American Museum of Natural History

Center for Biodiversity and Conservation
Network of Conservation Educators and Practitioners

Designing a Conservation Genomics Project Incorporating DNA from Museum Specimens

Author(s): Anna Penna, Megan Wallace, Lauren T. Clark, Lochlan Sife Krupa, Suzanne K. Macey, Luca Pozzi, and Mary E. Blair

Source: Lessons in Conservation, Vol. 14, Issue 1, pp. 160-176

Published by: Network of Conservation Educators and Practitioners, Center for Biodiversity and Conservation, American Museum of Natural History

DOI: https://doi.org/10.5531/cbc.linc.14.1.9

Supplementary materials:
https://doi.org/10.5531/cbc.ncep.0190

Designing a Conservation Genomics Project Incorporating DNA from Museum Specimens

Anna Penna[i,ii,iii], Megan Wallace[iv], Lauren T. Clark[iv,v], Lochlan Sife Krupa[vi], Suzanne K. Macey[vii], Luca Pozz[iii], and Mary E. Blair[vii]

[i]National Museum of Natural History, Smithsonian Institution, Washington, DC, USA; [ii]Department of Anthropology, University of Texas at San Antonio, San Antonio, TX, USA; [iii]Department of Biology, Lund University, Lund, Sweden; Department of Herpetology, American Museum of Natural History, New York, NY, USA; [iv]Institute of Comparative Genomics, American Museum of Natural History, New York, NY, USA; [v]Centre for Paleogenetics, Stockholm University, Stockholm, Sweden; [vi]Ecology, Evolution and Environmental Biology Department, Columbia University, New York, NY, USA; [vii]Center for Biodiversity and Conservation, American Museum of Natural History, New York, NY, USA

DOI: https://doi. org/10.5531/cbc.linc.14.1.9 | Supplementary: http://doi.org/10.5531/cbc.ncep.0190

ABSTRACT

In this exercise, students reflect critically on the research process and understand the different steps required to implement a research project. Specifically, students will develop a conservation museomics project using a set of hypothetical case studies that include the sampling of historical and modern specimens. Students will 1) use fundamental principles of conservation genetics to formulate research questions and hypotheses, 2) map out the components of their hypothetical experiment, and 3) create contingency and budget optimization plans to overcome common research challenges and a visual aid of their choosing to communicate conservation management strategies for hypothetical taxa.

LEARNING OBJECTIVES

In this exercise, students will apply conservation genetics concepts to develop and explore a hypothetical research project.

Specifically, students will:

1. Apply conservation genomics principles: students will apply principles of conservation genetics and museum genomics to design a comprehensive conservation project that addresses biodiversity conservation issues.
2. Formulate and interpret scientific hypotheses: students will formulate hypotheses based on their research questions and expected outcomes, contrasting their expectations and interpreting results.
3. Develop experimental design skills: students will identify key parameters for measurement and outline a study design encompassing the genetic sampling, processing, sequencing, and analysis of both modern and historical specimens.
4. Reflect on common research challenges: students will contemplate potential issues that can arise throughout the research process, including budget constraints. Based on these obstacles, students will create contingency plans and budget optimization plans.
5. Communicate conservation management strategies: students will brainstorm management strategies, practice communicating their findings, and express solutions and the importance of conservation to the public, a critical skill to develop.

PREREQUISITES

Please read this module's synthesis, "Applications of museum collections and genomics to biodiversity conservation" (https://doi. org/10.5531/cbc.linc.14.1.7) prior to completing this exercise.

This exercise assumes that students are familiar with the fundamental principles of conservation genetics and population genetics analyses. For example, students should be comfortable with major principles used in conservation genetics, such as genetic diversity, genetic drift, gene flow, evolutionary systematics, geographic isolation, and inbreeding.

If you would like to review background material on these concepts, we suggest you read the NCEP module "Conservation genetics" (https://doi.org/10.5531/cbc.ncep.0123).

This module also has a companion exercise: "The application of conservation museomics approaches to the protection of the Iberian lynx (*Lynx pardinus*)" (https://doi. org/10.5531/cbc.linc.14.1.8), which may be helpful for setting the foundations of designing a new conservation genetics project.

The authors of this exercise and their colleages have also developed a hands-on bioinformatics exercise of a population genetic study that combines historical and modern samples that may be a helpful extension to this exercise (https://github.com/sjgaughran/amnh-museumomics).

OVERVIEW

Natural history collections are like time capsules. In the US alone, they house almost a billion specimens preserved during different periods, including prehistoric times! With the recent advancements in DNA sequencing technologies and molecular biology protocols for the isolation of genetic material, it is now possible to obtain genetic data from these preserved specimens—a field of research often referred to as "museomics." Thanks to the collective dedication of various naturalists, field biologists, curators, and collection staff, natural history collections provide records that can be used as a temporal comparison to investigate population trends, including specimens that serve as testimony of populations or species that are now extinct in the wild.

In this exercise, you will be presented with a few hypothetical scenarios in which principles of conservation genetics can be combined with museomics to better understand population trends, connectivity, and evolutionary relationships. Upon completion of the exercise, you will have developed a conservation genetics study design and risk analysis of a project incorporating historical samples.

First: carefully read the hypothetical cases and the respective researchers' interests found in Appendix 1. Pick one example to focus on in this exercise.

PART 1: RESEARCH FOUNDATION

Before starting an experiment, it is important to have a clear research question in mind. This question will serve as a guide through testing hypotheses and determining what data to collect. Thus, the research question must be aligned with the researcher's interests (the project goal) and make clear connections with measurable or quantifiable variables. These connections are formalized as hypotheses that will then be tested using statistical methods.

If you have not already, read the hypothetical case studies in Appendix 1 and select one to be the focus of the following prompts. Alternatively, your instructor may encourage you to come up with your own research question/case study.

1. First things first. What is this study aiming to answer? Formulate a research question that aligns with the researcher's interest. A research question must have a clear goal that connects the

research interest and the elements of the study (e.g., species, populations, regions, time). *Write your research question in the worksheet (Appendix 2).*

2. Now, identify the independent and the dependent variables in this study. An independent variable is a parameter whose change affects the dependent variables but is not affected by those variables itself (e.g., time, total area, temperature, flow rate). The dependent variables are all other components influenced by the independent variable. In conservation genetics, the dependent variables will often be quantitative (i.e., a numerical data). Here, we suggest recalling the fundamental principles of conservation genetics and the main parameters used in population genetics research [genetic diversity, gene flow, runs of homozygosity, inbreeding coefficient, amount of deleterious genetic variation (i.e., genetic load), effective population size, allele frequency]. We provide an overview of these concepts in this module's synthesis document (https://doi. org/10.5531/cbc.linc.14.1.7).
   Identify one independent variable and at least two dependent variables that the researcher will need to measure to answer the research question you formulated in #1 above and write them on the worksheet.

3. Formulate your hypotheses connecting your research question and the experimental variables (what you will be quantifying). One strategy is to use "If… then…" statements. For example: "If x (independent variable) increases, I expect an increase in y (a dependent variable)." By formulating these statements, you can start making predictions about the relationship between variables that can then be tested through some formal statistical framework. Hypotheses serve as the starting point for any experiment. Therefore, any flaw in their formulation may cause a flaw in the design of an entire experiment. So, let's define the null (H_0) and alternative hypotheses (H_1). In the scientific method, we can only falsify hypotheses. Usually, H_0 assumes there is no change or relationship between the observations, which researchers try to disprove. H_1 is accepted if there is sufficient evidence to reject the null.
   Write your H_0 *and* H_1 *in the worksheet.*

4. Now let's make a graphical representation of your expectations, connecting the independent and dependent variables you identified in #2 above. Contrast the expectations for each hypothesis using a pair of illustrative graphs, one depicting the projected outcomes for H_0, and the other the projected outcomes for H_1. These could be a line graph, scatter plot, boxplot, etc. Don't forget to label your axes!
   Draw graphs contrasting the expectations for H_0 *and* H_1 *in the empty boxes provided on the worksheet. You may wish to draw multiple graphs depending on the number of dependent variables you identified in #2 above.*

5. Write a concise paragraph explaining how these contrasting results can be interpreted. In your explanation, be sure to mention the research question and conclusions the researcher might make depending on the observed relationship between the variables identified in #2 above.
   On the worksheet, write the conclusions that can be made based on the expectations depicted in #4 above. Refer back to the hypotheses formed in #3 above.

6. Before moving on to Part 2, pause and think critically: Have you considered the assumptions you are making about relationships or data? What covariates might be interesting to consider? Does correlation mean causation? How can you control for variation? Write a few sentences about your thought process on these questions (or others that you have)—while you might not have the answer, the critical thinking may inform how you design your experiment.

PART 2: EXPERIMENTAL DESIGN

Now that we have a clear idea of what our research goal is, what we must quantify, and what we will be testing, it is time to consider how we will accept or reject the null hypothesis. To this end, you will outline an experimental design that includes the genetic sampling of modern and historical samples to answer the research question and test the hypotheses you formulated in Part 1.

7. What would the ideal sample size be to test this hypothesis and answer the researcher's question and why? What are some challenges you might encounter when trying to acquire and sequence this number of samples?
   Complete the table provided in the worksheet (Appendix 2) and answer the questions above. In your answer, consider sample acquisition, sample quality, and budget.

8. What would be the ideal source of tissue and preparation type (e.g., bone, organ tissue, skins, dried whole specimen) most commonly available for your study organism? What lab work protocols are more appropriate in this case? What are some challenges you might encounter when working with said type of preparation?
   Complete the table provided in the worksheet (Appendix 2) and answer the questions above. In your answer, consider sample acquisition, sample quality, and preservation method (e.g., ethanol, formalin, liquid nitrogen.)

9. Organize the study design in a workflow. When preparing your workflow, think about the different phases of a conservation museomics project (e.g., sample collection, lab work, data analysis) and how to progress from one step to the next. More importantly, think about the relevant intermediate steps to take within each phase. Here are some questions to help you prepare the workflow:
   - A research experiment starts with gathering the data. Will you need to sample different localities or temporal windows?
   - What type of genetic data will you need to generate for this investigation? Do they require different laboratory steps?
   - How will these steps allow you to measure the dependent and independent variables needed to test the hypothesis you proposed?

   Below are some example workflows (Figures 1–4) to inspire the development of your own.
   Complete the flowchart containing the main steps required for the completion of the proposed research. You can add as much detail as you want, but be sure to cover the three major research phases (sample collection, lab work, and data analysis).

Specific details on data analysis go beyond the content of this lesson. The authors of this exercise have developed a bioinformatics activity that may be a helpful resource (https://github.com/sjgaughran/amnh-museumomics). Below are other resources that can be useful to guide the bioinformatics part:

- Schubert, M., Ermini, L., Sarkissian, C. D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Martin, K., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nature Protocols, 9(5), 1056–1082. https://doi.org/10.1038/nprot.2014.063
- Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome biology, 17, 60. https://doi.org/10.1186/s13059-016-0918-z
- Yates, J. A. F., Lamnidis, T. C., Borry, M., Valtueña, A. A., Fagernäs, Z., Clayton, S., Garcia, M. U., Neukamm, J., & Peltzer, A. (2021). Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. PeerJ, 9, e10947. https://doi.org/10.7717/peerj.10947

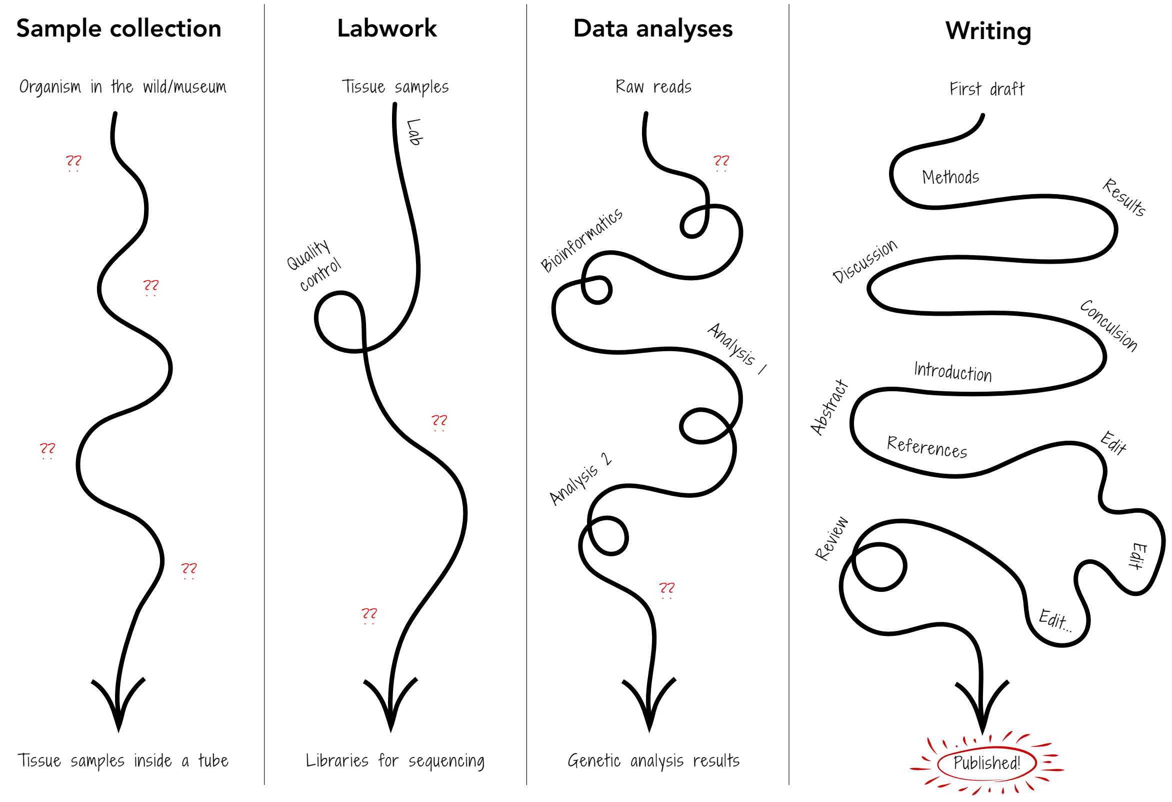

Figure 1: Each phase of the project will have many intermediate steps. Using a workflow can help you plan the study design and make sure you are prepared for each phase of the study. Image credit: Anna Penna.

Figure 2: Example workflow depicting the intermediate steps of specimen collection, sampling, and storage. Image credit: Anna Penna.

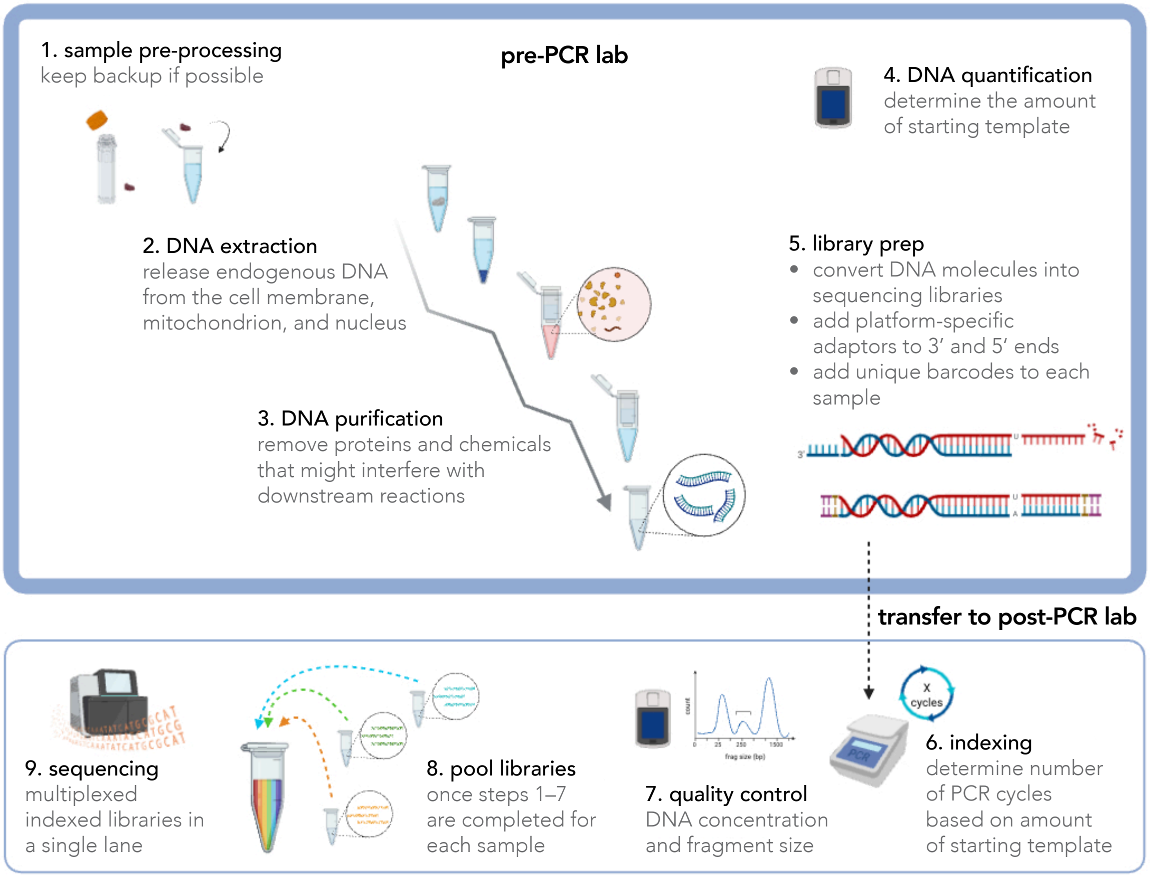

Figure 3: Example workflow depicting the intermediate steps of laboratory work and sequencing. Image credit: Modified from Penna et al., 2024.

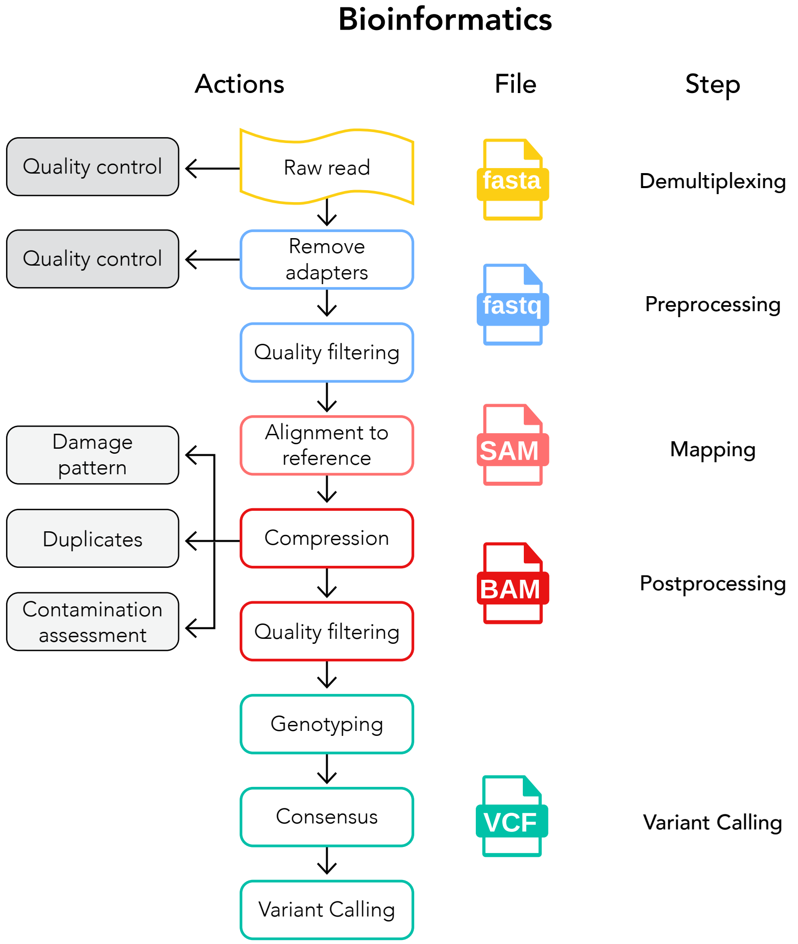

Figure 4: Example workflow depicting general steps taken in the bioinformatics pipeline. Image credit: Anna Penna.

PART 3: CRITICAL THINKING AND COMMUNICATION

Now that you have thought about all the main steps required to implement and complete this research project, it is time to reflect on some of the challenges you may encounter, as well as the potential applications of the data from your study towards conservation efforts. The considerations that emerge from this reflection can be used to create a contingency plan for research hiccups, a project budget optimization plan, and a conservation management plan.

10. What obstacles could arise when working with museum specimens compared to those recently collected in the field? Why do museum specimens pose a challenge? What is the likelihood of these problems occurring? List some strategies to overcome or avoid such obstacles.
Indicate one potential obstacle for each major research phase, briefly explain why this can be challenging, approximate the likelihood of this occurring, and list risk evasion/avoidance strategies in the table provided in the worksheet.

11. Which steps of this study design do you expect to require the most funding? Think about how you would acquire the samples and determine the sequencing strategy and desired sequencing coverage. For instance, while capture sequencing can help you increase the coverage of a target genomic region, the cost of developing and purchasing baits can be very expensive. Consequently, although the capture strategy can increase the coverage of certain regions, it might restrict the number of samples you can sequence. Fixed monetary sources for your research can impact how many samples you can collect, process, and sequence, so how would you optimize your budget to best accomplish each major phase of your study design?
Indicate one potential budget constraint for each major research phase, briefly explain why and how lack of funds would impact this stage of the study design, approximate the level of expense, and list cost-saving strategies in the table provided in the worksheet.

12. If the data from your study suggests significant change between historical and modern populations, think about how your findings could be applied to conservation management. How would you best communicate your results if you wanted to share them with management authorities? What are some tools and strategies that you could suggest implementing to aid modern populations in your hypothetical case study? An example conservation strategy is wildlife crossings, which can connect two disparate populations that have been separated by the construction of a highway.
In a separate document, create a pamphlet or poster aimed at communicating to both management authorities and the public the conservation issue that has affected your hypothetical taxa, the findings from your study, and proposed methods for rehabilitating these hypothetical populations or species. This final exercise should focus on applying science and science communication, so think outside the box and have fun with it!

ACKNOWLEDGMENTS

We would like to thank Mckenna Santiago Coyle, Melina Giakoumis, and Stephen Gaughran. A.P. was funded by the American Society of Mammalogy (ASM), Society for Systematic Biology (SSB), Systematics Association, Linnean Society of London, and National Science Foundation of the United States (NSF; Award No. BSC-2120691); A.P. was also funded by the Smithsonian Institution Graham Funds in Human Origins and the University of Texas at San Antonio (UTSA). A.T.S was funded by the U.S. National Science Foundation (DBI 2029955). L.T.C. was funded, in part, by the U.S. National Science Foundation (DBI 2029955). M.E.B. and S.K.M. were funded by the National Science Foundation of the United States (Awards No. BCS-1926105/1926215). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF.

APPENDIX 1

Designing a Conservation Genomics Project Incorporating DNA from Museum Specimens

Hypothetical Case Studies

Example 1

A recent study suggested that the distribution of a small alpine chipmunk species in the Sierra Nevada Mountains in California has shifted to higher altitudes every year, most likely due to increases in surface temperatures driven by climate change. Unfortunately, the mountains are not getting taller, so the populations that occupy the high-elevation areas are getting more and more confined to the very highest portions of the mountain range. If temperatures continue to rise, these populations might not have enough time to adapt and could go extinct. In the early 1900's, naturalists from a nearby museum collected a series of specimens of this species throughout its entire geographic and altitudinal range. You have a permit to sample 10 specimens per locality in this region for genetic studies.

Researcher's interest: Investigate the impact of this recent shift in distribution on the species' extinction risk.

Example 2

An archipelago on the coast of Aotearoa New Zealand has high levels of bird endemism, including an entire radiation of endemic species that evolved into different body sizes, feeding ecologies, and diverse niches. Unfortunately, ever since the accidental introduction of rats in the late 1800s by Europeans, the natural populations found on the islands have shown signs of drastic demographic changes. The rats feed on the birds' eggs and even prey on some adults in their breeding grounds. Consequently, some populations have gone through a rapid decline in population size over the past few decades. The most extreme case is that two out of the three species of flightless birds reported from the island have gone extinct, as they have not been seen for over a century. Lucky for you, a naturalist collected and preserved over 500 bird specimens from this archipelago before the introduction of rats, including the two now-extinct species of flightless birds and the other 20 species currently recognized in this endemic radiation.

Researcher's interest: Determine if the loss of flight evolved once or multiple times in this endemic bird radiation.

Example 3

The construction of a dam in the early 1930s completely changed the hydrography of a region. Streams that once were connected into a complex network flowing into the river have now turned into a big, flooded area. Another major consequence of such change was the discontinuity of the river flow. This river is home to an endemic salamander, which only reproduces in these waters. The interruption of the river flow is potentially impeding individuals from moving freely up and down the river, which might isolate the headwater populations from those in the river delta. Lucky for you, a famous herpetologist from the early 20th century collected over 300 individuals of this species of salamanders from this river before the dam's construction.

Researcher's interest: Evaluate the impact of the dam on the connectivity between populations from up and down the river.

Example 4

A species of butterfly only lays their eggs on the flowers of a particular plant species in the small (fictional) European country of Genovia. Unfortunately, the queen ruling Genovia in the 1960s was extremely allergic to the pollen produced by this particular flower, so she called for its eradication. A Genovian entomologist, foreseeing the consequences of this change, collected a set of male and female adults, juveniles, and pupae, and then added the preserved specimens to his private collection. Meanwhile, in the fields of Genovia, most of the butterflies die out, but those that survive shift to lay their eggs on the flowers of another plant species. The petals of this new host plant emit a similar fluorescence and conceal the butterfly eggs fairly well. However, there's a larger contrast between the colors of the petals and those of the butterflies, making the adult butterflies more susceptible to predators. Over time, the butterflies' coloration and wing patterns begin to change, and the butterflies become more susceptible to disease.

Researcher's interest: Determine the source(s) of increased disease susceptibility in the present-day butterfly population.

APPENDIX 1

Applications of Conservation Museomics to the Protection of the Iberian Lynx (*Lynx pardinus*)

Please read this module's synthesis, "Applications of museum collections and Genomics to biodiversity conservation" (https://doi. org/10.5531/cbc.linc.14.1.7), prior to completing this exercise.

This module also has a companion exercise: "The application of conservation museomics approaches to the protection of the Iberian lynx (*Lynx pardinus*)" (https://doi. org/10.5531/cbc.linc.14.1.8), which may be a helpful resource.

Date: ___________ Name/Group: ___________________________________

Part 1: Research Foundation Worksheet

Hypothetical Example: ___

1. First things first. What is this study aiming to answer? Formulate a research question that aligns with the researcher's interest. A research question must have a clear goal that connects the research interest and the elements of the study (e.g., species, populations, regions, time). Research Question:

____

____

____

2. Now, identify the independent and the dependent variables in this study. An independent variable is a parameter whose change affects the dependent variables but is not affected by those variables itself (e.g., time, total area, temperature, flow rate, etc.). The dependent variables are all other components influenced by the independent variable. In conservation genetics, the dependent variables will often be quantitative (i.e., a numerical data). Here, we suggest recalling the fundamental principles of conservation genetics and the main parameters used in population genetics research [genetic diversity, gene flow, runs of homozygosity, inbreeding coefficient, amount of deleterious genetic variation (i.e., genetic load), effective population size, allele frequency]. We provide an overview of these concepts in this module's synthesis document (01-ConservationMuseomics_SYN.docx).

Independent variable	Dependent variables
•	• •

3. Formulate your hypotheses connecting your research question and the experimental variables (what you will be quantifying). One strategy is to use "If… then…" statements. For example: "If x (independent variable) increases, I expect an increase in y (a dependent variable)." By formulating these statements, you can start making predictions about the relationship between variables that can then be tested through some formal statistical framework. Hypotheses serve as the starting point for any experiment. Therefore, any flaw in their formulation may cause a flaw in the design of an entire experiment. So, let's define the null (H0) and alternative hypotheses (H1). In the scientific method, we can only falsify hypotheses. Usually, H0 assumes there is no change or relationship between the observations, which researchers try to disprove. H1 is accepted if there is sufficient evidence to reject the null.

H0:

H1:

4. Now let's make a graphical representation of your expectations, connecting the independent and dependent variables you identified in #2 above. Contrast the expectations for each hypothesis using a pair of illustrative graphs, one depicting the projected outcomes for H0, and the other the projected outcomes for H1. These could be a line graph, scatter plot, boxplot, etc. Don't forget to label your axes!

5. Write a concise paragraph explaining how these contrasting results can be interpreted. In your explanation, be sure to mention the research question and conclusions the researcher might make depending on the observed relationship between the variables identified in #2 above.

Expectations and Conclusions:

6. Before moving on to Part 2, pause and think critically: Have you considered the assumptions you are making about relationships or data? What covariates might be interesting to consider? Does correlation mean causation? How can you control for variation? Write a few sentences about your thought process on these questions (or others that you have)—while you might not have the answer, the critical thinking may inform how you design your experiment.

Part 2: Experimental Design

7. What would the ideal sample size be to test this hypothesis and answer the researcher's question and why? What are some challenges you might encounter when trying to acquire and sequence this number of samples?

Specimen age	**Ideal sample size**	**Species or population**	**Locality**

Justification and challenges of determining an acceptable sample size:

8. What would be the ideal source of tissue and preparation type (e.g., bone, organ tissue, skins, dried whole specimen, etc.) most commonly available for your study organism? What lab work protocols are more appropriate in this case? What are some challenges you might encounter when working with said type of preparation?

Specimen age	**Ideal sample size**	**Species or population**	**Locality**

Justification and challenges of selecting specimens for your study:

9. Organize the study design in a workflow. When preparing your workflow, think about the different phases of a conservation museomics project (e.g., sample collection, lab work, data analysis) and how to progress from one step to the next. More importantly, think about the relevant intermediate steps to take within each phase. Here are some questions to help you prepare the workflow:
    - A research experiment starts with gathering the data. Will you need to sample different localities or temporal windows?
    - What type of genetic data will you need to generate for this investigation? Do they require different laboratory steps?
    - How will these steps allow you to measure the dependent and independent variables needed to test the hypothesis you proposed?

Sample collection
(modern and historical)

Nucleic acid isolation

Prepare samples for sequencing
(including quality control)

Sequencing method and platform

Bioinformatic analysis

Part 3: Risk Analysis and Budget Planning

10. What obstacles could arise when working with museum specimens compared to those recently collected in the field? Why do museum specimens pose a challenge? What is the likelihood of these problems occurring? List some strategies to overcome or avoid such obstacles.

<u>Risk Analysis</u>

Research phase	**Obstacle**	**Impact of obstacle**	**Likelihood of occurring** (low, medium, or high)	**Strategies to avoid or overcome obstacle**
Sample Collection				
Lab Work				
Data Analysis				

11. Which steps of this study design do you expect to require the most funding? Think about how you would acquire the samples and determine the sequencing strategy and desired sequencing coverage. For instance, while capture sequencing can help you increase the coverage of a target genomic region, the cost of developing and purchasing baits can be very expensive. Consequently, although the capture strategy can increase the coverage of certain regions, it might restrict the number of samples you can sequence. Fixed monetary sources for your research can impact how many samples you can collect, process, and sequence, so how would you optimize your budget to best accomplish each major phase of your study design?

Budget Planning

Research phase	**Obstacle**	**Impact of obstacle**	**Likelihood of occurring** (low, medium, or high)	**Strategies to avoid or overcome obstacle**
Sample Collection				
Lab Work				
Data Analysis				

12. If the data from your study suggests significant change between historical and modern populations, think about how your findings could be applied to conservation management. How would you best communicate your results if you wanted to share them with management authorities? What are some tools and strategies that you could suggest implementing to aid modern populations in your hypothetical case study? An example conservation strategy is wildlife crossings, which can connect two disparate populations that have been separated by the construction of a highway.