Prioritizing Populations for Conservation Attention Using Genomic SNP Data

by

Avneet Kaur Chhina

B.Ed., University of British Columbia, 2020
B.Sc., University of British Columbia, 2019

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Biological Sciences
Faculty of Science

SIMON FRASER UNIVERSITY
Fall 2024

Declaration of Committee

Name:	**Avneet Kaur Chhina**
Degree:	**Master of Science**
Thesis title:	**Prioritizing Populations for Conservation Attention Using Genomic SNP Data**
Committee:	**Chair:** Chelsea Little Assistant Professor, Biological Sciences and Environmental Science
	Arne Mooers Supervisor Professor, Biological Sciences
	Thomas Booker Committee Member Assistant Professor, Forest and Conservation Sciences University of British Columbia
	Jonathan Moore Committee Member Professor, Biological Sciences
	Mike Hart Examiner Professor, Biological Sciences

Abstract

The conservation of intraspecific genetic diversity is guided by the distribution of genetic variation across geographic space (i.e., spatial structure) and across the genome (i.e., genomic structure). One question linking both structures is whether genome-wide genetic variation and putatively adaptive genetic variation identify the same set of distinct populations within species. Many authors advocate to solely use adaptive genetic variation, but it is technically and conceptually challenging to identify adaptive genetic variation for conservation. Across 34 species of plants and animals, we find that genome-wide genetic variation, which is much easier to measure, generally but variably agrees with adaptive genetic variation on population prioritizations. Putatively adaptive SNPs do as well or show higher correlation with genome-wide SNPs compared to equal-sized random subsets of genome-wide SNPs. Overall, it generally seems sound to use genome-wide genetic variation for population prioritizations to protect intraspecific genetic diversity.

Keywords: Genome-wide genetic variation; Adaptive genetic variation; Spatial Organization; Conservation prioritization, Adaptive potential

Dedication

I dedicate this thesis to Navjot Chhina (my mom), Kulwinder Chhina (my dad), Harvinder Chhina, Balwinder Chhina (my uncles), Harkirat Chhina (my brother) and everyone who has supported me throughout my journey.

Acknowledgements

I am truly very grateful and fortunate to have an immense support of so many people. This thesis would not have been possible without the help of many.

First, I would like to thank Prof. Arne Mooers (my supervisor) for his exceptional support throughout the journey of my degree. Thanks for your kindness, thoughtfulness, encouragement, and guidance. I will always be grateful for your unwavering support, encouragement, and discussions. Thanks for always being there at every step. You have not only helped me become a better scientist, but also helped me grow as a person. No words can express how blessed I am to have you as my supervisor. It was one of the best decisions of my life to work with you and learn so much along the way.

My heart-felt thanks to my committee, Prof. Tom Booker and Prof. Jonathan Moore, for their tremendous support, guidance, discussions, ideas, and time.

I would like to thank Prof. Mike Hart for taking the time to be my examiner and Prof. Chelsea Little for agreeing to chair my master's thesis defence.

I would like to thank Dr. Jayme Lewthwaite, Niloufar Abhari, Philippe Fernandez-Fournier, Miranda Dennis, Daniel Pelletier, Pourya Sardari, Paige Stephanie Prince, Hanna Jackson, Emma Green, Elijah Reyes, and Matt for all their emotional and academic support. Thanks for all the laughs and fun times, for teaching me so much, and helping me grow as a person. Your friendships truly mean a lot to me. Thanks for all the cherishing moments and making this journey memorable.

I would like to thank Fab Lab and Prof. Bernie Crespi, Prof. Michael Collard, and Prof. Leithen M'Gonigle, Prof. Ailene Macpherson for useful discussions during Fab Lab and lab meetings. Your input greatly improved the project.

Thanks to the authors of studies who shared the hard-won data with us. This project would not have been possible without their help, support, collaboration, and data sharing.

Thanks to Dr. Ken Thompson and Prof. Dolph Schluter for inspiring me to pursue my research passion and giving me the opportunity to work in Schluter lab at University of British Columbia during my undergrad. I'll always be grateful for all the learning.

My heart-felt thanks to the staff and students of Princess Margaret Secondary School for their enormous support, inspiration, and motivation throughout the years.

I would like to thank my family for their unwavering emotional support. Thanks to Navjot Chhina, Kulwinder Chhina, Harvinder Chhina, Balwinder Chhina for being my

backbone. Thanks to my aunts (Archana and Amandeep Chhina) for all the love and support. Thanks to Harkirat, Deepinder, Mohjeet, Hargun, Sukhbal, Jasmeet, and Manbir for all the fun times.

Thanks to friends, family, relatives, colleagues, and well wishers for your kindness whose names are not mentioned here.

Table of Contents

List of Tables

List of Figures

Chapter 1

General Introduction

Conservation of intraspecific genetic variation is crucial to maintain the adaptive potential of species and populations [113, 206, 234]. However, the loss of genetic diversity within species is enormous and widespread across many taxonomic groups [82, 137]. So, there is an increasing urgency with which we must tackle this challenge of conserving this variation.

My aim is to contribute to the discussion on how genetic variation within species might best be conserved. Literature suggests that maintaining the adaptability of species and populations is guided by two main factors – the spatial distribution of genetic variation (i.e., its spatial structure) and the genomic distribution of genetic variation (i.e., its genomic structure).

Chapter 2 is a literature review linking the spatial and genomic structure of adaptive potential for conservation management. Prof. Arne Mooers and Dr. Jayme Lewthwaite contributed to the conception, writing, and editing. I conducted the extensive literature review and wrote the first draft and Niloufar Abhari contributed to editing. A version of Chapter 2 has been accepted by the Journal Genome and will be published as Chhina, A. K., Abhari N., Mooers A., and Lewthwaite J. (2024). Linking the spatial and genomic structure of adaptive potential for conservation management: a review. *GENOME*. https://doi.org/10.1139/gen-2024-0036.

Chapter 2 provides a detailed overview on some of the core concepts in conservation genetics and summarizes the current considerations relevant to the conservation of populations. It mainly connects the spatial structure which consists of within-and among-population genetic variation to the genomic structure which consists of genome-wide genetic variation and its subset that is currently adaptive. Within-population genetic variation refers to the variation in alleles among individuals within a particular population whereas among-population variation refers to the degree of genetic differentiation among populations. Genome-wide genetic variation refers to all kinds of variation such as nearly neutral, neutral, deleterious and adaptive whereas adaptive genetic variation consists of a defined subset of this genome-wide variation. One important takeaway is that it is critical to consider and harmonize both within-and among-population genetic variation aspects as solely

focusing on one aspect carries risks. I outline several methods in-detail that consider both aspects. Another crucial takeaway is that mixed evidence exists regarding the decades-long unresolved debate on whether prioritization should focus solely on adaptive genetic variation, or whether it is preferable to consider genome-wide genetic variation. I summarize the mixed theoretical and empirical evidence provided in the literature and suggest that more work is needed examining various species and the factors that may impact the correlation of genome-wide and putatively adaptive genetic variation.

While Chapter 2 provides a literature overview of two factors of genetic variation, Chapter 3 takes among-population genetic variation aspect of spatial structure and empirically links it to the genomic structure of genetic variation. Specifically, I investigate whether genome-wide genetic variation and putatively adaptive genetic variation identify the same set of distinct populations across a vast array of plant and animal species and examine the impact of potential covariates on the correlation. The idea is many authors advocate to solely use adaptive genetic variation, but there are many theoretical and conceptual limitations in identifying and using adaptive genetic variation for conservation, so we explore whether genome-wide genetic variation, which is much easier to measure, can act as a surrogate for adaptive genetic variation.

The main motivation and conception of Chapter 3 comes from the 2021 study "Do we need to identify adaptive genetic variation when prioritizing populations for conservation?" by Philippe Fernandez-Fournier, Dr. Jayme M. M. Lewthwaite, and Prof. Arne Mooers. This study found that standing genetic variation can act as a useful proxy for adaptive genetic variation in yellow warblers and lodgepole pine. Dr. Jayme Lewthwaite, Philippe Fernandez-Fournier and Prof. Arne Mooers conceived the idea to repeat and test this study on more than two species.

A version of Chapter 3 is in preparation for submission with co-authors Dr. Jayme Lewthwaite, Philippe Fernandez-Fournier, Prof. Arne Mooers, and Prof. Tom Booker. I collated genome-wide and adaptive SNPs data from already published studies on 34 species of plant and animals. All authors contributed to the analysis and interpretation of the data. I wrote the first draft of the manuscript.

One of the main takeaways from Chapter 3 is that genome-wide and putatively adaptive SNPs agree on population prioritizations for conservation, though their agreement varies a lot. Since putatively adaptive SNPs are a subset of genome-wide SNPs, as expected due to part-whole correlation, the agreement among both sets of SNPs is impacted by the proportion of putatively adaptive SNPs. Interestingly, putatively adaptive SNPs often show higher correlation with genome-wide SNPs compared to equally-sized random subsets of genome-wide SNPs. The overall message is that genome-wide genetic variation can act as a surrogate for adaptive genetic variation and may be a sound strategy for prioritizing populations for conservation to protect intraspecific genetic variation.

Although I use the first-person pronoun ("I" and "my") throughout the thesis for language consistency, the thesis (Chapter 2 writing and editing and Chapter 3 analysis, interpretation, and editing) reflects supervisory input of my supervisor (Arne Mooers), my committee and co-authors.

Chapter 2

Linking the spatial and genomic structure of adaptive potential for conservation

2.1 Abstract

We unified the recent literature with the goal to contribute to the discussion on how genetic diversity might best be conserved. We argue that this decision will be guided by how genomic variation is distributed among manageable populations (i.e. its spatial structure), the degree to which adaptive potential is best predicted by variation across the entire genome or the subset of that variation that is identified as putatively adaptive (i.e. its genomic structure), and whether we are managing species as single entities or as collections of diversifying lineages. The distribution of genetic variation and our ultimate goal will have practical implications for on-the-ground management. If adaptive variation is largely polygenic or responsive to change, its spatial structure might be broadly governed by the forces determining genome-wide variation (linked selection, drift, and gene flow), making measurement and prioritization straightforward. If we are managing species as single entities, then population-level prioritization schemes are possible so as to maximize future pooled genetic variation. We outline one such scheme based on the popular Shapley Value from cooperative game theory that considers the relative genetic contribution of a population to an unknown future collection of populations.

Keywords: Genetic variation, Adaptive potential, Spatial organization, Conservation prioritization, Shapley Value

2.1 Introduction

Intraspecific genetic diversity is generally considered an important facet of biodiversity [206, 113, 234], with international organizations now calling for transformative measures to protect it [55, 74, 82, 136, 137]. This call is coupled with comparative work suggesting that

the current loss of intraspecific genetic diversity may be large and taxonomically widespread [82, 137].

Genetic diversity encompasses variation in alleles within (for diploids) and among individuals, and among populations. The general argument is that genetic diversity predicts adaptive potential [71, 102, 157, 275], such that more of it increases the probability of lineage persistence over multiple temporal scales, particularly in the face of accelerating environmental change [47]. Beyond this general idea, however, much is unknown, and there is little guidance for policymakers [137, 155]. In the context of this paper, genome-wide genetic variation refers to all forms of variation, including neutral, nearly neutral, deleterious, and adaptive genetic variation sampled throughout the genome. Adaptive genetic variation is a defined subset of this genome-wide diversity that is due to past natural selection or that is posited to be relevant to ongoing or future natural selection [142, 259]. One common approach to finding such putatively adaptive variation (hereafter, we drop the "putative" modifier for readability) is to look for SNPs that are associated with particular populations or environments in genome scans and then, perhaps, look for genes near those SNPs with potential adaptive significance.

Infraspecific genetic variation (both genome-wide and its adaptive subset) typically exhibits a spatial pattern with variable levels of within-population diversity and among-population differentiation (population distinctiveness) due to the combined action of drift, gene flow, and selection. This spatial pattern necessitates specific descriptors that reflect these measures within and among (sets of) populations, be they subspecies, varieties, or constructs such as evolutionary significant units (ESUs), discrete population segments or designatable units. We can refer to all these generally as "entities", and the levels of diversity and differentiation among them may differ depending on which component of genomic variation is measured.

The distribution of genetic variation within and among populations immediately raises the question of how we classify genetic diversity and distinctiveness: the answer to these questions are central to the active management of species. To what extent should identifiable entities be managed individually and somewhat statically, and to what extent as collectives - when should we try to keep them separate, and when might we propose assisted gene flow (see, e.g. [4, 122, 300])? When and how should we explicitly rank infraspecific entities for conservation attention?

This review aims to contribute to the discussion on the conservation of genetic diversity by considering the spatial and genomic structure of populations. We begin by discussing how genetic variation may be distributed on the landscape, characterized as “within” vs. “among” population diversity (i.e. spatial structure). We then consider genomic structure, specifically contrasting genome-wide variation and its adaptive subset. We present arguments from the decades-long unresolved debate on whether or not prioritization efforts for conservation should focus on the subset of genetic variation that is adaptive, or whether

it is preferable to consider genome-wide variation. We consider how genome-wide genetic variation may act as a surrogate for its adaptive genetic components and discuss strategies to harmonize within-and among-population genetic variation. For the two axes structuring genetic variation – within- vs. among-population, and genome-wide vs. adaptive, we discuss how each is measured and maintained across populations and how each is relevant to conservation. We then discuss specific issues that we think are important to consider for each axis. Finally, we highlight an approach for on-the-ground management in the context of considering the adaptive potential of the collection of populations (e.g. the species as a whole), considering every populations' contribution to collective genetic diversity in the future (see Figure 1 and Table 1).

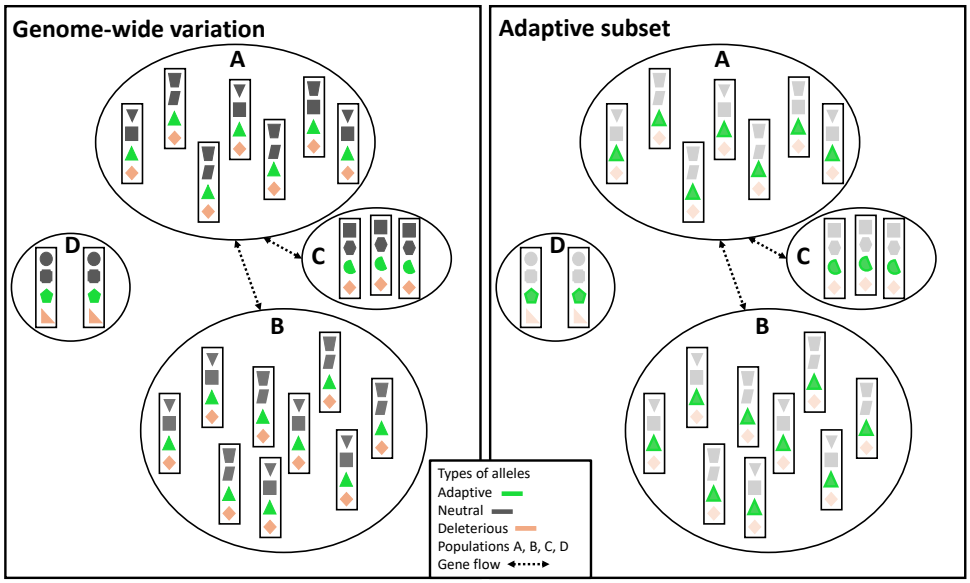

Figure 2.1: Spatial and genomic structure of genetic diversity. Left panel: Populations are elipeses, individual genomes are depicted as rectangles, and distinct alleles are distinct shapes. Loci can be associated with neutral, adaptive, or deleterious variation (depicted with colour). Populations A and B contain high within-population genetic variation, but share some of that variation due to gene flow. Populations C and D contain distinct alleles (both adaptive and non-adaptive). Some of those distinct alleles are not shared with other populations (e.g. Population D). The right panel highlights the adaptive subset of loci for comparison with overall variation.

Table 2.1: The structure of genetic diversity. The spatial and genomic structure of genetic diversity is characterized as within vs. among population variation and genome-wide vs. adaptive genetic variation, respectively. Each component of genetic diversity is described along with its relevance to conservation management schemes and potential problems.

	Spatial structure of genetic variation		**Genomic structure of genetic variation**	
	Within-population	**Among-population**	**Genome-wide**	**Adaptive**
Description	Variation in alleles among individuals within a given population, usually measured by A_r, π, H_o, or H_e.	The average degree of genetic differentiation of populations, usually measured by allele frequency differences (e.g. average pairwise F_{ST}) or with distance-based metrics.	All forms of variation, including neutral, nearly neutral, deleterious, and adaptive genetic variation, as assessed through genetic variants (e.g. SNPs).	Adaptive subset of genome-wide variation assessed through outlier and/or association analyses.
Conservation Relevance	May contribute to current population-level mean fitness, ongoing adaptation, and thus adaptive potential.	May contribute to species level adaptive potential by retaining distinct genetic variation needed for current and future adaptation.	May positively contribute to fitness, population viability, and prevent inbreeding depression. May maximize the information saved in the face of uncertain future conditions; some subset may become adaptive in the future. May preserve polygenic trait loci or genetic architecture that could otherwise go undetected.	May benefit prioritization schemes through its relatively direct link to function and adaptive capacity; offers the potential to use functional relationships to predict evolutionary responses and manage species accordingly.
Problems with focusing on only this component for prioritization schemes	May carry risks as genetically variable populations are often subject to high gene flow, which may introduce maladapted alleles and hinder population differentiation and local adaptation in some cases.	May carry risks as the extent to which genetic distinctness is due to local adaptation or drift is rarely measured: in some cases, distinct variation may be non-adaptive and counterproductive for species' survival.	There are limitations in reproducibility and accessibility of raw genetic data in a standardized format (i.e., data are expensive and difficult to collect) May incorporate currently deleterious variation along with adaptive and neutral variation. May not be good indicator of long-term conservation status if changes in this variation lags behind population decline. This may also hold for the adaptive subset of genetic variation.	Conceptual challenges: In the face of unpredictable future conditions, currently adaptive loci may not be adaptive in the future while putatively neutral variation may become adaptive in the future. Technical challenges: There is a frequent misidentification of true adaptive loci. Approaches are based on correlation and are biased in favour of detecting large effect loci, nonsynonymous mutations near or within coding regions, and overlooking polygenic trait loci. In addition, challenges in disentangling the complexity and interconnectedness of the genome may lead to high rate of both false-positives and false-negatives.

2.2 Spatial Organization of Genetic Variation

2.2.1 Within-Population Genetic Variation

How is within-population genetic variation measured?

As readers will well know, within-population genetic diversity can be measured in many ways – the degree of polymorphism, allelic richness, nucleotide diversity (π), or observed (H_o) or expected heterozygosity (H_e) (see, e.g., [101, 141]). To recap, the proportion of genetically variable loci (or polymorphisms) is simply the number of genetic loci with two or more alleles (at some minimum frequency) divided by the total number of loci assessed (see, e.g. [9, 191]). Allelic richness is usually measured as the average number of alleles per locus (see, e.g. [163, 213, 231]), however, it is less relevant for biallelic SNPs (the most commonly used data stream currently). Nucleotide diversity (π) is measured as the proportion of nucleotides differing between two random sequences [212]. Expected heterozygosity (or Nei's gene diversity) [209] is the expected probability that an individual will be heterozygous at a given locus (or, alternatively, the proportion of loci being assessed that are heterozygous):

$$H_e = 1 - \sum_{i}^{n} p_i^2$$

where p_i is the frequency of the ith of n alleles (see, e.g. [209, 214, 210]). This expected heterozygosity (H_e) is estimated from allele frequencies, whereas observed heterozygosity (H_o) is estimated from individual genotypes directly; both are scaled up from individual-based to population-based estimates by averaging across individuals and sites (see, e.g., [258]). Substantial differences between expected and observed heterozygosity can point towards inbreeding (see, e.g., [156]). We will return to expected heterozygosity in Section 4.

How is within-population genetic variation maintained?

As a review, all genetic variation arises from mutational input and subsequent genetic recombination and can be maintained within a population through (i) a large effective population size combating the eroding influence of drift, (ii) some forms of natural selection and genetic architecture and, commonly (iii) via gene flow among populations that have different allele frequencies due to (i) and (ii) [57, 136, 300].

Sewall Wright 1931 [315] conceptualized effective population size (N_e) as the number of breeding individuals in an ideal population that would show the same levels of genetic drift as the population under study. This is an important parameter in conservation genetics because while the census size of a population may be quite large, factors such as unequal sex ratios, high variability in the number of offspring per individual, non-random mating and fluctuating population sizes all result in the effective population size being much lower than the census size (see, e.g. [264]). Low N_e increases the rate of drift, and with it the likelihood

that any allele will be lost and potentially deleterious alleles will increase in frequency; maintaining a large effective population size is crucial to preventing the erosion of genetic variation (see [120]).

While stabilizing [148], background [48] and directional selection [295] at particular loci tend to reduce genetic variation in the long term [19] (both at those loci and in nearby linked genomic regions [131]), balancing selection (e.g. heterozygote advantage, negative frequency-dependent selection) [177] and disruptive selection can maintain genetic variation within populations [131, 229]. Importantly, more complex expectations are also possible; for example, pleiotropy can maintain genetic variation under stabilizing selection [107].

Gene flow from other populations aids in the production and maintenance of genetic diversity within a given population. Therefore, increasing isolation, e.g. via habitat fragmentation, which reduces gene flow, is expected to reduce within-population genetic diversity [11, 112]. For instance, Gomez et al. 2016 [108] found higher levels of genetic diversity in connected populations of a perennial shrub species (*Lepidium subulatum*) compared to fragmented populations.

Thus, within-population genetic variation can be maintained by large effective population sizes opposing the stochastic effects of drift, different forms of natural selection and its complex interactions, and gene flow among populations.

How important is within-population genetic variation?

Within-population genetic diversity is considered an essential component of global diversity due to its contribution to current population-level mean fitness, ongoing local adaptation, and, importantly, adaptive potential [71, 95, 202, 238].

Decades of research have established a positive association between average fitness (reflected through proxies such as average viability/survival, reproductive success, size, weight, growth rate, disease resistance, and gamete quality) and genetic diversity at the population level [71]. For example, across 34 sampled datasets, heterozygosity (along with and linked to quantitative genetic variation and population size) accounted for 15-20% of the variation in population fitness [238]. However, whether this relationship is causal vs. correlative is not always clear [118]. Theoretically, higher genetic diversity can itself be adaptive, e.g. via heterozygote advantage [132], but could also act as a proxy measure of low levels of past and deleterious genetic drift [238]. In this case, current genetic diversity is only indirectly linked to fitness. But regardless of the mechanism, this positive relationship has trickle-down consequences for conservation efforts: threatened taxa show, on average, 35% lower heterozygosity and/or allelic richness than non-threatened taxa [270, 312, 310]. Particularly small populations are at increased risk from this loss of genetic diversity if drift overpowers selection such that high numbers of mildly deleterious alleles go to fixation. This results in a positive feedback loop known as “mutational meltdown" whereby the initial reduction in fitness from fixed alleles leads to population declines, which leads to further accumulation

of deleterious variants by random genetic drift – this is also called the extinction vortex [187].

Theoretical work also supports the crucial role of genetic diversity as raw fuel for ongoing adaptation at the population level. According to Fisher's theorem, a population's rate of adaptation at any time is proportional to its additive genetic variation (see, e.g. [20, 112, 202]). Generally, population-level genetic diversity is considered a proxy for this additive genetic variance (to be precise, under conditions of strict additivity, and allele frequency and allele effect independence) [202]. Recent empirical work in vinous-throated parrotbill [167], rainbow trout [226] and threespine stickleback [18, 58] has indeed confirmed that existing genetic variation facilitates rapid adaptation.

Extensive theory also supports the idea that current standing genetic variation may be considered a useful proxy for future adaptive potential [96, 94, 120, 122, 202]. Current adaptations may not always be suitable under future conditions, and mounting evidence suggests that adaptation to novel environments proceeds from standing genetic variation more often than from new mutations in any case (reviewed in [12, 18, 239] for more recent examples). Additionally, within-population genetic diversity should predict population persistence in the face of anthropogenic change as it maximizes "options" upon which selection can act [154]. This may be especially important in endangered species in which genetic diversity has already been depleted, and emerging threats may exert strong selection.

These are the main arguments for why within-population genetic variation is considered an integral part of global diversity —it predicts population-level fitness, ongoing adaptation, and adaptive potential [71, 202, 238]. This is also at the basis of calls for assisted gene flow as a conservation intervention to increase population-level genetic variation [4].

2.2.2 Among-Population Genetic Variation

How is among-population genetic variation measured?

Among-population genetic variation is some measure of the average degree of genetic differentiation of populations (i.e., the average pairwise genetic distance between populations without correcting for within-population variation). This is measured using either allele frequency differentiation metrics such as F_{ST} or Nei's G_{ST}, or distance-based metrics such as principal component analysis on genetic variants.

The fixation index (F_{ST}), originally introduced by Wright as the "correlation between random gametes [or alleles], drawn from the same subpopulation, relative to the total population" [316] is a widely accepted and commonly used among-population metric (though some confusion remains on exactly how to estimate it as Wright did not define what the "total population" was; see [34]). F_{ST} indicates the similarity of individuals within populations relative to the similarity of individuals drawn randomly from some larger set of populations [144]. Though many definitions exist (see [147, 144, 301, 303]) it can be simply presented as:

$$F_{ST} = \frac{V(q)}{\bar{q}(1-\bar{q})}$$

where q is the frequency of an allele at a locus, $\bar{q}$ is the average across loci, and *V(q)* is the variance of allele frequencies over subpopulations [130]. Smaller F_{ST} values (close to zero) depict similar allele frequencies within each population whereas larger values (close to one) represent different allele frequencies among populations (see [144]). Some of the common programs used to estimate F_{ST} include Fdist2, Lositan, BayeScan, Bayenv, BayPass, FLK, and PCAdapt [138].

While F_{ST} was initially developed for biallelic loci, a later interpretation known as Nei's G_{ST} [209] expanded this concept to multiple alleles at a locus. As such, it is equivalent to Wright's 1949 [316] F_{ST} when there are two alleles at a locus, and when there are multiple alleles at a locus, it is equivalent to a weighted average of F_{ST} of all alleles [63, 209]. It can be calculated as

$$G_{ST} = \frac{H_T - H_S}{H_T}$$

where H_S is expected heterozygosity (the probability that two gene copies drawn randomly are different) within populations and H_T is the equivalent expected heterozygosity if all populations are considered together [110, 211].

When combining F_{ST} estimates across multiple loci (e.g., multiple single nucleotide polymorphism sites, or SNPs), several authors [34, 165] suggest that a "ratio of averages" method – separately averaging the numerator and denominator of the F_{ST} estimate across all SNPs and then producing a ratio from those averages – produces a less biased estimate than an "average of ratios" method – calculating an F_{ST} for each SNP and then averaging those F_{ST} values across all SNPs [56].

Distance-based metrics such as principal component analysis (PCA) [145, 227] on SNPs, originally intended to correct Genome-Wide Association Studies (GWAS) for the effects of population structure, represent a separate approach for measuring divergence [86, 104, 170, 223, 233, 288]. Genetic (SNP) data with multiple individuals (from one or more populations) are used as an input to graphically examine genetic distances, and clusters in multidimensional space. Whereas allele frequency differentiation metrics (including the F_{ST} statistic) require grouping individuals into populations *a priori*, which can be difficult when populations are poorly delineated from one another or population structure is unknown [73], in this approach, individuals are simply given coordinates in a graph. In addition, PCA is free from underlying population genetic models [169]. PCA data-reduction representation has multiple applications, including the discovery of population structure, ancestry, admixture, and, here, for investigating differentiation [100, 170, 282]. For example, if populations can be identified *a priori*, then the Euclidean distance between the centroids of each pair of populations in PCA space is a measure of pairwise divergence [86, 104].

Once pairwise differentiation among populations is measured (using, e.g. pairwise F_{ST}, or distances in PCA space), one can begin investigating what subset of this variation is unique vs. shared with other pairs, i.e. one can measure distinctiveness. Populations that are farther from others are considered more "distinctive". One fairly general measure of population-level distinctiveness is referred to as ED [151]. When the distances can be represented on a rooted tree, ED apportions the total diversity represented by that tree amongst its members by divvying up each branch along the path, from the root to each tip equally among the species subtending that branch, and then summing these values [151]. As such, it represents the amount of evolutionary history each taxon would contribute to future subsets and highlights genetically distinctive taxa that harbor a disproportionate amount of the total evolutionary history. A separately-derived metric known as the Shapley value [116, 263], is formally equivalent to ED [123], but can also be extended to networks [291], allowing for its use in population genetic analyses. We return to the Shapley value in Section 4.

Estimating the distance of a population to others is also the basis of approaches that measure the total diversity of population subsets. For instance, Weitzman [306] presented an approach that takes genetic distances among entities (say populations) as input and measures the total diversity of arbitrary subsets. The approach begins with the full set and removes populations, one by one, in turn [124], evaluating the marginal loss of diversity due to each population's removal from the set. This approach is taken up by Toro, Caballero and colleagues in their population genetics framework (see Section 4) [44, 284]. Weitzman used his approach to find subsets of populations that would retain the most diversity overall under particular constraints (e.g. when subset size is fixed). If one considers an unrooted phylogenetic tree as a representation of the pairwise distances among populations, then the unrooted version of the "phylogenetic diversity" metric [84], which measures the length of the minimum spanning tree that links subsets of entities, is a related measure of the total diversity represented by subsets of tips [274].

How is among-population genetic variation maintained?

Populations diverge primarily via mutational input, local adaptation, and/or drift, mediated by the effects of gene flow.

Local adaptation (where individuals harbor higher fitness in their local environmental conditions as opposed to elsewhere where conspecifics occur) maintains among-population variation via divergent selection in heterogeneous landscapes and/or via different mutational processes or histories [37, 88, 138]. This phenomenon is prevalent and has been well-documented [88, 159, 172, 249]. Gene flow may decrease interpopulation divergence directly via homogenization, or indirectly, by counteracting local adaptation via introducing maladaptive alleles that shift populations off their local optima [7]. However, gene flow can also facilitate adaptation through adaptive introgression, the process by which local

adaptation is accelerated through the introduction of increased within-population additive genetic variation (see [283] for empirical and theoretical examples).

On the other hand, random genetic drift leads to the divergence of populations through the fixation of alternative alleles [201]. Generally, this drift-induced population divergence and distinctiveness is expected to be maladaptive locally, since positively favoured alleles can be lost randomly as drift may overwhelm selection in small populations [2]. Furthermore, to the extent that drift-induced distinctiveness is caused by decreased variation within populations, it is unlikely to be globally beneficial [7, 14, 300].

How important is among-population genetic variation?

In 1988, Ehrlich advocated that "the loss of genetically distinct populations within species is, at the moment, at least as important a problem as the loss of entire species" [76], and many studies have since highlighted a need to identify and conserve distinct populations in order to preserve unique and potentially rare and unusual pools of variation [9, 57, 62, 84, 136, 175, 189, 195, 199, 279, 289, 291, 313, 297].

Conserving genetically distinct populations could in theory, achieve many of the same goals as conserving genetically diverse populations: distinct populations represent past evolution [291], would preserve contemporary processes that shape populations [240, 292] and provide raw material for future evolution [215]. Identifying and prioritizing the most distinctive of these distinct populations should, *ceteris paribus*, conserve more total genetic variation within species (see e.g., [86, 291]) and minimize the risk of genetic erosion at the species level. Again, this may be especially important for species at risk, where total population size may be small and where the extirpation of a given population could have long-term negative consequences [94]. For example, white nose syndrome (a fungal pathogen) has decimated many little brown bat populations (*Myotis lucifugus*). Importantly, studies have identified candidate genes associated with survival in different populations [15, 93, 106], pointing to the existence of variation in evolutionary responses and supporting the contention that preserving among population-level diversity could provide multiple pathways to adapting to future threats.

In general, distinct populations may have evolved in isolation for generations [205], which would explain why they contain non-redundant evolutionary and biological history [236, 298] representing the evolutionary and ecological processes that produced them [205]. We note that a genetically "distinct population" is key to the concept of ESUs in US law [119, 135, 205, 246, 297]. As such, they are often managed as separate legal units, not explicitly as components of the species of which they are part. We return to this in Section 4.

2.2.3 Conserving Within- vs. Among- Population Genetic Variation

When conservation biologists consider how genetic variation is partitioned among populations, several things should be kept in mind. The first is the role of gene flow, especially as it affects local adaptation, and the second is the extent to which populations may become distinct (and more or less distinctive) due to drift versus local adaptation.

Often the most genetically diverse populations are those that are largest and those that experience high incoming gene flow, such as at the center of the species' geographic range [75]. While a small amount of gene flow may aid local adaptation (see [283]) and prevent drift, high levels of gene flow can impede genetic and phenotypic differentiation [89, 174, 276]. A second and more serious consequence of the introduction of alleles from elsewhere in the species' range is the possibility that they may be independently maladaptive in the new populations, or may not be compatible with other local allele combinations [4, 89, 174, 266, 276]. This reduced fitness is referred to as migration load [4, 89, 174, 266, 276]. Over the medium term, excess migration can lead to the loss of local loci, and ultimately to the loss of local lineages through lineage swamping [4]; such loss would lead to reduced likelihood of future adaptation [174, 276]. Therefore, populations with high genetic variation may not always be of the highest conservation value, e.g. if that high variation is due to gene flow from elsewhere leading to migration load that erodes local adaptation.

In contrast, while the argument for prioritizing among-population genetic variation to maximize genetic diversity at the species level seems unassailable, it assumes that distinct populations harbor "useful" genetic variation, (and more distinctive populations contain even more), which may or may not be true. Small populations may become genetically distinctive through continuous drift such that their unique genetic information is both locally and globally non-adaptive. While small populations can undergo purging (the loss of deleterious alleles through selection), this takes time [287] and, because it involves alleles of small effect, it is generally ineffective [129, 157]. In addition, drift-dominated small populations may also be vulnerable to stochastic demographic events, such as bottlenecks and thus to episodic further reductions in genetic diversity [14, 300]. Additionally, inbreeding depression, increased genetic load, low genetic variation, and non-genetic Allee effects may contribute to an accelerating reduction of average individual fitness in some small populations, the oft-cited "extinction vortex" effect [140]; this trajectory would constrain the adaptive response to selection under changing environmental conditions [311]. In such cases, allocating specific conservation efforts towards these distinctive populations might be counterproductive for species' survival [300].

Generally, studies advocating for using genetic variation for conservation purposes rarely identify the extent to which that variation is due to local adaptation vs. drift, likely because it is difficult to do [86]. However, this information is critical for conservation managers and policymakers. For example, in the case of among-population variation, if the distinctiveness

of a given population is driven by significant adaptation to its environment, the solution might be to preserve it in its current state [97], such as is the case in locally adapted New Zealand Kiwi populations [28, 304]. However, if distinctiveness is driven by genome-wide genetic variation, the proposed solution is to restore gene flow via translocation [235, 300]. Therefore, in the following section, we discuss how distinguishing genomic structure (genome-wide genetic vs. adaptive genetic variation) is a second crucial facet to the spatial variation considerations we have presented.

2.3 Genomic Structure of Variation

2.3.1 Genome-Wide Genetic Variation

How is genome-wide genetic variation measured?

Conservation biologists have used many markers and methods to characterize genetic variation; we do not provide an exhaustive summary here, but rather point out the relative strengths and weaknesses of the most common ones. Allendorf 2017 [8] summarizes information on all major types of markers, their historical development, and their utility.

The original genetic data source was allozymes, or enzyme variants, coded by alleles at a single locus [8] that differ in structure but not function and discovered in the 1960s. Since these methods could be characterized using gel electrophoresis, they were a cost-effective way of surveying variation in large sample sizes [31]. However, they were criticized for being an indirect measure of variation in DNA, as they characterized protein variation rather than DNA polymorphism itself [256]. The shift from enzyme-based to DNA-based markers began with Restriction Fragment Length Polymorphisms (RFLPs) [111]. At sites recognized by restriction enzymes, individual sequence variation results in different fragment lengths produced after digestion with the restriction enzyme [40]. With this technique, for the first time, non-coding (or silent) changes in DNA variation could be identified.

Other key molecular markers used currently include microsatellites (randomly repeated sections of DNA around 100 base pairs long). These are highly polymorphic, making them an extremely popular marker of choice in population genetics in recent times [162]. However, they have complex mutation patterns, making modeling difficult. Also, scoring microsatellites is a challenge as allelic designations are highly laboratory and machine dependent [78]. Microsatellites can also be sparsely distributed throughout the genome. Amplified fragment length polymorphisms (AFLPs) emerged in the early 1990s and are similar to RFLPs [196]. They were useful alternatives to microsatellites in species with large or poorly characterized genomes, as they did not require *a priori* knowledge of the primer sequences for the focal species [256]. However, they show low reproducibility and so can be unreliable [256].

All of these approaches queried relatively few, and putatively neutral, positions in the genome, limiting the ability to estimate genome-wide parameters and the role of adaptive variation [99]. The advent of high-throughput sequencing in the early 2000s dramatically

increased the number of loci (up to thousands or millions), and was focused on the identification of SNPs: variation of a single base pair among individuals at a site. The main advantages of SNPs are that they allow the study of very many loci across the entire genome, which enables fine-scale mapping of variation and greatly improves estimates of genetic and demographic processes (e.g. gene flow). Additionally, although they have a relatively low mutation rate (meaning convergence is rare), they are abundant. There are a number of methods to identify SNPs, though we focus on the common 1) microarray-based and 2) sequencing-based approaches. Micro-array-based approaches (also known as "SNP chips") use short nucleotide sequences as probes to hybridize with the tested DNA sequences and confirm the presence of a specific allele at a SNP site. As such, these are targeted approaches in that they require *a priori* knowledge that variation is present at a known site, and thus can only query known SNP locations. This approach is especially challenging in lesser-studied species. Meanwhile, sequencing approaches across the most widely applied platforms sequence short DNA fragments (usually dozens to hundreds of base pairs). These "short reads" are then aligned either to a reference genome or de novo assembly e.g., RAD seq to allow for SNPs, genotypes, and sometimes haplotypes to be identified. Thus, whereas microarrays identify genotypes at a pre-defined set of SNPs, sequencing approaches are generally non-targeted and aim to provide information on any variants (including previously unknown variants). Sequencing approaches are more expensive and more labor-intensive to analyze.

Two of the main sequencing approaches are reduced- representation sequencing (RRS) and whole genome sequencing (WGS). WGS aims to sequence the vast majority of nucleotides in a genome (70-95%, depending on the situation; [99, 265]). Meanwhile, RRS techniques (such as genotype-by-sequencing and restriction site-associated DNA sequencing or RADseq), rely on restriction enzymes that cut at specific motifs in the genome, followed by sequencing beginning at these restriction cut sites. As such, repetitive regions of genomes can be avoided and lower copy regions can be targeted, which helps with aligning sequences to the reference genome, particularly in species with high levels of genetic diversity [79]. This makes the techniques much more cost-effective than WGS. However, although RRS approaches are "genome-wide", they usually sample only a small proportion of the genome (e.g. 1-5%), and result in much more missing data due to assembly methodology, sequencing errors and sampling biases [61, 98, 99, 184]. For either WGS or RRS, pooling DNA from individuals within a population (so called *Pool-seq*) may represent a compromise, as it can help reduce costs while maximizing the sample size and coverage of sequenced genomes. However, identifying rare variants can be challenging with this approach as they may be confounded with sequencing errors when that allele's frequency is very low in a pool [13], and of course, individual haplotypes and genotypes are lost during the process of pooling DNA [99].

Finally, structural variants (SVs) have recently emerged as another way to characterize genetic diversity. SVs include deletions, insertions, inversions, duplications, and large-scale copy number variants of large sections of DNA (typically 1 kb, so-called "long reads"; [251]). SVs are largely overlooked with short-read sequencing, but long-read sequencing has elucidated their important roles in population divergence and speciation [305]. Therefore, the question and budget at-hand will dictate how researchers choose to quantify genome-wide genetic diversity.

How important is genome-wide genetic variation?

Since genome-wide genetic variation incorporates neutral, nearly neutral, beneficial, and deleterious variation, it offers diverse insights for conservation genetics. Characterizing genome-wide variation allows conservation practitioners to assess population structure, demographic history (via, e.g. coalescent models), and both historical and current gene flow patterns [220, 90]. When it comes to prioritizing populations, we highlight two reasons for considering genome-wide variation.

First, as discussed in Section 2.1.3, there is extensive literature highlighting the positive relationship between genome-wide genetic diversity and current mean population fitness [71], either because it is a marker of past large population size and so less drift or by directly increasing population viability [157].

Second, and more critically, it is difficult to be certain which alleles at which loci may be useful in the future. Despite statistical and technical advancements, there are still challenges in accurately forecasting how the environment and (thus selective regime) will change in the future, including the unpredictable nature of climate change [102], and which loci and alleles are implicated in climate adaptation (see next section). Conserving populations with high levels of genome-wide variation may therefore act as insurance - with more variation overall increasing the probability that adaptive variation will be preserved [53, 71].

2.3.2 Adaptive Genetic Variation

How is adaptive genetic variation measured?

A subset of the pool of total genetic variation discovered with the above methods will be currently adaptive. A variety of methods exist to determine which variants are adaptive; these methods can be broadly grouped into outlier-type analyses vs. association-type analyses [138].

Outlier-type analyses (often called F_{ST} outlier approaches) search for variants that are highly differentiated among populations, more than expected under neutral models of drift alone [3, 26, 178, 183]. These outlier loci are then considered as putatively adaptive in their local environments - i.e. they are at higher than expected frequencies due to the effect of past

and ongoing selection. Arlequin, BayeScan, Fdist, Lositan, and OutFLANK are commonly used platforms for identifying such outlier loci [3].

Association-type analyses identify variants that are highly correlated with environmental or phenotypic variation across space (see e.g., [138]). Bayenv, BayPass, GWAS, LFMM, Logistic regression, RDA, Tassel, MatSAM, and the WZA [38] are commonly used for association-type analyses [3].

The motivation for using outlier-type vs. association-type analyses often relates to how much prior knowledge researchers have on the environmental axes or traits that are important for an individual species' adaptation to their local environment: when little is known, outlier-type analyses can be useful because they identify putatively adaptive loci independently of environment [138].

How important is adaptive genetic variation?

Identifying adaptive genetic variation has become an increasingly common approach in conservation biology [138, 259] with several advances in recent years for both outlier and association-type analysis [94]. In general, identifying adaptive variation will be useful for conservation practitioners because it will help characterize adaptive differentiation between populations, which in turn can inform both preservation and assisted gene flow (the movement of individuals from a source population with genetic variation that is predicted to be adaptive in the future) [90, 112]. As for prioritizing populations, several authors (see, e.g., [103, 281]) advocate for focusing on adaptive variation rather than genome-wide variation because: (i) such variation has a relatively direct link to function and thus adaptive capacity, and (ii) the functional relationship can be used to predict evolutionary responses and manage species accordingly [46, 94].

Teixeira and Huber (2021) quote [176]: "The question was never really how much genetic variation is there but rather what is the nature of genetic variation for fitness in a population" [176, 281] – in other words, high genetic diversity does not necessarily equate to suitable variation needed to adapt to a future changing environmental conditions. Teixeira and Huber (2021) offer the example of the Denisovans and Neanderthals: despite having low genetic diversity and low inferred effective population sizes, these populations survived in severe conditions and there is evidence that some alleles from each of these lineages introgressed into our lineage, potentially facilitating *Homo sapiens* adaptation to high altitude, cold climate, and local pathogen pressure. In some cases, conservation practitioners will have information on how a changing environment will affect a species, and what traits will be under strong selection. For example, in marine ectotherms, heat-resistance alleles have been linked to past persistence to high temperatures, and will likely continue to be important in warming oceans [23]. Based on this, focusing on preserving adaptive genetic variation directly might be useful to help populations adapt to changing conditions.

In addition, because there is a known link between adaptive alleles and selective regimes, this can provide valuable insights for predicting population responses to future change [83]. Much research (see [5, 46, 139]) is based on the assumption that climate change will disrupt some locally adapted populations by pushing them off their fitness peaks into maladaptive valleys. The level of disruption will depend on the distribution of adaptive alleles across a heterogenous landscape and the magnitude of environmental (primarily climate) change [46]. By combining approaches like common garden experiments, simulations, and Species Distribution Models, one can examine the mismatch between current allele frequencies in a population versus the allele frequencies required to maintain the current fitness levels under different climates [24, 46, 293]. One could then prioritize populations based on the magnitude of these mismatches: for example, depending on triage approaches, practitioners may choose to prioritize populations that are predicted to have small mismatches because they are more likely to succeed under future conditions; or conversely, they may prioritize highly-mismatched populations because they are most at risk of extirpation.

2.3.3 Limitations with Measures of Genetic Variation

Today, the genome of every imaginable species can, in theory, be probed [99]. However, reproducibility is still a problem. Results between studies can vary greatly and different sequencing techniques, such as ddRAD, pool-seq, target-chip, or whole-genome sequencing, have their own unique costs and benefits in terms of number and type of genetic markers identified, proportion of genome sampled, coverage, scalability and price [99]. Comparative work is also hampered by a lack of accessibility of raw genetic data in a standardized format, which impedes comparative analyses across approaches. Publicly-available meta-data attaching each individual to its geolocation and/or population, and lists of the subset of adaptive SNPs along with specific filters and comparable significance thresholds that were used to identify these SNPs would be especially beneficial towards this end. There are also some issues associated with each type of genetic variation (genome-wide vs. adaptive) that can complicate conservation decisions. Genome-wide genetic variation includes deleterious variation, and this will be of special concern in small populations where deleterious mutations may be at high frequency due to genetic drift or stochastic events. It may be difficult to identify deleterious variation, and if it spuriously inflates a population's diversity or distinctiveness scores such that it is prioritized over a genetically "healthier" population, it could misdirect limited conservation resources. In addition, there is only a weak correlation between genome-wide variation and conservation status [310], which some have suggested means genome-wide variation is uninformative of a species' or populations' true extinction risk [281]. This is likely due to a lag between an initial reduction in population size and the loss of genetic variation, as most threatened populations initially decline due to nongenetic factors [157]. Such a lag is problematic because a snapshot measurement of genome-wide

(or adaptive) variation within a population may be an unreliable predictor of future levels of variation.

Prioritizing adaptive variation in conservation decisions has its own unique challenges. We mention five here. First, adaptation may occur via at least three pathways: novel mutations, standing genetic variation, or adaptive introgression [283], so solely focusing on one of these pathways (e.g. adaptive loci from standing genetic variation) would be a gamble. And to the extent that new mutations contribute to future adaptive responses in the long term, the utility of prioritizing currently adaptive loci diminishes further [33]. Second, identifying local adaptation remains a challenge, with frequent misidentification of true adaptive loci. Both genotype-environment or genotype-phenotype (GEA or GPA) association tests or F_{ST} outlier approaches are correlational, not causal [87, 157]. Many factors can influence the results of these correlation approaches including current population structure and gene flow, demographic history, the spatial and temporal resolution of variables, and linkage disequilibrium [99, 157]. While we can disentangle some of these background variability effects [88], there is still *a priori* a degree of uncertainty in the process of identification, interpretation, and validation [94] of adaptive genetic loci and the inevitability of both false-positives and false-negatives [262, 67]. Due to these and other unforeseen factors, even well-designed studies may miss important adaptive SNPs [32, 225].

Third, these methods are biased towards detecting loci of large effects [241, 261] or nonsynonymous mutations near or within coding regions [46, 122, 126]. Current approaches have trouble detecting mutations that may influence regulatory regions, alternative splicing, and noncoding regions, all of which are likely to contain significant adaptive regulatory variation [122]. These biases are important as many (maybe a majority of) quantitative traits are controlled by alleles of small effect across many loci that will remain undetected [33, 126, 166, 241, 261]. Hayward and Sella's 2022 [126] recent modeling further emphasizes the ubiquity of polygenic adaptation by suggesting that large effect loci "almost never sweep to fixation" [126]. Even if we solely focus on loci of large effect, different methods may detect different SNPs or may miss others depending on the specific techniques and thresholds of each method. Overall, this bias against identifying the many loci of small effect may be a major impediment to the field.

Fourth, even if methods to identify adaptive loci were adept at isolating small-effect loci, preserving those loci without a full understanding of the underlying genetic architecture may be problematic. For example, selection at one locus can considerably impact patterns at other linked loci or genetic regions [314], and other genetic factors (ex. additivity, pleiotropy, epistasis, dominance, etc.) can make it entirely unfeasible to fully understand any particular variant's impact on fitness [16, 119], nor predict how it will behave in new genetic combinations or selective regimes. Genomes have not or will not evolve into a perfect adapted form as they are dynamic in space and time [200, 225], and ignoring this complexity could undermine conservation goals [200].

Finally, going through the additional steps of identifying which subset of genetic variation may be adaptive is so challenging, time-consuming, and expensive [53] that it is impractical for conservation practitioners with limited resources and time. Taken together, these factors suggest prioritization schemes based solely on adaptive variation may not be useful in all scenarios, and should also be interpreted cautiously.

2.3.4 Genome-Wide Genetic Variation as a Surrogate for Adaptive Genetic Variation

Given the logistical and technical challenges associated with identifying adaptive variation, an important question is whether it is necessary for conservation practitioners to do so. This debate extends back to at least 1986 [246] and has been frequently revisited in years since. The earlier discovery of allozyme variants [146, 177] ignited attempts to identify adaptive variation [267]. Geneticists, population biologists, and zoos aimed to identify and preserve gene pools with "adaptive genetic variation" [246]. Indeed, one of the ideas that gave rise to the conceptualization of ESUs came from attempts to determine which subspecies "actually represent[ed] significant adaptive variation" within species [246]. The conceptualization of ESUs emphasized the preferred use of ecological data, "significant adaptive variation" between populations or variation at "functionally divergent gene copies" in maintaining adaptive differences [119, 205, 60, 97].

However, it may be that genome-wide patterns of genetic variation capture the adaptive subset of this variation well, such that it can be used as a surrogate measure. Though genome-wide includes all genetic variation, much of the work in this area focuses on the (presumed large) neutral subset of that total genome-wide variation, and comparing that to the adaptive subset, and so we primarily focus on those two subsets below. Traditionally, the view was that neutral and adaptive variation may be only weakly, if at all, correlated [119, 128, 186, 205, 267] leading many researchers to recommend focusing directly on "genetic variation for traits that affect fitness" [119, 205]. In part, this stance may have been influenced by the possibility that many loci with "adaptive significance" [267] largely went undetected due to the technological limitations at the time. However, as techniques improved and more studies were conducted, a second view emerged that neutral and adaptive variation may indeed be positively correlated on the landscape. Recent literature has reignited this debate (see [71, 102, 157, 281]) and we now consider some theoretical arguments and empirical results as to whether genome-wide genetic variation (most presumed neutral) may capture and act as a surrogate for (the subset that is) adaptive, both within and among populations.

We begin with within-population genetic diversity, where neutral variation has often been dismissed as a proxy for adaptive variation because it is not directly targeted by selection [33, 142, 164]. However, population genetic theory does support a relationship between the two. As discussed in Section 2.1.3, a core tenet from theoretical genetics is

that adaptation requires additive genetic variance at loci underlying the selected trait. These loci are subjected to many of the same evolutionary processes as all other loci (drift, mutation, migration). As such, both neutral and adaptive variation are similarly affected by population size [157]. In small populations where selection is less effective, these processes will become more important than selection such that adaptive and neutral diversity may show similar patterns [204, 225]. Empirical work looking at the correlation between additive genetic variation and neutral diversity has produced mixed results. Mathur et al. 2023 [193] found that neutral and functional genomic diversity are correlated across populations of endangered Eastern Massasauga rattlesnake (*Sistrurus catenatus*). Bataillon et al. 1996 [21] carried out simulations to analyze phenotypic evolution in hermaphrodite plants and found a high correlation between diversity measured as allelic richness at neutral loci and loci under selection. However, other studies have found only weak correlations [202, 238], or no relationship at all between the two [27, 36, 232].

When it comes to among-population genetic variation (measured, e.g., by a population's distinctiveness), there is theoretical and empirical support for a correlation between adaptive and neutral variation. Divergent selection regimes between populations combined with low gene flow can lead to isolation-by-adaptation, and neutral genomic divergence can occur as a by-product [216, 217]. For instance, outlier loci under selection can impact other loci on the same chromosome in linkage disequilibrium with it, and this, through divergent selection, drift, and low gene flow, can lead to positive correlations between differentiation at adaptive loci and differentiation at neutral loci [197, 173, 216], though the relationship can be weak [142] (and see also [290]).

The degree to which genome-wide vs. adaptive genetic variation correlate with one another may vary between systems and organisms, but perhaps a more pressing question for practitioners is whether they diverge enough that selecting one over the other leads to measurable impacts on conservation decisions such as (i) determining how many populations there are; (ii) determining which populations to prioritize; and (iii) deciding which populations will be most resilient to anthropogenic change.

First, different markers (neutral or adaptive) often show distinct genetic structure patterns, leading to different population delineations [53]. For example, Pecoraro et al. 2018 [228] used data on Yellowfin tuna (*Thunnus albacares*) and detected higher number of stocks using adaptive variants compared to neutral variants. Sandoval-Castillo et al. 2018 [248] found that neutral and adaptive variants gave rise to different management units within the greenlip abalone (*Haliotis laevigata*). On the other hand, the two sets of markers largely captured similar population structure patterns in sage grouse [218], Atlantic salmon [204] and coho salmon [318].

Second, if genome-wide and adaptive genetic variation do not correlate with one another, different populations might be prioritized for conservation. Xuereb et al., 2021 [317] examined which sites (and therefore populations) were prioritized across the range of the

giant California sea cucumber (*Parastichopus californicus*) depending on whether neutral or adaptive SNPs were considered, and whether diversity vs. distinctiveness metrics were used. They found that priority areas varied greatly depending on both markers and metrics. In contrast, Fernandez-Fournier et al., 2021 [86] compared population prioritization rankings of yellow warblers in North America and lodgepole pines in western Canada using genome-wide vs. adaptive SNPs and found very similar population prioritization rankings using the two types of markers.

Third, if the two sets of markers do not positively co-vary, does one type of marker or metric better predict how well a species will adapt to a changing environment? Bertin et al. 2020 [33] found that genome-wide diversity better predicted a population's adaptive potential than climate-associated loci in high Andean wetland plant, *Carex gayana*. Similarly, Fitzpatrick et al. 2021 [87], using a gradient forests approach, found that sets of randomly selected SNPs better predicted climate adaptation in Balsam poplar than climate-associated SNPs in common garden experiments.

While some authors have recently dismissed the role of genome-wide genetic variation in conservation genetics [281], others have highlighted decades of research supporting its fundamental role [71, 157, 102]. The argument is that genome-wide variation may act as a surrogate metric of adaptive potential. We find that while the majority of empirical work generally supports a positive correlation between the two when examining population distinctiveness metrics, we agree with authors who point out that is not always a straightforward relationship [53], and this relationship is even less predictable for within-population diversity metrics [142]. Further work disentangling what factors mediate the covariance between genome-wide and adaptive variation (for example, number of adaptive loci, type of analysis used to identify adaptive variation, ecology) would provide important insight into which situations would most benefit from identifying adaptive variation in species of conservation concern.

2.4 Towards Conserving Relevant Genetic Variation

2.4.1 Genetic Variation and Conservation Urgency

Many genomic studies (see, e.g., [4, 53, 181, 220]) analyze factors such as evidence of population structure, gene flow, population levels of differentiation, historical distribution, demographic patterns, inbreeding and outbreeding depression, with the goal of offering conservation management recommendations. According to these studies, the extent to which entities should be managed separately (i.e., keeping them distinct without gene flow) or as collectives (i.e., encouraging or at least allowing gene flow) will depend on assessing these various factors and weighing genetic risks against possible advantages of each strategy on a case-by-case basis [4, 181].

However, while every conservation situation must be considered independently, for critically endangered species where extinction may be imminent, prioritizing and managing subpopulations separately from one another is likely not the best option: conserving as much genome-wide variation as possible may be the better strategy, perhaps achieved via active mixing [300]. This approach has been used successfully for the critically endangered Yellow-tufted honeyeater *Lichenostomus melanops cassidix* [121, 224] and Mountain pygmy possum *Burramys parvus* [299] in Australia, (and considered in others, like Arabian leopard *Panthera pardus nimr* [6]). The mixing approach has been associated with improved fitness and conservation outcomes. In situations where there is no immediate loss of genetic variation (e.g. where populations are still large though contracting), entities are managed separately, often under the assumption they are locally adapted. This is common in Canadian conservation, with infraspecific designatable units of many species being both defined based on presumed local adaptation, and subsequently managed separately, for example, the Halfmoon Hairstreak butterfly [59], pacific salmonids [298] and Atlantic salmon [171]. That said, we do not know of any specific examples where populations are *prioritized* based on genetic distinctiveness (genome-wide or adaptive) or within-population genetic diversity. More work is clearly needed here.

2.4.2 Approaches for Integrating Within and Among-Population Genetic Variation

As implied above, in scenarios where outbreeding depression is strong or there is strong local adaptation (most common in historically larger populations), managing distinct populations separately might be a beneficial strategy [4]; in small isolated populations that are experiencing inbreeding depression or when locally adapted populations are becoming maladapted, e.g. due to climate change, encouraging gene flow might introduce useful genetic variation, aid population growth and adaptive potential [4, 112]. The majority of cases will fall between these extremes.

Generally, conservation geneticists emphasize that both long-term survival and evolutionary potential of species require actively managing both within and among-population genetic diversity and suggest various approaches [54, 57, 119, 136, 179, 168, 231, 221, 230, 284, 279]. It is clear that, to the extent that a population's distinctiveness and standing genetic variation are negatively correlated [57, 294, 300], conservation management prioritization schemes must consider both facets explicitly and simultaneously [284]. What is less clear is whether the populations are conserved because they contribute to biodiversity conservation at the level of the evolving lineages that they are constituents of, i.e. species, or whether they are considered to be separately evolving lineages.

At the level of global conservation indicators of genetic diversity maintenance, Hoban et al. 2020 [136] and Laikre et al. 2020 [168] suggest that the number of species where genetic diversity data are being collected should be monitored, as well as within-species measures of

population sizes (e.g. the number of populations above and below the critical N_e=500 level [136]), and the rate at which genetically "distinct" populations within species are being lost [168].

Several authors have considered within- and among- population genetic diversity explicitly, with a goal to conserving total (species-level) variation. Petit et al. 1998 [231] suggest measuring the contribution of the k^{th} population to the total diversity of all populations pooled, using, e.g. Nei's diversity or allelic richness. They suggest that a population's contribution, which they call ($C_{k,T}$), is the change in the pooled diversity if the k^{th} population were removed. This change will be affected both by how diverse the k^{th} population is and by how different its genetic composition is from the mean of all populations, including, e.g. the presence of private alleles. These two components (within-population diversity and the difference from the mean) are exactly within- and among- population genetic diversity. Petit et al. 1998 [231] suggest comparing each population's contribution to the mean contribution across all populations to identify populations that are worthy of conservation attention because of their higher-than-average contribution. This measure was implemented by Taylor et al. 2011 [279] to rank *Oncorhynchus mykiss* populations for conservation using allelic richness.

Caballero and Toro in 2002 [44] and 2005 [284] offer another approach for conserving total genetic variation in a metapopulation in the context of the management of rare and endangered breeds of domesticated animals. Caballero and Toro also equate "genetic diversity" with species-level expected heterozygosity, and partition the total genetic diversity (GD_T) of a collection of populations into a within-population component (GD_{WS}) and a among-population genetic diversity (GD_{BS}) (note, Caballero and Toro call each population a "subpopulation" in keeping with their metapopulation framework):

$$GD_T = GD_{WS} + GD_{BS}.$$

GD_{WS} is the sum of two elements, the genetic diversity within individuals, and the genetic diversity between individuals within a population. GD_{BS} corresponds to the average pairwise genetic distance between all pairs of populations, measured as F_{ST}, i.e. the proportion of total genetic variation represented by divergence among subpopulations [284].

One important point that Caballero and Toro highlight is that over-emphasizing among-population diversity risks overlooking most of the global diversity, while focusing on within-population variation will favor larger populations and therefore, perhaps the less endangered ones. To overcome this issue they suggest explicitly weighting within-population and among-population diversity, scaling GD_{BS} by $\lambda \in (0, 1)$ (for more details see [284]).

Caballero et al. 2002 [44] and Toro and Caballero 2005 [284] present their derivations and indices in an expected co-ancestry and inbreeding framework, and [230] subsequently present the METAPOP software [230] as the support tool - like Petit et al. 1998 [231] approach, this

tool measures the contribution of each population to the pool by re-calculating the pool's global average co-ancestry [230] when the population in question is removed. Importantly, removing a population does not always result in a loss of total variation, depending on the population's contribution to the within and among components (i.e., total variation) [230, 284].

The framework of co-ancestry may not translate easily to the conservation biology of wild populations. Abhari et al., 2024 [submitted] defined the same measure from a different and simpler perspective and called it the Heterozygosity-Pooling ($\text{Het}_{\text{pooling}}$) method. They used $2\bar{p}(1-\bar{p})$ to measure the total genetic diversity of a group of populations (P), where $\bar{p}$ is the average allele frequency per locus, among all populations in P (assumed as a single meta-population), and the total genetic diversity is the sum over all such terms for all loci (see equation 2.1).

$$\text{Het}_{\text{pooling}}(P) = \sum_{i \in \text{all Loci}} [2\bar{p}(1-\bar{p})]_i \tag{2.1}$$

2.5 Conclusions

The adaptive potential of a species is influenced by how genetic diversity is distributed both spatially and across the genome. The spatial structure consists of within- and among-population genetic variation while the genomic structure considers how present and future adaptive variation is structured and how it covaries with overall genetic variation.

We present three conclusions from our review of the conservation-relevant patterns of genetic diversity across populations:

It is crucial to account for and harmonize within and among population genetic variation in conservation management schemes, particularly when they are negatively correlated (e.g. due to drift in small populations). Genetic variation within a given population is important due to its contribution to current population-level fitness and adaptive potential. Genetic differentiation among populations contributes to the species' total genetic variation, and so also to the adaptive potential. Prioritization schemes that focus on only one component carry risks. For instance, solely relying on within-population diversity may prioritize populations undergoing high gene flow, contributing to the homogenization of allele frequencies with other interacting populations. This may impede differentiation and local adaptation. Solely focusing on among-population genetic variation (i.e., the average pairwise genetic distance between populations without correcting for within-population variation) may prioritize some populations whose distinctiveness might be due to drift. This might make them locally non-adaptive and so vulnerable and might make their contribution of variation unhelpful.

The decades-long debate on whether prioritization should focus on genome-wide (and so at least partially neutral) or adaptive genetic variation, continues. The traditional view was

that neutral and adaptive variation are not correlated, whereas more recent, SNP-based work suggests that genome-wide variation may also capture adaptive genetic variation. Conserving genome-wide variation would then maintain current fitness and population viability as well as maximizing the potential for adaptation in an uncertain future. In addition, while the direct link between adaptive variation and function suggests it is a good predictor of adaptive capacity, identifying adaptive variation may be conceptually and technically challenging. Misidentification of loci based on correlational approaches or biases towards large effect loci or high rates of false-positive and false-negative pose problems. Conceptually, there is uncertainty as to whether currently adaptive loci will be adaptive in the future as global change accelerates.

Several methods have been proposed to balance within and among population variation that can be applied to genome-wide or adaptive genetic variation to prioritize populations within a species for conservation management. One possible approach, the pooled heterozygosity framework we highlight here (a simplification of that presented by Caballero and Toro 2002 [44] and Petit et al. 1998 [231]) explicitly considers the species as the primary locus of conservation - each population is evaluated in light of what it offers some present of future *collective* of populations.

2.6 Bridge to Chapter 3: Key Insights and Their Significance in Context

The unresolved debate as to whether to base prioritization on adaptive genetic variation or not extends back to at least 1986 [246], if not earlier, and has been reignited over years [246]. Literature offers mixed theoretical and empirical evidence, and implies the necessity of additional work. Further empirical work and exploration of factors impacting the correlation between genome-wide and adaptive variation would be useful given the current challenges and limitations in identifying and using adaptive genetic variation for conservation.

Fernandez-Fournier et al. 2021 [86] asked whether standing genetic variation can act as a reliable proxy for adaptive genetic variation using population distinctiveness as a metric. They tested this with two published datasets - yellow warbler and lodgepole pine (data from Bay et al., 2018 [22] and Mahony et al.,2020 [190]) and found that genome-wide and adaptive genetic variation produce similar population prioritizations for both test cases. In Ch-2, I propose to test the prediction from Fernandez-Fournier et al., 2021 [86], that genome-wide genetic distinctiveness captures a subset of adaptive variation: if so, genome-wide and adaptive genetic variation will produce similar population prioritizations.

One of the important insights from this review is to account and harmonize both within and among population genetic diversity. However, in Ch-2, I solely consider among-population genetic variation, without accounting for within-population variation since I am testing the prediction from Fernandez-Fournier et al., 2021 [86]. Populations identified as

genetically distinct are critical components in any management scheme whose goal is to preserve species-wide genetic diversity and attendant evolutionary potential, however, it is important to keep in mind that this approach of solely focusing on distinct populations may carry some risks.

2.7 Author Contributions

AOM and JL contributed to the conception, writing, and editing. AKC conducted the extensive literature review and wrote the first draft. NA contributed to editing.

2.8 Acknowledgements

We are grateful for feedback from the FAB lab at Simon Fraser University, and in particular P. Fernandez, B. Crespi, A. Macpherson, J. Moore and T. Booker. We thank the organizers of the Genomics and Conservation Symposium at the 2022 CSEE Annual Meeting for the opportunity to prepare this submission. AKC was supported by graduate scholarships from NSERC, the BC government, and Simon Fraser University. NA was supported by Simon Fraser University and NSERC grants to Caroline Colijn; AOM acknowledges ongoing support from the NSERC Discovery Grants programme; JL is supported by funding to Laura Melissa Guzman from USC and the Wrigley Innovator Award. No specific grant supported this review. We note that Simon Fraser University is located on the ancestral and unceded land of the Musqueam, Squamish, and Tsleil-Waututh Nations.

Chapter 3

Does genome-wide variation and putatively adaptive variation identify the same set of distinct populations?

3.1 Abstract

Identifying which populations within species to prioritize for conservation is a major challenge: one question is whether to prioritize populations based on adaptive variation versus considering genome-wide genetic variation. Many authors have advocated focusing solely on adaptive variation due to its direct connection to selection, function, and adaptive capacity. However, there are many limitations in identifying and using adaptive genetic variation for conservation. Patterns of genome-wide genetic variation may be congruent with patterns of adaptive genetic variation, and genome-wide variation is much easier to measure. However, evidence for congruence is mixed. We gather genome-wide and putatively adaptive SNP data across 34 species of plants and animals from published outlier and association studies to test congruence. We ask whether putatively adaptive subsets of genome-wide SNPs identify the same distinctive populations (measured using the Shapley Value of distinctiveness) as genome-wide SNPs. We find that genome-wide and putatively adaptive SNPs generally but variably agree on population prioritizations. As expected, the level of agreement is predicted by the proportion of putatively adaptive SNPs, and the agreement is lower when there is more overall population genetic structure. Interestingly, across our datasets, putatively adaptive SNPs do as well or better at predicting genome-wide population prioritization than sized-matched random subsets of SNPs. Taken together, using genome-wide genetic variation for population prioritization may be a generally sound and cost-effective strategy for prioritizing populations in order to safeguard species-level genetic variation.

Keywords: Genome-wide genetic variation, Adaptive genetic variation, Population prioritization, Population distinctiveness, Adaptive potential, Conservation genetics

3.1 Introduction

Intraspecific genetic diversity – differences in alleles and genotypes among individuals and among populations – is a critical aspect of biodiversity because it contributes to the adaptive potential of species and populations (i.e., populations' capacity to adapt to accelerating environmental change) [157, 206, 113]. While the United Nations post-2020 targets aim at maintaining "at least 90% of all species genetic diversity" [72, 81, 82], a recent study suggests we may already be falling short [82]. The widespread loss of genetic information across taxa increases the urgency with which we must tackle the issue of conserving intraspecific genetic variation [82]. However, due to limited funds, triage is a reality of conservation across various taxa [41], so identification and ranking of populations that contain substantial genetic information is crucial for conservation.

One strategy is to prioritize most genetically distinctive populations as these may maintain distinct genetic variation required for present and future adaptation, thus contributing to adaptive potential at the species level. Many studies [313, 199, 289, 297, 62, 84, 175, 279, 57, 195, 291, 189, 136, 9] have emphasized the importance of conserving genetically distinctive populations, including Ehrlich, in 1988, advocating that "the loss of genetically distinct populations within species is, at the moment, at least as important a problem as the loss of entire species" [76]. One common reason is that the consideration of distinctive populations is fundamental to the conceptualization of evolutionary significant units (ESUs), which are often managed as separate legal units under the Endangered Species Act in the USA [246, 297, 119, 205, 135]. Another argument is that these populations may harbor non-redundant genetic history compared to other populations [236, 298], perhaps because they may have been evolving in isolation for many generations [205]. Thus, conserving these distinctive populations should maximize the amount of total genetic information maintained for a species if genetic variation is distributed non-randomly across semi-discrete populations (see e.g., [291, 86]).

In the context of this paper, intraspecific genetic distinctiveness is classified into two types: genome-wide genetic distinctiveness and putatively adaptive genetic distinctiveness (hereafter, we do not use the "putative" modifier for readability purposes). Genome-wide genetic distinctiveness indicates "all forms of variation, including neutral, nearly neutral, deleterious, and adaptive genetic variation sampled throughout the genome" [51]. Theoretically, adaptive genetic distinctiveness refers to a defined subset of this genome-wide distinctiveness attributable to past natural selection or proposed to be linked to ongoing or future natural selection [142, 259]. Many authors might define "putatively adaptive" in very different ways depending on the type of analysis, study system, and study specifics used to identify adaptive SNPs, so this category may include very different definitions. Operationally, two common methods exist to identify currently putatively adaptive genetic variants, outlier-type analyses – identifying variants that exhibit higher differentiation among populations

than background F_{ST} (expected under neutrality) in genome scans [178, 26, 183, 3], and association-type analyses – detecting genetic variants that are associated with environmental or phenotypic variation across geographical space (see e.g. [138]). However, these common methods may generate high number of false-positives and false-negatives that may not always reflect true adaptive variants under selection [269].

The efforts to identify and base conservation prioritization solely on adaptive genetic distinctiveness extends back to at least 1986 [246], if not earlier, as biologists and zoos aimed at identifying populations (e.g. subspecies) within species that "actually represent[ed] significant adaptive variation" [246]. The theoretical reason for directly relying conservation on adaptive genetic variation dates back to 1974 when Lewontin advocated "the question was never really how much genetic variation is there but rather what is the nature of genetic variation for fitness in a population" [176, 281] – which means, high genetic variation may not strongly predict adaptation to future changes. Other reasons include the relatively direct association of adaptive genetic variation to selection, function and adaptive capacity and the usefulness of this association to anticipate future adaptive responses (see e.g., [5, 33, 46, 142, 139, 164, 103, 281, 293]).

However, there are several challenges and limitations in identifying and using adaptive variation for conservation. One challenge is current methods are based on correlation approaches and are biased in favour of identifying large effect loci, nonsynonymous mutations near or within coding regions. This is problematic in the face of ubiquity of polygenic adaptation as many of the small effect loci contributing to adaptation may go undetected [241, 261, 122, 46, 126, 33, 166]. Even if approaches could be refined to capture candidate loci contributing to polygenic adaptation, disentangling the strong interconnectedness of genomic background variability effects, genetic architecture, and linkage impact on fitness might be unfeasible [88, 314, 119, 16, 200, 225]. All these technical factors play a major role in frequent misidentification, false interpretation and validation of true adaptive loci, contributing to high rates of both false-positives and false-negatives [67, 99, 87, 94, 157, 262, 32, 225]. Even if all the technical challenges could be resolved, conceptually solely relying on currently adaptive loci would be risky as adaptation may occur via novel mutations, standing genetic variation, or adaptive introgression [283], depreciating the future usefulness of loci that are currently adaptive [33].

Given these current limitations in identification of adaptive variation, a critical question is whether it is necessary? Genome-wide genetic distinctiveness, which is much easier to measure, may act as a surrogate for adaptive genetic variation in some cases (see [71, 102, 157]). Because genome-wide genetic variation incorporates information about the future, present, and past, there are several theoretical and empirical reasons supporting that genome-wide variation provides useful data for conservation genetics in their own right (independent of their ability to track patterns of current adaptation). One future benefit is the possibility that current neutral variation may become adaptive in the uncertain future circumstances,

so conserving as much genetic variation as possible – high genome-wide variation – may act as an insurance as it increases the chances that current and future adaptive variation will be retained [53, 71, 102]. Presently, genome-wide variation is beneficial due to its positive relationship with current mean fitness [71]. This relationship might be due to high genome-wide variation being indicative of past larger population size (and so less past drift) or high population viability via current heterozygote advantage [157]. Because genome-wide genetic information contains historical information, it offers data to infer migration, gene flow, population structure, and demographic history patterns [220, 90]. For instance, Bertin et al., 2020 [33] used an Andean wetland plant, *Cares gayana*, to demonstrate that genome-wide variation better predicted population adaptive potential – in this case measured as an indicator of genomic response of climate-linked loci – compared to genetic variation at climate-associated loci. They suggested that this result might be influenced by the combined effects of neutral or demographic factors or small effect (polygenic) loci, many of which are part of genome-wide variation, but not adaptive variation due to detection bias towards large effect loci. Fitzpatrick et al., 2021 [87] conducted a common garden experiment and reported that randomly selected variation better predicted climate adaptation of Balsam poplar than did climate-associated variation (detected through gradient forest approach).

The decades-long debate on whether population prioritization efforts for conservation should solely focus on the adaptive subset of genetic variation, or whether it is preferable to consider genome-wide variation continues. The traditional view was that neutral and adaptive variation may hardly be correlated [119, 186, 128, 267, 205] whereas, more recent work on SNPs indicate that genome-wide variation may act as a surrogate for adaptive genetic variation (see [71, 102, 157]). Below we present theoretical arguments and mixed empirical results from literature for a positive correlation between the two types of genetic variation.

One argument is adaptation needs additive genetic variation at loci under selection. But there is weak empirical support for this argument as some estimates of neutral molecular variation may weakly or not correlate with quantitative genetic variation. For example, Reed and Frankham 2003 [238], Reed and Frankham 2001 [237], and Mittel et al. 2015 [202] show very weak correlation between neutral or nearly molecular markers and estimates of quantitative genetic variation (ex. heritability) whereas Podolsky 2001 [232] shows no relationship between estimates of putatively neutral markers and quantitative genetic variation.

Another theoretical reason for a correlation between random and adaptive genetic variation is that loci under weak selection may be subjected to many of the same evolutionary processes – drift, mutation, and migration – as all other loci [33, 204, 225] such that in small populations, genome-wide and adaptive variation may show similar patterns [204, 225]. This is supported empirically by Mathur et al. 2023 [193] examining small populations of endangered Eastern Massasauga rattlesnake (*Sistrurus catenatus*) populations and Bataillon et al. 1996 [21] examining hermaphrodite plants to simulate phenotypic evolution change in

germplasm collections found neutral and functional (non-neutral) genetic variation to be significantly positively correlated.

Another theoretical reason for expecting a positive correlation is that strong divergent selection regimes and lower gene flow can lead to neutral genomic divergence [216, 217] as loci under selection may impact non-selected loci through physical linkage or through collective or unique influence of divergent selection, drift, and low gene flow. This pattern may lead to positive correlation between differentiation of neutral loci and differentiation of both additive genetic variation and putatively adaptive loci [216]. Literature provides mixed empirical evidence. For example, Merila and Crnokrak 2001 [197] examined 18 studies of plants and animals and showed that differentiation at neutral markers and differentiation at quantitative traits markers are correlated. Leinonen et al. 2008 [173] analyzing data on 50 species found a positive correlation between Q_{ST} and F_{ST} values. On the other hand, Holderegger et al., 2006 [142] show a weak relationship between neutral and adaptive genetic data. Volis et al., 2005 [290] used regional Q_{ST} to imply divergent selection on quantitative traits and showed that F_{ST} and Q_{ST} differed at the regional scale, indicating a weak relationship between neutral and adaptive variation.

Another theoretical argument is that the degree of correlation between genome-wide and adaptive genetic variation depends on systems and organisms, and it is important for conservation practitioners to consider whether marker choice impacts the number, prioritization, or our estimate of resiliency of populations. Empirical studies again provide mixed results: in some cases, neutral and adaptive markers may provide different genetic patterns, such as detecting different stocks in Yellowfin tuna (*Thunnus albacares*) [228] and identifying different management units in greenlip abalone (*Haliotis laevigata*) [248] and other times, they may capture similar population structure patterns such as in sage grouse [218], Atlantic salmon [204], and coho salmon [318]. In some cases, sites (locations) prioritized across the range may vary greatly depending on the choice of both marker (neutral vs. adaptive) and metric as shown in giant California sea cucumber (*Parastichopus californicus*) [317]. In some cases, different markers provide similar population prioritization rankings as shown in yellow warblers and lodgepole pines [86] and in other cases, neutral and adaptive variation provide different population prioritization such as in an amphibian (*Rana temporaria*) and a plant (*Dracocephalum austriacum L.*) [36].

Recently, Fernandez-Fournier et al. 2021 [86] asked whether standing genetic variation can act as a reliable proxy for adaptive genetic variation using population distinctiveness as a metric. They tested this with two published datasets - yellow warbler and lodgepole pine (data from Bay et al., 2018 [22] and Mahony et al.,2019 [190]). They hypothesized that if population prioritization rankings are similar when using genome-wide and putatively adaptive genetic variation, it would support that genome-wide variation can be a useful proxy for adaptive genetic variation [86], while if different rankings are produced then true adaptive SNPs might need to be identified for conservation management. They

found that genome-wide and putatively adaptive genetic variation produced similar population rankings. I repeat, test and extend Fournier-Fernandez et al., 2021 [86] from two species to thirty-four species and test several covariates of the hypothesized proxy value of genome-wide SNPs to identify populations for conservation prioritization. I also compare the observed correlation between genome-wide SNPs to the expected correlation between genome-wide and random SNPs to see if there is any difference. Specifically, I ask, does genome-wide variation and putatively adaptive variation identify the similar set of distinct populations? We divide our overarching question into three sub-questions: Do genome-wide (all) SNPs and adaptive SNPs provide similar population prioritizations? Do adaptive SNPs behave like equal-sized samples of random SNPs? How do covariates impact the observed correlation of Shapley value distinctness scores using genome-wide and putatively adaptive SNPs (hereafter, referred to as prioritization agreement) and the correlation between genome-wide and adaptive SNPs relative to random expectation (hereafter, referred to as predictive power of putatively adaptive SNPs)?

3.2 Methods

3.2.1 Choosing the studies

Initial pass

I collated genome-wide and putatively adaptive SNPs data across various taxa from outlier-type and association-type published studies. F_{ST} outlier-type studies look for SNPs that are highly differentiated among populations, more than expected via drift [178, 26, 183, 3] whereas association-type studies scan for SNPs that are correlated to environmental or phenotypic variation across geographical space [138]. F_{ST} outlier-analyses use Arlequin, BayeScan, Fdist, Lositan, OutFLANK, and Random Forest whereas association-type analyses use Bayenv, BayPass, GWAS, LFMM, Logistic regression, RDA, Tassel, MatSAM platforms to identify putatively adaptive SNPs.

I did a literature search of genome-wide association studies on google scholar and web of science from September 2021 to February 2022 using terms such as "SNPs across entire range species", "SNPs associated with environment within species entire range", and "SNPs associated with climate across species range". From the literature search, I found three meta-analyses [3, 182, 82], each of which contained a list of outlier and/or association type analysis studies. Studies within three meta-analyses and recommendations from colleagues were evaluated based on a set criteria with thresholds, and the availability of full-data.

Criteria and thresholds

I selected published studies based on a set criteria that included incorporating only wild organisms, a full SNP profile per individual consisting of information attaching each individual to its geolocation and population as well as an identification of putatively adaptive SNPs

through F_{ST} outlier analyses, genotype-environment and/or genotype-phenotype association analyses from studies. Additionally, to meet the criteria thresholds, studies required a minimum of: 250 raw genome-wide SNPs, 10 raw putatively adaptive SNPs across genome or across many genes, 50 total number of individuals, 4 populations (i.e., number of geographic sites), 4 individuals per population (studies with less than 4 individuals per population were included, but, later on during the analysis stage, these specific populations with less than 4 individuals were removed), and at least 3-4% coverage of natural geographic range. These thresholds were decided arbitrarily. I used the same identified populations as authors assigned them in their studies. If I had multiple datasets of the same species, I chose the one with a higher geographic range covered. I also recorded information on the type of analysis (e.g., LFMM, GCTA, MLM, RDA, Bayenv2, Bayescan, MATSAM, FDIST2, etc), method type (F_{ST}, GEA (genotype-environment association), GPA (genotype-phenotype association)), type of sequencing method (ddRAD, genotype by sequencing, SNP arrays, etc), account for neutral population structure or demography, adaptive SNPs identification significance thresholds, range, biome, taxon, number of variables/traits, and proportion of adaptive SNPs for each study collected. A few candidate studies with big ($>$ 14.6 GB) data storage files were not used due to cumbersome operational analysis and time.

Outlier study removal

The major reason for exclusion of studies from three meta-analyses included the total number of SNPs falling below the threshold. A few studies that fit the criteria were not used for several reasons, including inaccessibility of the full data, files too big to run cumbersome operational analysis, and data with the same species already included. Additionally, I removed two outlier studies [68, 277] that passed all filters, including the initial criteria, thresholds, and reasons listed above. I removed De Kort et al., 2015 [68] because the raw data provided with 10% missing rate cutoff had 183 genome-wide SNPs and the proportion of putatively adaptive SNPs were unusually high (89% of the genome-wide SNPs). I removed Stuart et al., 2022 [277] because it was the only study with a very high missing rate. The missing rate that the study used for alleles per SNP was 0.54 and for SNPs per sample was 0.50. Essentially the majority of the data was removed after applying 10% missing rate cutoff (see section 2.2.4 for more information) with 819 SNPs remaining from 128217 total genome-wide SNPs and 3 putatively adaptive SNPs remaining from 1322 adaptive SNPs. The final datasets consisted of 34 species (see Table 1) that met the full criteria with full data availability. The datasets included 2 birds, 8 gymnosperms, 6 angiosperms, 9 fish, 4 mammals, 1 amphibian, and 4 arthropod species. The values of variables after applying 10% missingness cutoff across studies ranged between 187 to 105,000 number of genome-wide SNPs, 7 to 1609 number of putatively adaptive SNPs, 0.002 to 0.444 proportion of adaptive SNPs compared to genome-wide, 80 to 1473 number of individuals, 4 to 83 number of populations with 20 animal and 14 plant studies (see Table 1).

Table 3.1: Information on study number, study ID, organism examined, number of genome-wide and putatively adaptive SNPs, number of individuals and populations after applying 10% missing rate cutoff, along with the method used to identify putatively adaptive SNPs for each study.

Study number	Study ID	Organism	Number of genome-wide SNPs	Number of putatively adaptive SNPs	Number of individuals	Number of populations	Method type
1	Ruegg et al., 2018, 2021	*Empidonax traillii*	105000	179	144	6	GEA
2	Eimanifar et al., 2018	*Apis mellifera L.*	2449	140	471	29	FST
3	White et al., 2013	*Myodes glareolus*	4340	17	241	14	GEA
4	Schweizer et al., 2016	*Canis lupus*	13092	34	106	6	GEA
5	Mosca et al., 2016 *P. mugo*	*Pinus mugo*	615	89	622	20	FST, GEA
6	Mckown et al., 2014	*Populus trichocarpa*	29535	410	439	25	GPA
7	Royer et al., 2016	*Y. brevifolia* , *Y. jaegeriana* , hybrids	4603	35	197	6	GPA
8	Christmas et al., 2016	*Dodonaea viscosa*	7255	59	87	17	FST, GEA
9	Benestan et al., 2016	*Homarus americanus*	3395	7	421	19	FST
10	Roffler et al., 2016	*Ovis dalli dalli*	187	57	472	15	FST
11	Babin et al., 2017	*Anguilla rostrata*	12098	269	710	13	GEA
12	Guo et al., 2016	*Bufo andrewsi*	15577	586	264	11	FST, GEA
13	De kort et al., 2014	*Alnus glutinosa*	1714	16	295	23	FST, GEA
14	Hurel et al., 2021	*Pinus pinaster Aiton*	6074	96	515	33	GPA
15	Depardieu et al.,2021	*Picea glauca [Moench] Voss*	6153	359	1473	43	GPA, GEA
16	Chen et al., 2012	*Picea abies*	375	32	262	18	FST, GEA
17	Xuereb et al., 2022	*Oncorhynchus kisutch*	9683	119	836	26	GEA
18	Holliday et al., 2010	*Picea sitchensis*	437	35	407	13	GPA
19	Flanagan et al., 2021	*Syngnathus scovelli*	2738	312	235	7	FST, GEA
20	Bay et al., 2018	*Setophaga petechia*	104385	1609	173	19	GEA
21	Chavez-Galarza et al., 2013	*Apis mellifera iberiensis*	434	73	668	23	FST, GEA
22	Keller et al., 2018	*Populus balsamifera L.*	279	124	925	83	FST, GEA
23	Funk et al., 2016	*Urocyon littoralis*	2498	176	103	6	FST, GEA
24	Mosca et al., 2016 *P.cembra*	*Pinus cembra*	443	85	662	18	FST, GEA
25	Candy et al., 2015	*Thaleichthys pacificus*	3725	157	441	11	FST
26	Dallaire et al., 2021	*Salvelinus alpinus*	13596	732	545	23	FST, GEA
27	Hess et al., 2013	*Entosphenus tridentatus*	4439	164	513	21	FST, GPA
28	Milano et al., 2014	*Merluccius merluccius*	380	30	849	19	FST, GEA
29	Swaegers et al., 2015	*Coenagrion scitulum*	3470	566	161	10	FST, GEA
30	Moore et al., 2014	*Salmo salar*	3192	374	1079	50	FST
31	Mahony et al., 2020	*Pinus contorta*	26431	460	243	22	GPA
32	He et al., 2016	*Banksia attenuate*	5701	1049	80	9	FST
33	Li et al., 2021	*C. davidi* , *C. kishinouyei* , *C. longibarbatus*	5834	258	183	9	FST
34	Cullingham et al., 2014	*Pinus banksiana*	361	25	100	4	FST, GEA

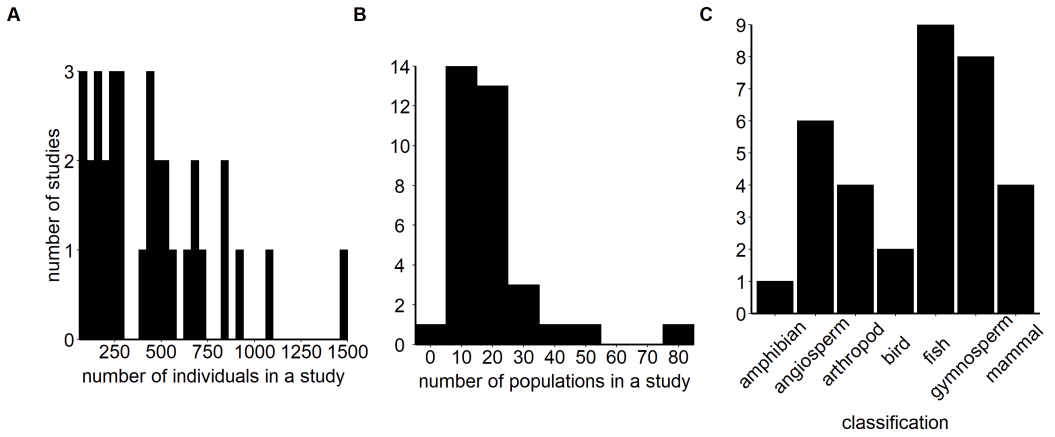

Figure 3.1: Distribution of studies showing the total number of individuals (A) and populations (B) in each study, along with their classification by taxa (C). Each study in the distribution is filtered according to the criteria, thresholds (outlined in section 3.2.1), and 10% missing rate cut-off (n= 34).

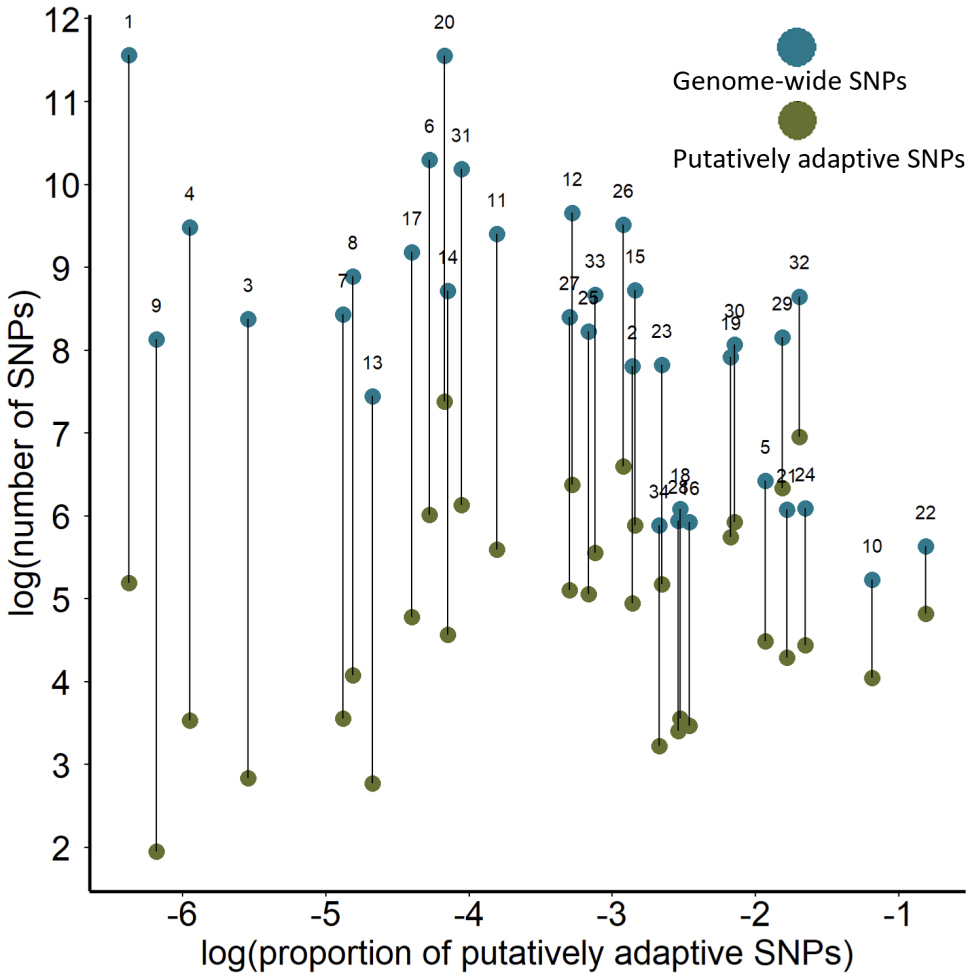

Figure 3.2: Log distribution of genome-wide (total) number of SNPs (blue) and putatively adaptive SNPs (green) for each study. The black line, for each study, shows the difference between the genome-wide (total) number of SNPs (blue) and putatively adaptive SNPs (green) indicating information regarding the proportion of putatively adaptive SNPs compared to the genome-wide (total) number of SNPs. Each study in the distribution is filtered according to the criteria, thresholds (outlined in section 2.2.1), and 10% missing rate cut-off (n = 34).

3.2.2 Data organization

File Formats

I retrieved genome-wide and putatively adaptive SNPs datasets online from Dryad, GitHub, main paper tables, supplemental files and/or files provided by the study authors upon request via email. I contacted authors via email and followed up regarding data files and/or analyses details. The genome-wide and/or adaptive SNPs data files were formatted in VCF, TXT, XLSX, GDS, and POPGEN and allele format varied from alphabetical to numerical

with haploid and diploid data. I used Notepad++, excel, 7-zip, Beluga and Cedar Compute Canada clusters (operated by the Digital Research Alliance of Canada) to view, unzip, or run analysis on big data SNP files. For each of the 34 studies, I performed the outlined analysis (see Figure 3.3 and subsections 3.2.2 to 3.2.3) separately for each type of SNPs dataset: genome-wide SNPs, putatively adaptive subset of genome-wide SNPs, and random subset of genome-wide SNPs (same number as putatively adaptive SNPs sampled from 1000 bootstrap runs).

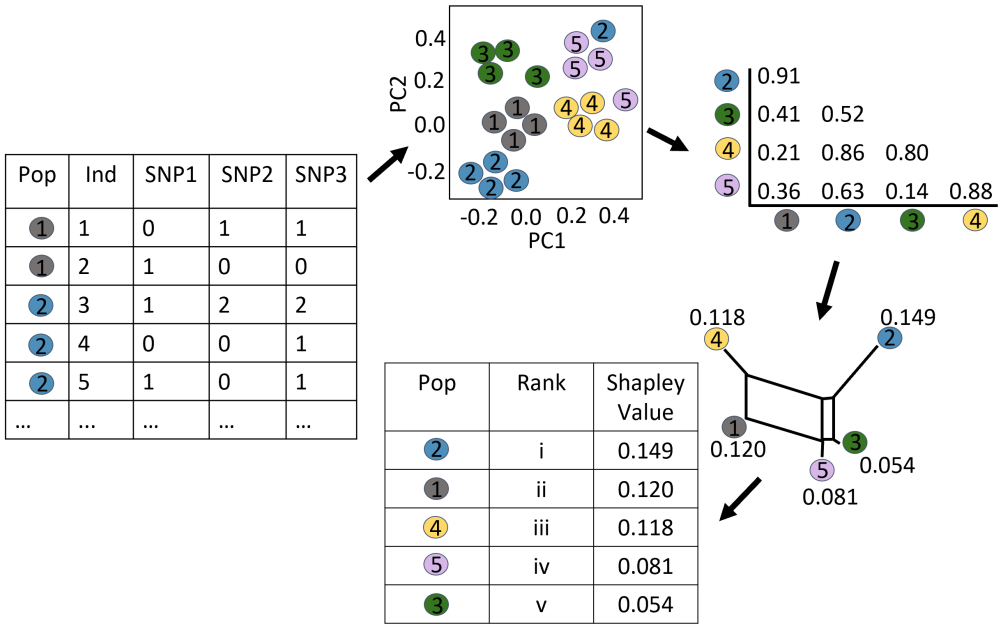

Figure 3.3: Overview of the analytical approach taken for data organization and data analysis. For each study and set of SNPs (genome-wide, putatively adaptive, random set of genome-wide SNPs (same number as putatively adaptive SNPs sampled randomly from 1000 bootstrap runs)), I obtained a SNP dataset for each individual attached to a population, and performed principal components analysis to measure pairwise genetic differentiation among populations. Split information from NeighborNet was used to calculate evolutionary distinctness scores (Shapley Values). Populations were ranked based on their distinctness scores. Then, I calculated the observed correlation (correlation between distinctness scores using genome-wide and putatively adaptive SNPs) and compared it with expected correlation (correlation between distinctness scores using genome-wide and random subset of genome-wide SNPs the same size as the adaptive set).

Standard File Format

I first converted the allele format of all files into a standard flat file format of 0,1,2,NA file where numbers (0,1,2) in each SNP column represented the copies of minority allele and NA represented missing data. This conversion was done in R version 4.1.1 (R Core Team, 2021 [280]) using tidyverse [308], dplyr [309], and adegenet [152, 153] packages prior to

conducting further analysis. The minority allele refers to an allele that is lower in frequency compared to a major allele in a SNP column. When there were multiple minor alleles in a SNP column, I recorded the majority allele as 0 and all other alleles as minority alleles. In each SNP column, zero represented zero copies of minor alleles (i.e., two copies of major alleles), one represented one copy of a minor allele (i.e., heterozygous individuals with one copy of major allele and one copy of minor allele) and two represented two copies of minor alleles. NA indicated missing data. SNP columns that had equal numbers of both alleles were randomly assigned 0 and 2. In some files where individuals were pooled (i.e., pool seq files), for each SNP column, we assigned 0 to the major alleles and 2 to the minor alleles and NA to missing alleles. In cases where haploid data was provided for individuals, for each SNP column, I assigned 0 to the major allele and 1 to the minor allele.

Missingness threshold

Different study authors accounted for missing data using different thresholds. While some studies imputed missing data using, e.g. linkImpute version 1.1.4 [203], others used different missingness thresholds for missing alleles per SNP and missing SNPs per sample. I chose to follow the missingness threshold of 10% because SVDImpute, used for imputing missing data in PCAMethods [273], is said to be tolerant of 10% missingness data [272, 286]. The 10% missingness threshold meant that I was removing missing alleles per SNP and missing SNPs per sample at the same time that have >= 10% missing values. If the number of individuals per population after accounting for the 10% missing rate (i.e., samples and SNPs) fell below four, I removed those populations at this stage. If studies did not require any 10% missingness cutoff, I still removed populations within each study with fewer than four individuals per population. I did this for both genome-wide and putatively adaptive SNPs datasets separately for each study.

3.2.3 Data analysis

Principal Components Analysis (PCA)

To ordinate all individuals in a multi-dimensional space, I performed a principal components analysis [227, 145] on the genetic variants of individuals. I did this for each study and each type of SNPs dataset. The main goal was to take the centroids of each population from PCA to calculate the Euclidean pairwise genetic distances among populations separately for genome-wide, putatively adaptive, and the random subset of genome-wide SNPs datasets. I chose PCA because it is one of the most commonly used dimensional reduction techniques in GWAS that assumes no population genetic model and is not constrained by many assumptions of population genetics [169]. PCA simply gives the coordinates to individuals in a graph to examine genetic distances. In this context, I used the same assignment of populations as authors identified them in their original studies. For each study,

datasets in a standard file format following cleaning (10% missingness threshold and 4 or more individuals per population) were used as input files for PCAmethods. Input data for PCAmethods included a numerical data matrix with individuals (i.e., samples) in rows and SNPs as columns [271]. Each SNP column in the input data was centered and scaled using the unit variance scaling method. Unit variance is

$$a = \frac{a}{\sigma_a}$$

, where a is variable (i.e., each SNP column) [271]. SVDImpute uses an algorithm to iteratively estimate linear combinations of a specific number of the most significant loadings by using the current estimate to predict the next estimate until the missing value estimate is below a certain threshold value [272, 286]. If a certain value is missing at a certain position in the matrix, then that value is not used to estimate the linear regression coefficient [272]. For each column, all the missing values are substituted with the mean of that SNP column. Stacklies and Redestig 2022 [272] suggest that SVDImpute appears to accept >10% missing data. For further analysis, I chose significant principal components (PC) using the common tracy-widom test [285] with a critical point of 0.9793, equivalent to a p-value of 0.05 (see [223, 222, 188]). Importantly, I didn't use all principal components because single SNPs may be heavily impacted by extended linkage disequilibrium patterns [85, 296, 319], and this can generate nuisance axes of principal components at lower eigenvalues, axes that capture localized LD instead of population structure [320, 319]. If I had zero principal components as significant, I simply used the first PC. I further used a threshold of PC 1 scores average +/- 4*standard deviation of PC1 scores for the identification of outlier individuals. I did this for Chavez-Galarza et al., 2013 [49] , Mosca et al., 2016 (P. mugo) [208], and Benestan et al., 2016 [29] studies. Outlier individuals were identified by using principal components 1 and 2.

Pairwise distances

After ordinating individuals using a principal components analysis with apriori identified populations, I calculated the Euclidean distance between the centroids of each population pair in a pairwise manner (following Fernandez-Fournier et al., 2021 [86]). The matrix of pairwise genetic distances among populations was used to build a split network ([150] using NeighborNet ([42]) via the phangorn package [255, 252] in R.

NeighborNet Split Networks

A split network refers to a set of splits with non-negative weighted edges depicting relationships among taxa [150, 1]. Splits network using NeighborNet takes a distance-based approach [43] to graph relationships among populations via pairwise distances and provides

a single network, accommodating for both tree-like and non-tree like patterns of relatedness [42, 43].

In short, NeighborNet takes the pairwise genetic distance between any two populations and represents that information in the weight of each split (i.e., edge length of the branch) [291]. A split partitions one taxon from all other non-overlapping taxa [43, 150]. The NeighborNet network is simply an assembly of these weighted splits of a set of populations [291].

Output from the phangorn package in R was compared to output from the Splitstree program written by the originators of NeighborNet for the Bay et al., 2018 [22], Mahony et al., 2019 [190], and Cullingham et al., 2014 [65] datasets using the SplitsTree Community Edition (CE) version 6.0.23-beta developed by Daniel H.Huson and David Bryant and phangorn version 2.11.1 (via ape) [254, 253]. Results were identical, and so phangorn was used due to its ease of integration with the R pipeline.

Shapley Value

Biologically, Shapley Value is the "expected contribution of a [population] to future subsets" of populations in unknown future scenarios [291]. It is a "measure of the contribution each [populations] brings to the diversity of a [species]" [116] and is equivalent to the Edge of Existence (EDGE of Existence :: Evolutionarily Distinct and Globally Endangered) evolutionary distinctiveness measure when calculated on a bifurcating tree. The most genetically distinct populations harbor a disproportionate amount of unique evolutionary history.

Shapley Value, which can be extended to split networks [192], is measured by averaging phylogenetic diversity complementarity of a population over all possibilities of equally weighted subset sizes [116, 291]. The sum of all split weights (branch lengths) in the NeighborNet network of underlying populations considering the minimum spanning path represents phylogenetic diversity (PD) [84], which is a conservation planning metric used to quantify evolutionary distinctness [291]. PD complementary measures the contribution of a population tip to the network by calculating the branch length (or pendant edge weights) connecting that population tip to the rest of the network [291]. Mathematically, equation (1) depicts the average of split weights from a set representing x|X, where |X are all unique possibilities of showing subsets of population X without including x [291].

$$\psi_x^{sh}(\sum, \lambda) = \sum_{S \in \sum} \frac{|\bar{S}(x)|}{|X| \cdot |S(x)|} \lambda(S)$$

[116, 192] where $(\sum, \lambda)$ is the total split set and their weights, |X| is the number of populations in total, |S(x)| is the split set size that includes the population x, $|\bar{S}(x)|$ is the complementary split set size that excludes the population x, and (S) is split weight that partitions S(x) from $\bar{S}(x)$ [291].

For each study and dataset type (i.e., genome-wide, putatively adaptive, or random subsets of genome-wide SNPs), I standardized Shapley values by dividing the raw Shapley values across all populations by the total sum of those raw Shapley Values to correct for the number of SNPs.

Population rankings

Based on the standardized Shapley Values, I ranked populations within each study for conservation prioritization using genome-wide SNPs, putatively adaptive, and random subsets of genome-wide SNPs. So, higher ranked populations are more distinct with higher Shapley Values.

Observed and expected correlation

The main goal is to compare the Spearman rho rank correlation across studies of Shapley distinctness scores using different sets of SNPs: genome-wide SNPs and putatively adaptive SNPs (which I refer to as the observed correlation) and genome-wide and random subset of genome-wide SNPs (which I refer to as the expected correlation). To calculate expected correlation for each study, I randomly sampled a subset of genome-wide SNPs that is the same number as putatively adaptive SNPs and tested the Spearman rho correlation between distinctness scores of ranked populations using genome-wide SNPs and a random set of genome-wide SNPs. I used the same procedure outlined above (in Figure 3.3) and repeated it 1000 times using bootstrap.

To calculate the 90% bootstrapped confidence interval of spearman correlation for each study, I used the 'spearman.ci' function in the "RVAideMemoire" package [133] using 1000 reps in R. I chose a Spearman correlation test because it is a non-parametric test that accounts for similar proportion of variability as a pearson correlation test [125] and is "more appropriate when there is less certainty about the reliability of close ranks" as it gives "greater weight to pairs of ranks that are further apart" whereas "Kendall weights each disagreement in rank equally" [268].

Statistical Tests

Our main goal is to address the question of how conservation prioritization differs when using genome-wide SNPs vs. subsets of SNPs that are putatively adaptive across 34 studies. To address this, we asked whether genome-wide (all) SNPs and putatively adaptive SNPs provide similar population prioritizations across 34 studies and whether putatively adaptive SNPs behave like equal-sized samples of a random set of all SNPs.

For the first question, I calculated the Spearman correlation coefficient of Shapley value distinctness scores using genome-wide SNPs and putatively adaptive SNPs for each study. The positive or negative signs of this observed correlation coefficients for each study and

the spread of correlation coefficient values across studies would indicate the extent to which genome-wide variation and adaptive variation seem to prioritize similar populations as distinctive (I refer to this genome-wide and putatively adaptive SNPs correlation as prioritization agreement). A higher positive observed correlation coefficient indicates a higher population prioritization agreement between genome-wide and putatively adaptive SNPs.

I compared the observed Spearman correlation of Shapley values calculated using genome-wide and putatively adaptive SNPs to the expected Spearman correlation calculated using genome-wide and random subsets of genome-wide SNPs for each study (I refer to this observed minus expected correlation as the predictive power of putatively adaptive SNPs) (see 2.2.3.6). The random subsets were of the same size as the subset of putatively adaptive SNPs: this tests whether putatively adaptive SNPs behave like an equal-sized sample of a random set of genome-wide SNPs.

I performed a two-tailed paired t-test to test for a difference between observed and expected Spearman correlation coefficients of each study (i.e., predictive power of putatively adaptive SNPs). If there is a statistically significant difference between the observed and expected Spearman correlations, this would mean that putatively adaptive SNPs do not behave like equal-sized samples of random SNPs. For interest purposes, I also performed a two-tailed paired t-test to test for a difference between observed and expected Kendall correlation coefficients of each study. I first tested for outliers using 'identify_outliers' function in rstatix package [158]. I state all results with and without removing outlier studies above Q3 + 1.5xIQR or below Q1 -1.5xIQR and extreme outlier studies above Q3 + 3xIQR or below Q1 - 3xIQR (if there are any) [158].

To further examine the predictive power of putatively adaptive SNPs, I compared allele frequency patterns for the putatively adaptive and equal sized random sets of genome-wide SNPs. To do this, I calculated the minor allele frequencies for adaptive and random sets of genome-wide SNPs (and also neutral sets of genome-wide SNPs after putatively adaptive SNPs were removed) for each study. A single draw of an equal number of these "random" SNPs as the number of adaptive SNPs were sampled using a set seed of 12. For each study, I used a Mann-Whitney U test to compare the median allele frequencies of random and adaptive SNPs and where applicable the median allele frequencies of neutral and adaptive SNPs. For four studies [91, 278, 66, 204], the SNP names in genome-wide and adaptive SNPs files didn't match, so due to this mismatch of names, I couldn't separate adaptive SNPs from non-adaptive SNPs only for the purposes of this particular analysis. Finally, to test whether the average median minor allele frequencies of adaptive subsets of genome-wide SNPs is greater than the average median minor allele frequencies of random/neutral subsets of SNPs across studies, I used a one-tailed paired t-test. To test if the difference between adaptive and random/neutral allele frequencies impact the predictive power of putatively adaptive SNPs, I did a spearman correlation between adaptive minus random median minor allele frequencies and observed minus expected correlation.

Tests of covariates

I tested the individual effects of six covariates on the prioritization agreement between putatively adaptive and genome-wide SNPs and on the predictive power of putatively adaptive SNPs: study-wide F_{ST}, correction for neutral population structure, number of genome-wide SNPs, number of putatively adaptive SNPs, proportion of putatively adaptive SNPs, and the method type (F_{ST} vs. non-F_{ST}) used to identify putatively adaptive SNPs. To examine the impact of continuous variables (study-wide F_{ST}, number of genome-wide SNPs, number of putatively adaptive SNPs, proportion of putatively adaptive SNPs), I used Spearman correlation between the observed correlation and variable of interest and observed minus expected correlation and variable of interest. To examine the effect of categorical variables (neutral population structure and method type- F_{ST} vs. non- F_{ST}), I used one-way anova. P value < 0.05 indicates a nominally statistical significant effect, and I report FDR corrected P value for multiple tests.

I examined the effect of study-wide F_{ST} because if a population is highly structured or if there is high differentiation among populations (i.e., high variance in Shapley values) it may be easier to detect the genes involved in local adaptation [37, 250]. This leads to the hypothesis that cases where study-wide F_{ST} is greater are also the ones where genome-wide SNPs and putatively adaptive SNPs may be less likely to prioritize similar sets of distinct populations. In other words, I expected a negative correlation between study-wide F_{ST} and prioritization agreement, and between study-wide F_{ST} and predictive power of putatively adaptive SNPs. I measured study-wide F_{ST} [301] using the hierfstat package [109, 302] in R version 4.1.1 (R Core Team, 2021 [280]) and log-transformed it. I used guidelines found at [257] to calculate study-wide F_{ST}.

I considered whether studies accounted for neutral population structure or demography in some way because non-independence among populations due to coancestry (reflecting shared population history or kinship effects) [35], spatial autocorrelation (ex. isolation-by-distance, environment or more), sharing common gene flow history or demographic history (see [183, 138, 37]) could all impact the correlation between genome-wide loci and loci under selection, inflating false-negatives as well as false-positives depending on the scenario [3, 183, 114]. Some types of analysis estimate and correct for evolutionary non-independence among samples during the process of identification of adaptive SNPs whereas others don't ([183, 65, 80, 26]; see Appendix A Table for more details). Based on literature, I attempted to score whether each type of analysis used to identify putatively adaptive SNPs across 34 studies accounted for the neutral population structure or demography. I categorized them into three groups: no, mixed, and yes, where no meant no tests accounted for it, mixed meant some tests did and yes meant all tests accounted for population structure in some way.

I log-transformed number of genome-wide SNPs and number of putatively adaptive SNPs and examined their impact on the prioritization agreement and the predictive power of putatively adaptive SNPs because it reflects aspects of sampling strategies that affect the power to differentiate adaptive from non-adaptive loci [3]. The number of genome-wide SNPs may impact various factors such as the proportion of putatively adaptive SNPs or false detection rate, or reflect aspects of species' genome-size [3] that could influence the genome-wide and putatively adaptive SNPs correlation. I had no prior prediction regarding the impact of these two covariates on the correlations but if these variables do impact the correlations, this might indicate whether caution should be applied with the interpretation of results as a function of marker number [3].

I chose to examine the impact of the proportion of putatively adaptive SNPs on the prioritization agreement and the predictive power of putatively adaptive SNPs because putatively adaptive SNPs are a subset of genome-wide SNPs, and so I expected a part-whole correlation. The proportion of putatively adaptive SNPs was calculated by dividing the number of genome-wide SNPs by the number of putatively adaptive SNPs and then log-transformed this ratio.

The impact of method-type (F_{ST} vs. non-F_{ST}) used to identify putatively adaptive SNPs on the prioritization agreement and predictive power of putatively adaptive SNPs was tested because of the suspicion that F_{ST} methods might produce more false positive adaptive SNPs [269], which could inflate agreement. I grouped F_{ST}, F_{ST} and GEA, F_{ST} and GPA into the F_{ST} category and GEA, GPA, GEA and GPA into the non-F_{ST} category.

I used beta regression to test the impact of proportion of putatively adaptive SNPs, and method type (F_{ST} vs. non-F_{ST}) used on the Spearman correlation between genome-wide and adaptive SNPs distinctness scores (i.e., the prioritization agreement). I did a t-test on the beta-regression model to produce nominal p-values. Since the Spearman correlation values were bounded between 0 and 1 and simple linear regression doesn't deal with the bounded nature because of violation of assumptions, so I applied beta-regression. I removed one study (Cullingham et al., 2014 [65]) when doing beta regression with the Spearman correlation as the response variable and proportion of putatively adaptive SNPs and method type as explanatory variables because it was an outlier residual in residual vs. linear predictor plot and generalized leverage vs. predicted values plot.

I used linear regression to test the impact of study-wide F_{ST} and the number of putatively adaptive SNPs used on the predictive power of putatively adaptive SNPs. I used anova to test the significance and p-values. All linear regression assumptions were met as values were restricted between 0 and 1.

3.3 Results

I set out to answer three clear questions with implications for the potential use of genetics in population prioritization in conservation: Do genome-wide (all) SNPs and adaptive SNPs provide similar population prioritizations? Do adaptive SNPs behave like equal-sized samples of random SNPs? How do covariates impact the prioritization agreement and the predictive power of putatively adaptive SNPs?

3.3.1 Do genome-wide (all) SNPs and putatively adaptive SNPs provide similar population prioritizations?

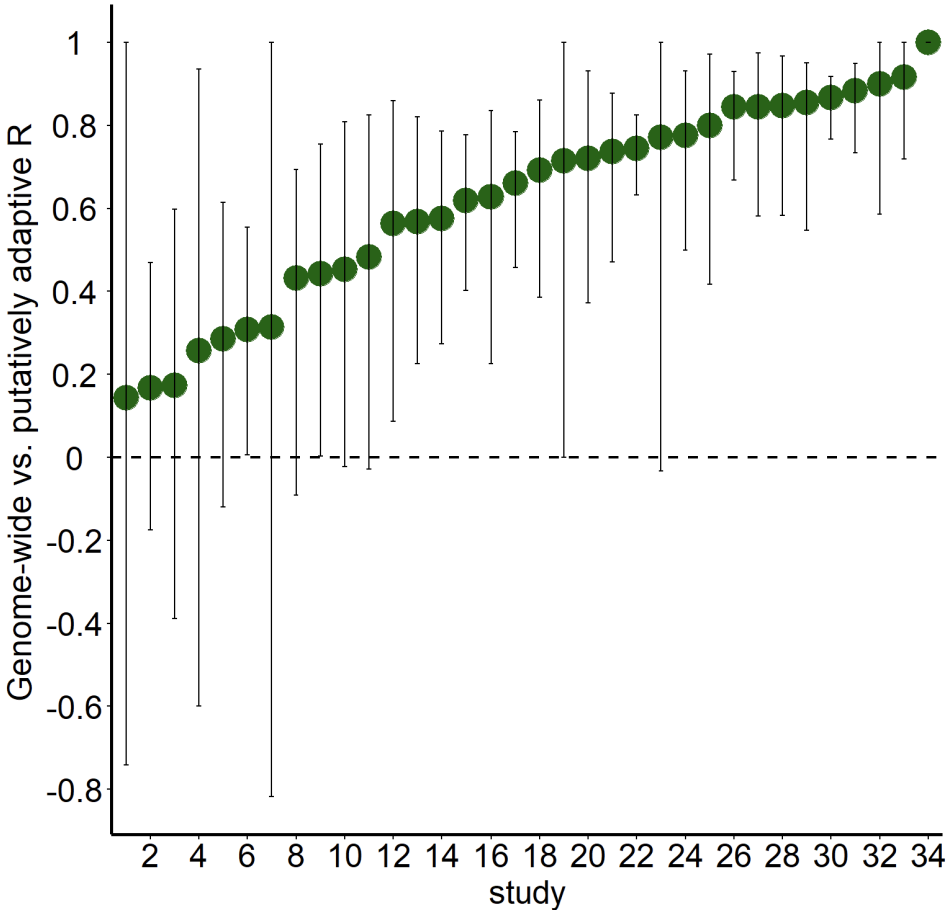

Figure 3.4: Prioritization agreement (i.e., observed Spearman correlation) of population distinctness scores calculated using each set of SNPs (genome-wide and putatively adaptive SNPs) indicated for each study (green points) (n=34). A more positive y-value indicates a higher prioritization agreement meaning that both sets of SNPs identify similar populations as distinct within a study. Black error lines indicate 90% bootstrapped confidence intervals.

I found that genome-wide (all) SNPs and putatively adaptive SNPs provide similar population prioritizations to some extent. The prioritization agreement (observed Spearman

correlation) coefficients of population distinctness scores using genome-wide and adaptive SNPs are all positive (Figure 3.4). However, these prioritization agreement coefficients vary a lot (from R +0.14 to +1.00) from across studies (Figure 3.4).

3.3.2 Do adaptive SNPs behave like equal-sized samples of random SNPs?

Overall, I found that the predictive power of putatively adaptive SNPs differs (Figure 3.5). In other words, adaptive subsets of genome-wide SNPs do not seem to behave the same as equally-sized samples of random subsets of genome-wide SNPs in terms of their correlation with genome-wide SNPs. A two-tailed paired t-test shows a significant difference between the observed and expected Spearman correlation; $t(33) = 2.45$, $p = 0.02$. The mean difference of the observed minus expected Spearman correlation coefficients is 0.09 (95% CI: 0.02 to 0.17). I found the mean difference of the observed minus expected Kendall correlation coefficients is 0.07 (95% CI: -0.0008 to 0.1340); $t(33) = 2.01$, $p = 0.05$. Here, I did not remove any outliers.

When I removed three outlier studies (Cullingham et al., 2014 [65]; Benestan et al., 2016 [30], and Roffler et al., 2016 [242]) that fell above the Q3 + 1.5xIQR or below Q1 -1.5xIQR from the Spearman correlation analysis, I observed a similar result: there is a significant difference between the observed minus expected Spearman correlation; $t(30) = 2.48$, $p = 0.02$ with a mean difference of 0.08. When I removed two outlier studies (Cullingham et al., 2014 [65] and Roffler et al., 2016 [242]) above Q3 + 1.5xIQR or below Q1 -1.5xIQR from Kendall tau correlation analysis there is, again, a significant difference between the observed minus expected Kendall correlation; $t(31) = 2.23$, $p = 0.03$ with a mean difference of 0.06.

The positive average difference between observed minus expected correlation from the two-tailed paired t-test suggests that adaptive SNPs subsets may show stronger correlation with genome-wide SNPs than random SNPs subsets do: as shown in Figure 3.5, approximately 2/3 of the time the observed correlation (green dots) lie above the expected correlation (orange dots).

Tom Booker (pers. comm.) hypothesized that one explanation for adaptive subsets seeming to correlate more strongly than random subsets with genome-wide SNPs is that the adaptive SNPs have higher allele frequencies on average than random or neutral subsets of genome-wide SNPs. Adaptive SNPs might have higher frequencies on average because they are involved in local adaptation and are favored by selection (especially positive selection) [117, 247]. If so, this could explain the predictive power of putatively adaptive SNPs because they would be less noisy compared to random subsets of (neutral) SNPs. If adaptive and random allele frequencies (patterns) are more exaggerated then it is easier to detect local adaptation (i.e., adaptive SNPs) [178, 25, 138], and this might also improve predictive power of putatively adaptive SNPs from noise.

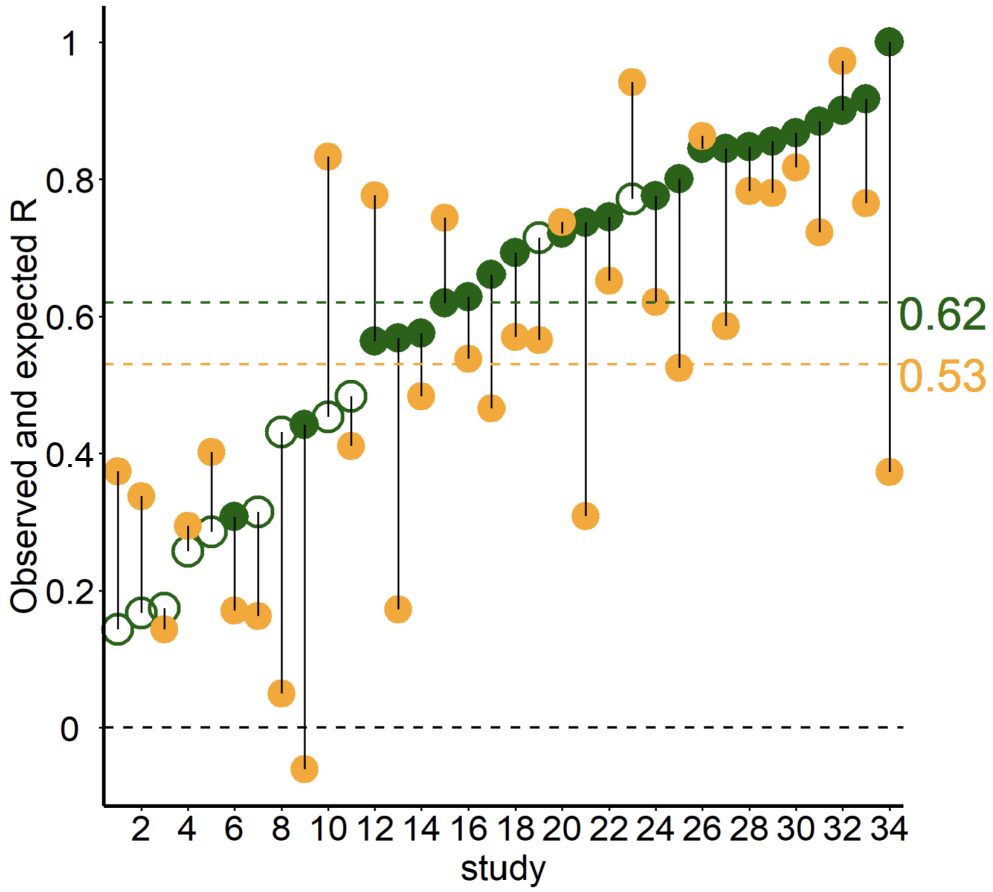

Figure 3.5: Observed (green) and expected (orange) Spearman correlation of population distinctness scores indicated for each study (n=34). Observed Spearman correlation coefficients of population distinctness scores are calculated using genome-wide and putatively adaptive SNPs and expected Spearman correlation coefficients are calculated using genome-wide and random subset of genome-wide SNPs (same number as putatively adaptive SNPs sampled from 1000 bootstrap replicates). The vertical black line is indicating the difference between the observed and expected Spearman correlation coefficients for each study (indicating the predictive power of putatively adaptive SNPs). The average observed Spearman correlation coefficient across all 34 studies is 0.62 (indicated by green dashed line) and the average expected Spearman correlation coefficient across all 34 studies is 0.53 (indicated by yellow dashed line). The hollow green points represent cases where 90% bootstrapped confidence intervals of observed Spearman correlation overlap zero.

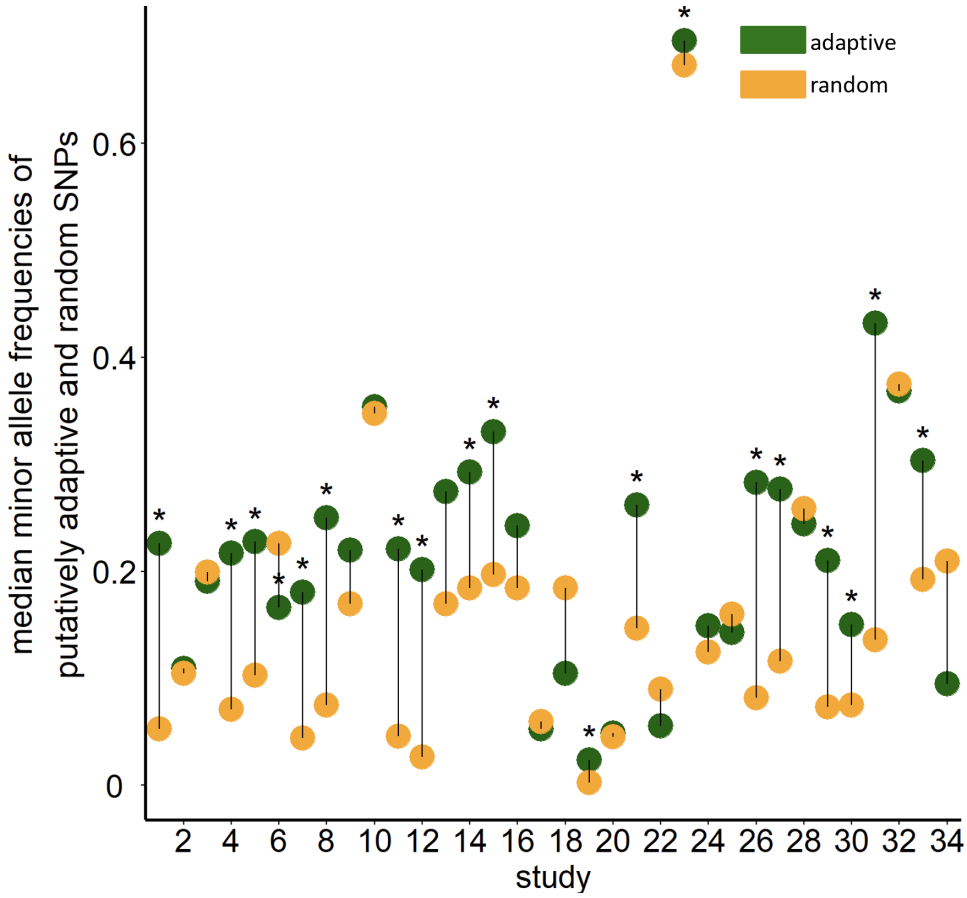

Figure 3.6: Median minor allele frequencies of putatively random (orange) and adaptive (green) SNPs for each study (n = 34). The black line is indicating the difference between median allele frequencies of adaptive and random subsets of genome-wide SNPs for each study. Asterisks for 19 out of 34 studies represent statistical significant differences between median minor allele frequencies of adaptive and random subsets of genome-wide SNPs indicated via Mann-Whitney U test.

I did find an average difference between neutral/random minor allele frequencies and adaptive minor allele frequencies across studies. I found that there is a difference (however small) in all of them (Figure 3.6). Using a Mann-Whitney U test, I found that 19 out of 34 studies showed differences of minor allele frequencies between random and adaptive subsets of genome-wide SNPs (Figure 3.6) and in 17 out of 30 studies there is a significant (across SNPs) difference in the minor allele frequencies of neutral and adaptive subsets of genome-wide SNPs, always (except one) with adaptive SNPs being at higher frequency. I found the mean difference between the median allele frequencies of adaptive SNPs and random SNPs across studies to be 0.07 (95% CI: 0.04 to Inf); $t(33) = 4.42$, $p = 5.00\text{x}10^{-5}$. I found the mean difference between the median allele frequencies of adaptive SNPs and neutral SNPs across studies to be 0.07 (95% CI: 0.04 to Inf); $t(29) = 4.16$, $p = 0.00013$.

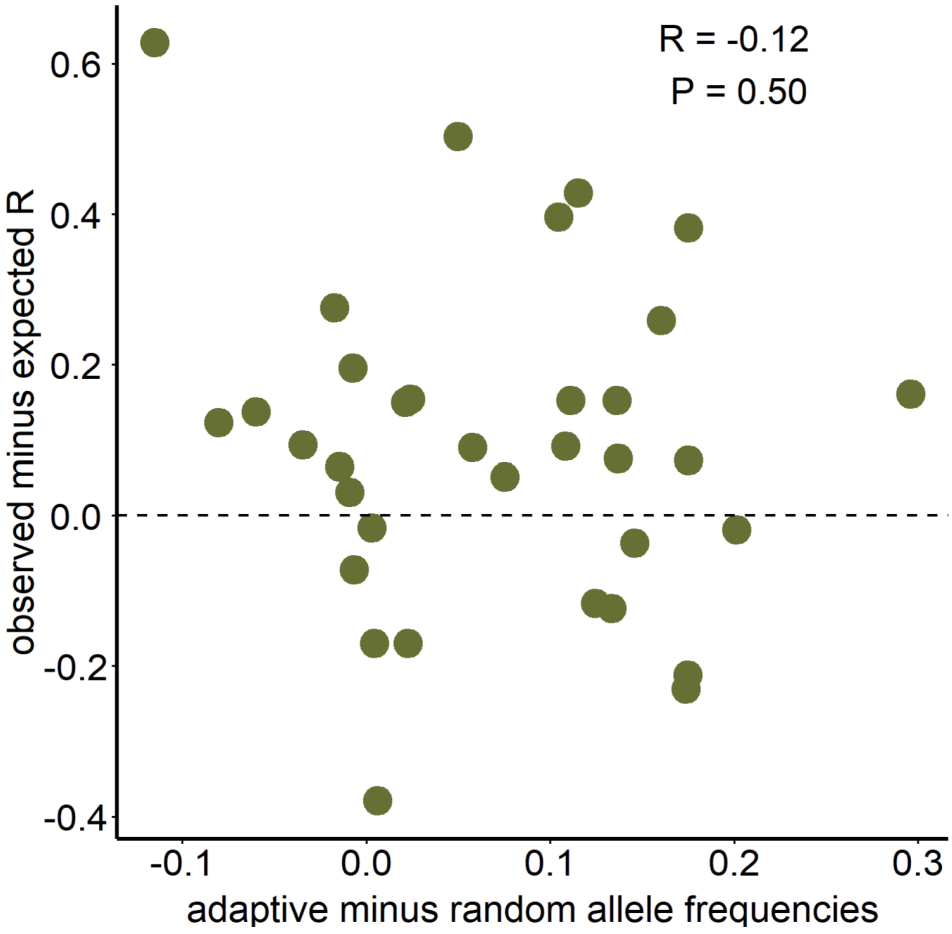

Figure 3.7: Scatterplot showing the predictive power of putatively adaptive SNPs (i.e., observed minus expected R) on y-axis and median adaptive minus random allele frequencies for each study (n= 34) on x-axis. There is no significant correlation between the two (Spearman rho = -0.12 with P = 0.50).

While I found statistical differences between adaptive and random/neutral median minor allele frequencies within and across studies, contrary to our expectation, there is no correlation between adaptive minus random median minor allele frequencies and the predictive power of putatively adaptive SNPs (observed minus expected correlation) (Spearman rho = -0.12, P = 0.50) (Figure 3.7).

3.3.3 How do covariates impact the prioritization agreement and the predictive power of putatively adaptive SNPs?

Table 3.2: Impact of six covariates on the prioritization agreement and the predictive power of putatively adaptive SNPs. Asterisk (*) indicates a P < 0.05 (uncorrected for FDR)

Covariates	Prioritization agreement	Predictive power of putatively adaptive SNPs
log(study-wide F_{ST})	R = -0.01; P = 0.94	R = -0.46, P = 0.007*
Account for neutral population structure or demography	P = 0.56	P = 0.45
log(number of genome-wide SNPs)	R = -0.20, P = 0.26	R = -0.17, P = 0.34
log(number of putatively adaptive SNPs)	R = 0.32, P = 0.06	R = -0.37, P = 0.03*
log(proportion of putatively adaptive SNPs)	R = 0.50, P = 0.003*	R = -0.17, P = 0.35
Method Type (F_{ST} vs. non-F_{ST})	P = 0.02*	P = 0.38

I tested the individual effects of six covariates – study-wide F_{ST}, account of neutral population structure, number of genome-wide SNPs, number of putatively adaptive SNPs, proportion of putatively adaptive SNPs, and method type (F_{ST} vs. non-F_{ST}) used to identify putatively adaptive SNPs – on the prioritization agreement and the predictive power of putatively adaptive SNPs (Table 3.2). To examine the impact of continuous variables (study-wide F_{ST}, number of genome-wide SNPs, number of putatively adaptive SNPs, proportion of putatively adaptive SNPs), I used Spearman correlation between the prioritization agreement (observed correlation) or predictive power of putatively adaptive SNPs (observed minus expected correlation) and variable of interest. To examine the effect of categorical variables (neutral population structure, method type- F_{ST} vs. non- F_{ST}), I used one-way anova. I applied arcsine transformation on the observed Spearman correlation (calculating using genome-wide and putatively adaptive SNPs) when examining how the covariates impact the observed Spearman correlation, and as expected I got the same results with and without transformation.

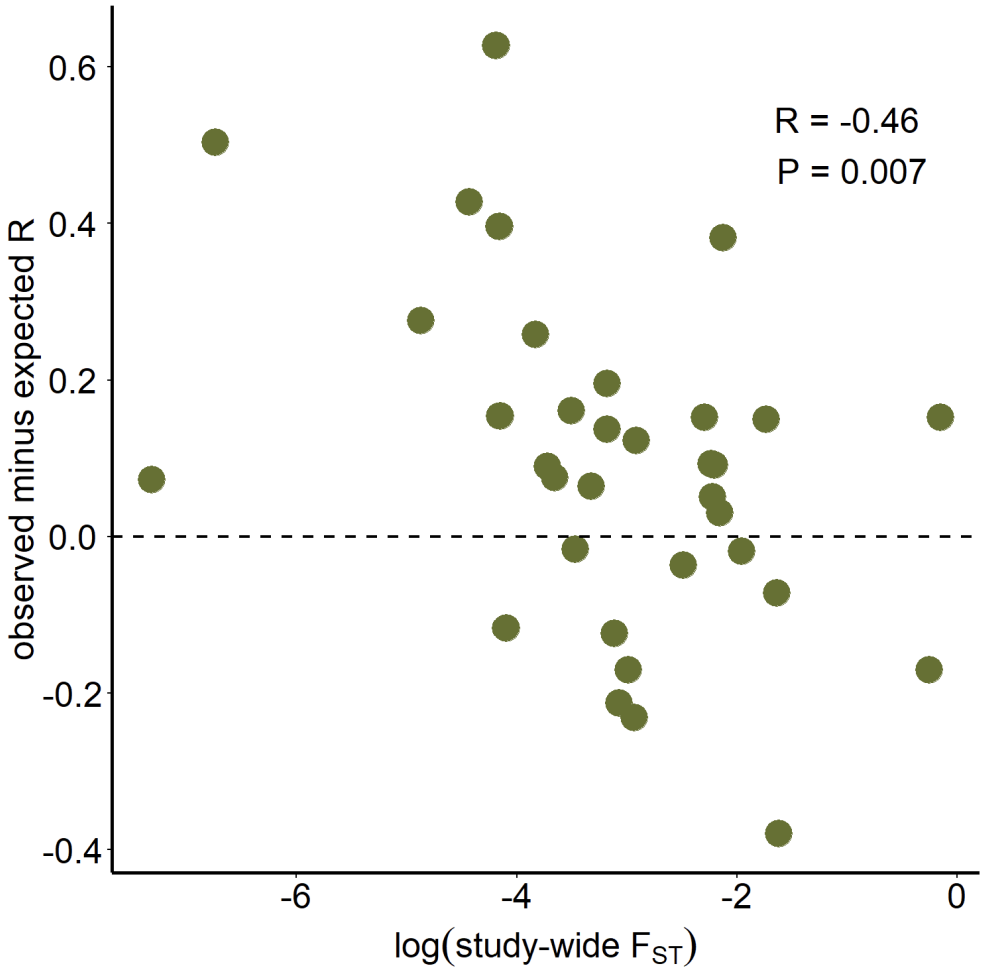

Figure 3.8: Increasing study-wide F_{ST} seems to decrease the predictive power of putatively adaptive SNPs relative to random expectation. Each green point represents a study (n = 34). The Spearman correlation is -0.46 with uncorrected P = 0.007; FDR corrected P = 0.04.

I expected a negative correlation between study-wide F_{ST} and prioritization agreement because if a population is highly structured (i.e., high study-wide F_{ST}), we predicted that it might be easier to detect the genes involved in local adaptation. However, I observed that study-wide F_{ST} had no effect on the prioritization agreement (Spearman rho = -0.01, uncorrected and FDR corrected P = 0.94). However, study-wide F_{ST} did impact the predictive power of putatively adaptive SNPs (Spearman rho = -0.46, uncorrected P = 0.007, FDR corrected P = 0.04), in other words, as study-wide F_{ST} increased, the correlation between genome-wide and adaptive SNPs decreased relative to random expectation (Figure 3.8).

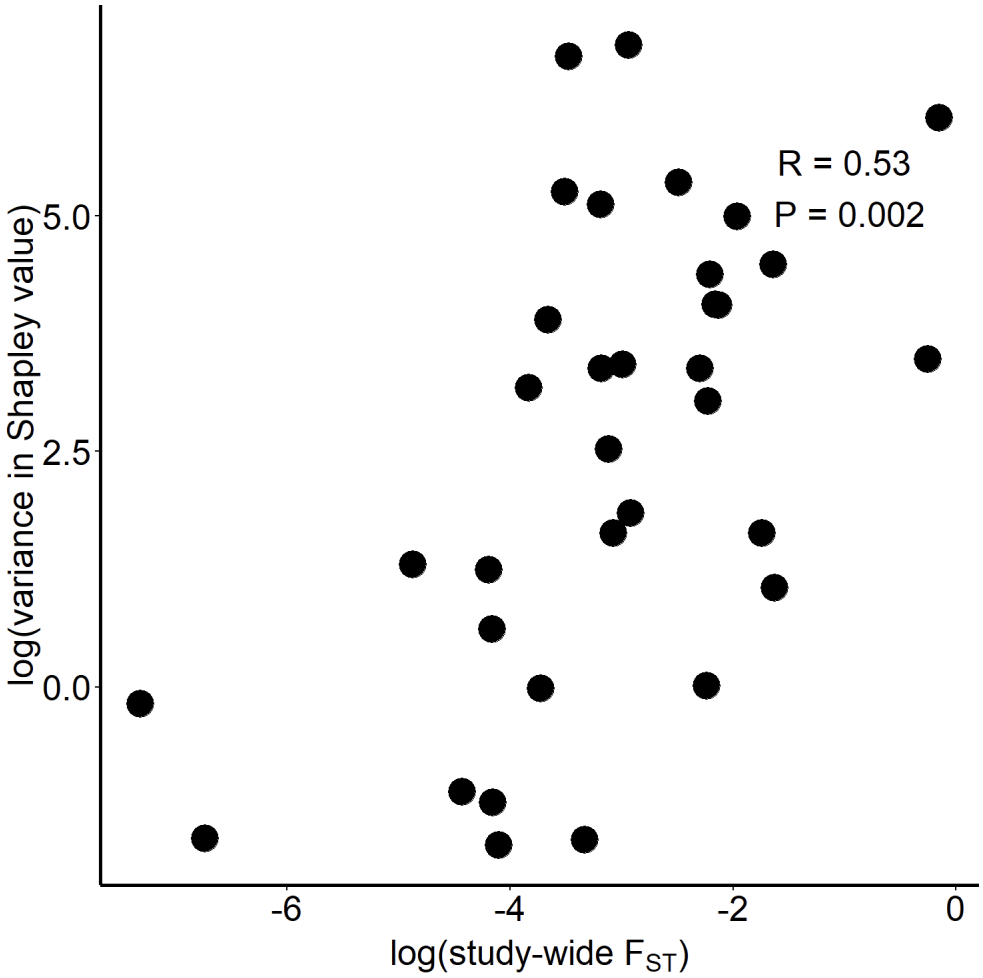

Figure 3.9: Positive correlation between log(study-wide F_{ST}) and log(variance in raw Shapley values) (n = 34) with R = 0.53 and P = 0.002.

I found a positive correlation between log(study-wide F_{ST}) and log (variance in raw genome-wide Shapley values) as expected (Figure 3.9). This means that species with low study-wide F_{ST}, reflecting low population structure show low variance in population distinctiveness (not high variability in genetic distinctiveness of populations), whereas species with high study-wide F_{ST}, reflecting high population structure, show higher variance in population distinctiveness (high variability in genetic distinctiveness of populations). I expected to find a negative relationship between variance in Shapley values and the predictive power of putatively adaptive SNPs, though, I do not find any statistically significant relationship between log of variance of raw genome-wide Shapley values and the predictive power of putatively adaptive SNPs using a Spearman correlation (S = 8154, P = 0.16, rho = -0.25) probably because of small sample sizes of data, I found that the data is suggestive of that expectation.

I did not have any prediction regarding the effect of accounting for neutral population structure on the prioritization agreement as well as on the predictive power of putatively adaptive SNPs because being conservative in accounting for it might increase false-negatives whereas not accounting for it at all might increase false-positives [3], both of which may contribute to increasing the prioritization agreement as adaptive set may contain random SNPs. Further limitations such as unresolved differences among methods, inconsistencies in application of methods preventing any apple-to-apple comparison across studies, and small sample sizes contributed to producing unsatisfactory rankings across studies. Despite these limitations hindering the analysis, I nonetheless attempted to categorize the account for neutral population structure across studies into three groups: no, mixed, and yes, where no meant none of the methods/tests used to identify adaptive SNPs accounted for it, mixed meant some tests did, and yes meant all tests accounted for it in some way. Using one-way anova, I found that the account for neutral population structure did not affect the prioritization agreement (uncorrected $P = 0.56$; FDR correction $= 0.67$) or the predictive power of putatively adaptive SNPs (uncorrected and FDR corrected $P = 0.45$) in the 34 species I analyzed.

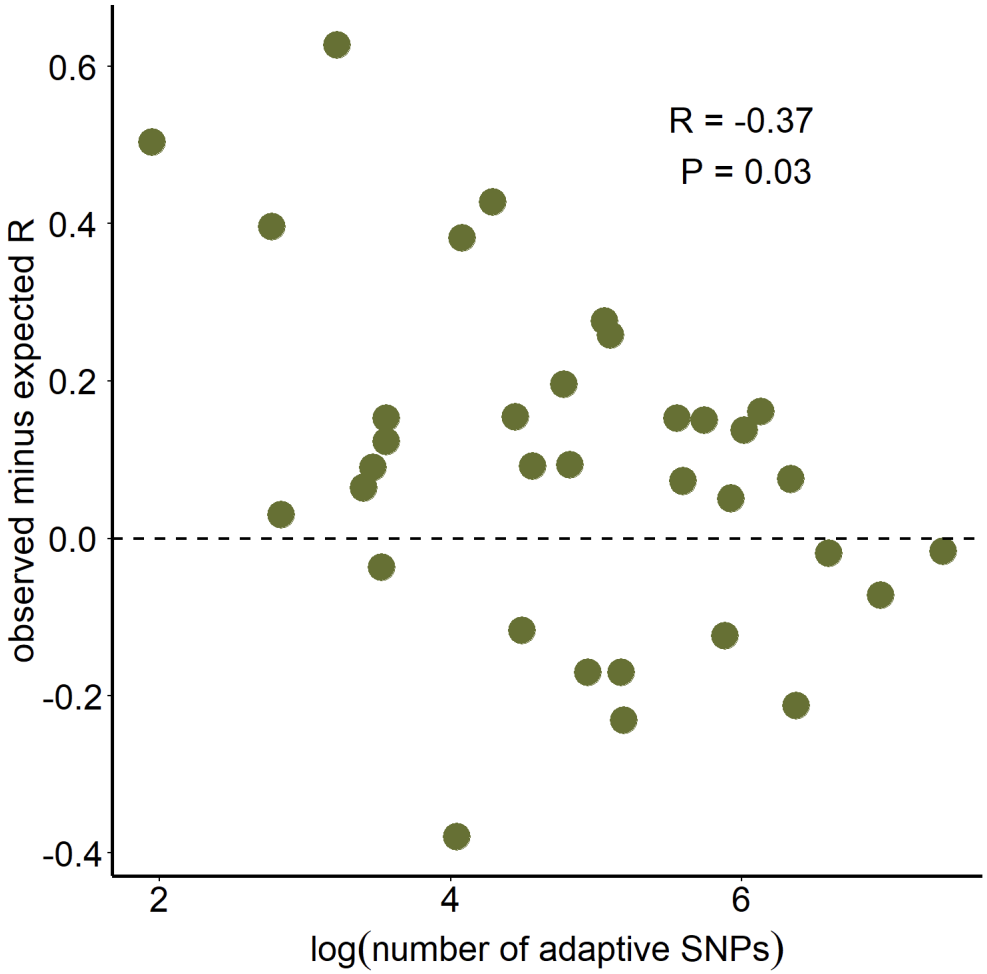

Figure 3.10: Increasing number of putatively adaptive SNPs seem to decrease the predictive power of putatively adaptive SNPs relative to random expectation. Each green point represents a study (n = 34). The spearman rho correlation is R -0.37, uncorrected P = 0.03, FDR corrected P = 0.09.

I examined the impact of the number of genome-wide SNPs and number of putatively adaptive SNPs on the prioritization agreement and the predictive power of putatively adaptive SNPs with no prior predictions. I found that the number of genome-wide SNPs did not impact the prioritization agreement (Spearman rho = -0.20, uncorrected P = 0.26; FDR corrected P = 0.39) and the predictive power of putatively adaptive SNPs (Spearman rho = -0.17, uncorrected P = 0.34, FDR corrected P = 0.45) and the number of adaptive SNPs did not impact the prioritization agreement (Spearman rho = 0.32, uncorrected P = 0.06, FDR corrected P = 0.13). However, the number of adaptive SNPs impacted the predictive power of putatively adaptive SNPs (Spearman rho = -0.37, uncorrected P = 0.03, FDR corrected P = 0.09). When the number of putatively adaptive SNPs increase, the predictive power of putatively adaptive SNPs (i.e., the observed minus expected correlation) seems to converge to zero (Figure 3.10). This might be suggesting that with higher numbers of

adaptive SNPs, equal-sized adaptive and random numbers of SNPs seem to equally correlate with genome-wide SNPs (i.e., seem to behave the same).

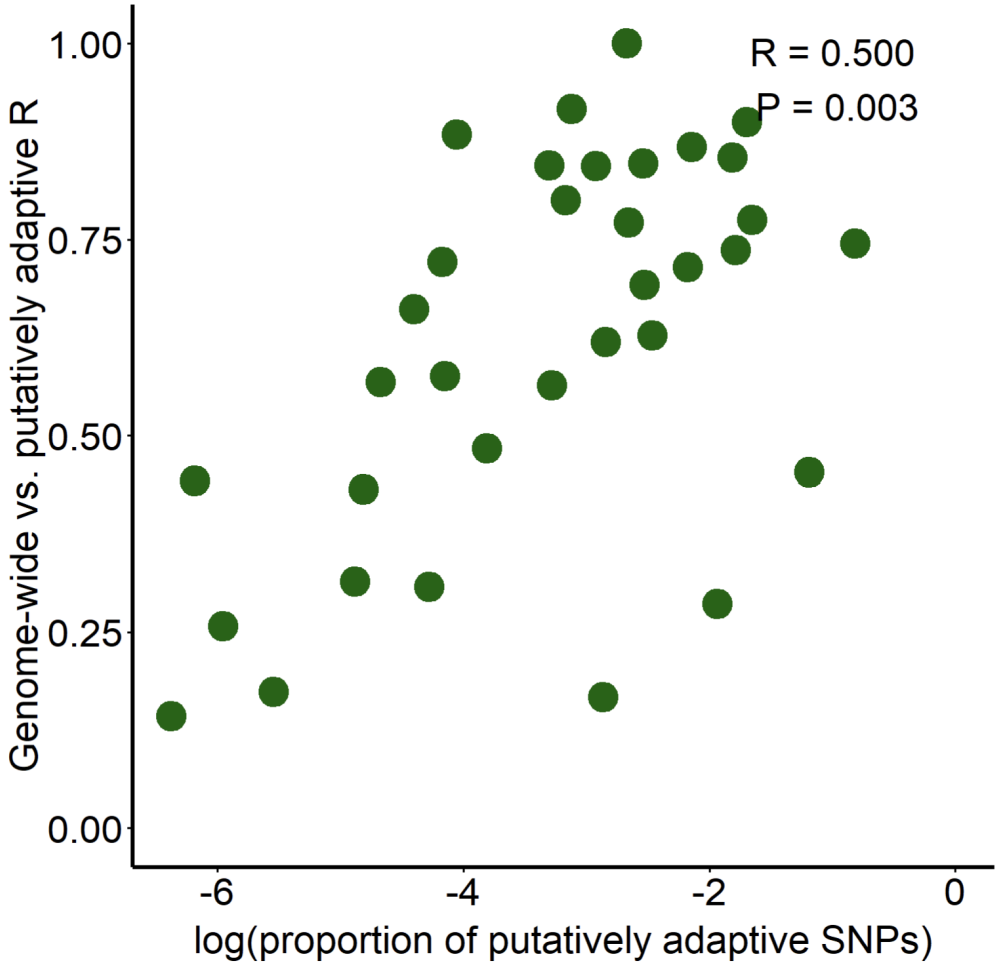

Figure 3.11: Proportion of putatively adaptive SNPs positively impacts the prioritization agreement (genome-wide vs. putatively adaptive R) as expected by the part-whole regression. Each dark green point represents a study (n = 34). The Spearman correlation between the prioritization agreement and log(proportion of adaptive SNPs) is 0.50 (uncorrected P = 0.003, FDR corrected P = 0.02).

In regards to the proportion of adaptive SNPs, I observed exactly what I predicted. Proportion of adaptive SNPs impacted the prioritization agreement (Spearman rho = 0.50, uncorrected P = 0.003, FDR corrected = 0.02) because putatively adaptive SNPs are a subset of genome-wide SNPs, so it provides information about the part-whole correlation (Figure 3.11). However, the proportion of adaptive SNPs did not impact the predictive power of putatively adaptive SNPs (Spearman rho = -0.17; uncorrected P = 0.35, FDR corrected P = 0.45).

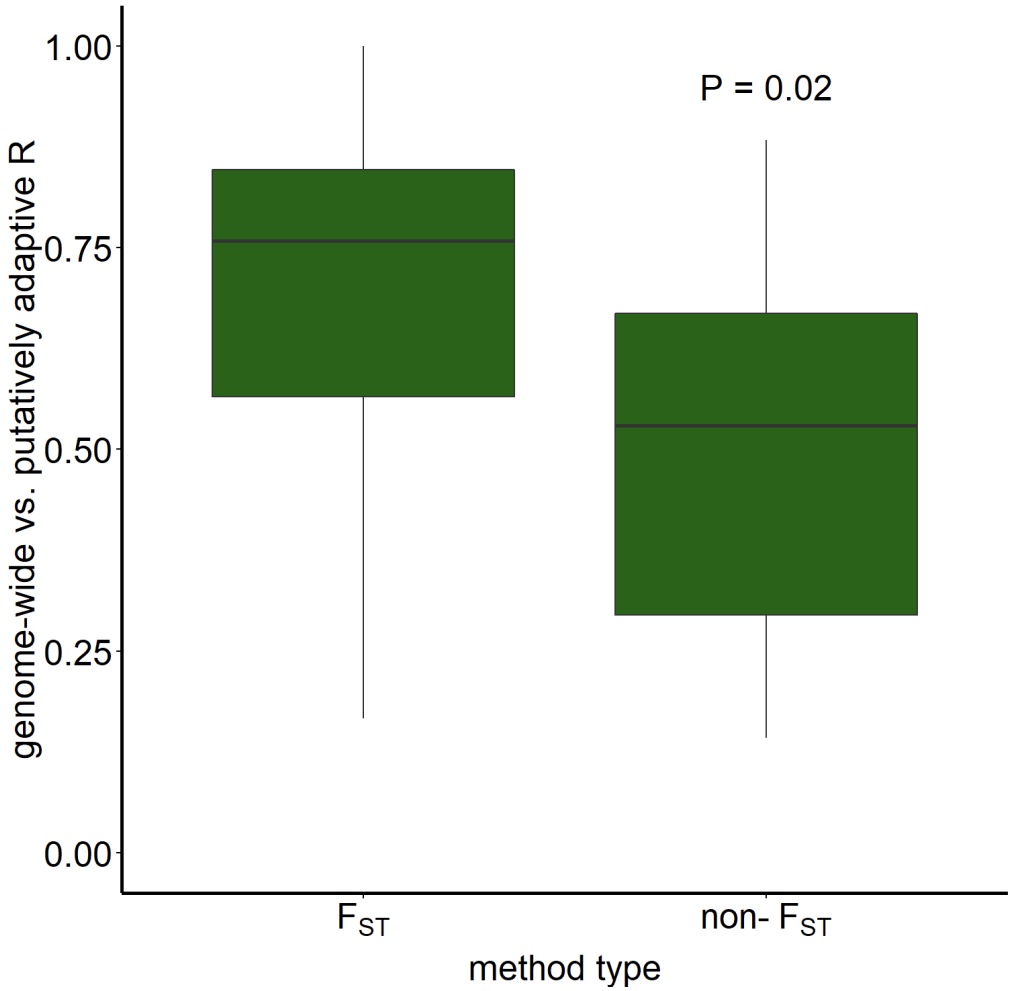

Figure 3.12: Method-type used to identify putatively adaptive SNPs seems to impact the prioritization agreement. One-way anova uncorrected P = 0.02, FDR corrected P = 0.05 (n = 34).

The method-type (F_{ST} vs. non-F_{ST}) used to identify putatively adaptive SNPs impacted the prioritization agreement (one-way anova uncorrected P = 0.02, FDR corrected = 0.05), with the F_{ST} method showing higher correlation between genome-wide and putatively adaptive SNPs than non-F_{ST} methods in these 34 studies (Figure 3.11). However, the method-type did not seem to impact the predictive power of putatively adaptive SNPs (one-way anova uncorrected P = 0.38, FDR corrected P = 0.45). Interestingly, the proportion of adaptive SNPs differed significantly between F_{ST} vs. non-F_{ST} methods (one-way anova P = 0.0005), but does not impact the number of adaptive SNPs (P = 0.74). Importantly, note that this result of method type impacting proportion of adaptive SNPs is not generalizable, as this is based on 34 different studies where each individual study had their own unique criteria and thresholds and each method decisions may be non-comparable across studies.

To predict factors driving the prioritization agreement, I performed a beta-regression to test the impact of proportion of adaptive SNPs, and method type (F_{ST} vs. non-F_{ST}) on non-transformed observed Spearman correlation. I found that the log of proportion of adaptive SNPs (coefficient = 0.33, P = 0.009) is mainly driving the observed correlation compared to method type (coefficient = -0.21, P = 0.53).

To predict factors driving the predictive power of putatively adaptive SNPs, I used linear regression to test the impact of study-wide F_{ST} and number of adaptive SNPs used on the observed minus expected correlation. I used anova to test the significance and p values. I found that both log of number of adaptive SNPs (estimate = -0.06, P = 0.009) and log of study-wide F_{ST} (estimate = -0.05, P = 0.03) significantly drive the predictive power of putatively adaptive SNPs.

3.4 Discussion

My goal is to contribute to the decades-long unresolved debate on whether prioritization when using genetics for conservation should focus on the subset of genetic variation that is adaptive, or whether it is preferable to consider genome-wide genetic variation. Given the current limitations and challenges in identifying and using adaptive variation for conservation, genome-wide variation, which is much easier to measure, may act as a surrogate for adaptive variation (see [71, 102, 157]). Currently, literature provides mixed theoretical and empirical evidence for congruence [51]. My motivation specifically comes from Fernandez-Fournier et al., (2021) [86] who asked whether genome-wide and putatively adaptive variation provide similar population prioritization rankings for conservation in yellow warbler and lodgepole pine. This study extends Fernandez-Fournier et al., (2021) [86] by examining the question on a vast array of species and taking a step further into disentangling factors impacting the relationship between genome-wide and adaptive genetic variation.

Specifically, I aim to answer three conservation questions: Do genome-wide (all) SNPs and adaptive SNPs provide similar population prioritizations? Do adaptive SNPs behave like equal-sized samples of random SNPs? How do covariates impact the prioritization agreement and the predictive power of putatively adaptive SNPs? Using the genetic data of 34 species representing a wide range of biodiversity (varying from plants, arthropods, amphibians, birds, fish, and mammals), population structures (from panmictic to subspecies), geographic regions, life histories, ecological conditions, and all major biomes, I compared population distinctness scores within species using genome-wide SNPs, and putatively adaptive, and random subsets of genome-wide SNPs. I found that genome-wide and putatively adaptive subsets of genome-wide SNPs generally agree on conservation prioritization, though agreement varies among studies (R varies from +0.1 to +1.0) and is mainly driven by the proportion of adaptive SNPs. Since putatively adaptive SNPs are a subset of genome-wide SNPs and part-whole correlation might impact the results, I decided to randomly sample an

equal number of random SNPs as putatively adaptive SNPs via bootstrap (repeated 1000 times). Overall, I found that putatively adaptive subsets of SNPs seem to correlate more strongly than random subsets of SNPs with genome-wide SNPs, with a modest but not negligible effect size. This pattern is explained by study-wide F_{ST}, and number of putatively adaptive SNPs.

3.4.1 Explanation and Interpretation of Results

Do genome-wide (all) SNPs and putatively adaptive SNPs provide similar population prioritizations?

I found that the observed Spearman correlation coefficients (i.e., prioritization agreement) of population distinctiveness scores using genome-wide and adaptive SNPs are all positive, but they vary a lot (from R +0.14 to +1.00) across studies. Overall, this indicates that genome-wide SNPs and adaptive SNPs may provide similar distinctness scores for ranked populations to some extent, however, the extent varies a lot, so should be interpreted and used with caution.

Do adaptive SNPs behave like equal-sized samples of random SNPs?

This study found that putatively adaptive SNPs subsets may generally show as good or stronger correlation with genome-wide SNPs, than the random subsets of SNPs do.

As expected, random and genome-wide SNPs correlate with each other probably due to part-whole correlation and/or because random SNPs are drawn from the same distribution as genome-wide SNPs. Since the difference between observed and expected correlation is not zero, this suggests that some properties or specific combinations of putatively adaptive and random SNPs (same size as putatively adaptive SNPs) may differ leading to differences in observed and expected correlations. These differences may lead to further downstream differences in PCA, NeighborNet, and Shapley Values among putatively adaptive and random SNPs. Tom Booker suggested investigating allele frequencies differences between adaptive and random SNPs.

As expected, I observed a statistical difference of adaptive and random median minor allele frequencies within and across studies, with adaptive allele frequencies being higher majority of the time. This is likely because adaptive SNPs are influenced by local adaptation and positive selection, leading them to be highly differentiated compared to non-adaptive SNPs that are mainly impacted by drift and demography [117, 247].

The hypothesis was that adaptive SNPs minor allele frequencies, if higher, might be easier to distinguish from noise, and this is why the observed correlation is higher than the expected correlation (i.e., higher predictive power of putatively adaptive SNPs). The thought was there is more noise in samples from low frequency alleles generally, and so, regardless of adaptation, high frequency SNPs will correlate with genome-wide (all) SNPs more strongly and the putatively adaptive SNPs may show stronger signals consistent with

genome-wide patterns. While many of the studies examined had higher adaptive minor allele frequencies and higher observed correlation compared to random, there was no clear link between the two, perhaps due to noise.

Because alleles involved in local adaptation are generally more differentiated from neutral/random alleles among populations making them easier to detect [178, 25, 138], the idea was studies with more exaggerated adaptive and random/neutral minor allele frequencies might be where adaptation has been very strong, so perhaps there are fewer false positives, contributing to a negative relationship between adaptive minus random allele frequencies and the predictive power of putatively adaptive SNPs (Figure 3.7). While we do see a very slightly negative trend (in Figure 3.7), it is clearly not significant.

Maybe both processes are going on and sometimes adaptation is polygenic and other times, it is not, but our data are too noisy to extract any links from adaptive and random allele frequency comparison. This might be because of lack of standardization of potential contributing factors across studies. We have not tested these empirically, but some other factors that may contribute to noise may include high gene flow among populations preventing differentiation [143] or the presence of high false-positives and false-negatives. High false-negatives and false-positives might arise due to either overcorrection of neutral population structure (or covariance of allele frequencies or demography) or not accounting for it or when neutral and selection gradients might be aligned with each other [3, 138, 37].

How do covariates impact the prioritization agreement and the predictive power of putatively adaptive SNPs?

It is extremely important to keep in mind that I analysed a very small sample size of studies with lots of noise and lack of standardization of factors impacting the prioritization agreement and the predictive power of putatively adaptive SNPs. The data (used from already published studies) varies substantially in terms of number of SNPs, individuals, and populations as well as sequencing/genotyping, filtering, and outlier detection approaches. Methods are not applied in a similar way across studies, so this prevents me from making apple-to-apple comparisons. Also, I only have power to detect rather large effects and given adaptation may mostly be polygenic, this exacerbates difficulties in interpreting results and patterns. Due to low power, I only analyzed a small number of covariates. I do not provide any generalizations due to lack of comparability, so based on literature, I set out expectations for covariates and the data are just suggestive or not of those expectations given that individual studies applied their own certain criteria, thresholds, and decisions.

I find that the proportion of putatively adaptive SNPs is mainly driving the prioritization agreement as expected by the part-whole correlation. This part-whole correlation suggests that since putatively adaptive SNPs are a subset of genome-wide SNPs, these sets of SNPs may provide similar correlations. This is one of the reasons supporting the consideration of using genome-wide genetic variation for prioritization when using genetics for conservation.

Interestingly, some of the covariates may co-vary with each other (see Appendix B Table S2). The proportion of putatively adaptive SNPs seem to correlate negatively with the total number of genome-wide SNPs (Spearman rho = -0.66, P = 3.13e-05). This means as the number of genome-wide SNPs increase, the proportion of putatively adaptive SNPs seem to decrease, a consistent result with Ahrens et al., (2018) [3]. One explanation for this (as suggested by Ahrens et al., (2018) [3]) is polygenic adaptation. In other words, studies with high numbers of genome-wide SNPs might have many small effect real associations aiding local adaptation, but this study might not be able to detect them. The proportion of putatively adaptive SNPs seem to be impacted by the method-type used to identify putatively adaptive SNPs, being higher for F_{ST} compared to non-F_{ST} methods (R = -0.57, P = 0.0005). However, since methods are not applied in a similar way across studies, I do not make any general inference regarding the relationship between proportion of putatively adaptive SNPs and the method type. The number of putatively adaptive SNPs is not correlated with the proportion of adaptive SNPs or the method type used to identify them, but they seem to be correlated with the number of genome-wide SNPs (Spearman rho = 0.57, P= 0.0004).

It is extremely challenging to account for the impact of neutral population structure, demography, covariance among alleles due to coancestry among populations, or neutral gradient aligning with selection gradient, or correlations among climatic & geographic distances with shared population history. There are mixed views in the literature regarding the preference on accounting for it or not, and if so, how. The methods are not sufficiently resolved enough to fully account for all these factors and to compare among different studies. Being conservative in accounting for neutral population structure may result in false-negatives (underestimating selection) and not accounting for it may contribute to false-positives [3, 269]. I have attempted to categorize the studies based on no, mixed, and yes categories in whether they somewhat accounted for some aspects of the neutral population structure, but due to small sample size, lack of power, and no standardization of methods applied across studies or even between different types of analysis used to detect putatively adaptive SNPs within each study contributing to noise, it is difficult to extract any clear patterns. More work is definitely needed to examine this further by resolving differences among different methods and applying methods in similar ways across studies.

Cases where genome-wide and putatively adaptive point in similar direction vs. in different directions compared to random expectation.

I find that the number of putatively adaptive SNPs and study-wide F_{ST} (reflective of population structure) seems to impact the predictive power of putatively adaptive SNPs.

When the number of putatively adaptive SNPs increase, the predictive power of putatively adaptive SNPs (i.e., the observed minus expected correlation) decreases and seems to converge to zero (Figure 3.10). This might be suggesting that with higher number of pu-

tatively adaptive SNPs, equal-sized putatively adaptive and random SNPs seem to equally correlate with genome-wide SNPs (i.e., seem to behave the same). Given small sample size, lack of standardization, weighing the pros and cons of each unique conservation situation, this result might not be generalizable. This result could be due to high number of false-positives.

As study-wide F_{ST} (reflective of population structure) increases, the predictive power of putatively adaptive SNPs (i.e. observed minus expected R) seems to decrease. Most studies seem to fall around the middle of study-wide F_{ST} axis with the predictive power of putatively adaptive SNPs ranging from +0.6 to -0.2 (Figure 3.8). In many of these cases, putatively adaptive SNPs and genome-wide SNPs seem to point in similar directions, so genome-wide SNPs may act as a surrogate for putatively adaptive variation.

Studies with lower to medium study-wide F_{ST} values (i.e., lower to medium population structure range of examined studies on Figure 3.8) are the cases where putatively adaptive subsets of genome-wide SNPs may show stronger correlation with genome-wide SNPs than random subsets of genome-wide SNPs do. In other words, putatively adaptive SNPs may point in the same direction as genome-wide SNPs compared to and better than random subsets of genome-wide SNPs. Interestingly, these (low study-wide F_{ST} studies) are also the cases where there seems to be lower variance in population distinctness scores (see Figure 3.9). In these cases, genome-wide SNPs may act as a surrogate for putatively adaptive variation as they both strongly correlate. One such example of low study-wide F_{ST} (point on the top very left of Figure 3.8) includes panmictic American eel with frequent random dispersal and mating across its range contributing to the generational persistence of individuals with maladaptive alleles, thus making it difficult for local adaptation to develop and detect [17]. Other examples of cases with low study-wide F_{ST} include *Homarus americanus* [30] and *Thaleichthys pacificus* [45] with high gene flow and weak genetic differentiation among populations where it might be difficult to detect genes involved in local adaptation. These low to medium study-wide F_{ST} might be cases where the identification of true adaptive SNPs is akin to finding a needle in a haystack. This will be true to the extent where adaptation may be highly polygenic such that we are missing many small-effect loci and also finding many false-positives.

Studies with high study-wide F_{ST} (i.e., high population structure range of examined studies on Figure 3.8) seem to be the cases where genome-wide and putatively adaptive SNPs do equally well and in many cases the correlation between genome-wide and putatively adaptive SNPs decreases relative to random expectation. In other words, putatively adaptive subsets of SNPs and genome-wide SNPs may point in different directions compared to random subsets of genome-wide SNPs. Interestingly, these (high study-wide F_{ST} or population structure) might be the cases where population distinctiveness scores vary considerably (i.e., high variance in Shapley values) (see Figure 3.9). Examples of these cases show strong local adaptation and high adaptive divergence with phenotypic & genetic dif-

ferentiation such as the *Chimarrichthys* fish complex [180], *Urocyon littoralis* (island fox) subspecies [101], and *Ovis dalli dalli* (Dall's sheep) populations [242]. It might be easier to identify locally adapted genes in these cases. Though more work is needed, it might be these high study-wide F_{ST} cases find more true adaptive SNPs and less hay (i.e. fewer false-positives and false-negatives).

Although it is difficult to make general statements, the data suggest that in low-to-medium population structure cases, investing efforts into conserving as much genetic variation as possible rather than identifying adaptive variation seems prudent. This is because in these cases, putatively adaptive and genome-wide SNPs point in a similar direction or might even show stronger correlation with genome-wide variation than random SNPs and prioritize similar distinct populations for conservation.

3.4.2 Considerations when identifying adaptive SNPs

Identification of true adaptive loci is akin to finding a needle in a haystack, so the studies I used would have misidentified and missed many of the true adaptive SNPs. Some of the reasons include limitations with current outlier differentiation and association analyses used to detect putatively adaptive loci. Firstly, the current methods might not consider individual effects of linkage disequilibrium, recombination, or quantifiable fitness effects and some may be biased towards identifying large effect loci within and near coding regions [138, 122], so we might have missed many small-effect loci and this might be problematic given the ubiquity of polygenic adaptation [39, 46, 86, 126, 241, 261, 122, 166]. Secondly, the signal of locally adapted loci at the population level might be difficult to detect due to high gene flow introducing non-local alleles that may swamp locally adapted alleles [4] or loci under past selection decaying its selection signal over time [219, 185, 86]. Thirdly, current approaches identify loci that are only currently adaptive. This is problematic since the future is unpredictable as adaptation could happen via new mutations, current standing genetic variation, or adaptive introgression [283], so it would be risky to rely solely on current adaptive loci [10, 33, 102, 262]. Fourthly, factors that we are unable to account for across studies due to lack of apple-to-apple comparison include the consideration of neutral population structure or different significance thresholds, different methods used or other methodological decisions made by individual studies might have contributed to false-negatives and false-positives. For example, there is mixed view in the literature on correction for neutral population structure and/or demography because consideration of differentiation or spatial structure may obscure identification of patterns at outlier loci leading to false-negatives and not accounting for neutral population structure or demographic patterns may increase false- positives [3, 66, 92]. Many times it may be difficult to untangle confounding effects of demographic changes such as contraction or expansion of populations from selection effects [3]. Further complications arise when assumed demographic history in simulation models may not match with the true observed neutral patterns in order to

correct for neutral population structure or demography [183]. Plus, the differences among methods may not be sufficiently resolved (for example, if one study used 2 latent factors while another study used 3 latent factors) to do apple-to-apple comparisons across studies of whether they accounted for neutral population structure or demography. Also, within a study, some methods used to identify adaptive loci might account for neutral population structure and/or demography while other methods might not.

3.4.3 Genome-wide variation as a surrogate for adaptive variation

It is crucial to consider a more realistic scenario where current methods in the literature used to detect adaptive SNPs are not perfect and give rise to false-negatives and false-positives. This study does not provide a statement on whether genome-wide genetic variation captures similar patterns as true adaptive genetic variation because of the wide variety of systems and data used (Tom Booker pers.comm.). This study acknowledges that the data and methods of ascertainment vary considerably across studies in terms of number of populations, individuals, and SNPs as well as sequencing/genotyping approaches, genome regions sampled, bioinformatic pipelines, filtering, analysis, and outlier detection (Tom Booker pers.comm.). This realistic approach is relevant for conservation geneticists and practitioners because identifying true adaptive genetic variation is challenging and the existing methods are extremely sensitive to sample size, population structure, etc. Because of this it becomes difficult to justify the blanket use of "adaptive" variation for conservation purposes and this study can only speak on the methodological and statistical reasons rather than the biological reasons.

Since the results of this study show that genome-wide variation and putatively adaptive genetic variation agree on population prioritization across such a wide variety of data and methods, prioritization may also consider genome-wide SNPs. Putatively adaptive SNPs are a subset of genome-wide SNPs, and the part-whole correlation supports the use of genome-wide SNPs for conservation. This study also shows that in certain situations (low to medium study-wide F_{ST}), putatively adaptive SNPs and genome-wide SNPs may show a stronger correlation compared to random subset of genome-wide SNPs.

There are several theoretical reasons that might explain the agreement among genome-wide and adaptive SNPs on population prioritization. Adaptation may be highly polygenic with a large amount of the genome under selection, so strong interconnectedness among complex genetic architecture might have contributed to the agreement between genome-wide and putatively adaptive SNPs. Furthermore, genome-wide and adaptive loci may be subjected to similar evolutionary processes showing similar patterns [157, 204, 225]. For example, divergent selection regimes and lower gene flow may lead to neutral genomic divergence as a by-product due to linkage among neutral and non-neutral loci [216, 217, 197, 173].

From an empirical standpoint, our study results agreed with other studies [204, 193, 21, 218, 318, 86] that found that genome-wide (or neutral) and putatively adaptive loci largely

provide similar patterns. Empirical evidence for genome-wide and adaptive congruence has been provided for sage grouse [218], Atlantic salmon [204], coho salmon [318], Eastern Massasauga rattlesnake (*Sistrurus catenatus*) [193], hermaphrodite plants [21], yellow warbler and lodgepole pine [86].

There are additional benefits of considering genome-wide variation as a surrogate for adaptive variation for conservation in certain situations. First, it is much easier to measure compared to adaptive variation. Second, loci that are currently neutral may become adaptive in the future, so conserving as much variation as possible acts as an insurance.

3.4.4 Limitations and Future Directions

Though genetics might play an important role in conservation decision-making, conservation practitioners would rarely solely utilize genetic information alone to prioritize their actions. Incorporating the impact of major non-genetic factors involved in conservation management decision-making such as threat status, risk assessment, and the fluctuation in population patterns through time in practical prioritization exercises would be useful [86]. While consideration of non-genetic factors is crucial, improvements can also be made to the genetic metric used to prioritize populations in this study by accounting for both within-and among-population genetic distinctiveness simultaneously (see Chapter 2), especially when, in a particular case, it is shown that they tend to be negatively correlated [57, 300]. Many studies [1, 44, 57, 54, 119, 136, 221, 231, 279, 284, 230, 168, 179] have advocated, developed or implemented methods to harmonize both within-and among-population genetic distinctiveness aspects for adaptive potential and long-term survival of species (summarized in [51]).

Potential cases where solely considering among-population distinctiveness and overlooking crucial factors such as within-population distinctness and threat status would be problematic are when there is only a single trait that is of management/conservation concern (e.g., run-timing in salmon) or when deciding whether to manage populations separately or to promote natural or assisted gene flow, especially in small distinct populations where distinctness may be adaptive or non-adaptive. It is challenging to measure the extent to which population distinctiveness is driven by local adaptation or genetic drift [86] and these have serious conservation management implications [181, 220]. While every conservation situation must be analyzed separately by weighing its costs and benefits, I offer some general considerations. If populations exhibit strong local adaptation and/or outbreeding depression and they are not in dire situations or at imminent extinction risk, then managing them separately might be suitable [4, 97]. Examples might include the Halfmoon Hairstreak butterfly [59], pacific salmonids [298], and Atlantic salmon [171]. However, if small populations' uniqueness is largely driven by genetic drift such that their genetic variation is non-adaptive and maladaptive, then allocating conservation efforts on them might be detrimental for species' survival [195] as that may promote fragmentation, reduce connectivity and gene flow and

that might make them vulnerable to stochastic demographic events, inbreeding depression, and allee effects [14, 51, 141, 291, 300]. In these cases, promoting natural or assisted gene flow might be beneficial rather than managing populations separately [4, 300, 235]. Another situation where preserving as much genome-wide genetic variation as possible via encouraging gene flow might be suitable is in dire situations where entities might be at imminent extinction risk [300]. Examples here might include the critically endangered Yellow-tufted honeyeater (*Lichenostomus melanops cassidix*) [121, 224], Mountain pygmy possum (*Burramys parvus*) [299], and Arabian leopard (*Panthera pardus nimr*) [6]. So, consideration of a metric that harmonizes both within- and among- population genetic variation along with consideration of threat status and extinction risk and then testing the correlation between population prioritizations using genome-wide SNPs and putatively adaptive SNPs would seem a fruitful future research endeavour.

Further extension of this study could investigate the geolocation of conservation important populations. The definition of conservation important populations highly varies on the context and case to case basis. However, depending on the context and study system, one potential question is whether high priority populations measured using genetic data are largely located at the periphery of species' range or idiosyncratically [14, 86].

Future studies would also benefit from standardization. Firstly, the accessibility of genomic data in a standardized format such as variant call format (VCF) or 0,1,2,NA SNP format where the number in each SNP column refers to the minority allele copies and NA refers to missing data with information on individuals, populations, and geolocation would greatly benefit the community [51]. Secondly, the identification of putatively adaptive SNPs along with the information on the type of analysis or method used to detect it with comparable (within and across studies) significance thresholds would be advantageous. The consistency among files on SNP names (for genome-wide, neutral and adaptive) and incorporating information on SNP positions would be useful. Further standardization on decisions like missing rate cut-off, agreement on the consideration of neutral population structure, minimum threshold on number of individuals, populations, genome-wide SNPs, putatively adaptive SNPs, percent of geographic range coverage would be valuable. Some of the challenges in reproducibility, interpretation, validation [94], and fair comparison within and across studies can be solved via standardization of study properties.

3.5 Summary

Identifying populations to prioritize for conservation is a challenge. One potential strategy is to prioritize populations that are most genetically distinct. Many authors suggest we should focus on adaptive variation directly when measuring distinctiveness because it is directly related to selection, function, and thus adaptive capacity. However, there are several limitations in identifying and using adaptive genetic variation for conservation. Also, if

adaptation is highly polygenic and finding true adaptive SNPs is like finding a needle in a haystack, then genome-wide genetic distinctiveness, which is much easier to measure, may be preferable. We gather genome-wide and putatively adaptive SNPs data from outlier and/or association studies across 34 species of plants and animals and ask whether putatively adaptive SNPs identify the same distinctive populations as measures based on genome-wide SNPs and whether adaptive SNPs behave like equal-sized samples of random subsets of genome-wide SNPs and what covariates impact these results. We find that genome-wide and putatively adaptive subsets of genome-wide SNPs agree on population prioritizations for conservation, though their agreement varies a lot (R varying from +0.14 to +1.00) and this correlation is impacted by the proportion of putatively adaptive SNPs. We find that putatively adaptive SNPs do not seem to behave the same as equal-sized random subsets of genome-wide SNPs. If anything, adaptive SNPs subsets may show stronger correlation with genome-wide than random SNPs subsets do. This pattern is explained by study-wide F_{ST} and number of putatively adaptive SNPs. The results from this study are directly relevant to climate change biologists, conservation biologists, and ultimately managers.

3.6 Details of studies included in the analysis

1. Ruegg et al., 2018 [245] and 2021 [244] generated a dataset of willow flycatcher (*Empidonax traillii*) across its breeding range and wintering locations. After applying 10% missingness cut-off, the dataset consisted of 105000 genome-wide SNPs, 179 putatively adaptive SNPs, 144 individuals, and 6 populations. Data is publicly available on GitHub [https://github.com/eriqande/ruegg-et-al-wifl-genoscapemake-the-genoscape.rmdv] (provided by original authors of the study).

2. Eimanifar et al., 2018 [77] generated a dataset on western honey bees (*Apis mellifera* L.) in the Republic of South Africa. After applying 10% missingness cut-off, the dataset consisted of 2449 genome-wide SNPs, 140 putatively adaptive SNPs, 471 individuals and 29 populations. Raw data is available on Dryad Digital Repository[https://doi.org/10.5061/dryad.98jh446].

3. White et al., 2013 [307] published a dataset on bank vole (*Myodes glareolus*) in Ireland. After applying 10% missingness cut-off, there were 4340 genome-wide SNPs, 17 putatively adaptive SNPs, 241 individuals, and 14 populations. Raw data is available on Dryad Digital Repository [https://doi.org/10.5061/dryad.fb782].

4. Schweizer et al., 2016 [260] published a dataset on grey wolf (*Canis lupus*) in Arctic and High Arctic six distinct ecotypes. After applying 10% missingness cut-off, 13092 genome-wide SNPs, 34 putatively adaptive SNPs, 106 individuals, and 6 populations. Raw data is available on Dryad Digital Repository [https://doi.org/10.5061/dryad.8g0s3].

5. Mosca et al., 2016 [208] sampled data on *Pinus mugo* within Italian Alpine regions. After applying 10% missingness cut-off, *Pinus mugo* data consisted of 615 genome-wide

SNPs, 89 putatively adaptive SNPs, 622 individuals, and 20 populations. Raw data is published on Dryad Digital Repository [https://doi.org/10.5061/dryad.tm33d]. This Dryad is stores data for Mosca et al., 2012 [207] study.

6. Mckown et al., 2014 [194] published data on black cottonwood (*Populus trichocarpa*) from its North American range. After applying 10% missingness cut-off, the dataset consisted of 29,535 genome-wide SNPs, 410 putatively adaptive SNPs, 439 individuals, and 25 populations. Population data came from supplementary files of Geraldes et al., 2014 [105].

7. Royer et al., 2016 [243] produced data on Joshua tree species (*Y.brevifolia*, *Y.jaegeriana*, and hybrids) from Tikaboo Valley, Nevada. After applying 10% missingness cut-off, the dataset consisted of 4603 genome-wide SNPs, 35 putatively adaptive SNPs, 197 individuals and 6 populations.

8. Christmas et al., 2016 [52] generated data on narrow-leaf hopbush (*Dodonaea viscosa* spp. *angustissima*) of Australia. After applying 10% missingness cut-off, the dataset consisted of 7255 genome-wide SNPs, 59 putatively adaptive SNPs, 87 individuals, and 17 populations. The data was obtained via email and supplementary files. Dr. Christmas and Dr. Lowe emailed the population data.

9. Benestan et al., 2016 [29] generated data on American lobster (*Homarus americanus*) sampled across the North American Atlantic Coast. The dataset consisted of 3395 genome-wide SNPs, 7 putatively adaptive SNPs, 421 individuals and 19 populations after applying 10% missingness cut-off. Raw data is available on Dryad Digital Repository [https://doi.org/10.5061/dryad.5vb8v].

10. Roffler et al., 2016 [242] sampled data on Dall's sheep *Ovis dalli dalli* in Alaska and Yukon Territory available publicly on Dryad Digital Repository [https://doi.org/10.5061/dryad.kk466]. After applying 10% missingness cut-off, the dataset consisted of 187 genome-wide SNPs, 57 putatively adaptive SNPs, 472 individuals, and 15 populations.

11. Babin et al., 2017 [17] sampled data on the American Eel (*Anguilla rostrata*) from Newfoundland to Florida. The authors emailed me their data and with permission we publish our conversion of their data on Dryad Digital Repository. After applying 10% missingness cut-off, the dataset had 12,098 genome-wide SNPs, 269 putatively adaptive SNPs, 710 individuals, and 13 populations.

12. Guo et al., 2016 [115] published data on Andrew's toad (*Bufo andrewsi*) from Tibetan Plateau's edge on Dryad Digital Repository [https://doi.org/10.5061/dryad.n70c7]. The data consisted of 15,577 genome-wide SNPs, 586 putatively adaptive SNPs, 264 individuals, and 11 populations.

13. De Kort et al., 2014 [69] published data on Black alder (*Alnus glutinosa*) from throughout Europe on Dryad Digital Repository [https://doi.org/10.5061/dryad.rg82f]. After applying 10% missingness cut-off, the dataset consisted of 1714 genome-wide SNPs, 16 putatively adaptive SNPs, 295 individuals and 23 populations.

14. Hurel et al., 2021 [149] sampled data on *Pinus pinaster* in southwestern France. The raw data files are available on Dryad Digital Repository [https://doi.org/10.5061/dryad.r4xgxd2df]. After applying 10% missingness cut-off, the dataset consisted of 6074 genome-wide SNPs, 96 putatively adaptive SNPs, 515 individuals and 33 populations.

15. Depardieu et al., 2021 [70] examined white spruce (*Picea glauca* [Moench] Voss) data across Quebec. The data is available on Dryad Digital Repository [https://doi.org/10.5061/dryad.6rd6f] and on Github [https://github.com/ClaireDepardieu/Genetic_basis_drought]. The dataset consisted of 6153 raw genome-wide data, 359 putatively adaptive SNPs, 1473 individuals, and 43 populations.

16. Chen et al., 2012 [50] sampled data on Norway spruce trees (*Picea abies*) in Germany, Russia, Finland, and Sweden. The data is available on Dryad Digital Repository [https://doi.org/10.5061/dryad.82201]. After applying 10% missingness cut-off, the dataset included 375 genome-wide SNPs, 32 putatively adaptive SNPs, 262 individuals, and 18 populations.

17. Xuereb et al., 2022 [318] published data on Coho salmon (*Oncorhynchus kisutch*) in British Columbia on Dryad Digital Repository [https://doi.org/10.5061/dryad.r4xgxd2gx]. I only analyzed Thompson River data for this study. After applying 10% missingness cut-off, the dataset included 9683 genome-wide SNPs, 119 putatively adaptive SNPs, 836 individuals, and 26 populations.

18. Holliday et al., 2010 [143] sampled data on Sitka spruce (*Picea sitchensis*) along the west coast of North America (from Alaska to California). The authors emailed me their data and granted permission to publish 0,1,2,NA (our version of their data) on the Dryad Repository. After removing populations with less than 4 individuals, the dataset incorporated 437 genome-wide SNPs, 35 putatively adaptive SNPs, 407 individuals, and 13 populations.

19. Flanagan et al., 2021 [91] published data on Gulf pipefish (*Syngnathus scovelli*) of southeastern USA. The data is publicly available on Dryad Digital Repository [https://doi.org/10.5061/dryad.12jm63xvh]. After applying 10% missingness cut-off, the dataset contained 2738 genome-wide SNPs, 312 putatively adaptive SNPs, 235 individuals, and 7 populations.

20. Bay et al., 2018 [22] sampled data on yellow warbler (*Setophaga petechia*) in the USA, Canada, Central America, and northern South America. The authors emailed us their data and granted permission to publish the converted version of their data on Dryad. After applying 10% missingness cut-off, the dataset consisted of 104,385 genome-wide SNPs, 1609 putatively adaptive SNPs, 173 individuals, and 19 populations.

21. Chavez-Galarza et al., 2013 [49] sampled data on Iberian honey bee (*Apis mellifera iberiensis*) in the Iberian Peninsula. The data is publicly available on Dryad Digital Repository [https://doi.org/10.5061/dryad.1kk2k] and Dr.Pinto also emailed me data files and granted permission to post our conversion of their data files on Dryad. After applying 10%

missingness cut-off, the dataset contained 434 genome-wide SNPs, 73 putatively adaptive SNPs, 668 individuals, and 23 populations.

22. Keller et al., 2018 [160] expanded on data from Keller et al., 2012 [161] by sampling across balsam poplar's (*Populus balsamifera L*)species range. The dataset is available on Dryad Digital Repository [https://doi.org/10.5061/dryad.gp78p]. After applying 10% missingness cut-off, the dataset contained 279 genome-wide SNPs, 124 putatively adaptive SNPs, 925 individuals, and 83 populations.

23. Funk et al., 2016 [101] published data on island fox (*Urocyon littoralis*) from California Channel Islands on Dryad Digital Repository [https://doi.org/10.5061/dryad.2kn1v]. After applying 10% missingness cut-off, the dataset contained 2,498 genome-wide SNPs, 176 putatively adaptive SNPs, 103 individuals, and 19 populations.

24. Mosca et al., 2016 [208] sampled data on *Pinus cembra* within Italian Alpine regions. After applying 10% missingness cut-off, *Pinus cembra* data consisted of 443 genome-wide SNPs, 85 putatively adaptive SNPs, 662 individuals, and 18 populations. Raw data is published on Dryad Digital Repository [https://doi.org/10.5061/dryad.tm33d]. This Dryad is stores data for Mosca et al., 2012 [207] study.

25. Candy et al., 2015 [45] published data on anadromous Pacific smelt (*Thaleichthys pacificus*, Osmeridae) on Dryad Digital Repository [https://doi.org/10.5061/dryad.1797v]. After applying 10% missingness cut-off, the dataset contained 3725 genome-wide SNPs, 157 putatively adaptive SNPs, 441 individuals, and 11 populations.

26. Dallaire et al., 2021 [66] sampled neutral and putatively adaptive data on anadromous Arctic Char (*Salvelinus alpinus*) across Nunavik, southern Baffin Island, and Labrador. Dr. Dallaire emailed me genome-wide SNPs, putatively adaptive SNPs, and population data and granted permission to publish our converted version of their dataset on dryad. After applying 10% missingness cut-off, the dataset comprised of 13,596 genome-wide SNPs, 732 putatively adaptive SNPs, 545 individuals, and 23 populations.

27. Hess et al., 2013 [134] published data on Pacific lamprey (*Entosphenus tridentatus*) on Dryad Digital Repository [https://doi.org/10.5061/dryad.nd853]. After applying 10% missingness cut-off, the dataset comprised of 4,439 genome-wide SNPs, 164 putatively adaptive SNPs, 513 individuals, 21 populations.

28. Milano et al., 2014 [198] published data on European hake populations (*Merluccius merliccius*) sampled from the Atlantic and Mediterranean on Dryad Digital Repository [https://doi.org/10.5061/dryad.7bn22]. After applying 10% missingness cut-off, the dataset comprised of 380 genome-wide SNP data, 30 putatively adaptive SNPs, 849 individuals, 19 populations.

29. Swaegers et al., 2015 [278] collected data on damselfly *Coenagrion scitulum*. The data is available on Dryad [https://doi.org/10.5061/dryad.n0hk7] and Dr. Swaegers also emailed us the data. After applying 10% missingness cut-off, the dataset consisted of 3470 genome-wide SNPs, 566 putatively adaptive SNPs, 161 individuals, and 10 populations.

30. Moore et al., 2014 [204] published data on anadromous Atlantic Salmon (*salmo salar*) covering the entire North American range. The data is available on Dryad [https://doi.org/10.5061/dryad.sb601]. After applying 10% missingness cut-off, the dataset included 3192 genome-wide SNPs, 374 putatively adaptive SNPs, 1079 individuals, 50 populations.

31. Mahony et al., 2020 [190] published data on lodgepole pine (*Pinus contorta*) sampled across western Canada on Dryad [https://doi.org/10.5061/dryad.56j8vq8]. I used a subset of 22 populations and a list of putatively adaptive SNPs from Fernandez Fournier et al., 2021 analysis [86][https://github.com/philippeff/PopulationPrioritization]. After applying 10% missingness cut-off, the dataset included 26431 genome-wide SNPs, 460 putatively adaptive SNPs, 243 individuals, and 22 populations.

32. He et al., 2016 [127] generated data on *Banksia attenuata* (Proteaceae) from south-western Australia. Dr. He kindly emailed me the datasets and granted permission to publish our converted version of their dataset on Dryad. The dataset included 5701 genome-wide SNPs, 1049 putatively adaptive SNPs, 80 individuals, 9 populations.

33. Li et al., 2021 [180] sampled data on Chimarrichthys fish complex, (*C. davidi*, *C. kishinouyei*, *C. longibarbatus*) found in the Hengduan Mountain Region of China. They collected fish from Jinsha River, Qingyi River, Dadu River and Yalong River. Dr.Peng emailed me the data and kindly provided permission to publish our converted version of their datasets on Dryad. After applying 10% missingness cut-off, the dataset contained 5834 genome-wide SNPs, 258 putatively adaptive SNPs, 183 individuals, and 9 populations.

34. Cullingham et al., 2014 [65] used data on lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*Pinus banksiana*) and their interspecific hybrids described in Cullingham et al., 2013 [64]. Dr. Cullingham kindly emailed me the data and granted permission to publish our converted conversion of their datasets on Dryad. I only used Jackpine data. After applying 10% missingness cut-off, the dataset contained 361 genome-wide SNPs, 25 putatively adaptive SNPs, 100 individuals, and 4 populations. Since two populations had the exact same Shapley Value (during data analysis), I used 3 populations for this. I got the same results using 3 and 4 populations.

3.7 Acknowledgements

I would like to thank all the authors of the studies who provided us their hard-won data. I am grateful for feedback from the FAB lab, Mooers Lab, and M'Gonigle Lab at Simon Fraser University and Jean-Sebastien Moore Lab at Laval University. Dr. Arne Mooers, Dr. Jayme Lewthwaite, Philippe Fernandez-Fournier, and Dr. Tom Booker are co-authors on a version of this chapter in preparation for submission. Thanks to M. Dennis, N. Abhari, D. Pelletier, B. Crespi, L. M'Gonigle, A. MacPherson, M. Collard, and J. Moore from SFU for all their feedback. Thanks to the scientific community of Canadian Society of Ecology and Evolution (CSEE), EcoEvo Retreat 2023, Biodiversity Genomics 2023, EVO WIBO 2023,

IDEAS 2023 for useful suggestions. Thanks to the graduate scholarships from NSERC, the BC government, and Simon Fraser University. I note that Simon Fraser University is located on the ancestral and unceded land of the Musqueam, Squamish, and Tsleil-Waututh Nations.

Chapter 4

Conclusions and Future Directions

4.1 Summary and Conclusions

My main goal is to contribute to the discussion of how to prioritize populations within species for conservation as it is crucial for the maintenance of adaptive potential of species and populations.

Chapter 2 is a literature review providing a detailed overview of the core concepts in conservation genetics and summarizing considerations applicable to the conservation of populations using genomic SNP data. Chapter 2 suggests that the conservation of intraspecific genetic diversity is guided by the distribution of genetic variation across the geographic space (i.e. spatial structure) and across the genome (i.e. genomic structure). The spatial structure consists of within-and among-population genetic variation whereas the genomic structure consists of genome-wide genetic variation and a subset of that variation that is currently putatively adaptive. There are two main key conclusions from Chapter 2. The first main conclusion is that it is important to consider and harmonize both within-and among-population genetic variation when prioritizing populations as solely focusing on one component may carry risks. The second conclusion is that mixed literature evidence exists regarding the unresolved debate of whether population prioritization should focus solely on adaptive genetic variation or whether it is preferable to consider genome-wide genetic variation.

Chapter 3 links one aspect of the spatial structure (among-population genetic variation component) to the genomic structure of genetic variation. I ask whether genome-wide genetic variation and putatively adaptive genetic variation identify the same set of distinct populations within species and provide empirical support by investigating 34 plant and animal species. Many authors advocate focusing solely on adaptive genetic variation, however, there are many technical and conceptual challenges in identifying and using adaptive genetic variation for conservation. So, if genome-wide genetic variation, which is much easier to measure, can act as a surrogate for adaptive genetic variation, then it will be a sound and cost-effective strategy for population prioritization for conservation. We find that

genome-wide genetic variation and putatively adaptive genetic variation agree generally, but variably (R varying from +0.14 to +1.00) on population prioritizations across 34 species. As expected, the proportion of putatively adaptive SNPs impacts the prioritization agreement among genome-wide and putatively adaptive SNPs. When compared to equal-sized random subsets of genome-wide SNPs, we find that putatively adaptive SNPs do as well or show higher correlation with genome-wide SNPs. This is impacted by study-wide F_{ST} and the number of putatively adaptive SNPs. Overall, the conclusion is genome-wide genetic variation seems like a sound strategy to use for population prioritization and to protect intraspecific genetic variation for conservation.

Based on the conclusion from Chapter 3, it seems that practitioners can consider using genome-wide genetic variation when using genetics for conservation. However, much more work is needed to be done in the area. A small sample size of 34 species in this study limits the ability to interpret some important patterns and greatly influences the power. Also, data and methods of ascertainment vary to a large degree across studies. While in some aspects, data varying considerably across studies acts as a strength because it represents a more realistic scenario where identifying true adaptive variation is challenging and provides evidence that despite such high variance, genome-wide genetic variation and putatively adaptive genetic variation identify similar populations as distinct. In other aspects, it acts as a weakness when assessing the biology of a wide variety of systems because it prevents fair comparisons among studies. Chapter 3 could greatly be improved by using a greater sample size and standardizing each of the study properties to make fair comparisons.

4.2 Future Research

The investigation on how to maintain intraspecific diversity for conservation using SNP genomic data is a promising area of research and can be extended out in various directions, depending on the specific perspective and focus area one chooses. More research is needed in the area of local adaptation, conservation management, and the development of new metrics to prioritize populations.

In this section, I focus on two main future research recommendations based on the findings from Chapter 2 and 3.

First, while we can probe the genome of many species, reproducibility and validation is still a problem and lack of raw genetic data in a standardized format is a key factor contributing to the issue. This can be resolved if raw genetic data can be made publicly available in one agreed upon format along with information attaching each individual to its geolocation and population. Also, availability of putatively adaptive SNPs along with specific methods and comparable significance thresholds used to identify them would be useful. Further agreements on data processing decisions such as on missing rate cut-off,

accounting for neutral population structure, or determining minimum thresholds for marker numbers would be useful.

Second, it is crucial to consider and harmonize both components of the spatial structure, within- and among-population genetic variation, along with the threat status of populations and species when prioritizing populations for conservation. In Chapter 3, I only consider the among-population genetic variation component by focusing on populations' distinctiveness using Shapley Value. However, as outlined in Chapter 2, small populations could become distinct due to drift, such that their unique genetic variation is non-adaptive. In such cases, allocating conservation efforts on these populations can be detrimental for species' survival as we might be conserving maladaptive alleles that are not useful for present or future adaptation. So, to maintain the genetic variation required for current and future adaptation, it is crucial to consider both within-and among-population genetic variation along with populations' threat status.

Bibliography

[1] Niloufar Abhari, Caroline Colijn, Arne Mooers, and Paul Tupper. Capturing diversity: Split systems and circular approximations for conservation. *Journal of Theoretical Biology*, 578:111689, 2024.

[2] Milo D Adkison. Population differentiation in Pacific salmons: local adaptation genetic drift, or the environment? *Canadian Journal of Fisheries and Aquatic Sciences*, 52(12):2762–2777, 1995.

[3] Collin W Ahrens, Paul D Rymer, Adam Stow, Jason Bragg, Shannon Dillon, Kate DL Umbers, and Rachael Y Dudaniec. The search for loci under selection: trends, biases and progress. *Molecular Ecology*, 27(6):1342–1356, 2018.

[4] Sally N Aitken and Michael C Whitlock. Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution, and Systematics*, 44:367–388, 2013.

[5] Sally N Aitken, Sam Yeaman, Jason A Holliday, Tongli Wang, and Sierra Curtis-McLane. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, 1(1):95–111, 2008.

[6] Hadi Al Hikmani, Cock van Oosterhout, Thomas Birley, Jim Labisko, Hazel A Jackson, Andrew Spalton, Simon Tollington, and Jim J Groombridge. Can genetic rescue help save Arabia's last big cat? *Evolutionary Applications*, 17(5):e13701, 2024.

[7] M Alleaume-Benharira, IR Pen, and Ophélie Ronce. Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *Journal of Evolutionary Biology*, 19(1):203–215, 2006.

[8] Fred W Allendorf. Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology*, 26(2):420–430, 2017.

[9] Fred W Allendorf, W Chris Funk, Sally N Aitken, Margaret Byrne, and Gordon Luikart. *Conservation and the Genomics of Populations*. Oxford University Press, 2022.

[10] Fred W Allendorf, Paul A Hohenlohe, and Gordon Luikart. Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11(10):697–709, 2010.

[11] Inger Greve Alsos, Dorothee Ehrich, Wilfried Thuiller, Pernille Bronken Eidesen, Andreas Tribsch, Peter Schönswetter, Claire Lagaye, Pierre Taberlet, and Christian Brochmann. Genetic consequences of climate change for northern plants. *Proceedings of the Royal Society B: Biological Sciences*, 279(1735):2042–2051, 2012.

[12] Joel M Alves, Miguel Carneiro, Jade Y Cheng, Ana Lemos de Matos, Masmudur M Rahman, Liisa Loog, Paula F Campos, Nathan Wales, Anders Eriksson, Andrea Manica, et al. Parallel adaptation of rabbit populations to myxoma virus. *Science*, 363(6433):1319–1326, 2019.

[13] Santosh Anand, Eleonora Mangano, Nadia Barizzone, Roberta Bordoni, Melissa Sorosina, Ferdinando Clarelli, Lucia Corrado, Filippo Martinelli Boneschi, Sandra D'Alfonso, and Gianluca De Bellis. Next generation sequencing of pooled samples: guideline for variants' filtering. *Scientific Reports*, 6(1):33735, 2016.

[14] Jorge Assis, Nelson Castilho Coelho, Filipe Alberto, Myriam Valero, Pete Raimondi, Dan Reed, and Ester Alvares Serrão. High and distinct range-edge genetic diversity despite local bottlenecks. *PLoS One*, 8(7):e68646, 2013.

[15] Giorgia G Auteri and L Lacey Knowles. Decimated little brown bats show potential for adaptive change. *Scientific Reports*, 10(1):3023, 2020.

[16] Alex N Nguyen Ba, Katherine R Lawrence, Artur Rego-Costa, Shreyas Gopalakrishnan, Daniel Temko, Franziska Michor, and Michael M Desai. Barcoded bulk QTL mapping reveals highly polygenic and epistatic architecture of complex traits in yeast. *Elife*, 11:e73983, 2022.

[17] Charles Babin, Pierre-Alexandre Gagnaire, Scott A Pavey, and Louis Bernatchez. RAD-seq reveals patterns of additive polygenic variation caused by spatially-varying selection in the American eel (*Anguilla rostrata*). *Genome Biology and Evolution*, 9(11):2974–2986, 2017.

[18] Rowan DH Barrett and Dolph Schluter. Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1):38–44, 2008.

[19] Nicholas H Barton and Peter D Keightley. Understanding quantitative genetic variation. *Nature Reviews Genetics*, 3(1):11–21, 2002.

[20] William F Basener and John C Sanford. The fundamental theorem of natural selection with mutations. *Journal of Mathematical Biology*, 76(7):1589–1622, 2018.

[21] Thomas M Bataillon, Jacques L David, and Daniel J Schoen. Neutral genetic markers and conservation genetics: simulated germplasm collections. *Genetics*, 144(1):409–417, 1996.

[22] Rachael A Bay, Ryan J Harrigan, Vinh Le Underwood, H Lisle Gibbs, Thomas B Smith, and Kristen Ruegg. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, 359(6371):83–86, 2018.

[23] Rachael A Bay and Stephen R Palumbi. Multilocus adaptation associated with heat resistance in reef-building corals. *Current Biology*, 24(24):2952–2956, 2014.

[24] Rachael A Bay, Noah H Rose, Cheryl A Logan, and Stephen R Palumbi. Genomic models predict successful coral adaptation if future ocean warming rates are reduced. *Science Advances*, 3(11):e1701413, 2017.

[25] Mark A Beaumont. Adaptation and speciation: what can F_{ST} tell us? *Trends in Ecology & Evolution*, 20(8):435–440, 2005.

[26] Mark A Beaumont and Richard A Nichols. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1377):1619–1626, 1996.

[27] Sarah Adine Bekessy, Richard A Ennos, Mark A Burgman, Adrian C Newton, and Peter K Ades. Neutral dna markers fail to detect genetic divergence in an ecologically important trait. *Biological Conservation*, 110(2):267–275, 2003.

[28] Jordan B Bemmels, Else K Mikkelsen, Oliver Haddrath, Rogan M Colbourne, Hugh A Robertson, and Jason T Weir. Demographic decline and lineage-specific adaptations characterize New Zealand kiwi. *Proceedings of the Royal Society B*, 288(1965):20212362, 2021.

[29] Laura Benestan, Brady K Quinn, Halim Maaroufi, Martin Laporte, Fraser K Clark, Spencer J Greenwood, Rémy Rochette, and Louis Bernatchez. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Molecular Ecology*, 25(20):5073–5092, 2016.

[30] Laura Marilyn Benestan, Anne-Laure Ferchaud, Paul A Hohenlohe, Brittany A Garner, Gavin JP Naylor, Iliana Brigitta Baums, Michael K Schwartz, Joanna L Kelley, and Gordon Luikart. Conservation genomics of natural and managed populations: building a conceptual and practical framework, 2016.

[31] Edward E Berg and JL Hamrick. Quantification of genetic diversity at allozyme loci. *Canadian Journal of Forest Research*, 27(3):415–424, 1997.

[32] L Bernatchez. On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*, 89(6):2519–2556, 2016.

[33] Angéline Bertin, Mara I Espinosa, Catalina A Bustamante, Alejandra J Troncoso, and Nicolas Gouin. Genome-wide genetic diversity yields insights into genomic responses of candidate climate-selected loci in an Andean wetland plant. *Scientific Reports*, 10(1):1–8, 2020.

[34] Gaurav Bhatia, Nick Patterson, Sriram Sankararaman, and Alkes L Price. Estimating and interpreting F_{ST}: the impact of rare variants. *Genome Research*, 23(9):1514–1521, 2013.

[35] Maxime Bonhomme, Claude Chevalet, Bertrand Servin, Simon Boitard, Jihad Abdallah, Sarah Blott, and Magali SanCristobal. Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, 186(1):241–262, 2010.

[36] Aurelie Bonin, Florence Nicole, Francois Pompanon, Claude Miaud, and Pierre Taberlet. Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conservation Biology*, 21(3):697–708, 2007.

[37] Tom R Booker. Spatial autocorrelation of the environment influences the patterns and genetics of local adaptation. *bioRxiv*, pages 2023–05, 2023.

[38] Tom R Booker, Sam Yeaman, James R Whiting, and Michael C Whitlock. The wza: A window-based method for characterizing genotype–environment associations. *Molecular Ecology Resources*, 24(2):e13768, 2024.

[39] Tom R Booker, Sam Yeaman, and Michael C Whitlock. Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Molecular Ecology*, 29(22):4274–4279, 2020.

[40] David Botstein, Raymond L White, Mark Skolnick, and Ronald W Davis. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3):314, 1980.

[41] Madeleine C Bottrill, Liana N Joseph, Josie Carwardine, Michael Bode, Carly Cook, Edward T Game, Hedley Grantham, Salit Kark, Simon Linke, Eve McDonald-Madden, et al. Is conservation triage just smart decision making? *Trends in Ecology & Evolution*, 23(12):649–654, 2008.

[42] David Bryant and Vincent Moulton. Neighbornet: an agglomerative method for the construction of planar phylogenetic networks. In *Algorithms in Bioinformatics: Second International Workshop, WABI 2002 Rome, Italy, September 17–21, 2002 Proceedings 2*, pages 375–391. Springer, 2002.

[43] David Bryant and Vincent Moulton. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, 21(2):255–265, 2004.

[44] Armando Caballero and Miguel A Toro. Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, 3:289–299, 2002.

[45] John R Candy, Nathan R Campbell, Matthew H Grinnell, Terry D Beacham, Wesley A Larson, and Shawn R Narum. Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources*, 15(6):1421–1434, 2015.

[46] Thibaut Capblancq, Matthew C Fitzpatrick, Rachael A Bay, Moises Exposito-Alonso, and Stephen R Keller. Genomic prediction of (mal) adaptation across current and future climatic landscapes. *Annual Review of Ecology, Evolution, and Systematics*, 51:245–269, 2020.

[47] Gerardo Ceballos, Paul R Ehrlich, and Rodolfo Dirzo. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences*, 114(30):E6089–E6096, 2017.

[48] Brian Charlesworth, Magnus Nordborg, and Deborah Charlesworth. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research*, 70(2):155–174, 1997.

[49] Julio Chávez-Galarza, Dora Henriques, J Spencer Johnston, João C Azevedo, John C Patton, Irene Muñoz, Pilar De la Rúa, and M Alice Pinto. Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Molecular Ecology*, 22(23):5890–5907, 2013.

[50] Jun Chen, Thomas Källman, Xiaofei Ma, Niclas Gyllenstrand, Giusi Zaina, Michele Morgante, Jean Bousquet, Andrew Eckert, Jill Wegrzyn, David Neale, et al. Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics*, 191(3):865–881, 2012.

[51] Avneet Chhina, Niloufar Abhari, Arne Mooers, and Jayme Lewthwaite. Linking the spatial and genomic structure of adaptive potential for conservation management: a review. *Genome*, in press.

[52] Matthew J Christmas, Ed Biffin, Martin F Breed, and Andrew J Lowe. Finding needles in a genomic haystack: targeted capture identifies clear signatures of selection in a nonmodel plant species. *Molecular Ecology*, 25(17):4216–4233, 2016.

[53] Mi Yoon Chung, Sungwon Son, Jordi López-Pujol, Kangshan Mao, and Myong Gi Chung. Plant conservation practitioners can benefit from neutral genetic diversity. *Diversity*, 13(11):552, 2021.

[54] Els Coart, Sabine Van Glabeke, Rémy J Petit, Erik Van Bockstaele, and Isabel Roldan-Ruiz. Range wide versus local patterns of genetic diversity in hornbeam (*Carpinus betulus L.*). *Conservation Genetics*, 6(2):259–273, 2005.

[55] David J Coates, Margaret Byrne, and Craig Moritz. Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution*, 6:165, 2018.

[56] William Gemmell Cochran. *Sampling techniques*. john wiley & sons, 1977.

[57] RA Coleman, AR Weeks, and AA Hoffmann. Balancing genetic uniqueness and genetic variation in determining conservation and translocation strategies: a comprehensive case study of threatened dwarf galaxias, *Galaxiella pusilla* (Mack)(pisces: Galaxiidae). *Molecular Ecology*, 22(7):1820–1835, 2013.

[58] Pamela F Colosimo, Kim E Hosemann, Sarita Balabhadra, Guadalupe Villarreal Jr, Mark Dickson, Jane Grimwood, Jeremy Schmutz, Richard M Myers, Dolph Schluter, and David M Kingsley. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, 307(5717):1928–1933, 2005.

[59] COSEWIC. COSEWIC assessment and status report on the Half-moon Hairstreak *Satyrium semiluna* Okanagan-Similkameen and population Waterton Lakes population in Canada, 2022.

[60] Keith A Crandall, Olaf RP Bininda-Emonds, Georgina M Mace, and Robert K Wayne. Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, 15(7):290–295, 2000.

[61] Marco Crotti, Christopher D Barratt, Simon P Loader, David J Gower, and Jeffrey W Streicher. Causes and analytical impacts of missing data in RADseq phylogenetics: insights from an African frog (*Afrixalus*). *Zoologica Scripta*, 48(2):157–167, 2019.

[62] Ross H Crozier. Genetic diversity and the agony of choice. *Biological Conservation*, 61(1):11–15, 1992.

[63] Theresa M Culley, Lisa E Wallace, Karla M Gengler-Nowak, and Daniel J Crawford. A comparison of two methods of calculating GST, a genetic measure of population differentiation. *American Journal of Botany*, 89(3):460–465, 2002.

[64] Catherine I Cullingham, Janice EK Cooke, and David W Coltman. Effects of introgression on the genetic population structure of two ecologically and economically important conifer species: lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*Pinus banksiana*). *Genome*, 56(10):577–585, 2013.

[65] Catherine I Cullingham, Janice EK Cooke, and David W Coltman. Cross-species outlier detection reveals different evolutionary pressures between sister species. *New Phytologist*, 204(1):215–229, 2014.

[66] Xavier Dallaire, Éric Normandeau, Julien Mainguy, Jean-Éric Tremblay, Louis Bernatchez, and Jean-Sébastien Moore. Genomic data support management of anadromous arctic char fisheries in nunavik by highlighting neutral and putatively adaptive genetic variation. *Evolutionary Applications*, 14(7):1880–1897, 2021.

[67] Benjamin Dauphin, Christian Rellstab, Rafael O Wüest, Dirk N Karger, Rolf Holderegger, Felix Gugerli, and Stéphanie Manel. Re-thinking the environment in landscape genomics. *Trends in Ecology & Evolution*, 2023.

[68] H De Kort, K Vandepitte, J Mergeay, KV Mijnsbrugge, and O Honnay. The population genomic signature of environmental selection in the widespread insect-pollinated tree species *Frangula alnus* at different geographical scales. *Heredity*, 115(5):415–425, 2015.

[69] Hanne De Kort, Katrien Vandepitte, Hans Henrik Bruun, Déborah Closset-Kopp, Olivier Honnay, and Joachim Mergeay. Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Molecular Ecology*, 23(19):4709–4721, 2014.

[70] Claire Depardieu, Sébastien Gérardi, Simon Nadeau, Geneviève J Parent, John Mackay, Patrick Lenz, Manuel Lamothe, Martin P Girardin, Jean Bousquet, and Nathalie Isabel. Connecting tree-ring phenotypes, genetic associations and transcriptomics to decipher the genomic architecture of drought adaptation in a widespread conifer. *Molecular Ecology*, 30(16):3898–3917, 2021.

[71] J Andrew DeWoody, Avril M Harder, Samarth Mathur, and Janna R Willoughby. The long-standing significance of genetic diversity in conservation. *Molecular Ecology*, 30(17):4147–4154, 2021.

[72] Sandra Díaz, Noelia Zafra-Calvo, Andy Purvis, Peter H Verburg, David Obura, Paul Leadley, Rebecca Chaplin-Kramer, Luc De Meester, Ehsan Dulloo, Berta Martín-López, et al. Set ambitious goals for biodiversity and sustainability. *Science*, 370(6515):411–413, 2020.

[73] Nicolas Duforet-Frebourg, Keurcien Luu, Guillaume Laval, Eric Bazin, and Michael GB Blum. Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 genomes data. *Molecular Biology and Evolution*, 33(4):1082–1093, 2016.

[74] Sandra Díaz, Noelia Zafra-Calvo, Andy Purvis, Peter H Verburg, David Obura, Paul Leadley, Rebecca Chaplin-Kramer, Luc De Meester, Ehsan Dulloo, Berta Martín-López, et al. Set ambitious goals for biodiversity and sustainability. *Science*, 370(6515):411–413, 2020.

[75] Christoper G Eckert, Karen E Samis, and Stephen C Lougheed. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, 17(5):1170–1188, 2008.

[76] P.R. Ehrlich. The loss of diversity: Causes and consequences, in.: Wilson, eo, peter, fm (eds.), biodiversity, 1988.

[77] Amin Eimanifar, Samantha A Brooks, Tomas Bustamante, and James D Ellis. Population genomics and morphometric assignment of western honey bees (*Apis mellifera L.*) in the republic of South Africa. *BMC Genomics*, 19:1–26, 2018.

[78] Jonathan S Ellis, John Gilbey, A Armstrong, Torveig Balstad, Eef Cauwelier, Corinne Cherbonnel, Sofia Consuegra, Jamie Coughlan, Thomas F Cross, W Crozier, et al. Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (*Salmo salar L.*). *Genetica*, 139:353–367, 2011.

[79] Robert J Elshire, Jeffrey C Glaubitz, Qi Sun, Jesse A Poland, Ken Kawamoto, Edward S Buckler, and Sharon E Mitchell. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6(5):e19379, 2011.

[80] Laurent Excoffier, Tamara Hofer, and Matthieu Foll. Detecting loci under selection in a hierarchically structured population. *Heredity*, 103(4):285–298, 2009.

[81] M Exposito-Alonso, TA Booker, L Czech, T Fukami, L Gillespie, S Hateley, CC Kyriazis, P Lang, L Leventhal, and DJb Nogues-Bravo. Quantifying the scale of genetic diversity extinction in the anthropocene. *bioRxiv*, 2021.

[82] Moises Exposito-Alonso, Tom R Booker, Lucas Czech, Lauren Gillespie, Shannon Hateley, Christopher C Kyriazis, Patricia LM Lang, Laura Leventhal, David Nogues-Bravo, Veronica Pagowski, et al. Genetic diversity loss in the anthropocene. *Science*, 377(6613):1431–1435, 2022.

[83] Moises Exposito-Alonso, François Vasseur, Wei Ding, George Wang, Hernán A Burbano, and Detlef Weigel. Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. *Nature Ecology & Evolution*, 2(2):352–358, 2018.

[84] Daniel P Faith. Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1):1–10, 1992.

[85] Qingfu Feng, Joseph Abraham, Tao Feng, Yeunjoo Song, Robert C Elston, and Xiaofeng Zhu. A method to correct for population structure using a segregation model. In *BMC Proceedings*, volume 3, pages 1–5. Springer, 2009.

[86] Philippe Fernandez-Fournier, Jayme MM Lewthwaite, and Arne Ø Mooers. Do we need to identify adaptive genetic variation when prioritizing populations for conservation? *Conservation Genetics*, 22:205–216, 2021.

[87] Matthew C Fitzpatrick, Vikram E Chhatre, Raju Y Soolanayakanahally, and Stephen R Keller. Experimental support for genomic prediction of climate maladaptation using the machine learning approach gradient forests. *Molecular Ecology Resources*, 21(8):2749–2765, 2021.

[88] Matthew C Fitzpatrick and Stephen R Keller. Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, 18(1):1–16, 2015.

[89] Sarah W Fitzpatrick and Brendan N Reid. Does gene flow aggravate or alleviate maladaptation to environmental stress in small populations? *Evolutionary Applications*, 12(7):1402–1416, 2019.

[90] Sarah P Flanagan, Brenna R Forester, Emily K Latch, Sally N Aitken, and Sean Hoban. Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary Applications*, 11(7):1035–1052, 2018.

[91] Sarah P Flanagan, Emily Rose, and Adam G Jones. The population genomics of repeated freshwater colonizations by gulf pipefish. *Molecular Ecology*, 30(7):1672–1687, 2021.

[92] Félix Forest, Justin Moat, Elisabeth Baloch, Neil A Brummitt, Steve P Bachman, Steffi Ickert-Bond, Peter M Hollingsworth, Aaron Liston, Damon P Little, Sarah Mathews, et al. Gymnosperms on the edge. *Scientific Reports*, 8(1):6053, 2018.

[93] Brenna Forester and Tanya Lama. The role of genomics in the future of ESA decision-making. 2022.

[94] Brenna R Forester, Erik A Beever, Catherine Darst, Jennifer Szymanski, and W Chris Funk. Linking evolutionary potential to extinction risk: applications and future directions. *Frontiers in Ecology and the Environment*, 20(9):507–515, 2022.

[95] Richard Frankham. Conservation genetics. *Annual Review of Genetics*, 29(1):305–327, 1995.

[96] Richard Frankham, Corey JA Bradshaw, and Barry W Brook. Genetics in conservation management: revised recommendations for the 50/500 rules, red list criteria and population viability analyses. *Biological Conservation*, 170:56–63, 2014.

[97] Dylan J Fraser and Louis Bernatchez. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, 10(12):2741–2752, 2001.

[98] James Friel, Aureliano Bombarely, Carmen Dorca Fornell, Francisco Luque, and Ana Maria Fernández-Ocaña. Comparative analysis of genotyping by sequencing and whole-genome sequencing methods in diversity studies of *Olea europaea L.* *Plants*, 10(11):2514, 2021.

[99] Angela P Fuentes-Pardo and Daniel E Ruzzante. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, 26(20):5369–5406, 2017.

[100] Matteo Fumagalli, Filipe G Vieira, Thorfinn Sand Korneliussen, Tyler Linderoth, Emilia Huerta-Sánchez, Anders Albrechtsen, and Rasmus Nielsen. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*, 195(3):979–992, 2013.

[101] W Chris Funk, Robert E Lovich, Paul A Hohenlohe, Courtney A Hofman, Scott A Morrison, T Scott Sillett, Cameron K Ghalambor, Jesus E Maldonado, Torben C Rick, Mitch D Day, et al. Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). *Molecular Ecology*, 25(10):2176–2194, 2016.

[102] Aurora García-Dorado and Armando Caballero. Neutral genetic diversity as a useful tool for conservation biology. *Conservation Genetics*, 22(4):541–545, 2021.

[103] B Gebremedhin, GF Ficetola, S Naderi, H-R Rezaei, C Maudet, D Rioux, G Luikart, Ø Flagstad, W Thuiller, and P Taberlet. Frontiers in identifying conservation units: from neutral markers to adaptive genetic variation. *Animal Conservation*, 12(2):107–109, 2009.

[104] Arthur Georges, Luis Mijangos, Andrzej Kilian, Hardip Patel, Mark Aitkens, and Bernd Gruber. Distances and their visualization in studies of spatial-temporal genetic variation using single nucleotide polymorphisms (SNPs). *bioRxiv*, 2023.

[105] Armando Geraldes, Nima Farzaneh, Christopher J Grassa, Athena D McKown, Robert D Guy, Shawn D Mansfield, Carl J Douglas, and Quentin CB Cronk. Landscape genomics of *Populus trichocarpa*: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution*, 68(11):3260–3280, 2014.

[106] Sarah A Gignoux-Wolfsohn, Malin L Pinsky, Kathleen Kerwin, Carl Herzog, MacKenzie Hall, Alyssa B Bennett, Nina H Fefferman, and Brooke Maslo. Genomic signatures of selection in bats surviving white-nose syndrome. *Molecular Ecology*, 30(22):5643–5657, 2021.

[107] A Gimelfarb. Pleiotropy as a factor maintaining genetic variation in quantitative characters under stabilizing selection. *Genetics Research*, 68(1):65–73, 1996.

[108] Alicia Gómez-Fernández, Irene Alcocer, and Silvia Matesanz. Does higher connectivity lead to higher genetic diversity? Effects of habitat fragmentation on genetic variation and population structure in a gypsophile. *Conservation Genetics*, 17:631–641, 2016.

[109] Jérôme Goudet. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1):184–186, 2005.

[110] H-R Gregorius. The relationship between the concepts of genetic diversity and differentiation. *Theoretical and Applied Genetics*, 74:397–401, 1987.

[111] Terri Grodzicker, J Williams, P Sharp, and J Sambrook. Physical mapping of temperature-sensitive mutations of adenoviruses. In *Cold Spring Harbor symposia on quantitative biology*, volume 39, pages 439–446. Citeseer, 1974.

[112] Jared A Grummer, Tom R Booker, Remi Matthey-Doret, Pirmin Nietlisbach, Andréa T Thomaz, and Michael C Whitlock. The immediate costs and long-term benefits of assisted gene flow in large populations. *Conservation Biology*, 36(4):e13911, 2022.

[113] F Gugerli, T Englisch, H Niklfeld, A Tribsch, Z Mirek, M Ronikier, Niklaus E Zimmermann, R Holderegger, Pierre Taberlet, IntraBioDiv Consortium, et al. Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation–a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics*, 10(4):259–281, 2008.

[114] Torsten Günther and Graham Coop. Robust identification of local adaptation from allele frequencies. *Genetics*, 195(1):205–220, 2013.

[115] Baocheng Guo, Di Lu, Wen Bo Liao, and Juha Merilä. Genomewide scan for adaptive differentiation along altitudinal gradient in the andrew's toad *Bufo andrewsi*. *Molecular Ecology*, 25(16):3884–3900, 2016.

[116] Claus-Jochen Haake, Akemi Kashiwada, and Francis Edward Su. The shapley value of phylogenetic trees. *Journal of Mathematical Biology*, 56(4):479–497, 2008.

[117] Jennafer AP Hamlin and Michael L Arnold. Neutral and selective processes drive population differentiation for *Iris hexagona*. *Journal of Heredity*, 106(5):628–636, 2015.

[118] Bengt Hansson and Lars Westerberg. On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology*, 11(12):2467–2474, 2002.

[119] J Hard, Jeffrey. A quantitative genetic perspective on the conservation of intraspecific diversity. 1995.

[120] Matthew P Hare, Leonard Nunney, Michael K Schwartz, Daniel E Ruzzante, Martha Burford, Robin S Waples, Kristen Ruegg, and Friso Palstra. Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology*, 25(3):438–449, 2011.

[121] Katherine A Harrisson, Alexandra Pavlova, Anders Gonçalves da Silva, Rebecca Rose, James K Bull, Melanie L Lancaster, Neil Murray, Bruce Quin, Peter Menkhorst, Michael JL Magrath, et al. Scope for genetic rescue of an endangered subspecies though re-establishing natural gene flow with another subspecies. *Molecular Ecology*, 25(6):1242–1258, 2016.

[122] Katherine A Harrisson, Alexandra Pavlova, Marina Telonis-Scott, and Paul Sunnucks. Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, 7(9):1008–1025, 2014.

[123] Klaas Hartmann. The equivalence of two phylogenetic biodiversity measures: the shapley value and fair proportion index. *Journal of Mathematical Biology*, 67:1163–1170, 2013.

[124] Klaas Hartmann and Mike Steel. Maximizing phylogenetic diversity in biodiversity conservation: greedy solutions to the noah's ark problem. *Systematic Biology*, 55(4):644–651, 2006.

[125] Jan Hauke and Tomasz Kossowski. Comparison of values of pearson's and spearman's correlation coefficients. *Quaestiones Geographicae*, 2011.

[126] Laura Katharine Hayward and Guy Sella. Polygenic adaptation after a sudden change in environment. *Elife*, 11:e66697, 2022.

[127] Tianhua He, Haylee D'Agui, Sim Lin Lim, Neal J Enright, and Yiqi Luo. Evolutionary potential and adaptation of *Banksia attenuata* (Proteaceae) to climate and fire regime in southwestern Australia, a global biodiversity hotspot. *Scientific Reports*, 6(1):26315, 2016.

[128] Philip V Hedrick. Techniques: A perspective. *Molecular genetic approaches in conservation*, page 459, 1996.

[129] Philip W Hedrick. Purging inbreeding depression and the probability of extinction: full-sib mating. *Heredity*, 73(4):363–372, 1994.

[130] Philip W Hedrick. A standardized genetic differentiation measure. *Evolution*, 59(8):1633–1638, 2005.

[131] Philip W Hedrick. Balancing selection. *Current Biology*, 17(7):R230–R231, 2007.

[132] Philip W Hedrick, Daniel R Stahler, and Dick Dekker. Heterozygote advantage in a finite population: black color in wolves. *Journal of Heredity*, 105(4):457–465, 2014.

[133] Maxime Hervé. Rvaidememoire: testing and plotting procedures for biostatistics. r package version 0.9-73. *The R Foundation*, 2020.

[134] Jon E Hess, Nathan R Campbell, David A Close, Margaret F Docker, and Shawn R Narum. Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, 22(11):2898–2916, 2013.

[135] Sean Hoban, Frederick I Archer, Laura D Bertola, Jason G Bragg, Martin F Breed, Michael W Bruford, Melinda A Coleman, Robert Ekblom, W Chris Funk, Catherine E Grueber, et al. Global genetic diversity status and trends: towards a suite of essential biodiversity variables (EBVs) for genetic composition. *Biological Reviews*, 97(4):1511–1538, 2022.

[136] Sean Hoban, Michael Bruford, Josephine D'Urban Jackson, Margarida Lopes-Fernandes, Myriam Heuertz, Paul A Hohenlohe, Ivan Paz-Vinas, Per Sjögren-Gulve, Gernot Segelbacher, and Cristiano Vernesi. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation*, 248:108654, 2020.

[137] Sean Hoban, Michael W Bruford, W Chris Funk, Peter Galbusera, M Patrick Griffith, Catherine E Grueber, Myriam Heuertz, Margaret E Hunter, Christina Hvilsom, Belma Kalamujic Stroil, et al. Global commitments to conserving and monitoring genetic diversity are now necessary and feasible. *Bioscience*, 71(9):964–976, 2021.

[138] Sean Hoban, Joanna L Kelley, Katie E Lotterhos, Michael F Antolin, Gideon Bradburd, David B Lowry, Mary L Poss, Laura K Reed, Andrew Storfer, and Michael C Whitlock. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, 188(4):379–397, 2016.

[139] Ary A Hoffmann and Carla M Sgrò. Climate change and evolutionary adaptation. *Nature*, 470(7335):479–485, 2011.

[140] J Hoglund. The extinction vortex, is genetic variation related to extinction. *Evolutionary Conservation Genetics*, 6:45–66, 2009.

[141] Jacob Höglund. *Evolutionary Conservation Genetics*. Oxford University Press, 2009.

[142] Rolf Holderegger, Urs Kamm, and Felix Gugerli. Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, 21:797–807, 2006.

[143] Jason A Holliday, Kermit Ritland, and Sally N Aitken. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist*, 188(2):501–514, 2010.

[144] Kent E Holsinger and Bruce S Weir. Genetics in geographically structured populations: defining, estimating and interpreting f st. *Nature Reviews Genetics*, 10(9):639–650, 2009.

[145] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417, 1933.

[146] John L Hubby and Richard C Lewontin. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics*, 54(2):577, 1966.

[147] Richard R Hudson, Montgomery Slatkin, and Wayne P Maddison. Estimation of levels of gene flow from dna sequence data. *Genetics*, 132(2):583–589, 1992.

[148] John Hunt, Mark W Blows, Felix Zajitschek, Michael D Jennions, and Robert Brooks. Reconciling strong stabilizing selection with the maintenance of genetic variation in a natural population of black field crickets (teleogryllus commodus). *Genetics*, 177(2):875–880, 2007.

[149] Agathe Hurel, Marina de Miguel, Cyril Dutech, Marie-Laure Desprez-Loustau, Christophe Plomion, Isabel Rodríguez-Quilón, Agathe Cyrille, Thomas Guzman, Ricardo Alía, Santiago C González-Martínez, et al. Genetic basis of growth, spring phenology, and susceptibility to biotic stressors in maritime pine. *Evolutionary Applications*, 14(12):2750–2772, 2021.

[150] Daniel H Huson and David Bryant. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2):254–267, 2006.

[151] Nick JB Isaac, Samuel T Turvey, Ben Collen, Carly Waterman, and Jonathan EM Baillie. Mammals on the edge: conservation priorities based on threat and phylogeny. *PLoS One*, 2(3):e296, 2007.

[152] T. Jombart. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24:1403–1405, 2008.

[153] T. Jombart and I. Ahmed. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 2011.

[154] Alistair S Jump, Rob Marchant, and Josep Peñuelas. Environmental change and the option value of genetic diversity. *Trends in Plant Science*, 14(1):51–58, 2009.

[155] Andrew N Kadykalo, Steven J Cooke, and Nathan Young. Conservation genomics from a practitioner lens: Evaluating the research-implementation gap in a managed freshwater fishery. *Biological Conservation*, 241:108350, 2020.

[156] KK Kanaka, Nidhi Sukhija, Rangasai Chandra Goli, Sanjeev Singh, Indrajit Ganguly, SP Dixit, Aishwarya Dash, and Anoop Anand Malik. On the concepts and measures of diversity in the genomics era. *Current Plant Biology*, 33:100278, 2023.

[157] Marty Kardos, Ellie E. Armstrong, Sarah W. Fitzpatrick, Samantha Hauser, Philip W. Hedrick, Joshua M. Miller, David A. Tallmon, and W. Chris Funk. The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences*, 118(48), 2021.

[158] Alboukadel Kassambara. Rstatix: Pipe-friendly framework for basic statistical tests. *Package 'rstatix'*, Feb 2023.

[159] Tadeusz J Kawecki and Dieter Ebert. Conceptual issues in local adaptation. *Ecology Letters*, 7(12):1225–1241, 2004.

[160] Stephen R Keller, Vikram E Chhatre, and Matthew C Fitzpatrick. Influence of range position on locally adaptive gene–environment associations in populus flowering time genes. *Journal of Heredity*, 109(1):47–58, 2018.

[161] Stephen R Keller, Nicholas Levsen, Matthew S Olson, and Peter Tiffin. Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera L. Molecular Biology and Evolution*, 29(10):3143–3152, 2012.

[162] Kyung Seok Kim and Thomas W Sappington. Microsatellite data analysis for population genetics. *Microsatellites: Methods and protocols*, pages 271–295, 2013.

[163] Motoo Kimura and James F Crow. The number of alleles that can be maintained in a finite population. *Genetics*, 49(4):725, 1964.

[164] Heather Kirk and Joanna R Freeland. Applications and implications of neutral versus non-neutral markers in molecular ecology. *International Journal of Molecular Sciences*, 12(6):3966–3988, 2011.

[165] Shuichi Kitada, Reiichiro Nakamichi, and Hirohisa Kishino. Understanding population structure in an evolutionary context: population-specific F_{ST} and pairwise F_{ST}. *G3*, 11(11):jkab316, 2021.

[166] Amanda Kowalczyk, Maria Chikina, and Nathan Clark. Complementary evolution of coding and noncoding sequence underlies mammalian hairlessness. *Elife*, 11:e76911, 2022.

[167] Yu-Ting Lai, Carol KL Yeung, Kevin E Omland, Er-Li Pang, YU Hao, Ben-Yang Liao, Hui-Fen Cao, Bo-Wen Zhang, Chia-Fen Yeh, Chih-Ming Hung, et al. Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences*, 116(6):2152–2157, 2019.

[168] Linda Laikre, Sean Hoban, Michael W Bruford, Gernot Segelbacher, Fred W Allendorf, Gonzalo Gajardo, Antonio González Rodríguez, Philip W Hedrick, Myriam Heuertz, Paul A Hohenlohe, et al. Post-2020 goals overlook genetic diversity. *Science*, 367(6482):1083–1085, 2020.

[169] Chen-Ruei Lee and Thomas Mitchell-Olds. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology*, 20(22):4631–4642, 2011.

[170] Chih Lee, Ali Abdool, and Chun-Hsi Huang. PCA-based population structure inference with generic clustering algorithms. *BMC Bioinformatics*, 10(1):1–13, 2009.

[171] Sarah J Lehnert, Ian R Bradbury, Brendan F Wringe, Mallory Van Wyngaarden, and Paul Bentzen. Multifaceted framework for defining conservation units: An example from Atlantic salmon (*Salmo salar*) in Canada. *Evolutionary Applications*, 16(9):1568–1585, 2023.

[172] Roosa Leimu and Markus Fischer. A meta-analysis of local adaptation in plants. *PLoS One*, 3(12):e4010, 2008.

[173] Tuomas Leinonen, Robert B O'HARA, José Manuel Cano, and Juha Merilä. Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, 21(1):1–17, 2008.

[174] Thomas Lenormand. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17(4):183–189, 2002.

[175] Peter Lesica and Fred W Allendorf. When are peripheral populations valuable for conservation? *Conservation Biology*, 9(4):753–760, 1995.

[176] Richard C Lewontin. *The genetic basis of evolutionary change*, volume 560. Columbia University Press New York, 1974.

[177] Richard C Lewontin and Jack L Hubby. A molecular approach to the study of genic heterozygosity in natural populations. ii. amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, 54(2):595, 1966.

[178] Richard C Lewontin and Jesse Krakauer. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 74(1):175–195, 1973.

[179] Shan Li, Xiaohong Gan, Hongyan Han, Xuemei Zhang, and Zhongqiong Tian. Low within-population genetic diversity and high genetic differentiation among populations of the endangered plant tetracentron sinense oliver revealed by inter-simple sequence repeat analysis. *Annals of Forest Science*, 75(3):1–11, 2018.

[180] Yanping Li, Christopher P Burridge, Yunyun Lv, and Zuogang Peng. Morphometric and population genomic evidence for species divergence in the *Chimarrichthys* fish complex of the Tibetan Plateau. *Molecular Phylogenetics and Evolution*, 159:107117, 2021.

[181] Erin Liddell, Paul Sunnucks, and Carly N Cook. To mix or not to mix gene pools for threatened species management? Few studies use genetic data to examine the risks of both actions, but failing to do so leads disproportionately to recommendations for separate management. *Biological Conservation*, 256:109072, 2021.

[182] Brandon M Lind, Mitra Menon, Constance E Bolte, Trevor M Faske, and Andrew J Eckert. The genomics of local adaptation in trees: are we out of the woods yet? *Tree Genetics & Genomes*, 14:1–30, 2018.

[183] Katie E Lotterhos and Michael C Whitlock. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, 23(9):2178–2192, 2014.

[184] David B Lowry, Sean Hoban, Joanna L Kelley, Katie E Lotterhos, Laura K Reed, Michael F Antolin, and Andrew Storfer. Breaking RAD: an evaluation of the utility of restriction site-associated dna sequencing for genome scans of adaptation. *Molecular Ecology Resources*, 17(2), 2017.

[185] Gordon Luikart, Phillip R England, David Tallmon, Steve Jordan, and Pierre Taberlet. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4(12):981–994, 2003.

[186] Michael Lynch. A quantitative-genetic perspective on conservation issues. *Conservation Genetics*, pages 471–501, 1995.

[187] Michael Lynch and Wilfried Gabriel. Mutation load and the survival of small populations. *Evolution*, 44(7):1725–1737, 1990.

[188] Jianzhong Ma and Christopher I Amos. Principal components analysis of population admixture. *PLoS One*, 7(7):e40115, 2012.

[189] Barbara K Mable. Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conservation Genetics*, 20(1):89–100, 2019.

[190] Colin R Mahony, Ian R MacLachlan, Brandon M Lind, Jeremy B Yoder, Tongli Wang, and Sally N Aitken. Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. *Evolutionary Applications*, 13(1):116–131, 2020.

[191] DR Marshall and SK Jain. Genetic polymorphism in natural populations of *Avena fatua* and *A. barbata*. *Nature*, 221(5177):276–278, 1969.

[192] Iain Martyn, Tyler S Kuhn, Arne O Mooers, Vincent Moulton, and Andreas Spillner. Computing evolutionary distinctiveness indices in large scale analysis. *Algorithms for Molecular Biology*, 7:1–7, 2012.

[193] Samarth Mathur, Andrew J Mason, Gideon S Bradburd, and H Lisle Gibbs. Functional genomic diversity is correlated with neutral genomic diversity in populations of an endangered rattlesnake. *Proceedings of the National Academy of Sciences*, 120(43):e2303043120, 2023.

[194] Athena D McKown, Jaroslav Klápště, Robert D Guy, Armando Geraldes, Ilga Porth, Jan Hannemann, Michael Friedmann, Wellington Muchero, Gerald A Tuskan, Jürgen Ehlting, et al. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytologist*, 203(2):535–553, 2014.

[195] Barry J McMahon, Emma C Teeling, and Jacob Höglund. How and why should we implement genomics into conservation? *Evolutionary Applications*, 7(9):999–1007, 2014.

[196] Tamra C Mendelson and Kerry L Shaw. Use of AFLP markers in surveys of arthropod diversity. In *Methods in Enzymology*, volume 395, pages 161–177. Elsevier, 2005.

[197] J Merilä and P Crnokrak. Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, 14(6):892–903, 2001.

[198] Ilaria Milano, Massimiliano Babbucci, Alessia Cariani, Miroslava Atanassova, Dorte Bekkevold, Gary R Carvalho, Montserrat Espiñeira, Fabio Fiorentino, Germana Garofalo, Audrey J Geffen, et al. Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, 23(1):118–135, 2014.

[199] Constance I Millar and William J Libby. Strategies for conserving clinal, ecotypic, and disjunct population diversity in widespread species. *Genetics and Conservation of Rare Plants*, 149:170, 1991.

[200] Philip S Miller and Philip W Hedrick. MHC polymorphism and the design of captive breeding programs: simple solutions are not the answer. *Conservation Biology*, 5(4):556–558, 1991.

[201] L Scott Mills and Fred W Allendorf. The one-migrant-per-generation rule in conservation and management. *Conservation Biology*, 10(6):1509–1518, 1996.

[202] Elizabeth A Mittell, Shinichi Nakagawa, and Jarrod D Hadfield. Are molecular markers useful predictors of adaptive potential? *Ecology Letters*, 18(8):772–778, 2015.

[203] Daniel Money, Kyle Gardner, Zoë Migicovsky, Heidi Schwaninger, Gan-Yuan Zhong, and Sean Myles. Linkimpute: fast and accurate genotype imputation for nonmodel organisms. *G3: Genes, Genomes, Genetics*, 5(11):2383–2390, 2015.

[204] Jean-Sébastien Moore, Vincent Bourret, Mélanie Dionne, Ian Bradbury, Patrick O'Reilly, Matthew Kent, Gérald Chaput, and Louis Bernatchez. Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci

identify the same patterns of population structure as neutral loci. *Molecular Ecology*, 23(23):5680–5697, 2014.

[205] Craig Moritz. Conservation units and translocations: strategies for conserving evolutionary processes. *Hereditas*, 130(3):217–228, 1999.

[206] Craig Moritz. Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, 51(2):238–254, 2002.

[207] Elena Mosca, AJ Eckert, EA Di Pierro, D Rocchini, Nicola La Porta, Piero Belletti, and DB Neale. The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular Ecology*, 21(22):5530–5545, 2012.

[208] Elena Mosca, Felix Gugerli, Andrew J Eckert, and David B Neale. Signatures of natural selection on *Pinus cembra* and *P. mugo* along elevational gradients in the alps. *Tree Genetics & Genomes*, 12:1–15, 2016.

[209] Masatoshi Nei. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12):3321–3323, 1973.

[210] Masatoshi Nei. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89(3):583–590, 1978.

[211] Masatoshi Nei. Definition and estimation of fixation indices. *Evolution*, 40(3):643–645, 1986.

[212] Masatoshi Nei and Wen-Hsiung Li. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76(10):5269–5273, 1979.

[213] Masatoshi Nei, Takeo Maruyama, and Ranajit Chakraborty. The bottleneck effect and genetic variability in populations. *Evolution*, pages 1–10, 1975.

[214] Masatoshi Nei and ANDA K Roychoudhury. Sampling variances of heterozygosity and genetic distance. *Genetics*, 76(2):379–390, 1974.

[215] Erica S Nielsen, Maria Beger, Romina Henriques, and Sophie von der Heyden. A comparison of genetic and genomic approaches to represent evolutionary potential in conservation planning. *Biological Conservation*, 251:108770, 2020.

[216] Patrik Nosil, Scott P Egan, and Daniel J Funk. Heterogeneous genomic differentiation between walking-stick ecotypes:"isolation by adaptation" and multiple roles for divergent selection. *Evolution*, 62(2):316–336, 2008.

[217] Patrik Nosil, Luke J Harmon, and Ole Seehausen. Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, 24(3):145–156, 2009.

[218] Kevin P Oh, Cameron L Aldridge, Jennifer S Forbey, Carolyn Y Dadabay, and Sara J Oyler-McCance. Conservation genomics in the Sagebrush Sea: population divergence, demographic history, and local adaptation in sage-grouse (*Centrocercus* spp.). *Genome Biology and Evolution*, 11(7):2023–2034, 2019.

[219] Taras K Oleksyk, Michael W Smith, and Stephen J O'Brien. Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537):185–205, 2010.

[220] Kym M Ottewell, Doug C Bickerton, Margaret Byrne, and Andrew J Lowe. Bridging the gap: A genetic assessment framework for population-level threatened plant conservation prioritization and decision-making. *Diversity and Distributions*, 22(2):174–188, 2016.

[221] John R Pannell and Brian Charlesworth. Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1404):1851–1864, 2000.

[222] Peristera Paschou, Elad Ziv, Esteban G Burchard, Shweta Choudhry, William Rodriguez-Cintron, Michael W Mahoney, and Petros Drineas. PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genetics*, 3(9):e160, 2007.

[223] Nick Patterson, Alkes L Price, and David Reich. Population structure and eigenanalysis. *PLoS Genetics*, 2(12):e190, 2006.

[224] Alexandra Pavlova, Sara Petrovic, Katherine A Harrisson, Karina Cartwright, Elizabeth Dobson, Laura L Hurley, Meagan Lane, Michael JL Magrath, Kimberly A Miller, Bruce Quin, et al. Benefits of genetic rescue of a critically endangered subspecies from another subspecies outweigh risks: Results of captive breeding trials. *Biological Conservation*, 284:110203, 2023.

[225] Devon E Pearse. Saving the spandrels? Adaptive genomic variation in conservation and fisheries management. *Journal of Fish Biology*, 89(6):2697–2716, 2016.

[226] Devon E Pearse, Michael R Miller, Alicia Abadía-Cardoso, and John Carlos Garza. Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. *Proceedings of the Royal Society B: Biological Sciences*, 281(1783):20140012, 2014.

[227] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[228] Carlo Pecoraro, Massimiliano Babbucci, Rafaella Franch, Ciro Rico, Chiara Papetti, Emmanuel Chassot, Nathalie Bodin, Alessia Cariani, Luca Bargelloni, and Fausto Tinti. The population genomics of yellowfin tuna (*Thunnus albacares*) at global geographic scale challenges current stock delineation. *Scientific Reports*, 8(1):13890, 2018.

[229] Christophe Pelabon, Thomas F Hansen, Ashley JR Carter, and David Houle. Evolution of variation and variability under fluctuating, stabilizing, and disruptive selection. *Evolution*, 64(7):1912–1925, 2010.

[230] A Pérez-Figueroa, M Saura, J Fernández, MA Toro, and Armando Caballero. Metapop—a software for the management and analysis of subdivided populations in conservation programs. *Conservation Genetics*, 10:1097–1099, 2009.

[231] Rémy J Petit, Abdelhamid El Mousadik, and Odile Pons. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, 12(4):844–855, 1998.

[232] Robert H Podolsky. Genetic variation for morphological and allozyme variation in relation to population size in *Clarkia dudleyana*, an endemic annual. *Conservation Biology*, 15(2):412–423, 2001.

[233] Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8):904–909, 2006.

[234] Allan Raffard, Frédéric Santoul, Julien Cucherousset, and Simon Blanchet. The community and ecosystem consequences of intraspecific diversity: A meta-analysis. *Biological Reviews*, 94(2):648–661, 2019.

[235] Katherine Ralls, Jonathan D Ballou, Michele R Dudash, Mark DB Eldridge, Charles B Fenster, Robert C Lacy, Paul Sunnucks, and Richard Frankham. Call for a paradigm shift in the genetic management of fragmented populations. *Conservation Letters*, 11(2):e12412, 2018.

[236] David W Redding and AØ Mooers. Can systematists help decide the relative worth of bits of biodiversity. *Systematist*, 32:4–8, 2010.

[237] David H Reed and Richard Frankham. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution*, 55(6):1095–1103, 2001.

[238] David H Reed and Richard Frankham. Correlation between fitness and genetic diversity. *Conservation Biology*, 17(1):230–237, 2003.

[239] Noah M Reid, Dina A Proestou, Bryan W Clark, Wesley C Warren, John K Colbourne, Joseph R Shaw, Sibel I Karchner, Mark E Hahn, Diane Nacci, Marjorie F Oleksiak, et al. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science*, 354(6317):1305–1308, 2016.

[240] LA Rocha, MT Craig, and BW Bowen. Phylogeography and the conservation of coral reef fishes. *Coral Reefs*, 26:501–512, 2007.

[241] Matthew V Rockman. The qtn program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, 66(1):1–17, 2012.

[242] Gretchen H Roffler, Stephen J Amish, Seth Smith, TED Cosart, Marty Kardos, Michael K Schwartz, and Gordon Luikart. SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate. *Molecular Ecology Resources*, 16(5):1147–1164, 2016.

[243] Anne M Royer, Matthew A Streisfeld, and Christopher Irwin Smith. Population genomics of divergence within an obligate pollination mutualism: Selection maintains differences between joshua tree species. *American Journal of Botany*, 103(10):1730–1741, 2016.

[244] Kristen Ruegg, Eric C Anderson, Marius Somveille, Rachael A Bay, Mary Whitfield, Eben H Paxton, and Thomas B Smith. Linking climate niches across seasons to assess population vulnerability in a migratory bird. *Global Change Biology*, 27(15):3519–3531, 2021.

[245] Kristen Ruegg, Rachael A Bay, Eric C Anderson, James F Saracco, Ryan J Harrigan, Mary Whitfield, Eben H Paxton, and Thomas B Smith. Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology Letters*, 21(7):1085–1096, 2018.

[246] OA Ryder. Species conservation and systematics: The dilemma of subspecies. *Trends in Ecology & Evolution*, 1:9–10, 1986.

[247] Pardis C Sabeti, Patrick Varilly, Ben Fry, Jason Lohmueller, Elizabeth Hostetter, Chris Cotsapas, Xiaohui Xie, Elizabeth H Byrne, Steven A McCarroll, Rachelle Gaudet, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164):913–918, 2007.

[248] Jonathan Sandoval-Castillo, Nick A Robinson, Anthony M Hart, Lachlan WS Strain, and Luciano B Beheregaray. Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient. *Molecular Ecology*, 27(7):1603–1620, 2018.

[249] Outi Savolainen, Martin Lascoux, and Juha Merilä. Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11):807–820, 2013.

[250] Outi Savolainen, Tanja Pyhäjärvi, and Timo Knürr. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, 38(1):595–619, 2007.

[251] Stephen W Scherer, Charles Lee, Ewan Birney, David M Altshuler, Evan E Eichler, Nigel P Carter, Matthew E Hurles, and Lars Feuk. Challenges and standards in integrating surveys of structural variation. *Nature Genetics*, 39(Suppl 7):S7–S15, 2007.

[252] Schliep, Klaus, Potts, Alastair J., Morrison, David A., Grimm, and Guido W. Intertwining phylogenetic trees and networks. *Methods in Ecology and Evolution*, 8(10):1212–1220, 2017.

[253] Klaus Schliep, Alastair J. Potts, David A. Morrison, and Guido W. Grimm. Intertwining phylogenetic trees and networks. *Methods in Ecology and Evolution*, 8(10):1212–1220, Apr 2017.

[254] Klaus Peter Schliep. phangorn: phylogenetic analysis in R. *Bioinformatics*, 27(4):592–593, 2011.

[255] K.P. Schliep. phangorn: phylogenetic analysis in r. *Bioinformatics*, 27(4):592–593, 2011.

[256] Christian Schlötterer. The evolution of molecular markers—just a matter of fashion? *Nature Reviews Genetics*, 5(1):63–69, 2004.

[257] Dolph Schluter. Rtips: Genomics.

[258] Thomas L Schmidt, Moshe-Elijah Jasper, Andrew R Weeks, and Ary A Hoffmann. Unbiased population heterozygosity estimates from genome-wide sequence data. *Methods in Ecology and Evolution*, 12(10):1888–1898, 2021.

[259] Sean D Schoville, Aurélie Bonin, Olivier François, Stéphane Lobreaux, Christelle Melodelima, and Stéphanie Manel. Adaptive genetic variation on the landscape: methods and cases. *Annual Review of Ecology, Evolution, and Systematics*, 43:23–43, 2012.

[260] Rena M Schweizer, Jacqueline Robinson, Ryan Harrigan, Pedro Silva, Marco Galverni, Marco Musiani, Richard E Green, John Novembre, and Robert K Wayne. Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Molecular Ecology*, 25(1):357–379, 2016.

[261] Guy Sella and Nicholas H Barton. Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annual Review of Genomics and Human Genetics*, 20:461–493, 2019.

[262] Aaron BA Shafer, Jochen BW Wolf, Paulo C Alves, Linnea Bergström, Michael W Bruford, Ioana Brännström, Guy Colling, Love Dalén, Luc De Meester, Robert Ekblom, et al. Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, 30(2):78–87, 2015.

[263] Lloyd S Shapley. A value for n-person games. 1953.

[264] Andrew J Shirk and Samuel A Cushman. Spatially-explicit estimation of Wright's neighborhood size in continuous populations. *Frontiers in Ecology and Evolution*, 2:62, 2014.

[265] David Sims, Ian Sudbery, Nicholas E Ilott, Andreas Heger, and Chris P Ponting. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2):121–132, 2014.

[266] Montgomery Slatkin. Gene flow and the geographic structure of natural populations. *Science*, 236(4803):787–792, 1987.

[267] Thomas B Smith and Robert K Wayne. *Molecular genetic approaches in conservation*. Oxford University Press, 1996.

[268] R.R. Sokal and F.J. Rohlf. *Biometry: The Principles and Practice of Statistics in Biological Research*. W. H. Freeman, 1981.

[269] VL Sork, SN Aitken, RJ Dyer, AJ Eckert, P Legendre, and DB Neale. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, 9:901–911, 2013.

[270] Derek Spielman, Barry W Brook, and Richard Frankham. Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences*, 101(42):15261–15264, 2004.

[271] Wolfgang Stacklies, Henrik Redestig, Michael Scholz, Dirk Walther, and Joachim Selbig. *pcaMethods: A Bioconductor package providing PCA methods for incomplete data*, 2016. R package version 1.62.0.

[272] Wolfram Stacklies and Henning Redestig. The pcamethods package, 2022.

[273] Wolfram Stacklies, Henning Redestig, Matthias Scholz, Dirk Walther, and Joachim Selbig. pcamethods—a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, 23(9):1164–1167, 2007.

[274] Mike Steel. Phylogenetic diversity and the greedy algorithm. *Systematic Biology*, 54(4):527–529, 2005.

[275] Tammy E Steeves, Jeff A Johnson, and Marie L Hale. Maximising evolutionary potential in functional proxies for extinct species: a conservation genetic perspective on de-extinction. *Functional Ecology*, 31(5):1032–1040, 2017.

[276] Andrew Storfer. Gene flow and endangered species translocations: a topic revisited. *Biological Conservation*, 87(2):173–180, 1999.

[277] Katarina C Stuart, William B Sherwin, Adam PA Cardilini, and Lee A Rollins. Genetics and plasticity are responsible for ecogeographical patterns in a recent invasion. *Frontiers in Genetics*, 13:824424, 2022.

[278] J Swaegers, J Mergeay, A Van Geystelen, L Therry, MHD Larmuseau, and R Stoks. Neutral and adaptive genomic signatures of rapid poleward range expansion. *Molecular Ecology*, 24(24):6163–6176, 2015.

[279] Eric B Taylor, Patrick Tamkee, Ernest R Keeley, and Eric A Parkinson. Conservation prioritization in widespread species: the use of genetic and morphological data to assess population distinctiveness in rainbow trout (*Oncorhynchus mykiss*) from british columbia, canada. *Evolutionary Applications*, 4(1):100–115, 2011.

[280] R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing. *Computing*, 2021.

[281] João C Teixeira and Christian D Huber. The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences*, 118(10):e2015096118, 2021.

[282] Chao Tian, Robert M Plenge, Michael Ransom, Annette Lee, Pablo Villoslada, Carlo Selmi, Lars Klareskog, Ann E Pulver, Lihong Qi, Peter K Gregersen, et al. Analysis and application of European genetic substructure using 300 k SNP information. *PLoS Genetics*, 4(1):e4, 2008.

[283] Anna Tigano and Vicki L Friesen. Genomics of local adaptation with gene flow. *Molecular Ecology*, 25(10):2144–2164, 2016.

[284] MA Toro and A Caballero. Characterization and conservation of genetic diversity in subdivided populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459):1367–1378, 2005.

[285] Craig A Tracy and Harold Widom. Level-spacing distributions and the Airy kernel. *Communications in Mathematical Physics*, 159:151–174, 1994.

[286] Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B Altman. Missing value estimation methods for dna microarrays. *Bioinformatics*, 17(6):520–525, 2001.

[287] Tom van der Valk, Marc de Manuel, Tomas Marques-Bonet, and Katerina Guschanski. Estimates of genetic load suggest frequent purging of deleterious alleles in small populations. *bioRxiv*, page 696831, 2019.

[288] Joost van Heerwaarden, TL Odong, and FA van Eeuwijk. Maximizing genetic differentiation in core collections by PCA-based clustering of molecular marker data. *Theoretical and Applied Genetics*, 126:763–772, 2013.

[289] Richard I Vane-Wright, Christopher J Humphries, and Paul H Williams. What to protect?—systematics and the agony of choice. *Biological Conservation*, 55(3):235–254, 1991.

[290] S Volis, B Yakubov, I Shulgina, D Ward, and S Mendlinger. Distinguishing adaptive from nonadaptive genetic differentiation: comparison of Q_{ST} and F_{ST} at two spatial scales. *Heredity*, 95(6):466–475, 2005.

[291] Logan Volkmann, Iain Martyn, Vincent Moulton, Andreas Spillner, and Arne O Mooers. Prioritizing populations for conservation using phylogenetic networks. *PLoS One*, 9(2):e88945, 2014.

[292] S von der Heyden. Why do we need to integrate population genetics into South African marine protected area planning? *African Journal of Marine Science*, 31(2):263–269, 2009.

[293] Sofie Vranken, Thomas Wernberg, Armin Scheben, Anita Alice Severn-Ellis, Jacqueline Batley, Philipp Emanuel Bayer, David Edwards, David Wheeler, and Melinda Ann Coleman. Genotype–environment mismatch of kelp forests under climate change. *Molecular Ecology*, 30(15):3730–3746, 2021.

[294] Lisette P Waits and Andrew Storfer. Basics of population genetics: quantifying neutral and adaptive genetic variation for landscape genetic studies. *Landscape Genetics*, pages 35–57, 2015.

[295] Bruce Walsh and Mark W Blows. Abundant genetic variation+ strong selection= multivariate genetic constraints: a geometric view of adaptation. *Annual Review of Ecology, Evolution, and Systematics*, 40:41–59, 2009.

[296] Dai Wang, Yu Sun, Paul Stang, Jesse A Berlin, Marsha A Wilcox, and Qingqin Li. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. In *BMC Proceedings*, volume 3, pages 1–6. Springer, 2009.

[297] Robin S Waples. Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. *Marine Fisheries Review*, 53(3):11–22, 1991.

[298] Robin S Waples and Steven T Lindley. Genomics and conservation units: The genetic basis of adult migration timing in Pacific salmonids. *Evolutionary Applications*, 11(9):1518–1526, 2018.

[299] Andrew R Weeks, Dean Heinze, Louise Perrin, Jakub Stoklosa, Ary A Hoffmann, Anthony van Rooyen, Tom Kelly, and Ian Mansergh. Genetic rescue increases fitness and aids rapid recovery of an endangered marsupial population. *Nature Communications*, 8(1):1071, 2017.

[300] Andrew R Weeks, Jakub Stoklosa, and Ary A Hoffmann. Conservation of genetic uniqueness of populations may increase extinction likelihood of endangered species: the case of Australian mammals. *Frontiers in Zoology*, 13:1–9, 2016.

[301] Bruce S Weir and C Clark Cockerham. Estimating f-statistics for the analysis of population structure. *Evolution*, pages 1358–1370, 1984.

[302] Bruce S Weir and Jérôme Goudet. A unified characterization of population structure and relatedness. *Genetics*, 206(4):2085–2103, 2017.

[303] Bruce S Weir and William G Hill. Estimating f-statistics. *Annual Review of Genetics*, 36(1):721–750, 2002.

[304] Jason T Weir, Oliver Haddrath, Hugh A Robertson, Rogan M Colbourne, and Allan J Baker. Explosive ice age diversification of kiwi. *Proceedings of the National Academy of Sciences*, 113(38):E5580–E5587, 2016.

[305] Matthias H Weissensteiner, Ignas Bunikis, Ana Catalán, Kees-Jan Francoijs, Ulrich Knief, Wieland Heim, Valentina Peona, Saurabh D Pophaly, Fritz J Sedlazeck, Alexander Suh, et al. Discovery and population genomics of structural variation in a songbird genus. *Nature Communications*, 11(1):3403, 2020.

[306] Martin L Weitzman. On diversity. *The Quarterly Journal of Economics*, 107(2):363–405, 1992.

[307] Thomas A White, Sarah E Perkins, Gerald Heckel, and Jeremy B Searle. Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology*, 22(11):2971–2985, 2013.

[308] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019.

[309] Hadley Wickham, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. *dplyr: A Grammar of Data Manipulation*, 2023. R package version 1.1.4, https://github.com/tidyverse/dplyr.

[310] Aryn P Wilder, Megan A Supple, Ayshwarya Subramanian, Anish Mudide, Ross Swofford, Aitor Serres-Armero, Cynthia Steiner, Klaus-Peter Koepfli, Diane P Genereux, Elinor K Karlsson, et al. The contribution of historical processes to contemporary extinction risk in placental mammals. *Science*, 380(6643):eabn5856, 2023.

[311] Yvonne Willi, Josh Van Buskirk, and Ary A Hoffmann. Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution, and Systematics*, 37:433–458, 2006.

[312] Janna R Willoughby, Mekala Sundaram, Bhagya K Wijayawardena, Steven JA Kimble, Yanzhu Ji, Nadia B Fernandez, Jennifer D Antonides, Maureen C Lamb, Nicholas J Marra, and J Andrew DeWoody. The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biological Conservation*, 191:495–503, 2015.

[313] Edward O Wilson et al. Biodiversity. 1988.

[314] Meike J Wittmann, Sylvain Mousset, and Joachim Hermisson. Modeling the genetic footprint of fluctuating balancing selection: From the local to the genomic scale. *Genetics*, 223(4):iyad022, 2023.

[315] Sewall Wright. Evolution in mendelian populations. *Genetics*, 16(2):97, 1931.

[316] Sewall Wright. The genetical structure of populations. *Annals of Eugenics*, 15(1):323–354, 1949.

[317] Amanda Xuereb, Cassidy C d'Aloia, Marco Andrello, Louis Bernatchez, and Marie-Josée Fortin. Incorporating putatively neutral and adaptive genomic data into marine conservation planning. *Conservation Biology*, 35(3):909–920, 2021.

[318] Amanda Xuereb, Quentin Rougemont, Xavier Dallaire, Jean-Sébastien Moore, Eric Normandeau, Bérénice Bougas, Alysse Perreault-Payette, Ben F Koop, Ruth Withler, Terry Beacham, et al. Re-evaluating Coho salmon (*Oncorhynchus kisutch*) conservation units in Canada using genomic data. *Evolutionary Applications*, 15(11):1925–1944, 2022.

[319] Huaqing Zhao, Nandita Mitra, Peter A Kanetsky, Katherine L Nathanson, and Timothy R Rebbeck. A practical approach to adjusting for population stratification in genome-wide association studies: principal components and propensity scores (PCAPS). *Statistical Applications in Genetics and Molecular Biology*, 17(6):20170054, 2018.

[320] Fei Zou, Seunggeun Lee, Michael R Knowles, and Fred A Wright. Quantification of population structure using correlated SNPs by shrinkage principal components. *Human Heredity*, 70(1):9–22, 2010.

Appendix A

Details on accounting for neutral population structure or false positives

This file contains information on whether the type of analysis used to detect putatively adaptive SNPs accounted for neutral population structure or impact of population demography or contributed to reduction in high false-positives. I categorized the type of analysis into three groups: no (indicated by 0), yes (indicated by 1), and depends (indicated by 2). I provided a reference for each categorization. This is a csv file named "Table S1. Details_on_accounting_for_neutral_population_structure_or_false_positives_file.csv".

Appendix B

Correlation among six covariates

Table S2 provides the correlation among all six covariates. Some covariates co-vary and show high correlation such as the proportion of putatively adaptive SNPs, method-type, and number of genome-wide SNPs.

	log(number of putatively adaptive SNPs)	log(proportion of putatively adaptive SNPs)	log(genome-wide SNPs)	log(study-wide F_{ST})	Account for neutral population structure or demography	Method Type (F_{ST} vs. non-F_{ST})
log(number of putatively adaptive SNPs)	1	0.29	0.54	0.24	0.24	0.06
log(proportion of putatively adaptive SNPs)	0.29	1	-0.65	0.18	-0.29	-0.57
log(number of genome-wide SNPs)	0.54	-0.65	1	0.03	0.45	0.55
log(study-wide F_{ST})	0.24	0.18	0.03	1	0.009	-0.06
Account for neutral population structure or demography	0.24	-0.29	0.45	0.009	1	0.55
Method Type (F_{ST} vs. Non-F_{ST})	0.06	-0.57	0.55	-0.06	0.55	1

Appendix C

Details on 34 studies

This file contains information on 26 variables for 34 studies and is used as the input file for 'Appendix D' R code file. Some variables included in this file are organism examined, its classification based on taxa, number of individuals, populations, genome-wide SNPs, putatively adaptive SNPs, and proportion of putatively adaptive SNPs along with information on the type of sequencing used for genome-wide SNPs and method used to identify putatively adaptive SNPs. Information on the observed and expected Spearman correlation as well as study-wide F_{ST} and allele frequencies is provided. The name of the file is "Details of 34 studies.csv".

Appendix D

R code file for figures, tables, and statistical analysis

This script contains the code for all figures and tables as well as meta statistical analysis performed on the data. The file name is "main_overall_analysis.R".

Appendix E

Data and scripts for each study

The data and R scripts for each study is available on Dryad [https://doi.org/10.5061/dryad.nvx0k6f1j]. It is titled "Data from: Does genome-wide variation and putatively adaptive variation identify the same set of distinct populations?". All details are in read.me file.