Bringing genomics to the field: An integrative approach to seed sourcing for forest restoration

Anoob Prakash and Stephen Keller

Department of Plant Biology, University of Vermont, Burlington, VT-05405, USA

INTRODUCTION

Atmospheric warming has shifted the treelines towards high elevations and increased outbreaks of insect pests, while changes in precipitation have played a major role in increasing drought-induced mortality (Pachauri et al., 2014).

ENVIRONMENTAL STRESSOR'S: Resulting in decline of abundance and extent of tree species due to the direct effects of higher temperature, drought stress and/or failure to meet chilling requirements (Kimmins & Lavender, 1987; McCreary et al., 1990).

ANTHROPOGENIC STRESSOR'S: Land use changes and deforestation have fragmented forests, reducing both the size and connectivity of remaining populations.

These problems are compounded in long-lived organisms like trees, which cannot keep pace with the changing climate through natural dispersal and whose long generation times make any loss of genetic diversity through mortality or demographic bottlenecks, a slow process to replenish.

SOLUTION: Integrating population genomic approaches with restoration planning to inform seed sourcing decisions with an eye towards capturing a representative fraction of the regional diversity.

MATERIALS AND METHODS

STUDY SPECIES: Red spruce (*Picea rubens* Sarg.) is a temperate coniferous tree species that has experienced widespread decline due to anthropogenic causes since the late 19th century and is a prime example of a tree species vulnerable to rapid changes in climate due to its isolated populations and limited within-population genetic diversity in the southern part of its range (Capblancq et al., 2020, 2021).

SEED SELECTION: We focused on two measures of genomic variation to evaluate candidate sources – genetic diversity (GD) and genetic load (GL). These two measures were chosen based on our previous work in red spruce, in which they were found to be significant predictors of early-life fitness of seedlings (Capblancq et al., 2021) (Fig 1).

GD was estimated based on expected heterozygosity (H_{exp}) across all single nucleotide polymorphisms (SNPs), which provides a fundamental measure of genetic variation based on the frequency of heterozygous genotypes expected under Hardy-Weinberg equilibrium (Nei, 1973).

GL measures the accumulation of deleterious mutations in the population due to genetic drift. To obtain an estimate of GL in populations, we calculated the number of non-synonymous SNPs (Pn) and the number of synonymous (silent) SNPs (Ps), weighted by their corresponding frequencies (fn and fs, respectively).

We then estimated GL as the ratio Pnfn/Psfs under the well-supported assumption that the majority of nonsynonymous mutations are deleterious (Willi et al., 2018).

Following functional categories from SNPEff (Cingolani et al., 2012) was categorized as non-synonymous for GL calculation: *missense variant, splice acceptor variant, splice donor variant, splice region variant, start lost, stop gained, stop lost.*
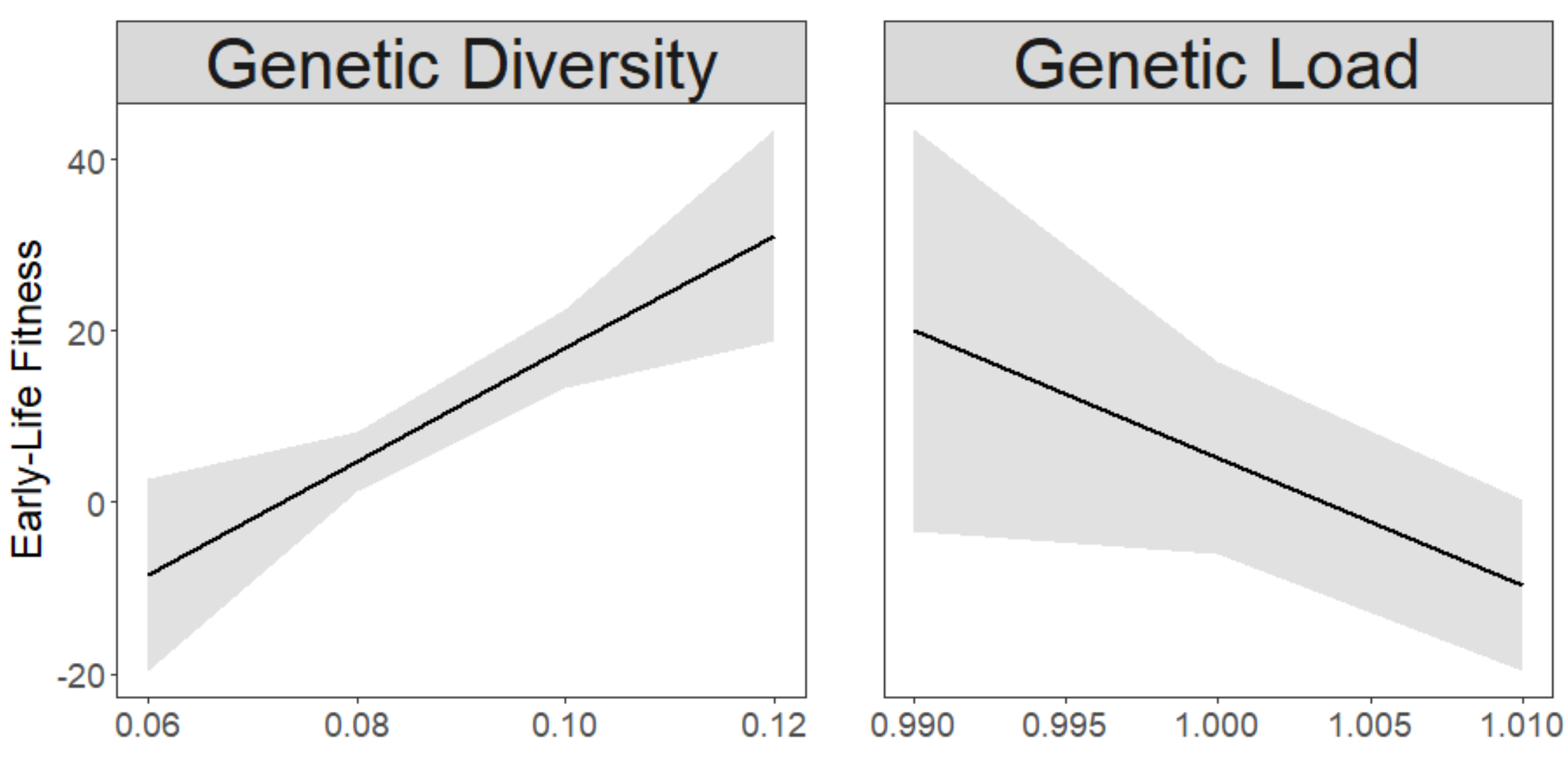

Figure 1. Marginal effects estimated from a multivariate linear model for Early-Life fitness for genetic diversity (GD) and genetic load (GL). Recreated from the results of (Capblancq et al., 2021) for the red spruce range edge.

DEFINING REGIONAL FOCUS: Regional admixture provenancing strategy is based on sourcing seeds from a set of populations within the region of the target restoration sites that share a common biogeographic history and patterns of local adaptation (Bucharova et al., 2019).

EXPERIMENTAL DESIGN FOR MONITORING RESTORATION: Monitoring protocol at planting to enable tracking of establishment success (survival and growth through the second year). The monitoring design consisted of replicate georeferenced experimental plots stratified by seed source so that we could assess single-source variability at each site and compare the effect of single vs. pooled sources on trait variance (Fig. 2). Plot locations were buffered by 18m (16m plot radius + 2m buffer) in ArcGIS to avoid overlapping adjacent plots. Data on establishment success was collected 1 year after planting at the beginning of the growing season (April 2022).
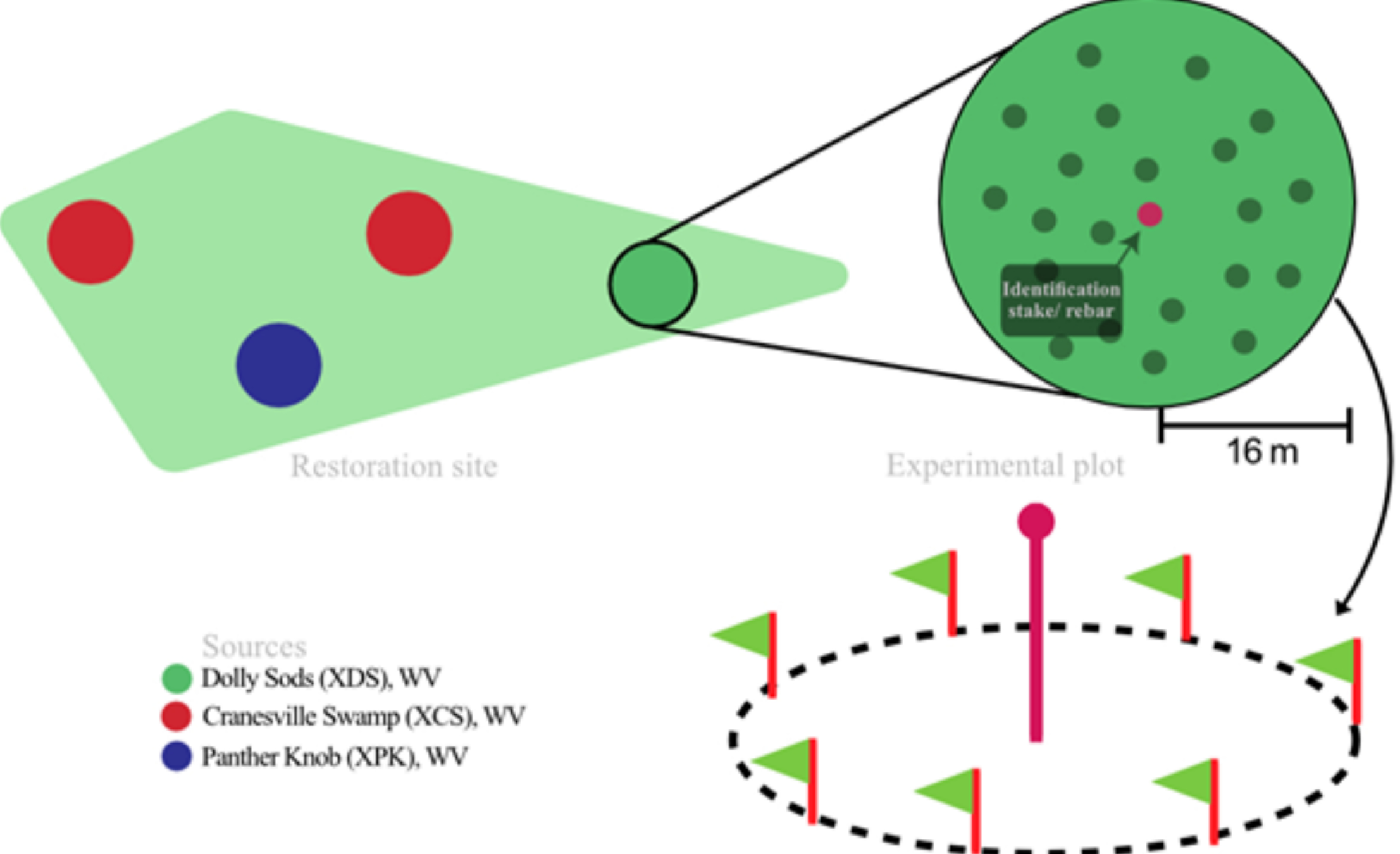

Figure 2. The experimental design for the red spruce monitoring effort. A circular experimental plot of radius 16m was established at the restoration site for reforestation monitoring and to assess the performance of individual sources. Each experimental plot consisted of forty plants from a single source to reflect the planting density of the reforestation activity. These single source experimental plots were marked with a central metal rod capped with a bright orange cap along with GPS coordinates for future monitoring efforts.

COEFFICIENT OF EVOLVABILITY (CV_G): To analyse the effect of single vs. pooled sources on evolutionary potential of trait variance in the restored populations. We used the seedling height data to calculate a scale-free measure of evolvability based on the genetic coefficient of variation (Houle, 1992)

$$CV_G = \sqrt{V_G / \bar{x}}$$

WHERE,
V_G *is the genetic component of trait variance*
$\bar{x}$ *is the trait mean*
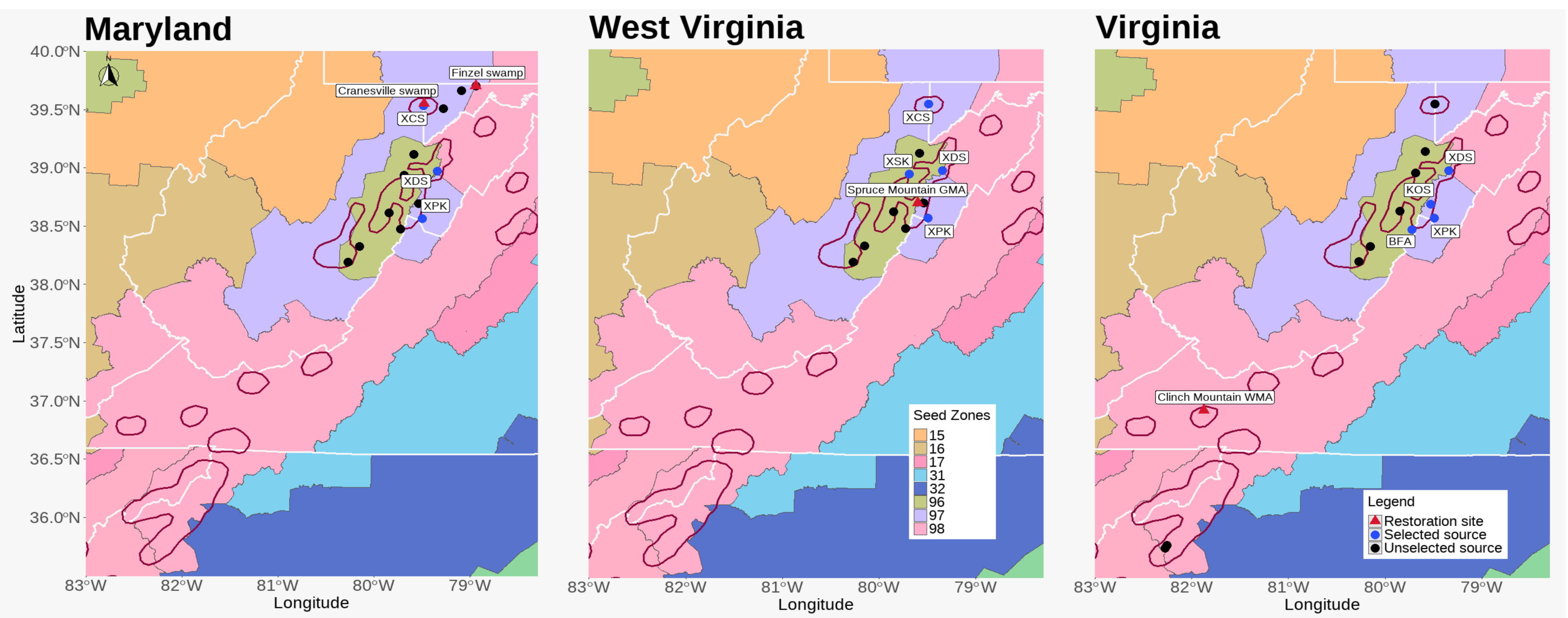
Maryland   West Virginia   Virginia

Figure 3. STUDY AREA AND SEED SOURCING: The location of the red spruce restoration site and the seed sources selected for planting sites in Maryland, West Virginia and Virginia. The dark red polygons on the map are the known range extent of red spruce. The map colouring is based on the Eastern Seed Zone database.

RESULTS

REGIONAL ADMIXTURE PROVENANCING: The seeds sourced were mostly within the same climate-based seed zone or one seed zone away from the restoration site (Pike et al., 2020), keeping with the idea of environmental similarity during regional admixture provenancing (Fig 3).

SOURCE POOLING: Across the sites, CV_G and GD:GL were highest for the pooled sources compared to any single sources (Fig. 4). High CV_G for functional traits such as height can increase the resilience of populations to future environmental perturbations by increasing the genetic variability the population has for traits to respond to novel selection pressures.
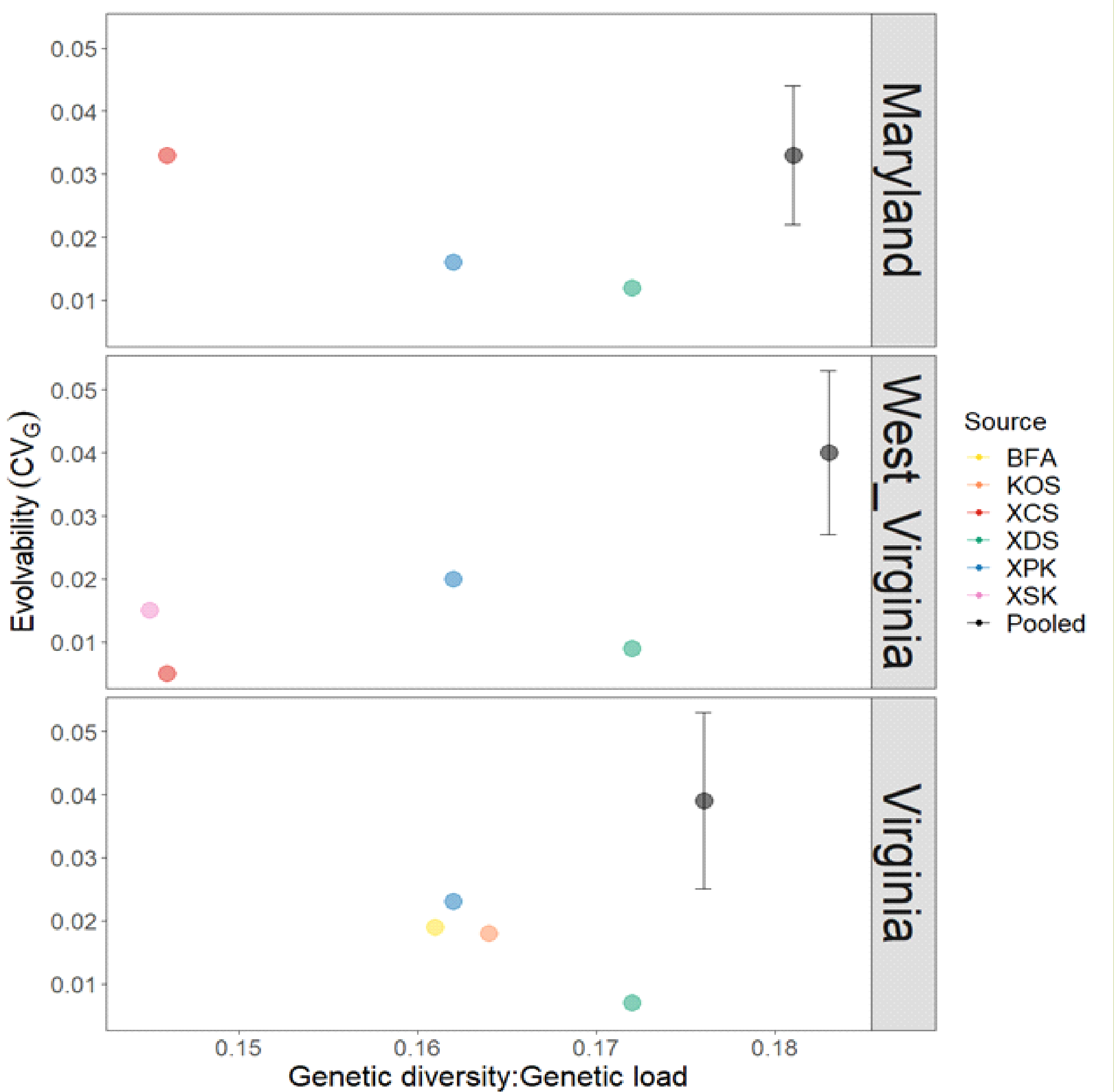

Figure 4. Evolvability (CV_G) for different sources planted at each restoration site and the bootstrapped pooled CV_G for source combinations at each site. The x-axis shows the GD:GL estimates for the singular and pooled sources.

DISCUSSION

MAJOR FINDINGS
No "super sources" found across the sites, with the pooled combination for each site having high GD:GL and evolvability compared to any singular sources (Fig 4).

We found that GD:GL increased with set size until around four to five sources, beyond which there were minimal gains in diversity by combining additional sources (Fig. 5).

This work sets up the blueprint for successful collaboration between science and practice using a novel approach using forest genetics for ecological restoration. This science-practice-monitoring should be seen as a feedback loop rather than a one directional transfer of knowledge and skills. The work presented here is just a snapshot in time and it marks the beginning of conservation efforts that hopefully spans for centuries to come.
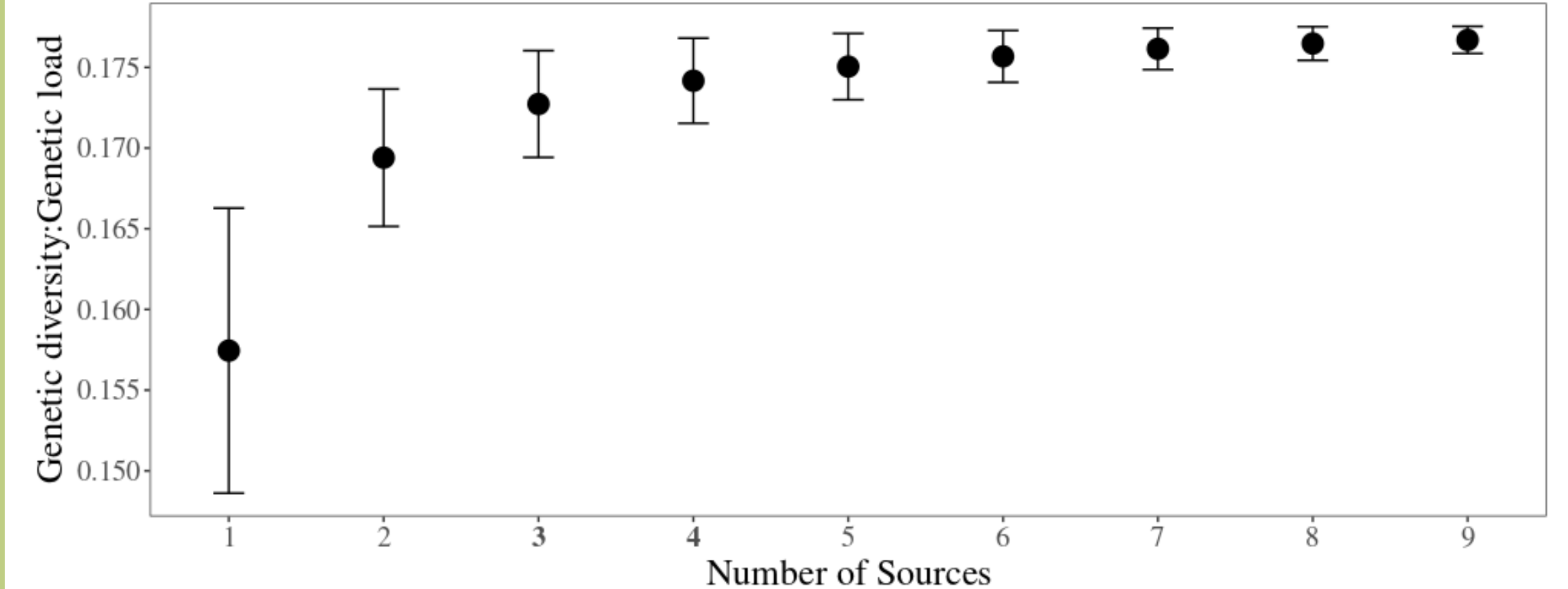

Figure 5. Total number of sources selected for restoration and their respective GD:GL estimate. The estimate for GD:GL flattens out around 4-5 sources with diminishing returns with increasing number of sources.

References

Bucharova, A., Bossdorf, O., Hölzel, N., Kollmann, J., Prasse, R., & Durka, W. (2019). Mix and match: Regional admixture provenancing strikes a balance among different seed-sourcing strategies for ecological restoration. Conservation Genetics, 20(1), 7–17. https://doi.org/10.1007/s10592-018-1067-6

Capblancq, T., Butnor, J. R., Deyoung, S., Thibault, E., Munsom, H., Nelson, D. M., Fitzpatrick, M. C., & Keller, S. R. (2020). Whole-exome sequencing reveals a long-term decline in effective population size of red spruce (Picea rubens). Evolutionary Applications, 13(9), 2190–2205. https://doi.org/10.1111/eva.12985

Capblancq, T., Munsom, H., Butnor, J. R., & Keller, S. R. (2021). Genomic predictions of early-life fitness in Picea rubens. Conservation Genetics. https://doi.org/10.1007/s10592-021-01378-7

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. SNPs in the genome of Drosophila melanogaster strain w^1118; iso-2; iso-3. Fly, 6(2), 80–92. https://doi.org/10.4161/fly.19695

Houle, D. (1992). Comparing evolvability and variability of quantitative traits. Genetics, 130(1), 195–204. https://doi.org/10.1093/genetics/130.1.195

Kimmins, J., & Lavender, D. (1987). Implications of climate change for the distribution of biogeoclimatic zones in British Columbia and for the growth of temperate forest species. Woody Plant Growth in a Changing Chemical and Physical Environment. Ed. DP Lavender. University of British Columbia Press, Vancouver, BC, 269–309.

McCreary, D., Lavender, D., & Hermann, R. (1990). Predicted global warming and Douglas-fir chilling requirements. Annales Des Sciences Forestières, 47(4), 325–330. https://doi.org/10.1051/forest:19900404

Nei, M. (1973). Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences, 70(12), 3321–3323.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., & Dasgupta, P. (2014). Climate change 2014: Synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change. Ipcc. https://www.ipcc.ch/site/assets/uploads/2018/02/SYR_AR5_FINAL_full.pdf

Willi, Y., Fracassetti, M., Zoller, S., & Van Buskirk, J. (2018). Accumulation of Mutational Load at the Edges of a Species Range. Molecular Biology and Evolution, 35(4), 781–791. https://doi.org/10.1093/molbev/msy003