**Hemoglobin Function and Variation in Wild Populations**

By Elizabeth J. Beckman and Kayce C. Bell

University of New Mexico

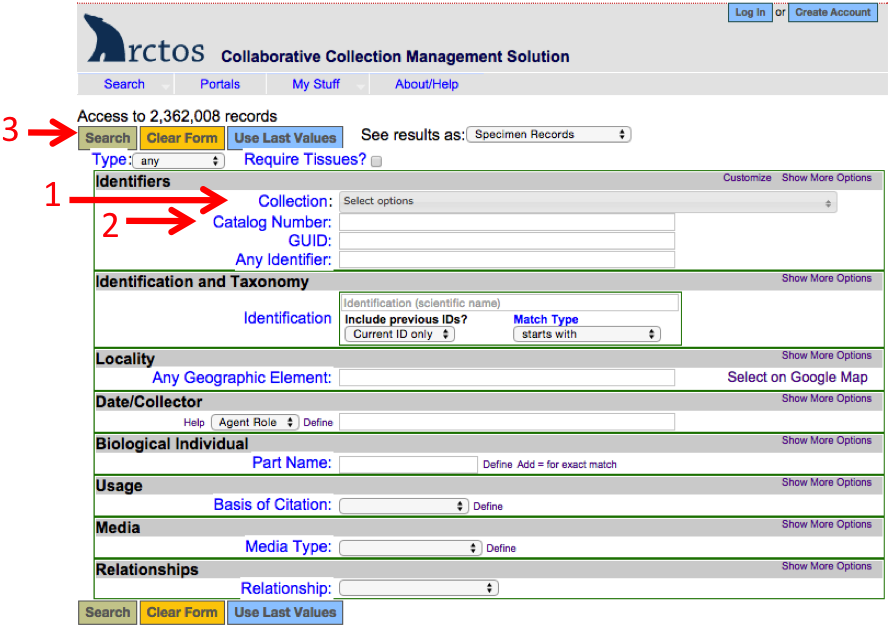
**Part I: Understanding Genetic Variation and Its Impact on Protein Function in Deer Mouse Hemoglobin α chains.**

Storz (2007, *J. Mammalogy*) described two types of hemoglobin α (HBA) subunit functional types: a1c1, found in low elevation deer mice, and a0c0, found in high elevation deer mice. In this section, you will investigate the genetic basis for these functional types and hypothesize how different amino acid substitutions impact hemoglobin function. **Use the Activity Sheet to record your answers.**

**Task I-I: Finding locality and sequence data on HBA in deer mice (*P. maniculatus*) using databases:**

**Arctos** (http://arctos.database.museum) is a vast, public online database of natural history information amassed from research collections. On Arctos, 93 natural history collections make their specimen records available to scientists, students, educators, and amateur naturalists. Specimen records may constitute: locality, date, mass, photos, recordings, skeletal measurements, habitat notes, parasite presence, a list of parts collected and curated (skin, skeleton, tissue) and more. Arctos also has links to other databases that host more information associated with specimens including **GenBank**, the public repository for all published DNA sequences in the USA. The information on Arctos is curated by museum professionals, experts in their fields. Because of its **(1)** public availability, **(2)** association with many collections and **(3)** the reliability of specimen-based data, Arctos is a powerful tool for both scientists and students.

You will use Arctos and GenBank to explore in-depth the genetic variation in the α subunit of hemoglobin in the deer mouse (*Peromyscus maniculatus*). First, you will search for specimen records of deer mice (*P. maniculatus*) used Storz et al. (2007, *PLOS Genetics*) to gather specific locality data. Then, using the link to GenBank in the Arctos record, you can find HBA sequences for each individual in order to evaluate the genetic variation in HBA.



**Fig. 1. Screenshot of Arctos search page.** Red arrows indicate (1) Collection dropdown menu, (2) Catalog Number window and (3) Search button.

With Arctos, you can search for museum specimens by species, locality, date, collector, and other search criteria. In this case, we know the museum and catalog numbers for the individuals Storz et al. (2007) used, and will use this information to search for the records. In the **Identifiers** section, on the **Collection** dropdown menu (Fig. 1), select Museum of Southwestern Biology (MSB): Mammal specimens (MSB:Mamm). Then, in the **Catalog Number** search box, enter 73952,73954,73958,73961,73966,73967 (no spaces between numbers or commas), then click the **Search** button at the top. Click on one of the catalog numbers on the left. This will take you to the detailed specimen record. In the record is information about locality, date, collector as well as the types of material that are available in the museum for the specimen, such as skeleton, skin and tissue samples.

**Activity sheet #1.** Record the state and specific locality for the six specimens.

On the right, in the Identifiers box, there is a list of GenBank numbers. The GenBank accession numbers EF369636-EF369651 are the hemoglobin DNA sequences for the specimens that were used in the Storz et al. 2007 study. If you click on the hyperlink, it will take you to the GenBank entry for that specimen. If you click on the PopSet hyperlink at the top, it will take you to the list and links for all of the hemoglobin specimens used in the study. These DNA sequences have been translated into amino acid sequences and a subset of these are provided to you below.

**Task I-II: Calculating allele frequencies for amino acid sites that vary among deer mice populations:**

To see how genetic variation of HBA changes across geographic space, you will calculate the **allele frequency**of an HBA gene for several polymorphic nucleotides and look at how the allele frequency differs among populations.

To understand what an allele frequency is, imagine you sequence the DNA for 10 deer mice from one population. Because mice are diploid like humans, each individual has 2 copies of a gene so in the end, you have 20 copies of the gene sequenced. 5 of these copies turn out to be the allele "A". The **allele frequency** of allele "A", then, is the number of "A" copies out of all copies sequenced, so 5 "A" alleles divided by the total of 20 sequenced allele copies equals an allele frequency of 0.25. You can also depict this answer in a pie chart (Fig. 2).



**Fig. 2. Example pie chart of genetic diversity showing proportion of allele "A".**

In groups, you will calculate the allele frequencies for 5 polymorphic amino acid sites in the translated gene HBA-T2 in a single population. You will then combine your results with the other groups to see how the allele frequency of these 5 amino acid sites differs among deer mice populations. Some of the differences in these polymorphic amino acid sites are likely the source of the functional differences Storz (2007) reported between high and low elevation deer mice populations.

You could gather the sequence data yourself from GenBank; however, to save time, the amino acid sequences, translated from DNA sequences, are available below. The sequences span from positions 46 to 75 in the total amino acid sequence for the hemoglobin α subunit. In your group, determine how many times each amino acid variant occurs in one population. There are 5 amino acid sites (positions 50, 57, 60, 64, and 71) that are polymorphic (more than one amino acid possible) in deer mice in this sequence of 30 amino acids.

For your **GROUP (1-6)**, fill in **Activity Sheet Table I** for the amino acid at each position from the population data provided in **Group\_Data\_for\_Activity\_I.pdf**. Use your table to draw a pie chart of proportion of each mutation (**Activity sheet #2)**. Draw the one specified at the top of your sequence list on the board. When completed, continue on; the class will discuss the results on the board later.

**Example for Group 1** – Draw your pie charts of positions 50 and 57 at the appropriate location on the board. Be sure you label the **position** and which **amino acid** is represented by each slice of each pie chart.

Low elevation population, Fort Larned National Monument, Kansas 620 m

**Amino acid position: 46 75**

↓ ↓

NK53266\_EF369651 FDVSHGSAQVKAHGGKVADALATAAGHLDD

NK53266\_EF369650 FDVSHGSAQVKAHGGKVADALATAAGHLDD

NK53265\_EF369649 FDVSHGSAQVKGHGAKVAGALATAASHLDD

NK53265\_EF369648 FDVSHGSAQVKGHGAKVADALATAAGHLDD

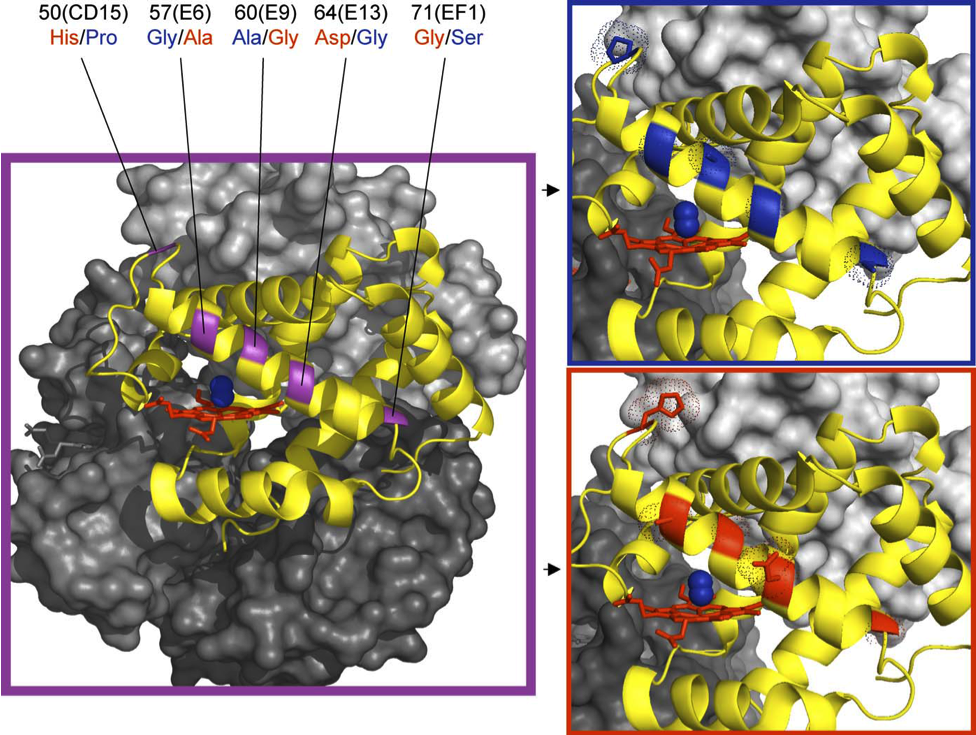
NK53256\_EF369647 FDVSHGSAQVKGHGAKVAGALATAASHLDD

NK53256\_EF369646 FDVSHGSAQVKGHGAKVADALATAAGHLDD

NK53256\_EF369645 FDVSHGSAQVKGHGAKVAGALATAASHLDD

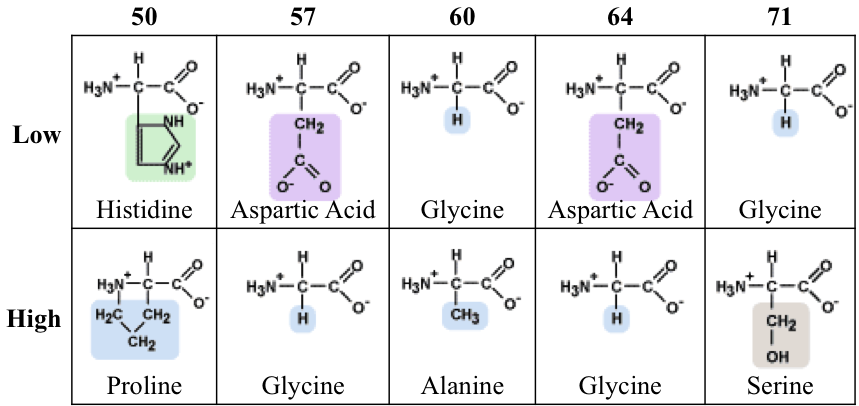
NK53252\_EF369644 FDVSHGSAQVKGHGAKVAGALATAASHLDD

**Task I-III: Understanding how HBA amino acid changes impact hemoglobin function:**



**Fig. 3. Structural model of deer mouse hemoglobin.** In all panels, bound oxygen depicted as blue spheres, α subunit depicted as yellow ribbon model, and the iron-containing *heme* group which binds O2 is depicted in red. Purple box: purple indicates the location of Five Amino Acid Replacement Polymorphisms. Blue box: Typical high elevation amino acid replacements, including amino acid R groups, highlighted in blue. Red box: Typical low elevation amino acid replacements, including R groups, highlighted in red. Modified from Storz et al. (2007).

Fig. 3 is an illustration of where the five polymorphic amino acid sites on the α subunit occur on the hemoglobin protein. The active site where hemoglobin binds oxygen which includes the iron-containing *heme* group, is visible in the figure. In Fig. 4, you can see the various side chains (R-groups) for the alternate amino acids, including how large the R group is for each amino acid and if it is polar, non-polar or charged.



**Fig. 4. Amino acid structure comparison at low and high *P. maniculatus* populations at 5 polymorphic sequence sites.** R-groups are colored green for basic, purple for acidic, blue for non-polar and tan for polar. Modified from http://www.personal.psu.edu/staff/m/b/mbt102/bisci4online/chemistry/chemistry8.htm

Use Figs. 3 and 4 to consider the location of these amino acid replacements and the relative size and polarity of amino acids found at high and low elevations in deer mice. Within your group, consider the following questions:

* What do you think the functional consequence might be for each amino acid variant?
* Would all of the amino acid differences cause a functional change in hemoglobin-O2 binding?
* Which of the 5 replacements is most likely to cause a functional change in Hb-O2 binding?

**Activity Sheet #3.** Make a hypothesis about how amino acid replacements impact hemoglobin function. In your hypothesis, be sure to address why the Hb-O2 binding affinity is different in *P. maniculatus* at different elevations as reported by Storz (2007, *J. Mammalogy*).

**Activity Sheet #4.** How would you test your hypothesis? (Hint: Hb-O2 dissociation curves!)

Get back together as a class and draw your assigned pie chart on the board. As a class, discuss the following questions:

* Do the allele frequencies of HBA-T1 vary across geographic space? If so, does the variation appear random or can you describe a pattern?
* What, if any, geographic features do you think are important in generating the allele distribution you see?
* Strong allelic differences among populations may be evidence of natural selection - do you think natural selection may be acting on HBA-T1?
* What amino acid replacement is likely the *most* important in determining Hb-O2 binding? Can the class reach a consensus?

**Task I-IV: Connecting chemical function to whole organism fitness:**

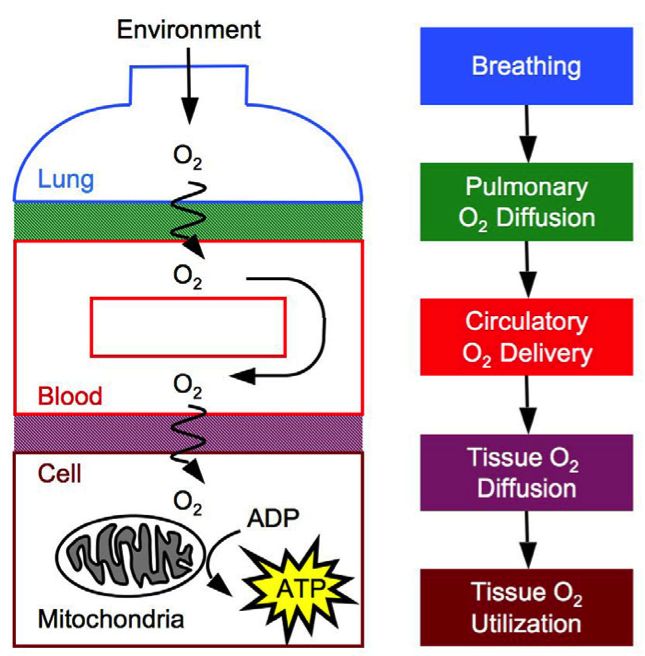
From your investigations above using Arctos, GenBank, data provided, and your reading (Storz, 2007, *J. Mammalogy*), you now:

* Know that the functional types a1c1 and a0c0 have a *genetic* basis.
* Have a mechanistic, testable hypothesis for how an amino acid replacement may affect hemoglobin function.
* Know how the key amino acid replacements change across geography
* Know how key amino acid replacements correlate with important geographic variables.

The functional differences HBA alleles and the geographic distribution of alleles are strong evidence that natural selection is occurring on HBA genes in deer mice. However, step back a moment and consider the case at the whole organism level using Fig. 5 with your group:

**Activity sheet #5.** If an organism has the mutations that you have found associated with high-elevation populations, which way (right, left) will that shift the oxygen binding curve? It may help to draw an Hb-O2 dissociation curve that shows Hb-O2 binding for both low and high elevation deer mice.

**Activity sheet #6.** At what stage in O2 transport within the body (Fig. 5) will the performance difference between the low elevation and high elevation HBA alleles be greatest? Put another way, at what stage in O2 transport will high elevation mice have an *advantage* over low elevation mice when both are at a low environmental partial pressure of O2.



**Fig. 5. Oxygen transport in the body.** O2 is transported from atmospheric air to the mitochondria of tissue cells along a pathway with several diffusive and convective steps. The steps in this O2 cascade are breathing, O2 diffusion across the blood–gas interface, circulation of O2 throughout the body, O2 diffusion across the blood–tissue interface to the mitochondria, and O2 utilization to generate ATP by oxidative phosphorylation.Modified from Storz et al. (2010).

**Part II: Understanding function and natural selection in deer mouse hemoglobin β chains.**

In this section, you will apply your understanding of HBA from Part I to a new set of genes, hemoglobin β subunit genes (HBB), and evaluate the evidence for functional differences and natural selection in HBB.

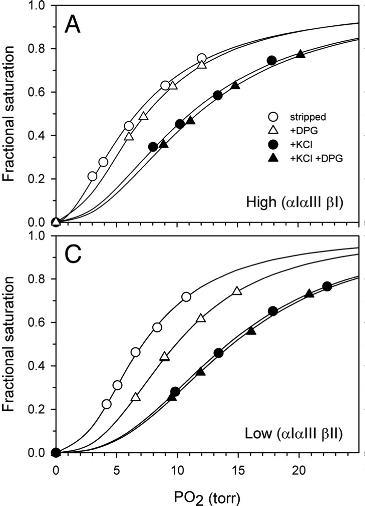
**Task II-I: What are the ways in which the activity level of hemoglobin can be modified?**

Recall that hemoglobin is a protein that includes an active site where it binds its substrate, oxygen. It may also include several binding sites for allosteric affectors. A bound allosteric affector will change the conformation of the protein which may impact the protein's affinity for its substrate. The hemoglobin protein is also a tetramer, meaning four subunits, two α chains and two β chains interact with one another to create a complete protein. This subunit interaction contributes to the cooperative binding property of hemoglobin, where the binding of a single O2 molecule results in a conformational change from the T (tense), low O2 affinity state to the R (relaxed), higher O2 affinity state of the molecule.

**Activity sheet #7.** Generate a list of the ways in which hemoglobin-oxygen binding affinity could be modified by changing the primary sequence of the molecule. Include if the modification is likely to result in a right or left shifted Hb-O2 binding curve. Where in Fig. 5 would your modification make the most biologically significant impact?

(Hint: What kind of modifications impacted Hb-O2 affinity in deer mouse HBA? In Andean camelids - discussed in Storz 2007? What happens when you exercise?).

**Task II-II. Evaluate HBB function in Colorado deer mice**.



**Fig. 6. Hemoglobin-O2 binding curves for *P. maniculatus* HBB alleles.** The Hb-O2 binding affinity is shown for the hemoglobin protein without any modifier (stripped, open circles), with the allosteric affector 2,3 diphosphoglycerate (DPG, open triangles), with the allosteric affector KCl (black circles) and both DPG and KCl (closed triangles). Both (A) and (C) graphs were constructed with hemoglobin proteins using the same HBA alleles (αI, αIII). However, graph (A) shows activity for hemoglobin with the HBB allele βI while graph (C) includes HBB allele βII. Modified from Storz et al. (2009).

DPG is an organic phosphate (OPH) found in mammals. Work with your group to answering the following questions about HBB function:

**Activity sheet #8.**

1. Report the P50 for HBB allele βII (Fig. 6C) stripped, DPG present, KCl present and DPG+KCl present. Recall that the P50 is the partial pressure of oxygen (PO2) for which 50% of the hemoglobin in the sample is saturated.
2. What are the independent and combined effects of adding the allosteric affectors DPG and KCl? Is affinity increased or decreased compared to the stripped protein? Where in the body would the impact of this shift have the greatest impact (lungs, capillaries, mitochondria)?
3. Compare the P50 for HBB allele βI (graph A) to the P50 for allele βII. Which state (stripped, DPG etc.) differs the most?
4. In terms of an organism, what is the consequence of having βI or βII? Using the result of HBA as an example, which allele is likely to be found in a high elevation animal?
5. Do you think that the differences between βI and βII oxygen binding capacity originate from just a few amino acid changes or a large number of amino acid changes in the β subunit primary sequence? Why? How would you test your hypothesis?

**Task II-III. Experimental design to test natural selection in HBB in *Peromyscus maniculatus*.**

Fig. 6 was generated by Storz et al. (2009) by comparing the common β Hb subunit allele (βI) found the summit of Mt. Evans in Clear Creek Co., CO (Fig. 6A) to the common β subunit allele (βII) in Yuma Co., CO (Fig. 6C). βI and βII differ from one another in four amino acids (Table 2).

**Table 2. HBB amino acid differences in *P. maniculatus* from Storz et al. (2009).**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Elevation (m)** | **HBB-62** | **HBB-72** | **HBB-128** | **HBB-135** |
| **βI** | 4347 | Gly | Gly | Ala | Ala |
| **βII** | 1158 | Ala | Ser | Ser | Ser |

For natural selection to operate, a trait must:

(1) be variable,

(2) variation must be heritable

(3) the trait must impact the fitness of an individual.

We have seen that HBB variation in deer mice is variable and heritable; hemoglobin with βI subunits have a lower P50 and higher affinity for oxygen under certain conditions than hemoglobin with βII subunits. From previous studies (reviewed in Storz et al., 2010), we also know that increasing the Hb-O2 binding affinity is *likely* to improve the fitness of animals at high elevations since they can acquire more oxygen at the air/blood barrier in the lungs (Fig. 5) despite the lower atmospheric partial pressure of oxygen. The elements for natural selection to act on HBB are present. How can we test if HBB is subject to natural selection?

Storz et al. (2009) elected to test natural selection by assessing the distribution of HBB alleles across geographic space. The rationale behind this approach is if the alleles found in the high population are *significantly different* from the alleles found in the low population, this may be evidence for natural selection, particularly when coupled with the functional differences of alleles βI and βII. Storz and his coauthors reported that the difference in alleles between the high and low populations in Colorado for HBB was statistically significant (Table 3), meaning the difference was statistically greater than expected in their genetic null model.

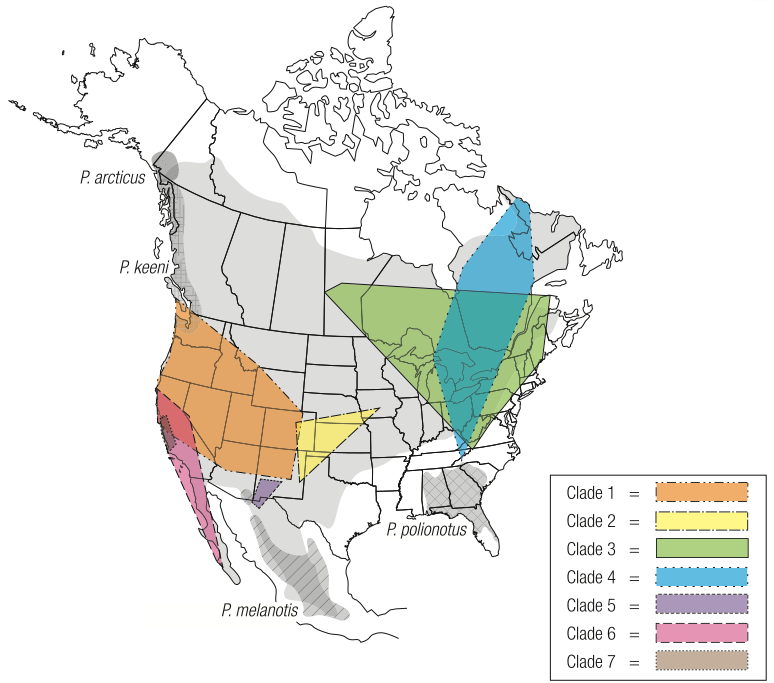
**Table 3. βI allele distribution summary from Storz et. al. (2009).** \* p-value is calculated from two genes, HBB-T1 and HBB-T2, that both produce the hemoglobin β subunit in deer mice.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Locality** | **Elevation** | **Sample size** | **% βI alleles in population** | **p-value\*** |
| Clear Creek Co., CO | 4347 m | 37 | 0.973 | 0.0001 |
| Yuma Co., CO | 1158 m | 38 | 0.263 | --- |

Ask yourself, what is the most exciting future direction for this work? Your task is to build on the findings of Storz et al. (2009) to assess the presence of natural selection acting on HBB in deer mice.

Consider:

1. What environmental variables could be most important to understanding HBB allele distributions in deer mice?
2. What biotic variables could you measure?
   1. Some possibilities include:
      1. Genetic diversity (alleles) in HBB in deer mice populations.
      2. Functional differences in Hb among deer mice populations.
      3. Physiological properties (like resting and maximum metabolism, breathing rates) of individuals within deer mice populations.
      4. Long term fitness (survival and offspring produced) of individuals.
      5. Functional differences due to specific mutations found in deer mice populations.
3. Could deer mice population history (i.e. some populations are isolated from one another) be important in your test? Put another way, what does it mean if the same pattern found in one group with shared genetic history, a clade, is also found in a different group (clade). Figure 3 illustrates 7 clades found in *P. maniculatus* based on mitochondrial DNA.



**Fig. 7. Genetic structure of *P. maniculatus* across range based on a mitochondrial gene.** Clades are populations that share a common genetic history. On the map, the occurrence of different clades are indicated by different colors (key). The solid light gray area is the total range of *P. maniculatus.* Gray with different hatching indicates the ranges of *Peromyscus* species that are close relatives to *P. maniculatus.* Modified from Natarajan et al., 2015.

1. How could understanding HBB in close relatives to deer mice (other *Peromyscus* species) help you understand the evolution of HBB in deer mice?
2. Across what geographic scope would you like to address HBB variation distribution? What are the pros and cons of a small-scale study? A large-scale study?

There all kinds of avenues you could follow to extend the work of Storz et al. (2009), but focus in on a **single idea** you could test using 6 sampling sites. Consider ample size and replication; they will, in part, determine the strength of your test.

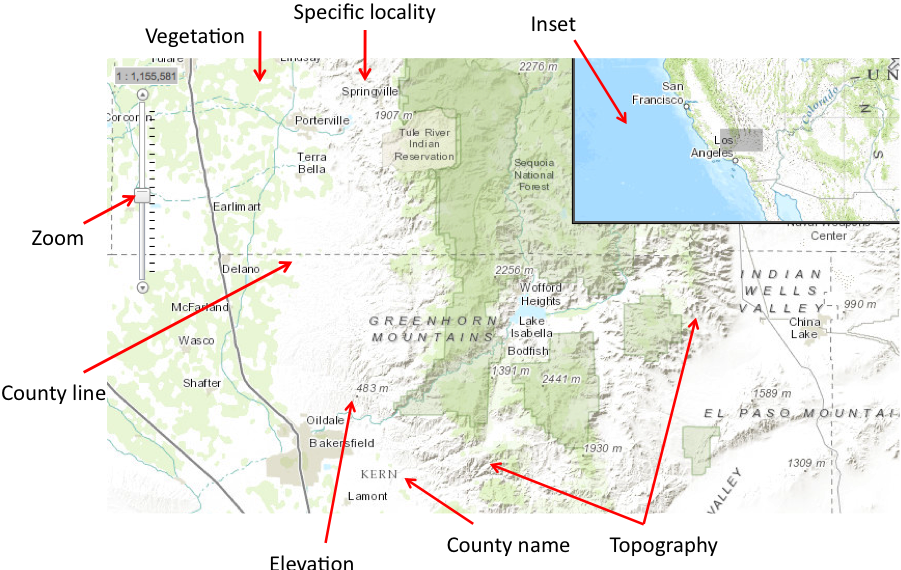
**Activity sheet #9.**

1. What is the essential question you will test?
2. Write a hypothesis that demonstrates how you will test your essential question.
3. What variables, environmental and biotic, will you measure in your study? Justify your selection of these variables.
4. For your study, you need to select 6 sampling localities where you will collect data on deer mice populations. Recall that for the Storz et al. (2009) study, they chose 2 sites that were 220 miles apart. What would you do? Remember to think carefully about what environmental variables are important to your study, how population history of the deer mice may be important and what you gain from dense sampling versus having a wide geographic scope.

Use the ERSI topographic interactive map to select your sample sites. For each sample site, you need a SPECIFIC LOCALITY, (*e.g.* town, peak), COUNTY and STATE as well as any other relevant details.

<http://bmproto.esri.com/WorldTopoMap/WorldTopographicMap.html>

For each of the specific localities you choose, keep the ESRI topographic interactive map open in a different tab (you will have 6 different tabs open in the internet browser window).



**Fig. 8. Features of ERSI interactive topographic map.**

**Helpful features on the ERSI map:**

**Zoom** - Use the zoom in the upper left hand corner of the screen to zoom in and out.

**Topography** - mountain ranges, lakes and rivers are illustrated. Mountains are indicated by rugged texture which shows shifts in elevation.

**Elevation** - Topographic features (marked by dot) are labeled with elevation in meters.

**County** - County lines are dashed, straight lines. County names are in all caps, no italics.

**Vegetation** - green indicates some natural vegetation (riparian corridor, forest, etc.).

**Inset** - Shadowed box in the upper right hand inset indicates region in main window.

Once you have a clear idea of your essential question, hypotheses, environmental and biological variables, and sampling requirements, turn to a neighboring group and share your approach. Explain what makes your approach powerful and what you will expect to see to accept your hypothesis. Do you have any suggestions for your neighboring team's project? Do they have any suggestions for you?

Modify your experimental design to incorporate useful suggestions from your discussion with your classmates. **You will submit a formal write up of this proposal as your homework.**

**Task II-IV: Testing the feasibility of your experimental design:**

If you were going to conduct this proposed study, one of the first things you would do is check to see if there are any biological resources already available for your work. This is particularly relevant if you plan to sequence HBB from different individuals since you could use previously collected tissues to extract DNA for sequencing.

To do this, you will use VertNet, a database you can use to search natural history museum catalogs for specimen data, including locality and the availability of preserved tissues. There are multiple sources of this type of data. It can come from an aggregator (such as [VertNet](http://vertnet.org/) or [GBIF](http://gbif.org/)) or from a museum database (such as [Arctos](http://arctos.database.museum/)). Note that it is always a good idea to verify data with the source, the institution housing the collection.

**VertNet** [**http://vertnet.org/**](http://vertnet.org/)

1. Open a new internet browser window. Navigate to the VertNet website and click on the “Search Options” link on the bottom left.
2. On the search page, click the “advanced search options” on the right side of the search box.
3. Then, on the top, click “Has tissues” and “Is mappable”.
4. In the Darwin Core Terms search boxes, enter the data relevant for your search to look for tissues to do your experiment: State, County, and, Specific Locality, or nearest landmark. enter “Peromyscus” in the Genus box and “maniculatus” in the SpecificEpithet box. Then click the blue search box on the top right. Fig. 9 demonstrates how to search for the specimens from Mount Evans in Activity I.

If you cannot find tissues using the search for a specific locality, just do a search for the **county** (don’t enter any specific locality). Look at the points on the map for the specimens in your results, look back at your ESRI map, how close are the samples to the sampling locality you chose?

**Activity sheet #10.** Are there any tissues for you to conduct your experiment? How many samples do you think you need from each locality?

****

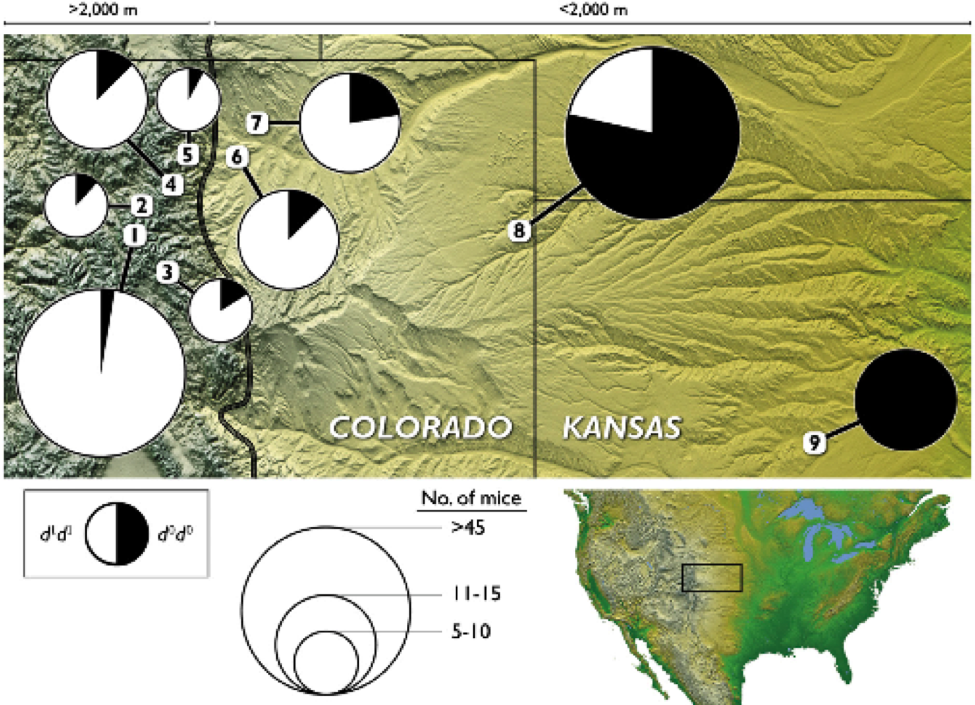
**Fig. 9. Advanced Search Options in VertNet.** Red boxes highlight key filters (has tissues, is mappable and essential Darwin Core terms.

**Task II-V: Assessing evidence for selection on HBB in *P. maniculatus*:**

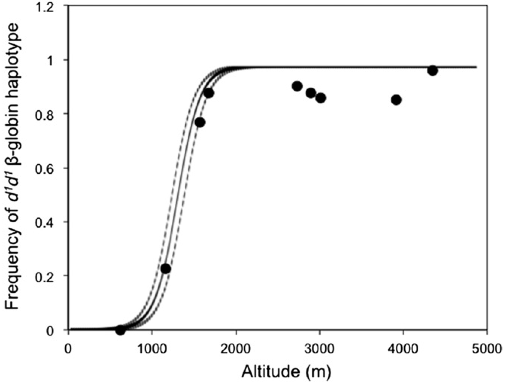
Storz and his colleagues performed Tasks II-III and Tasks II-IV that you have just completed and published their results in 2012. They chose 9 localities create an altitudinal transect from Colorado to Kansas. In these 9 *P. maniculatus* populations, they sequenced HBB-T1 and HBB-T2 to characterize two functional types (d1 is functionally equivalent to βI; d0 is functionally equivalent to βII from Storz et al. 2009). Note, because Storz et al. surveyed both HBB-T1 and HBB-T2, results include the functional types for both genes. Thus an individual with the haplotype d1d1 has the functional type d1 at both the HBB-T1 gene and the HBB-T2 gene.

**Activity sheet #11:**

1. Describe quantitatively how the distribution of the haplotype d1d1 changes across the 9 sample sites (Fig. 10). In which populations is d1d1 most abundant? Least abundant?
2. Clinal analyses demonstrate the rate of transition between one state to another across a transect, with steeper transitions indicating either natural selection or recent contact between two independent clades (groups with shared genetic history). In Fig. 11, the analysis shows the transition from no d1d1 present in a population (frequency = 0) to only d1d1 (frequency = 1) present across altitude. Describe how d1d1 frequency changes across altitude. Is the transition rapid or gradual?
3. Use Figs 6, 10 and 11 to evaluate the evidence for natural selection acting on the hemoglobin β subunit in deer mice. If natural selection is occurring, what do you think the *mechanism* is (what is the environmental pressure that results in selection?)? How does the sampling design (sample sizes per population and populations selected) of the Storz et al. 2009 and Storz et al. 2012 papers impact your conclusion?
4. What remaining questions do you have about natural selection on HBB in deer mice? Are you satisfied with the geographic scope of the studies presented? With the phylogenetic scope? Are you certain of the mechanism causing natural selection - if not, why not?



**Fig. 10. Distribution of HBB genotypes across an altitudinal transect.** Proportion of pie chart colored black is proportional to number of d0d0 copies found in the populations. From Storz et al. (2012).



**Fig. 11. Frequency of HBB haplotype d1d1 by altitude.** Dots indicate values from sample sites 1-9; solid line indicates results from clinal analysis fit.