	NA	0.0833	0.0009
F_{IS} filter >-0.2							
CHA15	0.0017						
ICO03	0.0204	0.0183					
ICO15	0.0197	0.0169	0.0002				
PER03	NA	NA	NA	NA			
PER15	0.0208	0.0182	0.0008	-0.0003	NA		
TAK03	0.056	0.0547	0.0782	0.076	NA	0.0781	
TAK15	0.0609	0.0594	0.0836	0.082	NA	0.0838	0.0008
F_{IS} filter >0							
CHA15	0.0003						
ICO03	0.0202	0.0183					
ICO15	0.0186	0.0161	-0.0002				
PER03	NA	NA	NA	NA			
PER15	0.0202	0.0183	0.0008	-0.0003	NA		
TAK03	0.0565	0.0532	0.0802	0.0778	NA	0.0796	
TAK15	0.0607	0.0572	0.0846	0.0832	NA	0.0844	0.0013

Note: Values in bold are significant differentiated ($p < .01$). Datasets are unfiltered data downloaded from DRYAD, (b) dataset including only individuals genotyped at >50% of loci, (c) dataset including only individuals genotyped at >50% of loci and loci with $F_{IS} > -0.2$ in the CHA samples, and (d) dataset including only individuals genotyped at >50% of loci and loci with $F_{IS} > 0$ in the CHA samples.

In summary, we found that the inadequate filtering and bioinformatic abnormalities are likely responsible for the major genetic patterns reported by Bowles et al. (2020). Specifically, the temporal genetic homogenization between southern populations that they reported was caused by missing data and the genetic differentiation and outlier loci that they observed in the CHA population was likely caused by variable over or undersplitting of loci across populations. In fact, when data were filtered properly, the reference TAK population displayed the largest differentiation between time points rather than any of the high exploitation populations they examined. These results illustrate the extreme importance of thorough data filtering.

5 | CONCLUSIONS

We re-analyzed the age/growth and genomic data presented in Bowles et al. (2020) and conclude that the data do not support the authors conclusion of phenotypic and genetic changes caused by harvest. The age/growth data suggest that walleye populations in Lake Mistassini may have decreased slightly in size-at-age over the study period, but this was true for all populations, including the reference population with lower exploitation. This result suggests that potential declines in walleye size-at-age is likely due to environmental factors that are impacting walleye populations in this

TABLE 2 Characterization of F_{ST} distributions for CHA, ICO, and TAK population pairs

Population pair	N 03 pop	N 15 pop	F_{ST}		F_{ST} range				
			Mean	Max	0–0.05	0.05–0.1	0.1–0.15	0.15–0.2	>0.2
CHA all samples	30	49	0.00003	0.27	2,967	116	21	3	1
ICO all samples	46	40	0.00006	0.16	3,628	69	4	1	0
TAK all samples	35	39	0.00100	0.18	3,008	109	14	2	0
CHA normalized	30	30	0.00030	0.21	2,840	142	35	5	2
ICO normalized	30	30	0.00006	0.18	3,486	138	23	4	0
TAK normalized	30	30	0.00100	0.23	2,946	118	29	4	1

Note: Data were generated with the dataset including only individuals genotyped at >50% of loci and loci with $F_{IS} > 0$ in the CHA samples. N is sample size, and populations in the “normalized” rows were randomly subsampled to 30, the minimum sample size in the dataset. Numbers in the columns beginning with F_{ST} are the number of loci in those categories.

lake regardless of exploitation rates, and/or that exploitation was similar across populations during the study period, meaning the Takwa River was not a viable reference population. Additionally, we discovered the genetic homogenization and divergence documented by the authors was the result of bioinformatic artifacts caused by inadequate data filtering. Specifically, when data were filtered properly using well-established methods, population structure between time points disappeared. Conclusively demonstrating that observed changes in fish populations are the results of harvest is difficult, as these changes can be caused or influenced by myriad biotic and abiotic factors. Our comment illustrates the potential pitfalls associated with studying harvest-induced changes in fish populations and highlights the importance of conducting robust analyses and data filtering before concluding that populations are influenced by harvest.

ACKNOWLEDGEMENTS

We thank two anonymous reviewers for editorial comments on a draft of this manuscript. Any use of trade, product, or company name is for descriptive purposes only and does not imply endorsement by the U.S. Government.

DATA AVAILABILITY STATEMENT

All scripts and input files used in this analysis are available as supplementary material or will be uploaded to DRYAD upon acceptance.

ORCID

Wesley A. Larson  <https://orcid.org/0000-0003-4473-3401>

Daniel A. Isermann  <https://orcid.org/0000-0003-1151-9097>

REFERENCES

- Bowles, E., Marin, K., Mogensen, S., MacLeod, P., & Fraser, D. J. (2020). Size reductions and genomic changes within two generations in wild walleye populations: Associated with harvest? *Evolutionary Applications*, 13(6):1128–1144, <https://doi.org/10.1111/eva.12987>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 1(1), 2017. <https://doi.org/10.18637/jss.v076.i01>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Coggins, L. G., Gwinn, D. C., & Allen, M. S. (2013). Evaluation of age-length key sample sizes required to estimate fish total mortality and growth. *Transactions of the American Fisheries Society*, 142(3), 832–840. <https://doi.org/10.1080/00028487.2013.768550>
- Dieckmann, U., & Heino, M. (2007). Probabilistic maturation reaction norms their history, strengths, and limitations. *Marine Ecology Progress Series*, 335, 253–270.
- Dunlop, E. S., Eikeset, A. M., & Stenseth, N. C. (2015). From genes to populations: How fisheries-induced evolution alters stock productivity. *Ecological Applications*, 25(7), 1860–1868. <https://doi.org/10.1890/14-1862.1>
- Dupont, P.-P., Bourret, V., & Bernatchez, L. (2007). Interplay between ecological, behavioural and historical factors in shaping the genetic structure of sympatric walleye populations (*Sander vitreus*). *Molecular Ecology*, 16(5), 937–951. <https://doi.org/10.1111/j.1365-294X.2006.03205.x>
- Feiner, Z. S., Chong, S. C., Fielder, D. G., Hoyle, J. A., Knight, C., Lauer, T. E., ... Höök, T. O. (2017). Sex-based trade-offs among growth, mortality, and maturation in Great Lakes yellow perch stocks. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(12), 2059–2072. <https://doi.org/10.1139/cjfas-2016-0173>
- Feiner, Z. S., & Höök, T. O. (2015). Environmental biology of percid fishes. In P. Kestemont, K. Dabrowski, & R. C. Summerfelt (Eds.), *Biology and culture of percid fishes* (pp. 61–100). Dordrecht: Springer. https://www.springer.com/us/book/9789401772266?gclid=Cj0KCQjw7s z6BRDYARIsAPHZrNKym4fJuLEBDT_TdLOCmBpBkoSIOGRVdxN nxLo6Asz_psXnafDesgQaAhAIEALw_wcB
- Feiner, Z. S., Shaw, S. L., & Sass, G. G. (2019). Influences of female body condition on recruitment success of walleye (*Sander vitreus*) in Wisconsin lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 76(11), 2131–2144. <https://doi.org/10.1139/cjfas-2018-0364>
- Haglund, J. M., Isermann, D. A., & Sass, G. G. (2016). Walleye population and fishery responses after elimination of legal harvest on Escanaba Lake, Wisconsin. *North American Journal of Fisheries Management*, 36(6), 1315–1324. <https://doi.org/10.1080/02755947.2016.1221002>
- Hansen, M. J., & Nate, N. A. (2014). Effects of recruitment, growth, and exploitation on walleye population size structure in northern Wisconsin Lakes. *Journal of Fish and Wildlife Management*, 5(1), 99–108. <https://doi.org/10.3996/092013-jfwm-065>
- Henderson, B. A., & Morgan, G. E. (2002). Maturation of walleye by age, size and surplus energy. *Journal of Fish Biology*, 61(4), 999–1011. <https://doi.org/10.1111/j.1095-8649.2002.tb01858.x>

- Hutchings, J. A., & Kuparinen, A. (2020). Implications of fisheries-induced evolution for population recovery: Refocusing the science and refining its communication. *Fish and Fisheries*, 21(2), 453–464. <https://doi.org/10.1111/faf.12424>
- Isermann, D. A. (2007). Evaluating walleye length limits in the face of population variability: Case histories from western Minnesota. *North American Journal of Fisheries Management*, 27(2), 551–568. <https://doi.org/10.1577/M06-047.1>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Lavigne, M., Lucotte, M., & Paquet, S. (2010). Relationship between Mercury Concentration and Growth Rates for Walleyes, Northern Pike, and Lake Trout from Quebec Lakes. *North American Journal of Fisheries Management*, 30(5), 1221–1237. <https://doi.org/10.1577/M08-065.1>
- Lecoutre, B. (1999). Beyond the significance test controversy: Prime time for Bayes. *Bulletin of the International Statistical Institute*, 58, 205–208.
- Lester, N. G., Shuter, B. J., Kushneriuk, R. S., & Marshall, T. R. (2000). Life history variation in Ontario walleye populations: implications for safe rates of fishing. Ontario Ministry of Natural Resources, Report from Percid Community Synthesis Population and Yield Characteristics Working Group.
- McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, 17(4), 656–669. <https://doi.org/10.1111/1755-0998.12613>
- O'Hagan, A. (2006). Science, subjectivity and software (comment on articles by Berger and by Goldstein). *Bayesian Analysis*, 1(3), 445–450.
- O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, 27(16), 3193–3206. <https://doi.org/10.1111/mec.14792>
- Pedersen, E. J., Goto, D., Gaeta, J. W., Hansen, G. J. A., Sass, G. G., Vander Zanden, M. J., ... Rypel, A. L. (2017). Long-term growth trends in northern Wisconsin walleye populations under changing biotic and abiotic conditions. *Canadian Journal of Fisheries and Aquatic Sciences*, 75(5), 733–745. <https://doi.org/10.1139/cjfas-2017-0084>
- Quist, M. C., Guy, C. S., Schultz, R. D., & Stephen, J. L. (2003). Latitudinal comparisons of walleye growth in North America and factors influencing growth of walleyes in Kansas reservoirs. *North American Journal of Fisheries Management*, 23(3), 677–692. <https://doi.org/10.1577/M02-050>
- R, Development Core Team, (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Rousset, F. (2008). GENEPOP '007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Sass, G. G., Hewett, S. W., Beard, T. D., Fayram, A. H., & Kitchell, J. F. (2004). The role of density dependence in growth patterns of ceded territory walleye populations of northern Wisconsin: Effects of changing management regimes. *North American Journal of Fisheries Management*, 24(4), 1262–1278. <https://doi.org/10.1577/M03-026.1>
- Sass, G. G., & Kitchell, J. F. (2005). Can growth be used as a surrogate measure of walleye (*Sander vitreus*) abundance change? *Canadian Journal of Fisheries and Aquatic Sciences*, 62(9), 2159–2168. <https://doi.org/10.1139/f05-129>
- Sass, G. G., & Shaw, S. L. (2018). Walleye Population Responses to Experimental Exploitation in a Northern Wisconsin Lake. *Transactions of the American Fisheries Society*, 147(5), 869–878. <https://doi.org/10.1002/tafs.10070>
- Sass, G. G., & Shaw, S. L. (2020). Catch-and-Release Influences on Inland Recreational Fisheries. *Reviews in Fisheries Science & Aquaculture*, 28(2), 211–227. <https://doi.org/10.1080/23308249.2019.1701407>
- Sinclair, A. F., Swain, D. P., & Hanson, J. M. (2002). Disentangling the effects of size-selective mortality, density, and temperature on length-at-age. *Canadian Journal of Fisheries and Aquatic Sciences*, 59(2), 372–382. <https://doi.org/10.1139/f02-014>
- Stan, Development Team, (2020). RStan: The R interface to Stan. R package version 2.19.3.
- Sullivan, M. G. (2003). Active management of walleye fisheries in Alberta: Dilemmas of managing recovering fisheries. *North American Journal of Fisheries Management*, 23(4), 1343–1358. <https://doi.org/10.1577/M01-232AM>
- Venturelli, P. A., Lester, N. P., Marshall, T. R., & Shuter, B. J. (2010). Consistent patterns of maturity and density-dependent growth among populations of walleye (*Sander vitreus*): Application of the growing degree-day metric. *Canadian Journal of Fisheries and Aquatic Sciences*, 67(7), 1057–1067. <https://doi.org/10.1139/F10-041>
- Wang, H.-Y., Einhouse, D. W., Fielder, D. G., Rudstam, L. G., Vandergoot, C. S., VanDeValk, A. J., ... Höök, T. O. (2012). Maternal and stock effects on egg-size variation among walleye *Sander vitreus* stocks from the Great Lakes region. *Journal of Great Lakes Research*, 38(3), 477–489. <https://doi.org/10.1016/j.jglr.2012.06.002>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population-structure. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.2307/2408641>
- Wuellner, M. R., Willis, D. W., Blackwell, B. G., & Lott, J. P. (2011). Empirical assessment of potential interactions between walleye and smallmouth bass. *Journal of Applied Ichthyology*, 27(5), 1173–1180. <https://doi.org/10.1111/j.1439-0426.2011.01789.x>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Larson WA, Isermann DA, Feiner ZS.

Incomplete bioinformatic filtering and inadequate age and growth analysis lead to an incorrect inference of harvested-induced changes. *Evol Appl*. 2021;14:278–289. <https://doi.org/10.1111/eva.13122>