Title

Landscape Genomics to Enable Conservation Actions: The California Conservation Genomics Project

Authors

Shaffer, Howard B
Toffelmier, Erin
Corbett-Detig, Russ B
et al.

Publication Date

2022-04-08

Peer reviewed

Journal of Heredity, 2022, **XX**, 1–12
https://doi.org/10.1093/jhered/esac020
Advance access publication 8 April 2022
Original Article

Original Article

Landscape Genomics to Enable Conservation Actions: The California Conservation Genomics Project

H. Bradley Shaffer*, Erin Toffelmier*, Russ B. Corbett-Detig, Merly Escalona, Bjorn Erickson, Peggy Fiedler, Mark Gold, Ryan J. Harrigan, Scott Hodges, Tara K. Luckau, Courtney Miller, Daniel R. Oliveira, Kevin E. Shaffer, Beth Shapiro, Victoria L. Sork, and Ian J. Wang

From the Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA (H. B. Shaffer, Toffelmier, Luckau, Miller, Oliveira, and Sork); La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095, USA (H. B. Shaffer, Toffelmier, Harrigan, Luckau, Miller, Oliveira, and Sork); Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA 95064, USA (Corbett-Detig, Escalona); U.S. Fish and Wildlife Service, Sacramento, CA 95825, USA (Erickson); Natural Reserve System, Office of the President, University of California, Oakland, CA 94607, USA (Fiedler); California Natural Resources Agency, 1416 Ninth Street, Suite 1311, Sacramento, CA 95814, USA (Gold); Center for Tropical Research, Institute for Environment and Sustainability, University of California, Los Angeles, CA 90095, USA (Harrigan); Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA 93106, USA (Hodges); California Department of Fish and Wildlife, Fisheries Branch, West Sacramento, CA 95605, USA (K. Shaffer); Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95064, USA (Shapiro); Howard Hughes Medical Institute, University of California, Santa Cruz, Santa Cruz, CA 95064, USA (Shapiro); Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA (Wang); Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720, USA (Wang)

*Co-first authors.
Address correspondence to e-mail: brad.shaffer@ucla.edu

Corresponding Editor: William Murphy

Abstract

The California Conservation Genomics Project (CCGP) is a unique, critically important step forward in the use of comprehensive landscape genetic data to modernize natural resource management at a regional scale. We describe the CCGP, including all aspects of project administration, data collection, current progress, and future challenges. The CCGP will generate, analyze, and curate a single high-quality reference genome and 100–150 resequenced genomes for each of 153 species projects (representing 235 individual species) that span the ecological and phylogenetic breadth of California's marine, freshwater, and terrestrial ecosystems. The resulting portfolio of roughly 20 000 resequenced genomes will be analyzed with identical informatic and landscape genomic pipelines, providing a comprehensive overview of hotspots of within-species genomic diversity, potential and realized corridors connecting these hotspots, regions of reduced diversity requiring genetic rescue, and the distribution of variation critical for rapid climate adaptation. After 2 years of concerted effort, full funding ($12M USD) has been secured, species identified, and funds distributed to 68 laboratories and 114 investigators drawn from all 10 University of California campuses. The remaining phases of the CCGP include completion of data collection and analyses, and delivery of the resulting genomic data and inferences to state and federal regulatory agencies to help stabilize species declines. The aspirational goals of the CCGP are to identify geographic regions that are critical to long-term preservation of California biodiversity, prioritize those regions based on defensible genomic criteria, and provide foundational knowledge that informs management strategies at both the individual species and ecosystem levels.

Key words: climate change, California Floristic Province, landscape genetics, non-model organism, whole-genome resequencing

Introduction

Conservation genomics can, and should, provide critical information for the future preservation of biodiversity. For decades, human activity has transformed and fragmented landscapes causing extirpations of local populations and disruptions of the natural metapopulation dynamics that sustain species and community assemblies (Barnosky et al. 2011). Adding to those alterations are the less immediately obvious, but equally disruptive effects of climate change. In many parts of the world, and particularly in western North America, climate change is altering ecosystems at rates that were previously unimaginable, threatening plants, animals, and the habitats in which they live (Shukla et al. 2019). Over the last decade these habitats have endured record heat and snowfall, severe drought, hurricane-associated flooding, and devastating wildfires that have wreaked havoc on natural resources (Schoennagel et al. 2017; Gershunov et al. 2019). An important goal for biological conservation is to establish habitat and species protection strategies that are most resilient to these environmental challenges. We need to catalog remaining

Received February 21, 2022; Accepted April 7, 2022

genetic variation as it exists within species. But we also need to understand which lineages and regions will be most, and least, resilient to the anthropogenic changes that are predicted for the remainder of the 21st century.

Traditionally, the discipline of conservation genetics has accomplished its goals by focusing on species that are in jeopardy of extinction, identifying strategies for the protection of contained populations and lineages (Frankham et al. 2017). In the US, the Endangered Species Act is the embodiment of this approach, and its success has been undeniable (Schwartz 2008). But individual species protection is also slow and incremental. To keep pace with the challenge, we must scale up, preserving the processes that will simultaneously protect many species and their communities, ensuring long-term ecosystem viability. A variety of strategies, including the establishment of protected parklands and wilderness areas, Habitat Conservation Plans, Natural Community Conservation Plans, and Marine Protected Areas, use habitat protection and management, in conjunction with state and federal species legislation, to protect species and the landscapes on which they depend.

However, the best way to achieve long-lasting multispecies protection is to identify those populations and landscapes that stand the best chance to persist without human intervention. The logic behind this approach is simple. Rather than exclusively focusing management activities on the recovery needs of individual species or populations, attention should also center on protecting and connecting areas that harbor the greatest multi-species genetic variation. This landscape genetic approach maximizes the potential for natural ecological and evolutionary processes to promote species persistence and alleviate our reliance on human intervention as the sole solution to species protection (Smith et al. 1997; Carroll et al. 2014; Scott et al. 2020). By identifying, protecting, and connecting landscapes that harbor genomically diverse individuals and populations of many species, we increase the likelihood of preserving both the individuals, and the evolutionary processes, that allow populations to rapidly adapt to changing conditions.

Implementing a genetically informed, geographically coherent strategy for species conservation requires three complementary approaches. The first uses the tools of landscape genetics to identify units within species that are genetically isolated and evolutionarily distinct, and therefore comprise the major units of conservation and management (Manel et al. 2003; Balkenhol et al. 2015). Such units are sometimes referred to as evolutionarily significant units, or ESUs, and their identification is both an essential step in species management and the primary way in which genetic data can contribute to species conservation. Examples from diverse systems in western North America include uncovering extreme genetic structure found in the stream-breeding foothill yellow-legged frog (McCartney-Melstad et al. 2018), detecting endangered and threatened salmon stock across the Sacramento-San Joaquin River system (Banks et al. 2014), and identifying population clusters of rare plants endemic to the Mojave Desert (Wolfe et al. 2016). A second critical element of conservation genetics is identifying regions of greatest genetic diversity within and across ESUs. By establishing relationships between patterns of genomic variation and natural and anthropogenic features of the environment, landscape genetic analyses can identify regions of low and high genetic diversity as well as their likely causal agents. While extreme cases can occur where successful purging of deleterious variants may reduce genetic load in particular taxa (Robinson et al. 2016), such events are apparently limited geographically and in their overall fitness consequences (Mathur and DeWoody 2021). As such, the generally accepted strategy for conservation action has been to prioritize and protect genetic variation within and between ESUs.

For single taxa, these two uses of genetic data have been recognized for decades as essential components of endangered species protection and recovery planning (https://ecos.fws.gov/ecp/report/species-listings-by-year-totals). However, when many species are analyzed across the same landscape, generalized regions of high and low genetic diversity, corridors of connectivity, and barriers to gene flow that define species assemblages may emerge, allowing for more strategic multispecies conservation actions (Thomassen et al. 2011). When many loci are analyzed, landscape genetics expands to landscape genomics, enabling genome-wide analyses that include the quantification of demographic history, gene flow, and natural selection (Sork et al. 2013; Funk et al. 2019; Teixeira and Huber 2021; DeWoody et al. 2021).

A third and final type of analysis examines the spatial patterns of adaptive, rather than neutral, genomic variation. Identifying "genes that matter" is a non-trivial task that typically requires landscape-level sampling, a well-annotated reference genome, and whole-genome resequencing of many individuals. The identification of strongly selected loci and their functionally important allelic variation can aid activities including post-fire forest replanting efforts (Browne et al. 2019) and genomic rescue of genetically depauperate populations (Frankham et al. 2017). These "outlier loci" may be specific to individual populations or species, or they may characterize collections of related taxa that have evolved in extreme environments. Whichever is the case, a deeper understanding of the taxonomic and geographic distributions of such loci is widely viewed as a critical component of conservation genomics and the discipline's contribution to ameliorating the negative impacts of climate change on biodiversity.

Many landscape genomics studies have used reduced representation approaches that sequence a small percentage of the genome, such as restriction site associated DNA sequencing, genotyping by sequencing, and target capture. Particularly for analyses that depend on neutral variation, studies using reduced representation techniques can provide important data that define ESUs and quantify overall genetic variation, effective population size, and recent and historical demography (Bay et al. 2018). Other large-scale projects use targeted DNA sequencing (DNA barcoding) of single loci to characterize community composition and biodiversity, sometimes at massive scales (Hobern 2021; International Barcode of Life Consortium, http://ibol.org). However, with advances in DNA sequencing, largely driven by the human genomics research community, conservation and landscape genomics can now be scaled up to the whole-genome level. It is now feasible to assemble reference genomes for hundreds to thousands of taxa (Formenti et al. 2022; Lewin et al. 2022) and embark on whole-genome resequencing to deliver data at the spatial scales relevant to conservation decision makers. This landscape genomics approach can deliver insights into the evolutionary and ecological histories of populations distributed throughout species' ranges, corridors and barriers to gene movement, and climate-resilient patterns of adaptive genetic

variation. Critically, this information provides a foundation for evidence-based resource management to protect self-sustaining populations that have the best chances to survive and adapt to rapid climate change with minimal human intervention. Lacking other information and faced with unprecedented levels of climate-change-driven selection, the best bet for many taxa may be to protect populations harboring the greatest levels of genomic variation. This general approach, which is an extension of the fundamental breeder's equation in quantitative genetics (Lande and Arnold 1983; Falconer and Mackay 1996), has long been advocated both to avoid inbreeding depression (Frankham 2005) and to maximize "adaptive potential" (Carroll et al. 2014; Smith et al. 2020) and can lead to enhanced survivorship of translocated endangered species (Scott et al., 2020).

In this paper, we describe the California Conservation Genomics Project (CCGP), the first effort to produce a comprehensive, landscape genomic analysis for conservation planning and implementation at the level of a large and complex biodiversity hotspot. We briefly review the administration and oversight of this publicly funded, multi-investigator effort, describe our framework for collecting and processing samples, discuss progress and challenges, and provide details on our goals and expected outcomes. In so doing, we provide a model for how the global quest to sequence every living species can be combined with the power of landscape genomics to bring conservation genomics to a new level of relevance in California, across the United States, and the planet.

The California Conservation Genomics Project

CCGP overview

In 2019, California launched an ambitious project to characterize the genomic variation of at-risk and ecologically significant species within its borders. Consistent with the 2018 California Biodiversity Initiative (https://www.californiabiodiversityinitiative.org/pdf/california-biodiversity-action-plan.pdf), the California Conservation Genomics Project (https://www.ccgproject.org) was created to enhance and modernize actionable conservation and wildlife management by adding landscape genomics to our existing conservation toolkit.

The fundamental goal of the CCGP is to generate a comprehensive database of genomic variation and associated georeferenced environmental data, and to use that database to help guide the protection of species and ecosystems that may be vulnerable to climate change and other anthropogenic threats. We accomplish this goal by identifying regions and landscapes that harbor high genetic diversity and large effective population sizes, or that serve as potential or realized corridors connecting such habitat patches. We consider such local populations that retain genetic diversity to be genetically resilient with a high probability of responding to future environmental change. By identifying landscapes with the greatest levels of standing genetic variation across a large set of diverse taxa, California can capitalize on the ability of species to evolve and adapt to anthropogenic changes on their own, without additional human interventions.

California is the ideal landscape for this bold, forward-looking approach to natural resource management. It is the most populous state in the United States of America, accommodating approximately 12% of the nation's human population in 5% of its continental land area; at a global scale, California ranks 59th in size compared with the roughly 200 generally recognized countries. The state harbors extremely high levels of native biodiversity and a similarly high number of at-risk, declining, or listed species. California has the greatest number of documented and possibly extinct species of vascular plants in the United States of America (Kartesz 2015) and almost twice as many federally protected plant and animal species (total of 287) as any other state in the continental United States of America (although Hawaii has 503; U.S. Fish and Wildlife Service 2021). As a consequence, the California Floristic Province was the only North American ecoregion to qualify for the first global assessment of 25 Biodiversity Hotspots, defined as regions that are both species rich and at greatest threat of species loss; it remains on the expanded list of 36 global hotspots (Mittermeier et al. 2004). California also has a rich history of research on the genetics of natural populations (Beninde et al., 2022). However, these earlier studies, while individually useful, lack the standardized sampling, data consistency and completeness, and landscape analyses required for drawing broadly comparative conclusions that are critical for broad scale conservation.

The CCGP will inform strategic geographic infrastructure investments that maximize ecosystem health and future wildlife management. As part of the state's actions for biodiversity conservation, California Governor Newsom issued the biodiversity Executive Order (EO-N-82-20) in 2020. It includes a commitment to protect biodiversity and to conserve 30% of California land and waters by 2030 (the "30 × 30" conservation framework). California is in the process of completing a pathway document for 30 × 30 that includes conserving biodiversity in a changing climate by creating climate refugia and greater climate resilience in state-conserved lands and coastal waters. As the initiative is implemented, the CCGP will provide a genomic roadmap that will help identify those critical landscapes. This project will also provide genomic information on species and ESUs that are genetically depleted and no longer resilient, but would benefit from assisted migration and genetic rescue efforts. This latter outcome should be particularly important for the many endangered plants with small, isolated populations that characterize many of the most endangered elements of the California flora (Vu et al. 2021).

CCGP Goals and Deliverables

Our overarching goal is to develop a landscape-level understanding of regions of greatest (and least) genomic variation, corridors of potential connectivity for the flow of critical genetic variants, and natural barriers to gene flow across species and ecoregions. To do so, we are generating data from a large number of "species projects" covering vascular plants, cryptogams, lichens, vertebrates, and metazoan invertebrates across California's 19 terrestrial ecoregions and inshore marine habitats. Our objective is to identify the major geographical units of conservation and management for each species, as well as the most resilient patches of habitat across taxonomic scales ranging from individual species to ecologically similar multi-species assemblages, to a full complement of terrestrial and marine species. Throughout this paper, we refer to a species project as the unit of analysis for the CCGP. A species project may be a single species, with or without contained subspecies, or it may be a set of closely related, geographical replacement species that are ecologically and

genetically similar. An example of the former is valley oak (*Quercus lobata*), and an example of the latter is the whip snake genus *Masticophis* (*Masticophis lateralis lateralis*, *M. l. euryxanthus*, *Masticophis flagellum piceus*, *M. f. ruddocki*, *Masticophis fuliginosus*, and *Masticophis taeniatus*).

During the now-completed first phase of the CCGP, the Scientific Executive Committee (SEC) decided on a group of species projects and a concrete set of achievable goals for all projects. We identified and funded 153 species projects encompassing 235 species, many of which contain subspecies of high conservation concern. We prioritized species that are broadly distributed across the state, and therefore inform statewide assessments of genomic variation, but also contain at-risk or declining segments that would benefit from the insights of landscape genomic analysis (see Supplementary Information for additional details). This number also includes 13 agricultural pest species identified by the California Department of Food and Agriculture as high-priority taxa whose control would benefit from a deeper understanding of their sources of introduction across the state.

A species project represents a collaboration between CCGP scientists and the principal investigator. For each project, the CCGP generates a high-quality reference assembly with chromosome-length scaffolds based on a fresh, flash-frozen sample provided by that project's principal investigator. For the landscape genomics resequencing effort, the principal investigator is responsible both for sample collections and the generation of genomic data following a uniform set of sampling and genomic protocols. For each species project, landscape genomic inferences are based on 100–150 individuals sampled from across their California distribution (including occasional out of state locations), with the genome of each individual resequenced to an average 10x coverage, regardless of genome size. These data are run through standardized bioinformatic and landscape genomic pipelines developed and implemented by the CCGP, and constitute our core data. The raw data are also immediately made available to each principal investigator. In this way, a standardized set of analyses across taxa to characterize California genomic diversity is developed, published, and made publicly available by the entire CCGP consortium, while individual researchers can pursue more specialized, species and question-specific analyses at their own pace. Given the broad phylogenetic coverage of our species projects (Figure 1), our assembled reference genomes are also being incorporated into larger projects under the Earth BioGenome Project umbrella (Lewin et al. 2022) and are complementary to reference genome projects in other geographic regions like the Darwin Tree of Life Project (The Darwin Tree of Life Project Consortium 2022).

An Administrative Framework to Enable Actionable Conservation Science

Establishing a Fair and Equitable Approach to Decision-Making

A major priority for the CCGP is fair and equitable project management across researchers and species. Given our mandate from the state, we have worked primarily with University of California principal investigators, encouraging them to include additional academic, agency, and NGO collaborators at their discretion.

Our first step was to consult a large and diverse group of stake-holders, and establish a decision-making framework. Upon receiving notice of funding from the State of California ($12M USD), we held the CCGP Design Planning Meeting on September 5, 2019, at UCLA. In attendance were faculty representatives from each UC campus with expertise in conservation and/or genomics, and state, federal and NGO practitioners who would benefit from the data generated by the project. An important consideration for this meeting was recognizing that it was not limited to "conservation genomicists", and thus to potential recipients of future funding. Rather, we sought input from taxonomic, regional, and conceptual leaders who could help us shape the direction of our future genomic efforts. Through a consensus-based process, the group of 55 attendees developed recommendations and guiding principles covering project administration, strategies for broadly distributing research funds to the most qualified teams, criteria for selection of species and preliminary recommended species, sampling designs, types of genomic data to generate, and data management. In addition to building consensus on how the research would be done across essentially the entire plant and metazoan tree of life, meeting participants also arrived at an administrative structure that would ensure equitable decision-making and high-quality research across all of the included projects.

Project Administration

CCGP is led by an administrative leadership team that makes decisions about the conservation science approach of the project, with several additional small-group teams to guide research and outreach decision making (Figure 2). The Reference Genome Group establishes protocols for consistent data generation and assembly. To coordinate and help troubleshoot these efforts, a Core Facility Technical Coordination Group works with the CCGP administration and collaborating UC core sequencing facilities. Two additional groups, the Bioinformatics and Landscape Genomics Analysis Groups, are tasked with developing uniform bioinformatic and landscape genomic statistical analysis pipelines. The latter two groups are led by UC faculty with support from CCGP-funded postdoctoral scholars and staff. For additional details on other advisory and technical support groups, see Supplementary Information.

Moving to Action

A critical decision from the CCGP Design Planning Meeting was that research be conducted by separate research teams, but also coordinated into a single, unified research product. Based on these discussions, the group decided that each species project includes: a single high-quality reference genome, comprehensive sampling across landscapes within species projects, only whole-genome resequencing, standardized informatic pipelines for genotyping, and standardized landscape genomic statistical analyses.

Choosing and Funding Species Projects

California is simultaneously one of the most biodiverse and ecologically threatened regions in the world and choosing a relatively small set of species for analysis by the CCGP was a non-trivial task. We established a proposal review process (see Supplementary Information for additional details) and outlined 4 criteria for species selection:

1) Conservation prioritization: We placed a high priority on species that were important conservation targets, as

Figure 1. Phylogenetic and taxonomic distribution of all species for which a reference genome is being assembled under the California Conservation Genomics Project. The tree was generated with the CommonTree tool from NCBI, and it is based on the classification in the NCBI taxonomy database. The tree encompasses a total of 9 phyla, 25 classes, 74 orders, 114 families, 146 genera, and 158 species.

identified by state and/or federally defined endangerment status, level of threat due to commercial exploitation, and general importance to conservation in California.

2) Ecological breadth: Species, or species complexes that are wide-ranging across California both ecologically and geographically, and therefore contribute to an overall picture of genetic variation across the state's 19 USDA ecoregions (Figure 3), were strongly favored over more geographically restricted taxa.
3) Taxonomic breadth: We strived to have reasonable representation from across the major clades represented in California (see Figure 1).
4) Feasibility and technical expertise: We favored teams with strong taxonomic and methodological expertise, and projects for which samples were already available and the proposed work could be completed on time and within budget.

Given our goal of producing consistent and comparable data sets across all included projects, we established a relatively narrow set of technical and sampling standards. These included range-wide sampling across California with a single individual per site (occasionally more), sample sizes of 100–150 individuals/sites per species project, species with an estimated genome size of less than 5 GB, and projects for which sampling was complete or for which sampling could be completed within 1 year of the award. To ensure consideration of the widest range of taxa, we solicited project proposals from the

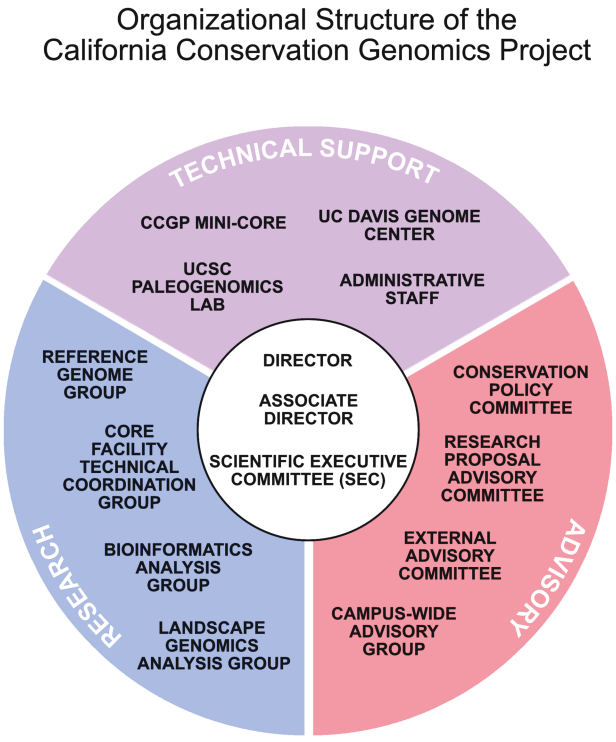

Figure 2. Organizational structure of the California Conservation Genomics Project. The project staff and leadership are organized into three main groups: Technical Support, Advisory, and Research, with a central leadership that includes a Director/Principal Investigator, Associate Director, and Scientific Executive Committee. The organizational structure is depicted in circular format, reflecting the co-equal importance and collaborative working relationships among groups and the central role of the directorship team in all aspects of the project.

entire University of California research community. In total, the CCGP funded 78 projects comprising 148 unique genera and 235 generally recognized species spanning the tree of life as represented in California. Our final sampling includes 29 marine species (13 invertebrates, 4 plants, 12 vertebrates) and 206 terrestrial/freshwater species (54 invertebrates, 68 plants, and 84 vertebrates; see Supplementary Table 1). Because our strategy for sampling and sequencing was consistent across projects, funding was also relatively standard for each project, with small deviations based on the genome sizes of target taxa.

Producing the Genomic Data

Reference genomes are assembled following a protocol adapted from Rhie et al. (2021). Assemblies are comprised of PacBio HiFi long read data, which is scaffolded using Omni-C (Dovetail Genomics) chromatin conformation data. Our minimum target reference genome quality is 6.7.Q40, and in most cases we expect to reach 7.C.Q50 or better (see Table 1 in Rhie et al. 2021). For five species (*Quercus lobata*, valley oak (Sork et al. 2022); *Strongylocentrotus purpuratus*, purple sea urchin (Sodergren et al. 2006), *Bactrocera dorsalis*, oriental fruit fly (ASM2028386v1); *Ceratitis capitata*, Mediterranean fruit fly (Ward et al. 2021); and threespine stickleback (Nath et al. 2021)), we were able to use an existing reference genome that reached or exceeded this standard (see Supplementary Table 1). Genome annotations using transcriptomic data from up to 7 different tissues per species will be added as those data become available (see Supplementary Information for additional details).

We expect that the final data set for the CCGP will be ~20 000 resequenced genomes. The choice of library preparation type and benchwork approach is left to the discretion of project primary investigators. While many lab groups possess the technical expertise to generate whole genome resequencing libraries, the CCGP also established a dedicated DNA extraction and library preparation facility, the CCGP Mini-Core, which PIs may utilize at their discretion (for additional details, see Supplementary Information).

Sampling the Landscape

To cover as much of California as possible with a collection of 100–150 samples per species, the CCGP utilizes an individual-based, rather than population-based, sampling scheme. This approach overcomes the need to delineate populations *a priori* and maximizes geographic coverage for a given sample size (Manel et al. 2003; Prunier et al. 2013; Sork et al. 2013; Seaborn et al. 2019). By analyzing individuals distributed across the landscape, this sampling strategy also maximizes the spatial extent of each species project, reduces gaps between samples, provides finer spatial resolution for inferences of gene flow and population structure (Manel et al. 2003; Balkenhol et al. 2015), captures more of the environmental variation across the state, and provides greater power to identify environmental correlates of genetic variation (Manel et al. 2012; Selmoni et al. 2020).

Our sampling design follows many general recommendations for landscape genomic analyses (Manel et al. 2012; Hall and Beissinger 2014; Wang and Bradburd 2014). Stratified sampling across environmental space is most efficient for estimating response curves that capture the relationships between species and the environment, and outperforms both random and uniform geographic sampling for many analyses (Schwartz and McKelvey 2008; Albert et al. 2010; Manel et al. 2012). Further informing our sampling schemes with "biological space" (Manel et al. 2012)—including population demography, evolutionary history, and phenotypic variation—sometimes provides additional power to capture ecologically important patterns of genetic variation.

Given the critical importance of sampling for our project outcomes, sampling schemes were designed to maximize coverage across geographic and environmental space while minimizing spatial autocorrelation. As part of the application process, principal investigators submitted a list of georeferenced localities for each species project. We characterized environmental space by performing a principal components analysis on rasters of bioclimatic variables from the WorldClim database (www.worldclim.org). The first 2PCs resulting from this analysis explain 99.2% of bioclimatic variation across California (Figure 4). High variable loadings for PC1 (90.3% of the total variation) were almost entirely based on temperature seasonality (bio4), ranging from the seasonally stable high desert basins to the variable Sierra Nevada mountains. PC2 largely reflected several aspects of annual (bio12) and seasonal (bio16, bio19) precipitation, ranging from the hot and dry deserts in the southeast to the cooler Pacific temperate rainforest along the northwest coast.

Minimizing spatial autocorrelation is essential for providing the statistical power to disentangle geographic and environmental factors that can influence the spatial distribution

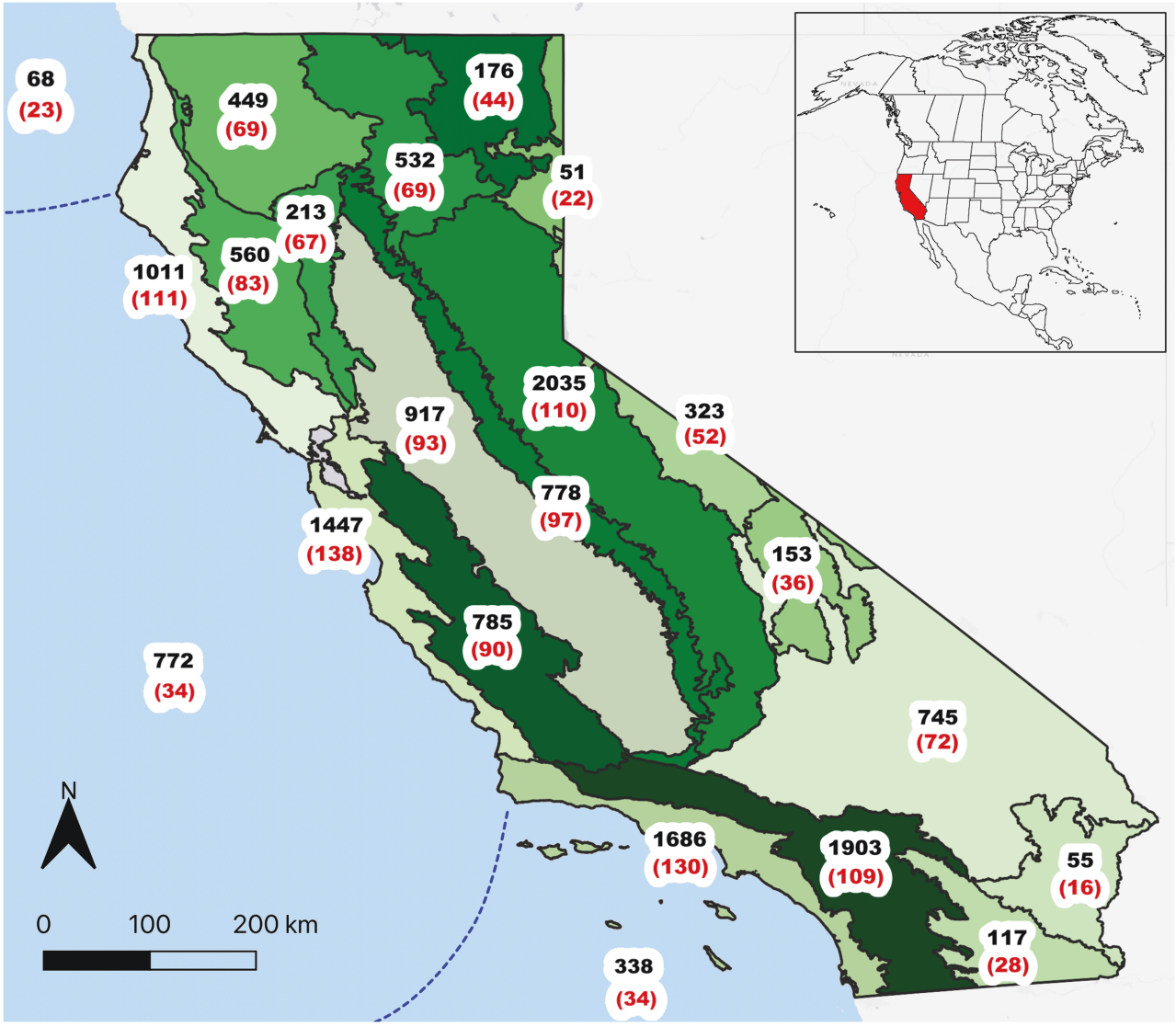

Figure 3. Summary of sampling across California and adjacent marine ecoregions for the California Conservation Genomics Project. Polygons represent the 19 USDA Ecoregions (downloaded from DataBasin, https://databasin.org/datasets/81a3a809a2ae4c099f2e495c0b2ecc91) as reported by Goudey and Smith (1994) and are colorized for visual clarity only (Cleland et al. 2007; Goudey and Smith 1994). Numbers for each ecoregion represent the number of samples (black) and the number of species (red, in parentheses) that have been collected or are anticipated to be collected and sequenced in that ecoregion. Marine regions were separated (blue-dashed lines) based on demarcations by Spalding et al. (2007).

of genetic variation (Manel et al., 2012; Wang and Bradburd, 2014) and sampling schemes were designed to reduce collinearity between environmental variables and to minimize the correlation of geographic and environmental distances between samples (see Supplementary Information for additional details).

Data Analysis and Hosting

Assembling and analyzing approximately 150 reference genomes and 20 000 resequenced individuals in a consistent manner that optimizes the speed and accuracy of data acquisition and management has required a team effort across multiple labs and investigators. Here, we simply emphasize our overarching objectives:

- Reference genomes should be produced, assembled, and stored to maximize efficiency, availability, and quality. This entails both optimizing data acquisition pipelines and minimizing human curation.
- Resequencing data, for genomes ranging from less than a megabase to roughly 5 GB should be sequenced, mapped, and called for variants in a single pipeline that produces compatible data across taxa and individuals.
- Data storage should focus on the subset of output files that are maximally useful to the CCGP community, minimizing costs associated with hosting and accessing large datasets.

We provide details on many aspects of data analysis and hosting in the Supplementary Information.

Bioinformatic Challenges for WGS Data

Handling this many species and datasets creates two key workflow challenges. First, the scale of data generation and analysis vastly exceeds most typical bioinformatic workflows and would

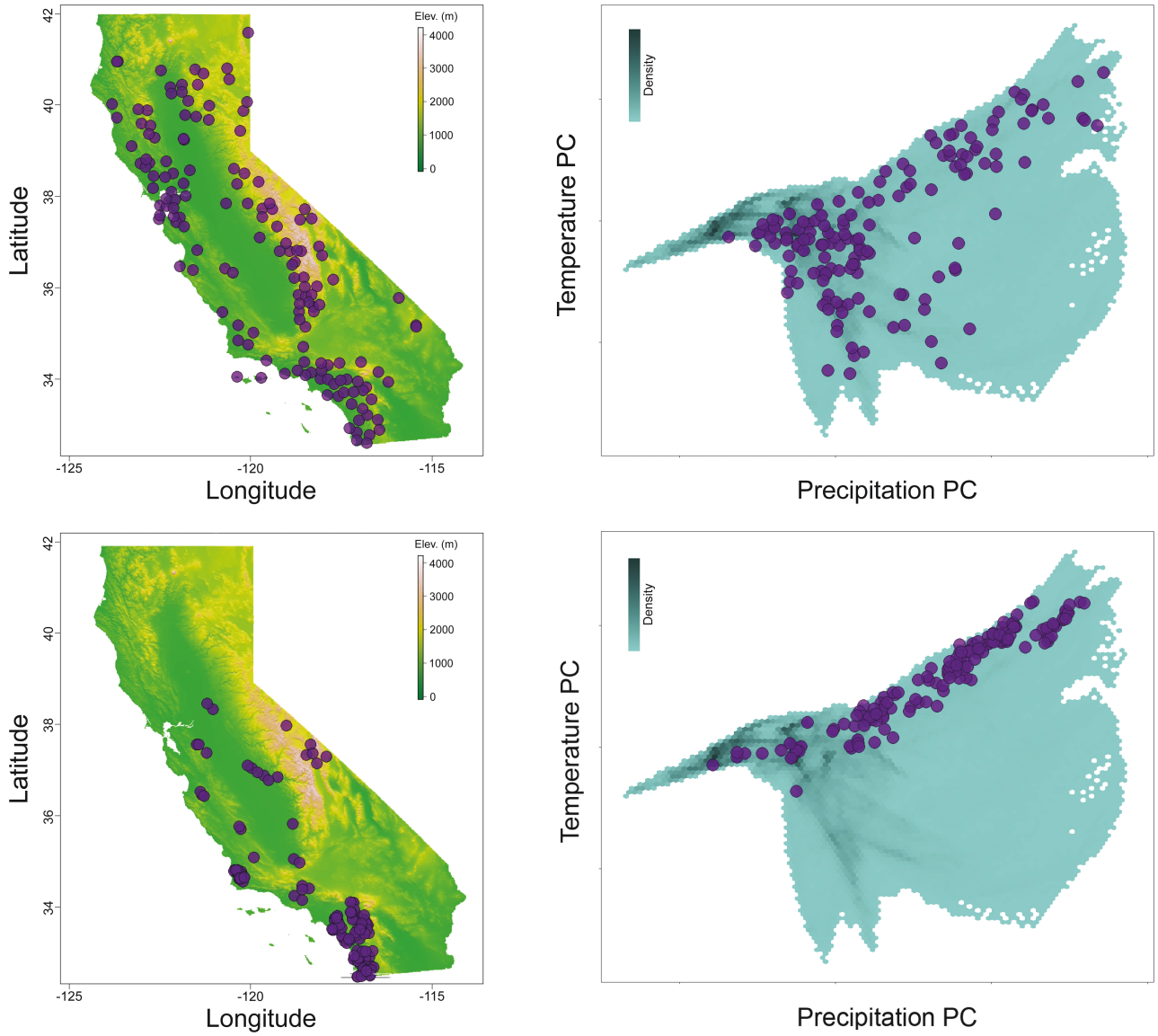

Figure 4. Sample locations for a habitat generalist (top two panels, the Western fence lizard, *Sceloporus occidentalis*) and a vernal pool specialist (bottom two panels, the Western spadefoot *Spea hammondii*) plotted as purple dots in the environmental space of California (right) and on a topographic map of the state (left). We characterized environmental space by conducting a spatial PCA on 19 bioclimatic variables from the WorldClim Database (www.worldclim.org). The first two PCs, plotted here, were loaded primarily by variables representing temperature seasonality (bio4) on PC1, and precipitation (bio12, bio16, and bio19) on PC2. Depicted in teal colors are the values for the temperature and precipitation PCs from across the state (the environmental background), with darker colors representing more frequent occurrences of those values.

overwhelm many local lab servers and many high-performance computer clusters. We resolved this with cost effective cloud computing services (several are available, we utilized the Google Cloud Platform). Second, although we will use some typical sample quality control metrics (e.g., the rate that raw reads are successfully mapped to the reference genome), the geographically widespread sampling design and unique features of species projects imposes important challenges for some standardized population genetic QC metrics. For example, although non-conformity to Hardy–Weinberg equilibrium is often a convenient way to identify problematic variant positions in the genome, many species projects include anticipated population structure and we expect deviations from these simple expectations. Similarly, individual populations may be inbred and therefore display unusual patterns of variation. Given that many of our study species consist of a combination of endangered and relatively healthy populations, we cannot simply eliminate outliers as might typically be done—such outliers may be the exact populations that we seek to identify and understand as part of our conservation mission. The ecological and natural history expertise of the individual CCGP labs will be essential to determining whether apparently unusual variation and population structure is consistent with what is known about the biology of a species, and therefore whether these data are reliable.

Current Progress

Sampling and Genome Progress

Given the constantly changing status of 153 separate projects, we provide weekly updates on the CCGP website (see

Supplementary Information for additional details). Reference genome material has been submitted for the majority of species projects (139 as of this writing). HiFi and Omni-C extractions are underway for about two thirds, and contig-level assemblies have been generated for about a third of species projects. Our first two completed reference genomes, the big berry manzanita (*Arctostaphylos glauca*; Huang et al. 2021) and the northwestern pond turtle (*Actinemys marmorata*; Todd et al. 2022), had excellent assembly statistics (manzanita: scaffold N50 of 45Mb, BUSCO complete score of 98.2%; northwestern pond turtle: scaffold N50 of 146 Mb, BUSCO complete score of 96.7%), and we anticipate similar results for most species. Our aspirational goals call for reference genome sequencing to be completed by July 2022, and resequencing by October 2022. Landscape sampling for whole-genome resequencing across species projects is currently underway, and virtually all projects have half or more of their sampling completed (Figure 5).

Looking Forward

The fields of genetics and conservation biology have come a long way in the last half century (Allendorf 2016). From the early allozyme-driven heyday of the 1970s to our current emphasis on genomic data sets, our ability to improve the capacity of resource managers to protect, steward, and relocate individuals and/or populations in the face of climate change (Supple and Shapiro 2018) has been remarkable. Nowhere has this progress been more pronounced than California (Beninde et al. 2022). In the last few years, the integration of remote sensing products with genomic data

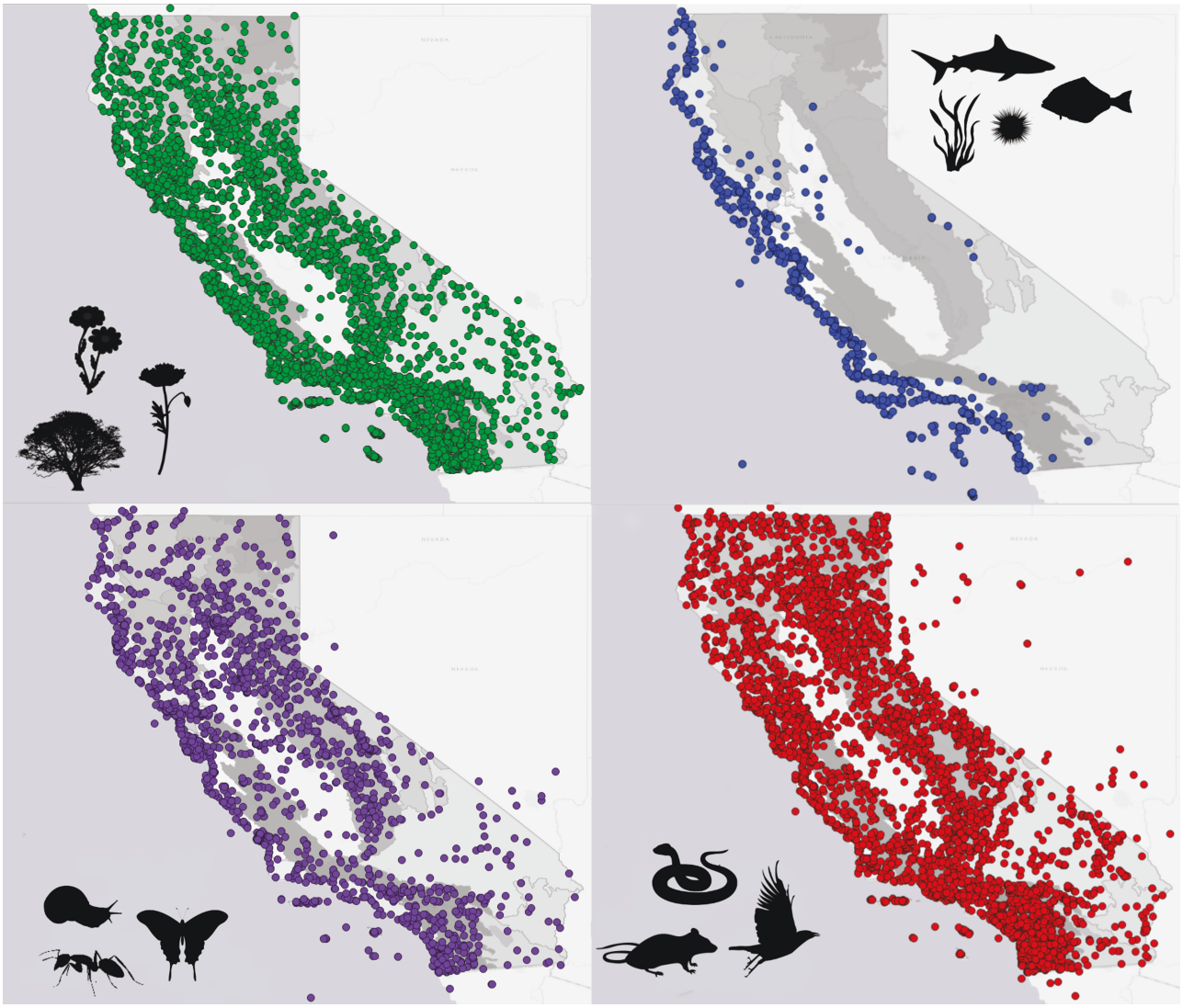

Figure 5. Anticipated and completed sampling for the California Conservation Genomics Project across four major sets of species. Black silhouettes indicate our representation (for visual clarity) of four functional groups: Upper left, terrestrial/freshwater plants in green; upper right, marine taxa, including plants, invertebrates and vertebrates, in blue; Lower left, terrestrial invertebrates, in purple; Lower right, terrestrial/freshwater vertebrates, in red. A total of 15 121 unique sampling locations are mapped, representing the completed and anticipated sampling strategy of the CCGP. Not mapped are 139 (<1% of the total) additional unique sampling locations that fell outside of the California and adjacent regions, several non-native pest species derived from globally distributed locations being sequenced in collaboration with the California Department of Food and Agriculture, and sampling locations of the black abalone, an extremely sensitive federally listed species. See Supplementary Table 1 for a complete list of all species and their categorization into these four sets of mapped species.

has led to qualitative shifts in understanding biodiversity evolution and predicting the consequences for biological conservation as our climate warms and dries (Yamasaki et al. 2017; Thackeray and Hampton 2020). The roots of this scientific partnership—genetics and conservation, with a healthy dose of natural history—have been growing for nearly a century, but the fruits of this liaison in the last few years include novel analytical methods and meaningful, extraordinary data sets that provide real hope for effective biodiversity stewardship.

What is next for the CCGP and its collaboration of scientists? For California, ambitious goals include genomic coverage for every imperiled and ecologically foundational species within the state, with the ultimate goal of complete taxonomic coverage for the California biota. A waystation may be genomic coverage for the California Floristic Province (a more natural biogeographic unit than its political boundaries), or more modestly, those regions with the state predicted by the California Fourth Climate Change Assessment (https://www.climateassessment.ca.gov/state/) to suffer the greatest likelihood of extinctions due to our rapidly warming and drying climate. An achievable shorter-term goal is to expand the CCGP partnership to include the state's 30 × 30 initiative (Exec. Order No. N-82-20, 2020). Providing landscape genomic data to help delimit the final configuration of the 30% of California's land and marine areas to be protected should help maximize the probability of "getting it right", including the provision of powerful and unequivocal data to bolster protected area selection, with decisions based on sound science. Equally important to the technical achievements made over the last 50 years are the cultural and societal advances in conservation that we hope to contribute to. Our goal, for current and future work in the state, is to share our results with the Indigenous tribes and communities of California, integrating their knowledge, values and landscapes into the CCGP community.

Expanding the CCGP model of conservation genomics to other geographies and collectives may be viewed as lofty, but we believe this will happen. For example, the International Cooperative for the Management of Mediterranean-Climate Ecosystems (https://incomme.org) collective of scientists and practitioners in the five Mediterranean-type ecosystems (MTE) of the world, in launching a CCGP-type of project, would be able to address not just landscape-scale questions about the genome evolution of MTE biodiversity, but also multiple spatio-temporal evolutionary phenomena that extend across a singularly defined ecosystem type around the globe that is renowned for its exceptionally rich, often threatened biodiversity. Strategically, such an addition would build upon the CCGP, seamlessly adding insights not only to the state's biodiversity, but to the ecological and evolutionary effects of global change. Data from the landscape genomic assessments in these five biodiversity hotspots would represent a unique dataset that could provide an efficiency in conservation planning not achievable in any single region, and for a suite of hotspots particularly suffering under our changing climate. Similarly, professional societies focused on tropical ecosystems (e.g., Society for Tropical Ecology), or on particular taxa (e.g., Society for Study of Amphibians and Reptiles) could pool data for a deeper understanding of their organism(s) of interest—adding knowledge not only for the sake of understanding organismal evolution, but for informing environmental policy at a global scale. Funding such endeavors, particularly for regions and countries with limited financial resources, may require international collaborative efforts, but the same is true for much of global conservation. The only thing holding us back is our limited imagination: we need to identify how, and with whom, to partner in the service of biodiversity conservation.

The CCGP also will provide a platform to explore some of the more vexing questions in conservation biology, such as whether species with limited ranges should be managed or protected in the same way as widely distributed ones, whether species formerly widespread but now limited in geographic scope should be stewarded differently than species that have always had small ranges (Robinson et al. 2016), and whether, and when, genome-level data are worth collecting for species that have been previously analyzed with more restricted data sets (Beninde et al. 2022; Gallego-García et al. 2021). With the rate of extinction in the last century estimated to be 22 times faster than the historical baseline rate (Ceballos et al. 2015) rapidly gathering large-scale landscape genomic datasets, complemented by high-quality reference genomes, is one positive step that can and should be made, to understand biodiversity as it now exists, and doing what we can to ensure resilience in the face of a rapidly changing world.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding

Funding for the CCGP provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224]. The CCGP Design Planning Meeting was co-funded by UCLA's Vice Chancellor of Research with logistical support from the UCLA Sustainable LA Grand Challenge and the UCLA La Kretz Center for California Conservation Science.

Acknowledgments

We particularly thank Duane Muller and the Sustainable Los Angeles Grand Challenge, Harris Lewin and the Earth Biogenome Project, and Dovetail Genomics and Illumina for logistical and additional financial support. We also thank the support staff and leadership at our partner genome centers and facilities, particularly the California Institute for Quantitative Biosciences (QB3) at UC Berkeley, the DNA Technologies and Expression Analysis Core Laboratory at UC Davis, and the Paleogenetics Laboratory at UC Santa Cruz for guidance, advice, and support. Finally, a special thanks to Ed Green for his help and advice in moving our reference genome production process to the finish line. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

Data Availability

Data generated for the CCGP are available under NCBI BioProject PRJNA720569.

References

Albert CH, Yoccoz NG, Edwards TC, Graham CH, Zimmermann NE, Thuiller W. 2010. Sampling in ecology and evolution—bridging the gap between theory and practice. *Ecography*. 33:1028–1037.

Allendorf FW. 2016. Genetics and the conservation of natural populations: allozymes to genomes. *Mol Ecol*. 26:420–430.

Balkenhol N, Cushman S, Storfer A, Waits L. 2015. *Landscape genetics: concepts, methods, applications*. Hoboken (NJ): John Wiley & Sons.

Banks MA, Jacobson DP, Meusnier I, Greig CA, Rashbrook VK, Ardren WR, Smith CT, Bernier-Latmani J, Van Sickle J, O'Malley KG. 2014. Testing advances in molecular discrimination among Chinook salmon life histories: evidence from a blind test. *Anim Genet*. 45:412–420.

Barnosky AD, Matzke N, Tomiya S, Wogan GO, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, et al. 2011. Has the Earth's sixth mass extinction already arrived? *Nature*. 471:51–57.

Bay RA, Harrigan RJ, Underwood VL, Gibbs HL, Smith TB, Ruegg K. 2018. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*. 359:83–86.

Beninde J, Toffelmier E, Andreas A, Nishioka C, Slay M, Soto A, Bueno J, Gonzalez G, Pham H, Pace J, et al. Forthcoming 2022. CaliPopGen: a genetic and life history database for the fauna and flora of California. *Sci Data*.

Browne L, Wright JW, Fitz-Gibbon S, Gugger PF, Sork VL. 2019. Adaptational lag to temperature in valley oak (Quercus lobata) can be mitigated by genome-informed assisted gene flow. *Proc Natl Acad Sci USA*. 116:25179–25185.

California Executive Order. No. N-82-20. October 7, 2020. *State of California*.

Carroll SP, Jørgensen PS, Kinnison MT, Bergstrom CT, Denison RF, Gluckman P, Smith TB, Strauss SY, Tabashnik BE. 2014. Applying evolutionary biology to address global challenges. *Science*. 346:1245993.

Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. 2015. Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Sci Adv*. 1:e1400253.

Cleland DT, Freeouf JA, Keys JE, Nowacki GJ, Carpenter CA, McNab WH. 2007. *Ecological subregions: sections and subsections for the conterminous United States*. General Technical Report WO-76D, 76D.

DeWoody JA, Harder AM, Mathur S, Willoughby JR. 2021. The long-standing significance of genetic diversity in conservation. *Mol Ecol*. 30:4147–4154.

Falconer DS, Mackay TFC. 1996. *Introduction to quantitative genetics*. Harlow (UK): Prentice Hall.

Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini A, Godoy JA, Höglund J, et al.; European Reference Genome Atlas (ERGA) Consortium. 2022. The era of reference genomes in conservation genomics. *Trends Ecol Evol*. 37:197–202.

Frankham R. 2005. Genetics and extinction. *Biol Conserv*. 126:131–140.

Frankham R, Ballou JD, Ralls K, Eldridge M, Dudash MR, Fenster CB, Lacy RC, Sunnucks P. 2017. *Genetic management of fragmented animal and plant populations*. Oxford (UK): Oxford University Press.

Funk WC, Forester BR, Converse SJ, Darst C, Morey S. 2019. Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. *Conserv Genet*. 20:115–134.

Gallego-García N, Caballero S, Shaffer HB. 2021. Are genomic updates of well-studied species worth the investment for conservation? A case study of the critically endangered magdalena river turtle. *J Hered*. 112:575–589.

Gershunov A, Shulgina T, Clemesha RES, Guirguis K, Pierce DW, Dettinger MD, Lavers DA, Cayan DR, Polade SD, Kalansky J, et al. 2019. Precipitation regime change in Western North America: the role of atmospheric rivers. *Sci Rep*. 9:9944.

Goudey CB., Smith DW, editors. 1994. *Ecological units of California: subsections (map)*. San Francisco, CA: U.S. Department of Agriculture, Forest Service. Pacific Southwest Region.

Hall LA, Beissinger SR. 2014. A practical toolbox for design and analysis of landscape genetics studies. *Landsc Ecol*. 29:1487–1504.

Hobern D. 2021. BIOSCAN: DNA barcoding to accelerate taxonomy and biogeography for conservation and sustainability. *Genome*. 64:161–164.

Huang Y, Escalona M, Morrison G, Marimuthu MPA, Nguyen O, Toffelmier E, Shaffer HB, Litt A. 2021. Reference genome assembly of the big berry Manzanita (*Arctostaphylos glauca*). *J Hered*. 113:xx–xx.

Kartesz JT. 2015. *The biota of North America Program (BONAP)*. Chapel Hill (NC): Taxonomic Data Center.

Lande R, Arnold SJ. 1983. The measurement of selection on correlated characters. *Evolution*. 37:1210–1226.

Lewin HA, Richards S, Aiden EL, Allende ML, Archibald JM, Bálint M, Barker KB, Baumgartner B, Belov K, Bertorelle G, et al. 2022. The earth biogenome project 2020: starting the clock. *Proc Natl Acad Sci USA*. 119:e2115635118.

Manel S, Albert CH, Yoccoz NG. 2012. Sampling in landscape genomics. *Methods Mol Biol*. 888:3–12.

Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol*. 18:189–197.

Mathur S, DeWoody JA. 2021. Genetic load has potential in large populations but is realized in small inbred populations. *Evol Appl*. 14:1540–1557.

McCartney-Melstad E, Gidiş M, Shaffer HB. 2018. Population genomic data reveal extreme geographic subdivision and novel conservation actions for the declining foothill yellow-legged frog. *Heredity (Edinb)*. 121:112–125.

Mittermeier RA, Gil PR, Hoffman M, Pilgrim J, Brooks T, Mittermeier CG, Lamoreux J, da Fonseca GAB. 2004. *Hotspots revisited: earth's biologically richest and most endangered terrestrial ecoregions*. Washington (DC): Conservation International.

Nath S, Shaw DE, White MA. 2021. Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 (Bethesda, MD)* 11:jkab007.

Prunier JG, Kaufmann B, Fenet S, Picard D, Pompanon F, Joly P, Lena JP. 2013. Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: towards a better assessment of functional connectivity using an individual-based sampling scheme. *Mol Ecol*. 22:5516–5530.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 592:737–746.

Robinson JA, Ortega-Del Vecchyo D, Fan Z, Kim BY, vonHoldt BM, Marsden CD, Lohmueller KE, Wayne RK. 2016. Genomic flatlining in the endangered island fox. *Curr Biol*. 26:1183–1189.

Schoennagel T, Balch JK, Brenkert-Smith H, Dennison PE, Harvey BJ, Krawchuk MA, Mietkiewicz N, Morgan P, Moritz MA, Rasker R, et al. 2017. Adapt to more wildfire in western North American forests as climate changes. *Proc Natl Acad Sci USA*. 114:4582–4590.

Schwartz MW. 2008. The performance of the endangered species act. *Ann Rev Ecol Evol Syst*. 39:279–299.

Schwartz MK, McKelvey KS. 2008. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conserv Genet*. 10:441–452.

Scott PA, Allison LJ, Field KJ, Averill-Murray RC, Shaffer HB. 2020. Individual heterozygosity predicts translocation success in threatened desert tortoises. *Science*. 370:1086–1089.

Seaborn T, Hauser SS, Konrade L, Waits LP, Goldberg CS. 2019. Landscape genetic inferences vary with sampling scenario for a pond-breeding amphibian. *Ecol Evol*. 9:5063–5078.

Selmoni O, Vajana E, Guillaume A, Rochat E, Joost S. 2020. Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach. *Mol Ecol Resour*. 20:154–169.

Shukla PR, Skeg J, Buendia EC, Masson-Delmotte V, Pörtner H-O, Roberts DC, Zhai P, Slade R, Connors S, Diemen S, et al. 2019. Climate change and land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. *Intergovernmental Panel on Climate Change*.

Smith TB, Fuller TL, Zhen Y, Zaunbrecher V, Thomassen HA, Njabo K, Anthony NM, Gonder MK, Buermann W, Larison B, et al. 2020. Genomic vulnerability and socio-economic threats under climate change in an African rainforest bird. *Evol Appl*. 14:1239–1247.

Smith TB, Wayne RK, Girman DJ, Bruford MW. 1997. A role for ecotones in generating rainforest biodiversity. *Science*. 276:1855–1857.

Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, Angerer LM, Arnone MI, Burgess DR, Burke RD, et al. 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science*. 314:941–952.

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

Sork VL, Cokus SJ, Fitz-Gibbon ST, Zimin AV, Puiu D, Garcia JA, Gugger PF, Henriquez CL, Zhen Y, Lohmueller KE, et al. 2022. High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. *Nat Commun*. 13:2047.

Spalding MD, Fox HE, Allen GR, Davidson N, Ferdaña ZA, Finlayson M, Halpern BS, Jorge MA, Lombana A, Lourie SA, et al. 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience*. 57:573–583.

Supple MA, Shapiro B. 2018. Conservation of biodiversity in the genomics era. *Genome Biol*. 19:131.

Teixeira JC, Huber CD. 2021. The inflated significance of neutral genetic diversity in conservation genetics. *Proc Natl Acad Sci USA*. 118:e2015096118.

Thackeray SJ, Hampton SE. 2020. The case for research integration, from genomics to remote sensing, to understand biodiversity change and functional dynamics in the world's lakes. *Glob Chang Biol*. 26:3230–3240.

The Darwin Tree of Life Project Consortium. 2022. Sequence locally, think globally: The Darwin Tree of Life Project. *Proc Natl Acad Sci. USA* 119:e2115642118.

Thomassen HA, Fuller T, Buermann W, Milá B, Kieswetter CM, Jarrín-V P, Cameron SE, Mason E, Schweizer R, Schlunegger J, et al. 2011. Mapping evolutionary process: a multi-taxa approach to conservation prioritization. *Evol Appl*. 4:397–413.

Todd BD, Jenkinson TD, Escalona M, Beraut E, Nguyen O, Sahasrabudhe R, Scott PA, Toffelmier E, Wang IJ, Shaffer HB. 2022. Reference genome of the northwestern pond turtle, *Actinemys marmorata*. *J Hered*. 113:xx–xx.

U.S. Fish and Wildlife Service. 2021. ECOS environmental conservation online system; listed species believed to or known to occur in each State. https://ecos.fws.gov/ecp/report/species-listings-by-state-totals?statusCategory=Listed

Vu JP, Vasquez MF, Feng Z, Lombardo K, Haagensen S, Bozinovic G. 2021. Relative genetic diversity of the rare and endangered Agave shawii ssp. shawii and associated soil microbes within a southern California ecological preserve. *Ecol Evol*. 11:1829–1842.

Wang IJ, Bradburd GS. 2014. Isolation by environment. *Mol Ecol*. 23:5649–5662.

Ward CM, Aumann RA, Whitehead MA, Nikolouli K, Leveque G, Gouvi G, Fung E, Reiling SJ, Djambazian H, Hughes MA, et al. 2021. White pupae phenotype of tephritids is caused by parallel mutations of a MFS transporter. *Nat Commun*. 12:491.

Wolfe AD, Necamp T, Fassnacht S, Blischak P, Kubatko L. 2016. Population genetics of *Penstemon albomarginatus* (Plantaginaceae), a rare Mojave Desert species of conservation concern. *Conserv Genet*. 17:1245–1255.

Yamasaki E, Altermatt F, Cavender-Bares J, Schuman MC, Zuppinger-Dingley D, Garonna I, Schneider FD, Guillén-Escribà C, van Moorsel SJ, Hahl T, et al. 2017. Genomics meets remote sensing in global change studies: monitoring and predicting phenology, evolution and biodiversity. *Curr Opin Environ Sustain*. 29:177–186.