



Working Dog Project Annual Update

Principal Investigator Dr. Elinor Karlsson

August 1, 2018

In its first year, the Working Dog Project has made substantial progress. We are on schedule to complete all work currently funded within the next year, and are poised to rapidly scale up the project when additional funding is secured.

In September 2017, Dr. Jessica Hekman (DVM, PhD) joined the Working Dog Project to coordinate sample collection, computational infrastructure development and data analysis.

During the first year, we focused on three main components: (1) collecting the DNA and phenotype information needed for the work currently funded; (2) developing the technology needed to analyze this data; (3) establishing new relationships with working dog organizations, including the International Working Dog Registry, American Humane, Southeastern Guide Dogs, The Seeing Eye, Diggity Dogs Service Dogs Inc, and Leader Dogs.

The working dog community is enthusiastic about the project, and the majority of organizations we've spoken to are aligned with our goal of building a large, shared data resource as the best way to make rapid progress. Currently, our growth is constrained primarily by the need for additional funding to implement collaborations with all the interested organizations. We have presented the project at the Penn Vet Working Dog Conference in April and participated in the stakeholder's meeting immediately

following. Dr. Jessica Hekman is invited to speak at the Assistance Dogs International Conference (Minnesota, August 2018). Dr. Elinor Karlsson will be a keynote speaker at the International Guide Dog Federation conference and the subsequent breeder's workshop (Sydney, Australia, September 2018).

We now have substantial buy-in from the guide and service dog organizations. Over the next year, we will focus on broadening our collaborations with law enforcement organizations. We're also interested in developing new ways to reach individuals breeding diverse types of working dogs, potentially through members of the Theriogenology community.



Project collaborators include (A) NEADS, (B,C) Guiding Eyes for the Blind and (D) Canine Performance Services at Auburn University.

Component 1. Collect detailed behavioral phenotypes and DNA samples.

Proposed work: We will include dogs from three different populations: (1) We will collect saliva samples and

behavioral assessments from 100 NEADS assistance dogs, including both dogs in training and dogs that have been placed; (2) the Canine Performance Sciences group at Auburn University, led by Dr. Paul Waggoner, has already assembled a databank with DNA samples and full performance evaluations for 140 dogs from their breeding program, and will be a partner on this project; (3) Our Darwin's Ark citizen science project (darwinsdogs.org) has enrolled over 13,000 dogs and collected detailed owner-reported behavioral phenotypes. We are currently generating dense genotype data for 600 of these dogs through an NIH R21 grant, and we will include these dogs in our project at no additional cost.

Status: Underway. ~75% complete.

Scope. Taking advantage of our more cost-efficient sequencing technology (see Goal 2), we were able to expand this component to include 96 German Shepherd dogs from Guiding Eyes for the Blind (GEB)

Behavioral phenotypes. We have complete phenotype information for the Guiding Eyes for the Blind dogs and the Darwin's Ark project dogs. NEADS has provided phenotype records, and we are currently processing these into an electronic format. We are negotiating a data usage agreement with Auburn.

DNA. All DNA samples have been collected.

Detailed description of work

NEADS: NEADS continues to be an exceptionally responsive collaborator. **DNA.** We have exceeded our initial goal of collecting 100 saliva samples from NEADS; in total, we have collected 180 saliva samples. **Phenotypes.** NEADS has shared behavioral phenotype data on these dogs. We are working with an outside contractor and the International Working Dog Breeders Association to develop a system for importing the phenotype data (which is not in an electronic format that can be automatically processed) into the International Working Dog Registry (IWDR) database. With this collaborative approach, the data will be accessible to us, and also allow NEADS to access services provided by the IWDR. This integration with IWDBA will facilitate phenotype collection from other organizations in the future.

Guiding Eyes for the Blind: Guiding Eyes for the Blind (GEB) is an extremely supportive and responsive collaborator. GEB has given us access to their exceptionally well maintained database, which includes detailed records on every dog they have bred for nearly 20 years. **DNA.** GEB has provided 96 DNA samples from German Shepherd guide dogs, providing the critical second breed needed to develop our cross-breed meta-analysis approach. These samples will be sent for sequencing within the next month. **Phenotypes.** We are collaborating with GEB and statistician Eldin Leighton (IWDBA) to analyze heritabilities for traits important to guide dog success. We are currently analyzing the heritability of Behavior Checklist (BCL) phenotypes (the guide dog industry standard) for fifteen generations of guide dogs (5,338 dogs total).

Auburn CPS group: We have launched the collaboration with the Auburn CPS group, providing a critical *second type* of working dog for our pilot project. **DNA.** Auburn has sent 198 DNA samples from their performance scent dog breeding program. **Phenotypes.** We are currently negotiating a data use license they have requested prior to sharing phenotype information.

Darwin's Ark citizen science project: We now have both genetic data and behavior phenotypes for 768 pet dogs (an increase from 350 dogs at our last report). This includes 332 dogs with genome sequence data and over 400 with dense genotype array data (our older approach). We anticipate this will increase to over 1000 within the next 3 months. We also launched the new Darwin's Ark website (darwinsark.org), which offers owners the opportunity to fund the sequencing of their own dogs,

enabling us to expand even more rapidly. This work, done using funds from other sources, will significantly benefit the Working Dog Project.

New relationships.

We have established new relationships with working dog groups who are enthusiastic to share both DNA and phenotype information. Because some of the groups, such as American Humane, do not currently rigorously phenotype their dogs, we are building capacity to collect working dog phenotype information into the Darwin's Ark website. American Humane is serving as our pilot project in this capacity, encouraging eight of their working dog handlers to report on their dog's behavior through the Darwin's Ark site. Based on these results, we will assess what additional, working dog specific surveys we need to build into Darwin's Ark to ensure we collect robust phenotype data. All results will be shared with American Humane to support their training program.

Next steps.

We anticipate completing this component within the next 3-6 months. This will include collecting DNA and behavioral phenotype information from our existing collaborators, and processing that data into an electronic format. We will continue to develop the relationships and systems needed to expand the project in the future.

Component 2. Generate low-cost, full genome sequence information on each dog using Broad Institute multiplex sequencing and whole genome imputation.

In component 2, we propose to generate full genome sequence information on each dog by applying genome imputation to low-coverage, whole genome sequencing data. At a price point comparable to using genotyping arrays, we will produce a dataset with far more statistical power, and one in which we can detect interesting genetic variation missed in the smaller panel of markers chosen for the array.

Status: Underway. ~75% complete

Scope. We are now sequencing through a new relationship with a company called Gencove. Together, we have developed the computational pipelines needed for low cost, low-pass, highly accurate full genome sequencing of dogs.

Detailed description of work

Reference dataset: We have developed a new low-pass sequencing approach that combines 1x sequencing with imputation. To do this we are using a reference panel, with a large number of deep-sequenced dogs, provided by Elaine Ostrander's research group at NIH. We are currently processing additional data provided by Dr. Ostrander and others, and anticipate the reference set will exceed 700 dogs by the end of 2018.

Low-pass sequencing: The quality of the low-pass sequencing approach has far exceeded our expectations. With just 2ng of DNA (1/350 of that required for a genotyping array), we get over 8.8 million data points per dog and capture nearly all common genetic variation. For comparison, a genotyping array costs more per dog, requires 750ng of DNA, and gives information on fewer than 700,000 data points per dog. To assess the quality of the data, we measured the concordance of low-pass sequencing data with "gold standard" deep-sequencing data, and found it was as accurate as genotyping (~99%), yet provided 12-15x more data. As very little DNA is required for low-pass

sequencing, we can collect just a single swab per dog and be confident we will have enough DNA, lowering costs further.

DNA sequencing status. We have now shifted entirely to low-pass sequencing. We have sequenced 110 NEADS dogs (exceeding the goal of 100 dogs). The GEB dogs are in process, and the CPS dogs await the data sharing agreement. We have 312 Darwin's Ark dogs with low-pass sequenced data, as well as more than 400 with genotyping data (predating the shift to sequencing).

Next steps

We anticipate completing this component within the next 3-6 months. Methods development is complete, and Gencove low-pass sequencing takes approximately 2 months to complete.

Component 3: Develop and apply new statistical methods for meta-analysis of both working dog populations and a cohort from Darwin's Ark.

Using the data from component 2, we will develop the methods for applying meta-analysis to large, very dense genomic datasets for dogs. We will adapt tests used in human populations to account for factors unique to dogs, including the limited genetic diversity within breeds, the large genetic differentiation between dog breeds, and the mixing of breeds seen in many pet dogs.

We will apply the dog specific methods we develop to the data from component 2 and search for genetic variants associated with scenting and/or retrieving ability.

To facilitate development, we will first assess our overall statistical power to detect any causal variants using the genome-wide distribution of association scores. We will examine in depth the top loci and their effect on phenotype. As we have detailed information for each dog, we will integrate environmental factors that might influence the behaviors of interest into the analysis through a joint meta-analysis for main and interaction effects. We will also use cross-phenotype meta-analysis to check for multiple associations at a single marker for both retrieving and scenting ability, to assess whether they share a common genetic background or are entirely distinct abilities. Finally, we will carry out exploratory cross-phenotype checks, to see if any of our significant markers are also associated with other phenotypes.

Status: Underway. ~25% complete.

Detailed description of work

Data. We currently have low-pass DNA sequence data from 312 Darwin's Ark dogs and 110 NEADS samples. By the end of September, we will also have data for 96 GEB dogs.

Pilot GWAS and data analysis. Using the current Darwin's Ark data, we have started developing the analytical methods needed in the Working Dog Project. We began with a genetic study of dog height. This trait is driven by large-effect genetic variants, making it a perfect "proof-of-principle" test (because behavioral traits are more genetically complex, we anticipate they will require many more dogs). The results of this study of 312 dogs were exceptionally strong. The low-pass sequencing data was very high quality, and our owner reported trait (dog height relative to adult human) was sufficiently accurate. We not only found genes already linked to dog size in previous studies (*IGF1* and *HMG2*), we found entirely new genes (including *LCORL*, a gene associated with size in horses) not detected in studies using less complete genetic data and less sophisticated analytical approaches.

In collaboration with Guiding Eyes for the Blind, we are also currently doing a pedigree analysis to identify behavioral phenotypes with the highest heritability. These highly heritable traits are most likely to yield results with a relatively small number of dogs.

Next steps

We will run our genetic analysis on our four study populations: the NEADS dogs, the Guiding Eyes for the Blind dogs, the Auburn dogs and the Darwin's Ark dogs independently, and then combine the results using a meta-analysis. We will focus on four phenotype categories: (1) scenting ability; (2) retrieving ability; (3) toy drive; (4) any additional highly heritable traits. We will first develop standardized metrics for measuring those phenotypes in each of our four study populations. We will then implement the statistical analyses planned for this component, including meta-analysis across all four study populations for each phenotype.

We anticipate the total number of dogs in the meta-analysis will exceed 1000. This may yield significant results, if large effect variants are responsible, but larger datasets will be needed to fulfill the goals of the Working Dog Project. We will develop algorithms that are easily scaled up as our available data increases.

Thank you!

We deeply appreciate the support we have received from the Theriogenology Foundation, and look forward to working together to grow the Working Dog Project in the future.