Project Summary

**Title: A conservation genomics program to identify climate adapted genes in the eastern hemlock (*Tsuga canadensis*)**

**Dr. Karl C Fetter**
*Postdoctoral Applicant*
University of Connecticut
Department of Ecology and Evolutionary Biology
75 N. Eagleville Road, Unit 3043
Storrs, CT 06269

**Dr. Jill Wegrzyn**
*Assistant Professor & Primary Mentor*
University of Connecticut
Department of Ecology and Evolutionary Biology
75 N. Eagleville Road, Unit 3043
Storrs, CT 06269

I am pleased to submit a proposal to the USDA AFRI EWD for a postdoctoral research fellowship (A7201). Trees are facing a global extinction crisis as we enter the Earth's sixth mass extinction. Globally, 30% of trees are threatened and 11% in the United States. Trees form forests which provide important agricultural services as habitat for pollinators, ecosystem services in nutrient cycles, and facilitate numerous ecological relationships.

Climate change and invasive pests are dual, interacting threats to the viability of tree populations. Development of resilient forests is a key priority area for conservation and sustainable agroecological landscapes. Resiliency can be developed through diverse methods, such as tree breeding, seed banking, assisted migration, or genetic engineering. Conservation genomics provides key information to make informed decisions. **I propose to develop the eastern hemlock as a system for conservation genomics by generating a reference genome and applying population genetic methods to identify climate adapted genomic variation.**

This proposal supports the following AFRI Farm Bill Priority Areas: Plant health and production pathways (PHPPP). The proposal most closely matches NIFA AFRI Foundational and Applied Science Program Area 1e Plant Breeding for Agricultural Production (A1141). A priority of the project is to dissect the genetic basis of climate adapted genomic variation, and to identify germplasm for future breeding projects to generate locally adapted cultivars.

The proposal supports the goals of the program through training a science professional (PD) to enter research and education. My career goal is to become a professor or government scientist focused on plant breeding for resilience to abiotic and biotic stress. The proposal is focused on addressing agricultural climate adaptation and conservation of forest trees for healthy agroecological systems. The proposal supports the overarching goal of the EWD to advance science by training post-graduate researchers in food and agriculture disciplines.

# A conservation genomics program to identify climate adapted genes in the eastern hemlock (*Tsuga canadensis*).

## RATIONALE & SIGNIFICANCE

### Importance of trees

Trees are crucially important for healthy ecosystems, economies, and cultures. Through their physiology, trees facilitate global elemental and nutrient cycles, storing vast amounts of carbon combating climate change. Some biogeographic regions are defined by trees and they directly provide habitat and food to countless organisms. Trees are key links in food webs of pollinators, fungi and even other plants. As such, trees are central ecological players that maintain functioning ecosystems and support biodiversity. Humans rely on trees to provide raw material for economies, shade to cool cities and neighborhoods, medicinal products, and cultural inspiration, to name only a few of their contributions.

During the last century, trees have increasingly faced a global biodiversity crisis. Habitat loss and over-harvesting are the two largest threats to trees (BGCI, 2021), but biological and ecological threats are constant and urgent. Declines in pollinator and seed disperser populations limit the fecundity of trees and alter forest structure (Caughlin *et al.,* 2015). Invasive pests and pathogens are major threats to biodiversity and have already pushed tree species into functional extinction (e.g., American chestnut, Ellison *et al.,* 2005). Climate change is altering range dynamics by displacing or eliminating habitat (Verrico *et al.,* 2020), but also interacts with biotic threats (Parmesan, 2006). Warmer winters allow tree pests and pathogens to migrate and survive in areas where they historically could not (Ellison *et al.,* 2018; Gustafson *et al.,* 2022). Taken together, climate change represents dual threats by directly increasing mortality due to abiotic stress, and indirectly through biotic interactions.
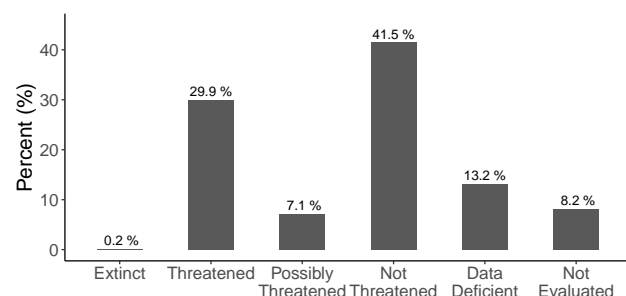


**Figure 1: Conservation status of the world's trees.** Adapted from BGCI, 2021.

An increased focus on conserving biodiversity of trees is needed to meet the challenges of the 21$^{\text{st}}$ century. Globally, 30% of trees are threatened under IUCN red list definitions (Fig. 1). In the United States, 11% of trees are threatened (Carrero *et al.,* 2022). Theoretical and methodological advances in ecology and evolution now provide a road map for applying powerful molecular techniques to identify and conserve adaptive genomic variation for conservation and reforestation.

### The conservation genomics program

Conservation genomics seeks to bridge basic and applied research to create the best available science to inform management decisions. The conservation genomics program is increasingly seen as one of the best viable options for understanding, identifying, and conserving genomic resources (Paez *et al.,* 2022). A strength of conservation genomics is that the field spans multiple domains of expertise, and brings conservationists together who may otherwise only communicate through

the literature, white papers, or the media (Rossetto *et al.,* 2021).

Populations which cannot replace themselves will eventually become extinct. The conservation genomics program begins with a community recognizing a problem (Fig. 2). Generating a reference genome is a critical step to deploy population genetic methods, enabling more effective use of genotyping resources, refined discovery of genes, and the ability to work with chromosome-specific genetic methods (Rossetto *et al.,* 2021). In the case of conifers, their massive genomes (regularly >5 Gb) require a combination of long and short read sequencing along with chromosome confirmation capture. Population genetic methods to identify, for example, climate adaptive genetic variants can enable futher testing and development of genetic management plans. The ultimate goal of conservation genomics is to deploy selected germplasm into the field to achieve long-term                  conservation                  goals                  (Paez                  *et                  al.,*                  2022).
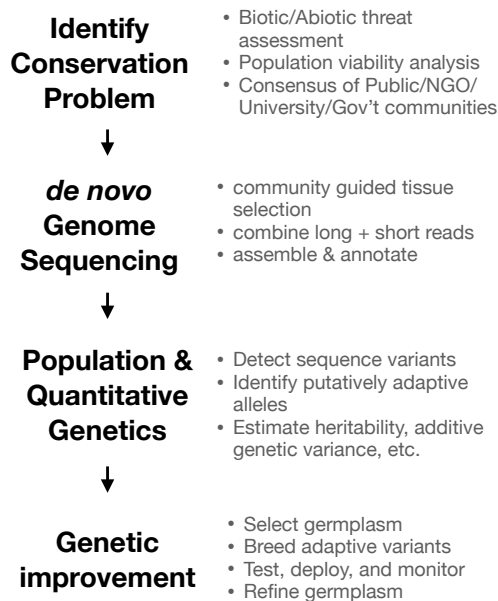
**Identify Conservation Problem**
- Biotic/Abiotic threat assessment
- Population viability analysis
- Consensus of Public/NGO/ University/Gov't communities

↓

***de novo* Genome Sequencing**
- community guided tissue selection
- combine long + short reads
- assemble & annotate

↓

**Population & Quantitative Genetics**
- Detect sequence variants
- Identify putatively adaptive alleles
- Estimate heritability, additive genetic variance, etc.

↓

**Genetic improvement**
- Select germplasm
- Breed adaptive variants
- Test, deploy, and monitor
- Refine germplasm

**Figure 2: The conservation genomics program** suited to forest trees.

## Introduction

### The hemlock system

*Tsuga* (Endlicher) Carrière, hemlocks, are a genus temperate conifers with a disjunct distribution in eastern Asia and North America. Hemlocks are one of eleven genera in the Pinaceae, the largest family of conifers with over 230 species (Ran *et al.,* 2018). The Pinaceae diverged approximately 206 Mya in the Late Triassic and contains two clades: the Pinoids and Abietoids. *Tsuga* is a derived lineage within the Abietoid clade having diverged from its monotypic sister genus *Nothotsuga* approximately 43 Mya. *Tsuga* is distinguished from other Pinaceae by their flattened leaves born on a petiole with a single resin duct (eFloras, 2008). *Tsuga* contains ten species (Feng *et al.,* 2021), with *T. ulleungensis* being the most recently described species from the Korean island of Ulleungdo (Holman *et al.,* 2017).

The biogeography of the genus underwent major revision in the last decade. Early efforts to unravel the evolutionary history relied on a single autosomal marker and plastid sequences (Havill *et al.,* 2008). The biogeographic interpretation indicated multiple migrations across the Bering land bridge had occurred. A decade later, phylotranscriptomic analysis of 41 *Tsuga* accessions overturned the multiple migration hypothesis, supporting a single migration event of the ancestral population of the Carolina hemlock (*T. caroliniana*) into Asia (Feng *et al.,* 2021). North American hemlocks are at the base of the tree and form a grade, with the eastern hemlock (*T. canadensis*) sister to the Carolina and Asian hemlocks.

Phylogenetic relationships are relevant to understand potential solutions to conservation problems in *Tsuga*. Hybridization is common within the Asian hemlock clade (Bentz *et al.,* 2002), complicating the phylogeny, while also creating the potential for adaptive introgression or guided

breeding. However, in North America, the western hemlocks (*T. heterophylla* and *T. mertensiana*) do not hybridize with each other (Taylor, 1972), nor will the eastern hemlock receive pollen from the sympatric Carolina hemlock (in a breeding study, 1 out of 10 genotyped offspring were crosses from a *T. canadensis* father to a *T. caroliniana* mother, Bentz *et al.,* 2002). All available population genetic data indicate natural hybridization between the eastern and Carolina hemlocks does not naturally occur (Potter *et al.,* 2012; Potter *et al.,* 2017). The barriers to reproduction among North American hemlocks removes adaptive introgression or breeding as a means for introducing novel genetic variation to solve conservation problems. Within species genetic variation must be surveyed and understood to attempt to solve current and future challenges.

Covering over 1.3 million hectares, the eastern hemlock is an integral part of eastern North American forests (Ellison, 2014). Throughout its range, *T. canadensis* is recognized as a keystone species, providing important ecosystem services for water quality, soil stabilization, wildlife habitat, and economic and aesthetic assets for humans (Ellison *et al.,* 2005). When stands of *T. canadensis* are lost, negative impacts to the abiotic (Domec *et al.,* 2013) and biotic environments (Ellison *et al.,* 2005), as well as to human economic systems (Li *et al.,* 2014) have been demonstrated. In 2013, the IUCN listed *T. canadensis* as near threatened (Farjon, 2013). Local populations across its range have been lost in recent decades (Ellison *et al.,* 2005), and more populations are expected to be lost as a result of an invasive pest (Ellison, 2014).

**Biotic threats to the eastern hemlock**

Novel biotic and abiotic environments pose significant challenges to the survival of the eastern hemlock. In the early 20th century, ornamental hemlocks were imported from Japan and planted in Maymont Park in Richmond, Virginia (Askins, 2014). Clinging to the branches of the ornamentals, a small, sap-feeding Hemiptera was introduced that would profoundly change the structure of eastern North American forests. The hemlock woolly adelgid (HWA, *Adelges tsugae*) is a relatively benign pest on Asian *Tsuga*. The eastern and Carolina hemlocks are susceptible to HWA, potentially pushing both species towards extinction (Ellison *et al.,* 2018). Since its first observation from Virginia in 1951 (Stoetzel *et al.,* 2002), the HWA has expanded its range to nearly every southern and mid-Atlantic population of *T. canadensis*. All US hemlock populations are projected to be infested by 2050 (Ellison *et al.,* 2018).

The exact mechanism of mortality to the HWA is unknown (Domec *et al.,* 2013). When HWA encounters a host tree, it will preferentially settle on young growth (Preisser *et al.,* 2014). The host plant will initiate a local, and then global hypersensitive response (HR) which results in leaf drop and cessation of growth at the site of feeding (Radville *et al.,* 2011). Hypersensitive responses that kill or drop tissue are typically effective against sessile pests/pathogens. Unfortunately, the HWA is mobile and responds to the host's hypersensitive response by migrating into the tree's canopy and settling on older leaves (McClure, 1991). False growth rings form in the wood of infected branches after HWA feeding, interfering with water and solute transport within the host and limiting photosynthesis (Domec *et al.,* 2013). Prolonged water stress, decreased photosynthesis, and the loss of leaves eventually kills the host within four to ten years (McClure, 1991).

The adelgid has been present in the eastern hemlock's core range (the Northeast and southern New England) for four to five decades (Fitzpatrick *et al.,* 2012). During that time, populations of remnant trees were identified in areas with high HWA-induced mortality through citizen science monitoring (Ingwell *et al.,* 2009). Limited studies suggest selection for HWA-resistance may

have occurred (Kinahan *et al.,* 2020). Differing terpenoid profiles between putatively resistant and susceptible hemlocks indicate foliar or twig chemical profiles may provide resistance (McKenzie *et al.,* 2014), although other mechanisms of resistance or tolerance have not been studied (e.g. a reduced HR in tolerant genotypes).

## Climate change effects

*Tsuga* are sub-alpine temperate trees and climate change represents direct and indirect threats to their viability. While western US forests are becoming hotter and drier, eastern forests are projected to become wetter and experience more extreme disturbances and drought events (Melillo *et al.,* 2014). The species composition of eastern forests is becoming unstable, with southeastern temperate montane forests projected to experience high rates of change in response to increased precipitation variability (Knott *et al.,* 2020). Early season moisture is critical for seed germination, and uneven precipitation can select for drought tolerant species, competitively excluding species like hemlocks (Pederson *et al.,* 2015). Already, *T. canadensis* in the northern region is being replaced by oak and drought tolerant conifers, having decreased by up to 18% in some forests (Nowacki & Abrams, 2015). In extreme southern populations of the Alabama Cumberland plateau, drought and extreme weather are reducing population densities (Myers *et al.,* 2015). Overall, increased competition from drought-adapted trees and increased physiological drought stress from HWA infestation may converge to accelerate the population viability crisis.

Temperature has an uneven effect on population viability in *T. canadensis* (Pederson *et al.,* 2015). Over the course of the entire growing season, warming temperatures have a weak positive effect on growth rates in southern populations, and a stronger positive effect in the north. However, growth is negatively correlated with May temperatures in northern populations (Vaganov *et al.,* 2011) and summer temperatures in southern populations (Hart *et al.,* 2010). The largest effect of temperature on population viability is almost certainly due to the indirect effects from the HWA. Globally, pests and pathogens are increasing their range under climate change (Parmesan, 2006). In the Northeast, milder winters are releasing the HWA, expanding the killing zone into new hemlock populations (Ellison *et al.,* 2018). HWA abundance is negatively correlated with hemlock regeneration (Preisser *et al.,* 2011), which is historically low (Sullivan & Ellison, 2006; Alverson *et al.,* 2019).

## Reference genome-enabled conservation

A reference genome is a high quality set of contiguous sequences assembled into chromosomes that is complete, or nearly complete, and is annotated with functional information (Paez *et al.,* 2022). Reference genomes enable an enormous array of population and comparative genomic analyses, and, increasingly, targeted modification of genes and individual sites (Formenti *et al.,* 2022). The utility of reference genomes for conservation was questioned in their early applications. Arguments were made that basic molecular tools can answer questions about genetic structure, migration, effective population size, and genetic diversity (McMahon *et al.,* 2014; Shafer *et al.,* 2015). Demographic information can (and still does) provide significant insights for conservation managers; however, the impact of reference genomes is their enabling of analyses that fundamentally rely on positional, homology, or functional annotation information. This knowledge addresses basic questions in organismal biology which relate directly to fitness and population viability.

The list of application of truly genome-wide methods in conservation is now large and the subject of reviews (Brandies *et al.,* 2019; Formenti *et al.,* 2022; Paez *et al.,* 2022). Applications of reference genomes in conservation can be broadly summarized into two sets of analyses: comparative genomics, and population-based inference. Comparative methods permit documenting the genomic diversity of life. This is an important basic research goal that is relevant during the current mass extinction crisis (Ceballos *et al.,* 2015). Comparative analyses of gene families and genomic structures identify gene expansions, chromosomal inversions, or chromosomal fusions whose variation is driven by evolution linked to fitness (Zhang *et al.,* 2020). Reference genomes yield better homology estimates between genes, enabling superior phylogenetic analyses crucial to identify conservation units (Formenti *et al.,* 2022).

The impact of reference genomes is perhaps greatest for population-based inference. A precise understanding of introns, exons, chromosomes, and functional information enhances the value of reduced representation sequencing (e.g. ddRADseq), yielding a larger quantity of high quality SNPs, copy number variants, and structural variants (Formenti *et al.,* 2022). Reference genomes enable more accurate estimates of classical population genetic parameters (e.g. $N_e$, Willi *et al.,* 2022), refined estimates of inbreeding depression and deleterious mutations within chromosomes (Wilder *et al.,* 2022), and identification and classification of outbreeding depression and Bateson-Dobzhansky-Muller genetic incompatibilities (Robinson *et al.,* 2021). Many tests of natural selection require known physical distances between genetic loci (Willi *et al.,* 2022). Reference genomes enable tests for selection, and provide evolutionary context to understand variation of fitness. Perhaps most importantly, genotype-phenotype and gene-environment analyses using loci with known genomic regions enables genomic selection. Breeding for conservation using genomic selection is a powerful tool that can accelerate species conservation and is now being realized in forest trees (e.g. the American chestnut, Westbrook *et al.,* 2020).

Additionally, reference genomes enable more precise modification of genes through transgenic technologies (Park *et al.,* 2017). These approaches are just beginning to be applied in conservation systems, but are having large impacts on the discussion of future solutions to complex issues (Nawaz & Satterfield, 2022). Embryogenesis is possible in eastern hemlocks (Merkle *et al.,* 2014), raising the possibility of introducing transgenes or genetic modifications for conservation. Indeed, recent leaders in the field of tree conservation have argued for increased collaboration between tree breeding guided by genomic selection and transgenic approaches (Nelson, 2022).

**Identifying climate adapted genomic diversity for gene conservation**

Identifying climate adapted sites and demonstrating a genetic offset between current and predicted climates follows a methodology that merges association genetics and species distribution modeling (Capblancq *et al.,* 2020). The motivation of genetic offset modeling is to identify populations at risk of decline due to climate change, and how the risk varies across populations. Genetic offset risk models have been applied in plants (Gougherty *et al.,* 2021), birds (Chen *et al.,* 2022), and mammals (Razgour *et al.,* 2019), and received scrutiny (Rellstab *et al.,* 2021; Láruson *et al.,* 2022) and support (Fitzpatrick *et al.,* 2021) from the conservation modelling community.

The general procedure to model genetic offsets begins with (1) identifying candidate genes for climate adaptation, (2) a generalization of the current climate-candidate locus relationship, (3) prediction of future climate-candidate locus relationship, and (4) calculation of the offset between current and future climate niches (Fitzpatrick & Keller, 2015). Genomic maladaption of popu-

lations is identified by spatially mapping the genetic offset onto the landscape (Capblancq *et al.,* 2020). Validation of the statistical results is recommended, which can take multiple forms, including common garden, growth chamber, or genetic manipulation experiments (Fitzpatrick *et al.,* 2021). The implications from genetic offset analyses for gene conservation are large. Populations at greatest risk for climate genetic offsets can be identified for seed banking, breeding, or assisted migration through this method.

Given the uncertainty of climate change effects on eastern forests (Knott *et al.,* 2020), the rapidly declining population sizes of eastern hemlocks (Nowacki & Abrams, 2015), and overall loss of forested habitats (BGCI, 2021), there is clearly a pressing need for conservation genomics in the eastern hemlock system. I propose to develop the eastern hemlock as a system for conservation genomics and to test applied population genomic hypotheses to aid management decisions.

## APPROACH

The proposed project has two major components aimed at conservation genomics in the eastern hemlock:

*Objective 1: Generate a chromosome-scale reference genome for the eastern hemlock.*
*Objective 2: Identify climate adapted genomic variation for gene conservation.*

### Objective 1: Generate a chromosome-scale reference genome for the eastern hemlock

### Creating the reference genome

An accession from the Harvard University's Arnold Arboretum has been selected to sequence. The tree (#1509-1*A) was collected in 1882 from wild provenance. Leaf tissue was collected and high molecular weight DNA was extracted using an Circulomics Nanobind Plant Nuclei Big DNA Kit. DNA quality was assessed, then fragmented to approximately 20 Kb. After passing quality control, we sequenced four libraries using a Oxford Nanopore PromethION to a coverage of 20x for the 20 Gb genome. To account for a higher error rate of Nanopore data, we created a set of Illumina short reads. Genomic DNA was extracted from the haploid megagametophyte of a seed of the reference tree. 350 bp and 550 bp paired-end libraries were prepared and sequenced on an Illumina NovaSeq S4 for 300 cycles to 80x coverage. This work was funded by a grant from The Nature Conservancy.

Using short and long reads, we will create a draft hybrid assembly using MaSuRCA (Zimin *et al.,* 2013). The quality of the genome assembly will be assessed with BUSCO (Simão *et al.,* 2015) and QUAST (Gurevich *et al.,* 2013). To create a high-quality, chromosome-scale assembly, we will contract with Dovetail genomics to use chromosome confirmation capture methods to generate a linkage map for each chromosome. We will provide a draft assembly and libraries prepared with their Omni-C kit. The libraries will be sequenced with Illumina short reads. The final assembled genome will merge our draft with the linkage map. This method represents the current state-of-the-art for genome assembly.

Genome annotation provides critical structural and functional information to an assembly. We will annotate the genome using an in-house pipeline developed for eukaryotic non-model organisms. The pipeline, called EASEL (Webster *et al.,* 2022), is a genome annotation tool that leverages machine learning, RNA folding, and functional annotations to enhance gene prediction accuracy. EASEL utilizes AUGUSTUS (Stanke & Morgenstern, 2005) with parameters optimized for pre-

diction of gene models incorporating extrinsic evidence from transcripts and protein alignments. The EASEL pipeline aligns transcript reads from RNA-seq and assembles putative transcripts via Stringtie2 (Kovaka *et al.,* 2019) and PSIClass (Song *et al.,* 2019). Open reading frames are predicted through TransDecoder (Brian & Papanicolaou, 2020) utilizing a gene family database (EggNOG, Huerta-Cepas *et al.,* 2019) for refinement. Transcriptome and protein hints are generated by aligning the frame-selected sequences to the genome. These hints are independently used to train AUGUSTUS, and the resulting predictions are combined into a single gene set. Implicated gene structures are further refined through machine learning based on a set of primary and secondary sequence features (e.g. RNA folding, Kozak consensus, splice sites, functional annotation). Annotations are combined and filtered from a feature matrix of output and scored to generate the highest supported annotation. Descriptive statistics of the final assembly will be reported.

### Objective 2: Identify climate adapted genomic variation for gene conservation

Genetic variation that increases fitness of an individual in a population in its home range is said to be locally adaptive. Finding locally adaptive genetic variation is difficult for several reasons: allele frequency of neutral sites across a genome can mimic sites under selection; tests to detect sites under selection may lack statistical power to yield true positives, and fitness effects may be small or undetectable, to name a few challenges (Savolainen *et al.,* 2013). The objective here is to identify a set of loci that demonstrate allele frequencies indicative of associations to climate, while accounting for background levels of allele frequency variation due to ancestry, kinship, or both. To achieve this objective I will conduct a gene-environment analysis. Genetic offsets will be identified using current and future climate predictions with gradient forest models, and compared to generate climate risk assessments.
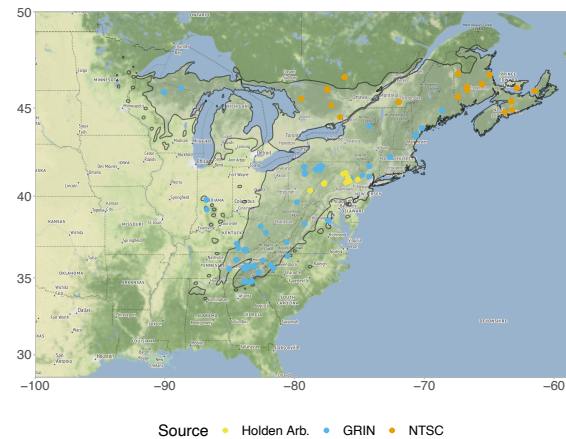


**Figure 3:** *Tsuga canadensis* **range map and seed bank source sites.**

### Creating the SNP set

In total, tissue from 288 unique genotypes will be collected from two sources: seed banks (N=144) and field collections (N=144). Seeds will be ordered from GRIN, the Canadian NTSC, and the Holden Arboretum representing a wide range of geographic diversity (Fig. 3). Field collections of leaves will be made in critical portions of the species' distribution not represented in seed banks, namely, the range core (sensu Hampe & Petit, 2005) in New England and New York, and populations in the Great Lakes basin. Both regions represent the largest population sizes of hemlock (in terms of basal area) (Fitzpatrick *et al.,* 2012), and are likely to harbor adaptive genetic variation.

Seeds will be stratified for 10 weeks and germinated following (Jetton *et al.,* 2014), grown in petri dishes until the six leaf stage. Whole plants will be placed in 96 well microcollection tubes and lysed for DNA extraction. Leaf tissue will be collected in the field and preserved in silica gel.

Approximately 250 mg of leaf tissue will be packed into a microcollection tube, lysed, and DNA extracted. Staff at CGI will perform DNA extractions. Libraries for ddRADseq will be prepared using SphI-MluCI RE following Johnson *et al.* (2017) and checked for quality control. 150 bp PE reads will be sequenced on an Illumina NovaSeq 6000 for 300 cycles across 6 lanes at CGI.

Read quality will be checked with fastp (Chen *et al.,* 2018) and adapters trimmed with sickle (Joshi & J, 2011). To detect variants, a combination of bioinformatic tools will be used to align reads to the reference genome (bwa, Li & Durbin, 2009), sort, de-duplicate, and index reads (samtools, Li *et al.,* 2009), and to call SNPs (bcftools, Danecek *et al.,* 2021). Sites will be filtered with bcftools to retain biallelic sites without indels, for site-wise missingness $<20\%$, genotype quality $>90$, and a site depth appropriate to the coverage. SNPs that show excess heterozygosity in tests of Hardy Weinberg Equilibrium will be removed.

### Climate adaptation and genetic offsets

Neutral population genetic structure will be described using ADMIXTURE (Alexander *et al.,* 2009) to obtain a Q-matrix, and with DAPC (Jombart *et al.,* 2010). Bioclimatic variables from each site will be downloaded from WorldClim (Fick & Hijmans, 2017) and summarized using PCA. Climate adapted candidate sites will be identified from a gene-environment analysis in a latent-factor mixed model using LFMM2 (Caye *et al.,* 2019). LFMMs are regression models fit for each genetic site that includes explanatory variables for climate (PC1) and a correction for ancestry (i.e. the Q-matrix). Predictive models generalizing genetic-environment relationships of current and future climate will be fit with gradient forests to yield genetic importance values (Ellis *et al.,* 2012). Future climate will be projected 50 years into the future using two different emission scenarios (RCP 4.5 and 8.5). Local genetic offsets will be calculated as Euclidean distances between genetic importance values of current and future climates for each sample location following Fitzpatrick and Keller (2015). Genetic offsets will be visualized and spatially mapped for the entire species' range using kriging interpolation with the gstat package (Gräler *et al.,* 2016). Areas with large genetic offsets are priority areas for gene conservation.

### Project relevance

The proposal supports the Plant Health and Production Pathways Priority area (PHPPP), specifically the Plant Breeding for Agricultural Production (A1141) area. A key objective of the project is to dissect the genetic basis for climate adaptation to identify genes for conservation. More broadly, intact forest ecosystems provide measurable benefits to agricultural systems. The diversity and abundance of bee pollinators and the number of pollinator visits per flower increases when agricultural fields are adjacent to forests (Nicholson *et al.,* 2017). Overall, crop yields are higher when farm lands are adjacent to intact patches of forest (Garibaldi *et al.,* 2013). The proposed work lays the foundation for applied conservation genomics in an important forest tree of the eastern United States. This fellowship will serve as a springboard upon which I can build decades of research, enabling me to become an independent and productive scientist.

## PROJECT PLAN

The project will last 24 months spanning three calendar years (Table 1). The project represents a new area of research for the PD (genome assembly, annotation, developing SNP sets, and genetic offset models) and for the co-PI (population genomics of climate adaptation). Experimental approaches employ routine molecular techniques. During exploratory wet-lab work, the PD encountered some difficulty extracting long fragments of high molecular weight DNA from hemlocks. Upon searching the literature it became apparent that hemlocks have flavonols embedded within chromatin (Feucht *et al.,* 2011). Increasing PVP-360 and BME during the tissue lysis solved this problem. I have not attempted to extract DNA for high-throughput sampling, but I am confident I can collaborate with the sequencing facility on campus to develop a method. Genome assembly of conifers can be a time consuming process, and I have access to high-memory nodes on Xanadu. A strength of the proposal is mapping the population read set to the reference genome to maximize variant detection. Main project deliverable will be the reference genome, population read set, and one or more publications.

**Table 1:** Project timeline.

| Objective | Year 1 (2023) | Year 2 (2024) | Year 3 (2025) |
|---|---|---|---|
| Develop reference genome | X | X | |
| Develop population set | X | X | |
| Analyze ref. genome | | X | X |
| Analyze population data | | X | X |
| Conference | | X | |
| Publish | | | X |
| Execute DMP | X | X | X |
| Teach | X | | |

An advisory committee composed of stakeholders and experts will be formed. Dr. Dana Nelson (USDA Forest Service) is a leader in tree improvement of forest trees and focused on forest health. He will provide guidance to maintain project relevancy to resource managers. Dr. Rachel Kappler (Holden Arboretum) is an expert in public outreach and citizen science, and will provide seed and collections-based knowledge. Dr. Stephen Keller developed genetic offset modeling and will provide expert guidance for gene conservation. We will meet semi-annually on zoom.

## EVALUATION PLAN

Progress will be evaluated through a set of milestones related to research and teaching. Research objectives will follow the timeline: *Objective 1* - deliver tissue to Dovetail; assemble draft genome; receive final genome assembly; and annotate the final assembly; *Objective 2* - obtain tissue for population set; deliver tissue to CGI; generate SNP set. Both objectives will merge during presentation of the work at Botany 2024 and during manuscript publication. My teaching objective is to teach a 1 credit conservation genomics seminar to 12 undergraduate or graduate students in Fall 2023. The dissemination plan is to present the work in at least one conference and manuscript.

## TRAINING/CAREER DEVELOPMENT

My career goal is to become an independent scientist employed as a professor in academia or as a government scientist. In either role, I plan to work in tree improvement and breeding to develop germplasm that may help solve complex environmental and societal problems. To achieve these goals, I have focused my research on population genetics (Fetter & Weakley, 2019), quantitative

genetics (Fetter *et al.,* 2021), and genomic methods (Fetter, 2019). My education and training have broadly prepared me to enter the field of conservation genomics. Reference genomes are an important tool for species conservation, and my current position is focused on developing genome annotation tools specialized for non-model eukaryotic organisms. The proposed work will integrate my current training with new knowledge for assembling a genome from short and long reads. Of the population genomic methods I propose to use, some are new and some are old to me. However, integrating over the complete cycle of reference genome assembly to population genomic analysis will enable me to work at a larger scale of science than ever before.

Mentoring is an important component of my career development and I have consistently engaged myself with students of all levels. In my current position as a PDA at UConn, I am a program mentor for the Biodiversity and Conservation Genomics program at the Center for Genome Innovation (CGI). I am a front-line mentor for four undergraduate students. The program is designed to introduce them to the entire process of genome assembly and annotation in a conservation setting. With the help of USDA Forest Service collaborators (Dr. Jennifer Koch), we are sequencing the genome of the endangered butternut (*Juglans cinerea*). I have mentored two undergraduate students who went on to graduate school J. Waterman (UVM) and N. Dietz (UGA).

To obtain teaching experience, I am currently a lead instructor for a 1 credit seminar in conservation genomics for the Fall 2022 semester. I designed the seminar, selected readings, and organized three introductory lectures for each module of the seminar. As part of the fellowship, I will lead the seminar again in Fall 2023 and deliver all lectures and lead discussions independently.

This fellowship will provide me a holistic experience as a principle investigator. I have numerous ideas for research topics, and this program will provide me the detailed introduction to federal funding of research that I need to progress to become an independent scientist.

## MENTORING PLAN

The mentor for the PD is Dr. Jill Wegrzyn, a leading expert in genome assembly and annotation, particularly in conifers and angiosperm trees. Dr. Wegrzyn's work was essential to publishing the loblolly pine (Neale *et al.,* 2014; Zimin *et al.,* 2014), sugar pine (Stevens *et al.,* 2016), (Neale *et al.,* 2017), giant sequoia (Scott *et al.,* 2020), coast redwood (Neale *et al.,* 2022), and cycas (Liu *et al.,* 2022) genomes. The mentoring and education by Dr. Wegrzyn will add a significantly new component to the PD's portfolio, which is predominantly in population genetics and quantitative genetics. Dr. Wegrzyn is also a leader in the department and in the classroom. She has trained 5 graduate students, 5 postdocs, in addition to 5 current graduate students and 1 current postdoc. Dr. Wegrzyn is a co-PI for the Biodiversity and Conservation Genomics program at UConn and is providing significant opportunity for the PD to mentor students. Dr. Wegrzyn is committed to training and educating the PD during weekly meetings, through review of code and manuscripts, and with support to conferences and professional development functions.

# Literature Cited:

1. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19,** 1655–1664 (2009).

2. Alverson, W. S., Lea, M. V. & Waller, D. M. A 20-year experiment on the effects of deer and hare on eastern hemlock regeneration. *Canadian Journal of Forest Research* **49,** 1329–1338 (2019).

3. Askins, R. A. *Saving the World's Deciduous Forests* (Yale University Press, 2014).

4. Bentz, S., Riedel, L. G., Pooler, M. R. & Townsend, A. Hybridization and self-compatibility in controlled pollinations of eastern North American and Asian hemlock (*Tsuga*) species. *Journal of Arboriculture* **28,** 200–205 (2002).

5. BGCI. *State of the World's Trees* tech. rep. (Botanical Gardens Conservation International, 2021).

6. Brandies, P., Peel, E., Hogg, C. J. & Belov, K. The value of reference genomes in the conservation of threatened species. *Genes* **10,** 846 (2019).

7. Brian, H. & Papanicolaou, A. *Transdecoder (Find Coding Regions Within Transcripts).* http://transdecoder.github.io (2022).

8. Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M. & Keller, S. R. Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Annual Review of Ecology, Evolution, and Systematics* **51** (2020).

9. Carrero, C., Beckman Bruns, E., Frances, A., Jerome, D., Knapp, W., Meyer, A., Mims, R., Pivorunas, D., Speed, D., Treher Eberly, A. & Westwood, M. Data sharing for conservation: A standardized checklist of US native tree species and threat assessments to prioritize and coordinate action. *Plants, People, Planet* **n/a** (2022).

10. Caughlin, T. T., Ferguson, J. M., Lichstein, J. W., Zuidema, P. A., Bunyavejchewin, S. & Levey, D. J. Loss of animal seed dispersal increases extinction risk in a tropical tree species due to pervasive negative density dependence across life stages. *Proceedings of the Royal Society B: Biological Sciences* **282,** 20142095 (2015).

11. Caye, K., Jumentier, B., Lepeule, J. & François, O. LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular biology and evolution* **36,** 852–860 (2019).

12. Ceballos, G., Ehrlich, P. R., Barnosky, A. D., Garcıa, A., Pringle, R. M. & Palmer, T. M. Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science advances* **1,** e1400253 (2015).

13. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34,** i884–i890 (2018).

14. Chen, Y., Jiang, Z., Fan, P., Ericson, P. G., Song, G., Luo, X., Lei, F. & Qu, Y. The combination of genomic offset and niche modelling provides insights into climate change-driven vulnerability. *Nature communications* **13,** 1–15 (2022).

15. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M. & Li, H. Twelve years of SAMtools and BCFtools. *GigaScience* **10** (Feb. 2021).

16. Domec, J.-C., Rivera, L. N., King, J. S., Peszlen, I., Hain, F., Smith, B. & Frampton, J. Hemlock Woolly Adelgid (*Adelges tsugae*) infestation affects water and carbon relations of Eastern Hemlock (*Tsuga canadensis*) and Carolina hemlock (*Tsuga caroliniana*). *New Phytologist* **199,** 452–463 (2013).

17. eFloras. *Flora of China: Tsuga (Endlicher) Carrière, Traité Gen. Conif. 185. 1855.* `http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=133931` (2012).

18. Ellis, N., Smith, S. J. & Pitcher, C. R. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93,** 156–168 (2012).

19. Ellison, A., Orwig, D., Fitzpatrick, M. & Preisser, E. The Past, Present, and Future of the Hemlock Woolly Adelgid (*Adelges tsugae*) and Its Ecological Interactions with Eastern Hemlock (*Tsuga canadensis*) Forests. *Insects* **9,** 172 (2018).

20. Ellison, A. M. Experiments are revealing a foundation species: a case study of eastern hemlock (*Tsuga canadensis*). *Advances in Ecology* (2014).

21. Ellison, A. M., Bank, M. S., Clinton, B. D., Colburn, E. A., Elliott, K., Ford, C. R., Foster, D. R., Kloeppel, B. D., Knoepp, J. D., Lovett, G. M., *et al.* Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. *Frontiers in Ecology and the Environment* **3,** 479–486 (2005).

22. Farjon, A. *Tsuga canadensis. The IUCN Red List of Threatened Species 2013: e.T42431A2979676.* `https://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T42431A2979676.en` (2022).

23. Feng, Y.-Y., Shen, T.-T., Shao, C.-C., Du, H., Ran, J.-H. & Wang, X.-Q. Phylotranscriptomics reveals the complex evolutionary and biogeographic history of the genus *Tsuga* with an East Asian-North American disjunct distribution. *Molecular Phylogenetics and Evolution* **157,** 107066 (2021).

24. Fetter, K. C. *Natural selection for disease resistance in hybrid poplars targets stomatal patterning traits and regulatory genes.* PhD thesis (University of Vermont, 2019).

25. Fetter, K. C., Nelson, D. M. & Keller, S. R. Growth-defense trade-offs masked in unadmixed populations are revealed by hybridization. *Evolution; International Journal of Organic Evolution* (2021).

26. Fetter, K. C. & Weakley, A. Reduced gene flow from mainland populations of *Liriodendron tulipifera* into the Florida peninsula promotes diversification. *International Journal of Plant Sciences* (2019).

27. Feucht, W., Schmid, M. & Treutter, D. Nuclei of *Tsuga canadensis*: Role of flavanols in chromatin organization. *International journal of molecular sciences* **12,** 6834–6855 (2011).

28. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology* **37,** 4302–4315 (2017).

29. Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y. & Keller, S. R. Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Molecular Ecology Resources* **21,** 2749–2765 (2021).

30. Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology letters* **18,** 1–16 (2015).

31. Fitzpatrick, M. C., Preisser, E. L., Porter, A., Elkinton, J. & Ellison, A. M. Modeling range dynamics in heterogeneous landscapes: invasion of the hemlock woolly adelgid in eastern North America. *Ecological Applications* **22,** 472–486 (2012).

32. Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J. A., Höglund, J., *et al.* The era of reference genomes in conservation genomics. *Trends in ecology & evolution* (2022).

33. Garibaldi, L. A., Steffan-Dewenter, I., Winfree, R., Aizen, M. A., Bommarco, R., Cunningham, S. A., Kremen, C., Carvalheiro, L. G., Harder, L. D., Afik, O., *et al.* Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *science* **339,** 1608–1611 (2013).

34. Gougherty, A. V., Keller, S. R. & Fitzpatrick, M. C. Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nature Climate Change* **11,** 166–171 (2021).

35. Gräler, B., Pebesma, E. J. & Heuvelink, G. B. Spatio-temporal interpolation using gstat. *R J.* **8,** 204 (2016).

36. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1075 (2013).

37. Gustafson, E. J., Miranda, B. R., Dreaden, T. J., Pinchot, C. C. & Jacobs, D. F. Beyond blight: Phytophthora root rot under climate change limits populations of reintroduced American chestnut. *Ecosphere* **13,** e3917 (2022).

38. Hampe, A. & Petit, R. J. Conserving biodiversity under climate change: the rear edge matters. *Ecology letters* **8,** 461–467 (2005).

39. Hart, J. L., van de Gevel, S. L., Sakulich, J. & Grissino-Mayer, H. D. Influence of climate and disturbance on the growth of *Tsuga canadensis* at its southern limit in eastern North America. *Trees* **24,** 621–633 (2010).

40. Havill, N. P., Campbell, C. S., Vining, T. F., LePage, B., Bayer, R. J. & Donoghue, M. J. Phylogeny and biogeography of *Tsuga* (Pinaceae) inferred from nuclear ribosomal ITS and chloroplast DNA sequence data. *Systematic botany* **33,** 478–489 (2008).

41. Holman, G., Del Tredici, P., Havill, N., Lee, N. S., Cronn, R., Cushman, K., Mathews, S., Raubeson, L. & Campbell, C. S. A new species and introgression in eastern Asian hemlocks (Pinaceae: *Tsuga*). *Systematic Botany* **42,** 733–746 (2017).

42. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research* **47,** D309–D314 (2019).

43. Ingwell, L., Brady, J., Fitzpatrick, M., Maynard, B., Casagrande, R. & Preisser, E. Intraspecific variation in *Tsuga canadensis* foliar chemistry. *Northeastern Naturalist,* 585–594 (2009).

44. Jetton, R. M., Mayfield III, A. E. & Powers, Z. L. Development of a rain down technique to artificially infest hemlocks with the hemlock woolly adelgid, *Adelges tsugae*. *Journal of Insect Science* **14** (2014).

45. Johnson, J. S., Gaddis, K. D., Cairns, D. M., Konganti, K. & Krutovsky, K. V. Landscape genomic insights into the historic migration of mountain hemlock in response to Holocene climate change. *American Journal of Botany* **104,** 439–450 (2017).

46. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics* **11,** 1–15 (2010).

47. Joshi, N. & J, F. *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33).* `https://github.com/najoshi/sickle.`.

48. Kinahan, I. G., Grandstaff, G., Russell, A., Rigsby, C. M., Casagrande, R. A. & Preisser, E. L. A four-year, seven-state reforestation trial with eastern hemlocks (*Tsuga canadensis*) resistant to hemlock woolly adelgid (*Adelges tsugae*). *Forests* **11,** 312 (2020).

49. Knott, J. A., Jenkins, M. A., Oswalt, C. M. & Fei, S. Community-level responses to climate change in forests of the eastern United States. *Global Ecology and Biogeography* **29,** 1299–1314 (2020).

50. Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L. & Pertea, M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome biology* **20,** 1–13 (2019).

51. Láruson, Á. J., Fitzpatrick, M. C., Keller, S. R., Haller, B. C. & Lotterhos, K. E. Seeing the forest for the trees: Assessing genetic offset predictions from gradient forest. *Evolutionary applications* **15,** 403–416 (2022).

52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25,** 1754–1760 (2009).

53. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

54. Li, X., Preisser, E. L., Boyle, K. J., Holmes, T. P., Liebhold, A. & Orwig, D. Potential social and economic impacts of the hemlock woolly adelgid in southern New England. *Southeastern Naturalist* **13,** 130–147 (2014).

55. Liu, Y., Wang, S., Li, L., Yang, T., Dong, S., Wei, T., Wu, S., Liu, Y., Gong, Y., Feng, X., *et al.* The Cycas genome and the early evolution of seed plants. *Nature Plants* **8,** 389–401 (2022).

56. McClure, M. S. Density-dependent feedback and population cycles in *Adelges tsugae* (Homoptera: Adelgidae) on *Tsuga canadensis*. *Environmental Entomology* **20,** 258–264 (1991).

57. McKenzie, E. A., Elkinton, J. S., Casagrande, R. A., Mayer, M., *et al.* Terpene chemistry of eastern hemlocks resistant to hemlock woolly adelgid. *Journal of Chemical Ecology* **40,** 1003–1012 (2014).

58. McMahon, B. J., Teeling, E. C. & Höglund, J. How and why should we implement genomics into conservation? *Evolutionary applications* **7,** 999–1007 (2014).

59. Melillo, J. M., Richmond, T., Yohe, G., *et al.* Climate change impacts in the United States. *Third national climate assessment* **52** (2014).

60. Merkle, S. A., Montello, P. M., Reece, H. M. & Kong, L. Somatic embryogenesis and cryostorage of eastern hemlock and Carolina hemlock for conservation and restoration. *Trees* **28,** 1767–1776 (2014).

61. Myers, J. A., Hart, J. L. & Cox, L. E. Decadal changes in disjunct eastern hemlock stands at its southern range boundary. *Castanea,* 171–182 (2015).

62. Nawaz, S. & Satterfield, T. On the nature of naturalness? Theorizing 'nature' for the study of public perceptions of novel genomic technologies in agriculture and conservation. *Environmental Science & Policy* **136,** 291–303 (2022).

63. Neale, D. B., McGuire, P. E., Wheeler, N. C., Stevens, K. A., Crepeau, M. W., Cardeno, C., Zimin, A. V., Puiu, D., Pertea, G. M., Sezen, U. U., *et al.* The Douglas-fir genome sequence reveals specialization of the photosynthetic apparatus in Pinaceae. *G3: Genes, Genomes, Genetics* **7,** 3157–3167 (2017).

64. Neale, D. B., Wegrzyn, J. L., Stevens, K. A., Zimin, A. V., Puiu, D., Crepeau, M. W., Cardeno, C., Koriabine, M., Holtz-Morris, A. E., Liechty, J. D., *et al.* Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome biology* **15,** 1–13 (2014).

65. Neale, D. B., Zimin, A. V., Zaman, S., Scott, A. D., Shrestha, B., Workman, R. E., Puiu, D., Allen, B. J., Moore, Z. J., Sekhwal, M. K., *et al.* Assembled and annotated 26.5 Gbp coast redwood genome: a resource for estimating evolutionary adaptive potential and investigating hexaploid origin. *G3* **12,** jkab380 (2022).

66. Nelson, C. D. Tree breeding, a necessary complement to genetic engineering. *New Forests,* 1–18 (2022).

67. Nicholson, C. C., Koh, I., Richardson, L. L., Beauchemin, A. & Ricketts, T. H. Farm and landscape factors interact to affect the supply of pollination services. *Agriculture, Ecosystems & Environment* **250,** 113–122 (2017).

68. Nowacki, G. J. & Abrams, M. D. Is climate an important driver of post-European vegetation change in the Eastern United States? *Global Change Biology* **21,** 314–334 (2015).

69. Paez, S., Kraus, R. H., Shapiro, B., Gilbert, M. T. P., Jarvis, E. D., Group, V. G. P. C., Al-Ajli, F. O., Ceballos, G., Crawford, A. J., Fedrigo, O., *et al.* Reference genomes for conservation. *Science* **377,** 364–366 (2022).

70. Park, D., Park, S.-H., Ban, Y. W., Kim, Y. S., Park, K.-C., Kim, N.-S., Kim, J.-K. & Choi, I.-Y. A bioinformatics approach for identifying transgene insertion sites using whole genome sequencing data. *BMC biotechnology* **17,** 1–8 (2017).

71. Parmesan, C. Ecological and evolutionary responses to recent climate change. *Annual review of ecology, evolution, and systematics,* 637–669 (2006).

72. Pederson, N., D'Amato, A. W., Dyer, J. M., Foster, D. R., Goldblum, D., Hart, J. L., Hessl, A. E., Iverson, L. R., Jackson, S. T., Martin-Benito, D., *et al.* Climate remains an important driver of post-European vegetation change in the eastern United States. *Global Change Biology* **21,** 2105–2110 (2015).

73. Potter, K. M., Campbell, A. R., Josserand, S. A., Nelson, C. D. & Jetton, R. M. Population isolation results in unexpectedly high differentiation in Carolina hemlock (*Tsuga caroliniana*), an imperiled southern Appalachian endemic conifer. *Tree Genetics & Genomes* **13,** 1–20 (2017).

74. Potter, K. M., Jetton, R. M., Dvorak, W. S., Hipkins, V. D., Rhea, R. & Whittier, W. A. Widespread inbreeding and unexpected geographic patterns of genetic variation in eastern hemlock (*Tsuga canadensis*), an imperiled North American conifer. *Conservation Genetics* **13,** 475–498 (2012).

75. Preisser, E. L., Miller-Pierce, M. R., Vansant, J. & Orwig, D. A. Eastern hemlock (*Tsuga canadensis*) regeneration in the presence of hemlock woolly adelgid (*Adelges tsugae*) and elongate hemlock scale (*Fiorinia externa*). *Canadian Journal of Forest Research* **41,** 2433–2439 (2011).

76. Preisser, E. L., Oten, K. L. & Hain, F. P. Hemlock woolly adelgid in the eastern United States: What have we learned? *Southeastern Naturalist* **13,** 1–15 (2014).

77. Radville, L., Chaves, A. & Preisser, E. L. Variation in plant defense against invasive herbivores: evidence for a hypersensitive response in eastern hemlocks (*Tsuga canadensis*). *Journal of Chemical Ecology* **37,** 592–597 (2011).

78. Ran, J.-H., Shen, T.-T., Wu, H., Gong, X. & Wang, X.-Q. Phylogeny and evolutionary history of Pinaceae updated by transcriptomic analysis. *Molecular phylogenetics and evolution* **129,** 106–116 (2018).

79. Razgour, O., Forester, B., Taggart, J. B., Bekaert, M., Juste, J., Ibáñez, C., Puechmaille, S. J., Novella-Fernandez, R., Alberdi, A. & Manel, S. Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proceedings of the National Academy of Sciences* **116,** 10418–10423 (2019).

80. Rellstab, C., Dauphin, B. & Exposito-Alonso, M. Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications* **14,** 1202–1212 (2021).

81. Robinson, N. M., Rhoades, C., Pierson, J., Lindenmayer, D. B. & Banks, S. C. Prioritising source populations for supplementing genetic diversity of reintroduced southern brown bandicoots *Isoodon obesulus obesulus*. *Conservation Genetics* **22,** 341–353 (2021).

82. Rossetto, M., Yap, J.-Y. S., Lemmon, J., Bain, D., Bragg, J., Hogbin, P., Gallagher, R., Rutherford, S., Summerell, B. & Wilson, T. C. A conservation genomics workflow to guide practical management actions. *Global Ecology and Conservation* **26,** e01492 (2021).

83. Savolainen, O., Lascoux, M. & Merilä, J. Ecological genomics of local adaptation. *Nature Reviews Genetics* **14,** 807–820 (2013).

84. Scott, A. D., Zimin, A. V., Puiu, D., Workman, R., Britton, M., Zaman, S., Caballero, M., Read, A. C., Bogdanove, A. J., Burns, E., *et al.* A reference genome sequence for giant sequoia. *G3: Genes, Genomes, Genetics* **10,** 3907–3919 (2020).

85. Shafer, A. B., Wolf, J. B., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., *et al.* Genomics and the challenging translation into conservation practice. *Trends in ecology & evolution* **30,** 78–87 (2015).

86. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31,** 3210–3212 (2015).

87. Song, L., Sabunciyan, S., Yang, G. & Florea, L. A multi-sample approach increases the accuracy of transcript assembly. *Nature communications* **10,** 1–7 (2019).

88. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic acids research* **33,** W465–W467 (2005).

89. Stevens, K. A., Wegrzyn, J. L., Zimin, A., Puiu, D., Crepeau, M., Cardeno, C., Paul, R., Gonzalez-Ibeas, D., Koriabine, M., Holtz-Morris, A. E., *et al.* Sequence of the sugar pine megagenome. *Genetics* **204,** 1613–1626 (2016).

90. Stoetzel, M. B., Onken, B., Reardon, R. & Lashomb, J. *History of the introduction of Adelges tsugae based on voucher specimens in the Smithsonian Institute National Collection of Insects* in *Proceedings: Hemlock Woolly Adelgid in the Eastern United States Symposium* **12** (2002).

91. Sullivan, K. A. & Ellison, A. M. The seed bank of hemlock forests: implications for forest regeneration following hemlock decline1. *The Journal of the Torrey Botanical Society* **133,** 393–402 (2006).

92. Taylor, R. J. The relationship and origin of *Tsuga heterophylla* and *Tsuga mertensiana* based on phytochemical and morphological interpretations. *American Journal of Botany* **59,** 149–157 (1972).

93. Vaganov, E., Anchukaitis, K. & Evans, M. Dendroclimatology: progress and prospects. *How well understood are the processes that create dendroclimatic records,* 37–75 (2011).

94. Verrico, B. M., Weiland, J., Perkins, T. D., Beckage, B. & Keller, S. R. Long-term monitoring reveals forest tree community change driven by atmospheric sulphate pollution and contemporary climate change. *Diversity and Distributions* **26,** 270–283 (2020).

95. Webster, C., Shrestha, S., Zaman, S., Vuruputoor, V., Bennett, J., Monyak, D., Richter P amd Bhattarai, A., Fetter, K. & J, W. *EASEL (Efficient, Accurate, Scalable Eukaryotic modeLs), an improved eukaryotic genome annotation tool* Paper presented at the North American Forest Genetics Society meeting, Monterrey California. 2022.

96. Westbrook, J. W., Zhang, Q., Mandal, M. K., Jenkins, E. V., Barth, L. E., Jenkins, J. W., Grimwood, J., Schmutz, J. & Holliday, J. A. Optimizing genomic selection for blight resistance in American chestnut backcross populations: A trade-off with American chestnut ancestry implies resistance is polygenic. *Evolutionary applications* **13,** 31–47 (2020).

97. Wilder, A. P., Dudchenko, O., Curry, C., Korody, M., Turbek, S. P., Daly, M., Misuraca, A., Wang, G., Khan, R., Weisz, D., *et al.* A chromosome-length reference genome for the endangered Pacific pocket mouse reveals recent inbreeding in a historically large population. *Genome biology and evolution* **14,** evac122 (2022).

98. Willi, Y., Kristensen, T. N., Sgrò, C. M., Weeks, A. R., Ørsted, M. & Hoffmann, A. A. Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. *Proceedings of the National Academy of Sciences* **119,** e2105076119 (2022).

99. Zhang, P., Zhao, Y., Li, C., Lin, M., Dong, L., Zhang, R., Liu, M., Li, K., Zhang, H., Liu, X., *et al.* An Indo-Pacific humpback dolphin genome reveals insights into chromosome evolution and the demography of a vulnerable species. *Iscience* **23,** 101640 (2020).

100. Zimin, A., Stevens, K. A., Crepeau, M. W., Holtz-Morris, A., Koriabine, M., Marçais, G., Puiu, D., Roberts, M., Wegrzyn, J. L., de Jong, P. J., *et al.* Sequencing and assembly of the 22-Gb loblolly pine genome. *Genetics* **196,** 875–890 (2014).

101. Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L. & Yorke, J. A. The Ma-SuRCA genome assembler. *Bioinformatics* **29,** 2669–2677 (2013).