AFRL-RH-FS-TR-2011-0013

GENOME-WIDE ASSOCIATION MAPPING FOR INTELLIGENCE IN MILITARY WORKING DOGS:

Canine Cohort, Canine Intelligence Assessment Regimen, Genome-Wide Single Nucleotide Polymorphism (SNP) Typing, and Unsupervised Classification Algorithm for Genome-Wide Association Data Analysis

Victor T. Chan
Camilla A. Mauzy
Armando Soto
Jessica A. Wagner
Bioeffects Division
Molecular Bioeffects Branch

Amy D. Walters
Jeanette S. Frey
Tiffany M. Hill
Henry M. Jackson Foundation
For the Advancement of Military Medicine
2729 R Street, Wright-Patterson AFB OH 45433-5707

Karen L. Overall
Soraya Juarbe-Diaz
Donna Dyer
Penn Med Translation Research Laboratory
125 S. 31th St. Philadelphia PA 19104-7051

Richard M. Wolfe
Lonnie R. Welch
School of Electrical Engineering & Computer Science
329 Stocker Center, Ohio University
Athens, OH 45701-2979

Technical Report - September 2011

Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Bioeffects Division
Molecular Bioeffects Branch

NOTICE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RH-FS-TR-2011-0013

THIS REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//	//SIGNED//
CAMILLA A. MAUZY, Work Unit Manager Molecular Bioeffects Branch	GARRETT D. POHLHAMUS, DR-IV, DAF Chief, Bioeffects Division Human Effectiveness Directorate 711th Human Performance Wing Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
September 2011	Final Report	April 2009 – September 2011

4. TITLE AND SUBTITLE
GENOME-WIDE ASSOCIATION MAPPING FOR INTELLIGENCE IN MILITARY WORKING DOGS: Canine Cohort, Canine Intelligence Assessment Regimen, Genome-Wide Single Nucleotide Polymorphism (SNP) Typing, and Unsupervised Classification Algorithm for Genome-Wide Association Data Analysis.

5a. CONTRACT NUMBER

5b. GRANT NUMBER
NA

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
*Victor T. Chan; *Camilla AS. Mauzy; *Armando Soto; *Jessica A. Wagner; Amy D. Walters; Jeanette S. Frey; Tiffany Ml Hill; Karen L. Overall; Soraya Juarbe-Diaz; Donna Dyer; Richard M. Wolfe; Lonnie R. Welch

5d. PROJECT NUMBER
ODA

5e. TASK NUMBER
WP

5f. WORK UNIT NUMBER
ODAWP001

8. PERFORMING ORGANIZATION REPORT NUMBER AFRL-RH-FS-TR-2011-0013

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command*
Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Bioeffects Division
Molecular Bioeffects Branch
Wright-Patterson AFB OH 45433-5707

10. SPONSOR/MONITOR'S ACRONYM(S)
711 HPW/RHDJ

11. SPONSOR/MONITOR'S REPORT
AFRL-RH-FS-TR-2011-0013

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution A: Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This seedling project aimed to genetically map intelligence in the military working dog (MWD) population. A total of 199 canine subjects were recruited from United States working dog contractors. Of the recruited subjects, 153 were tested using the Canine Intelligence Testing Protocol (CITP), developed by Dr. Karen Overall (UPENN) to specifically analyze canine intelligence. CITP allows quantitative assessment of intelligence in individual dogs using a scoring system based on the latency to response, success-in-effort time, attentiveness, interest in novelty exploration, response to signaling and showing, observational learning, problem solving/boldness, and handedness. Blood samples were collected from the canines, and genomic DNA prepared. A total of 117 dogs, belonging to three breeds (German Shepherds, Belgian Malinois, Labrador Retrievers) were down-selected and successfully genotyped for whole genome (WG) single nucleotide polymorphism (SNP) markers by means of the Affymetrix Canine SNP Array v2. A 'proof-of-concept' advanced data mining algorithm for unsupervised analysis of genome-wide association study (GWAS) dataset was successfully developed. Using this algorithm, canine subjects were successfully clustered into the correct breeds with an accuracy ranging from 89 – 100%, solely based on the WG SNP profiles. The details of the algorithm are described in TR AFRL-RH-WP-TR-2011-0081. While not initially part of the seedling proposal, this project did receive IACUC permission to test DoD MWDs for follow-on studies, a unique and significant accomplishment.

15. SUBJECT TERMS
Military working dog genome-wide association study genetic marker intelligence

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Camilla Mauzy
U	U	U	SAR	39	19b. TELEPHONE NUMBER *(include area code)* NA

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

THIS PAGE INTENTIONALLY LEFT BLANK.

TABLE OF CONTENTS

LIST OF TABLES

PREFACE

This research was conducted at the Applied Biotechnology Branch (711 HPW/RHPB), Human Effectiveness Directorate of the 711th Human Performance Wing of the Air Force Research Laboratory, Wright-Patterson AFB, OH, under Dr. John J. Schlager, Branch Chief. As of 1 October 2011, this branch is now the Molecular Bioeffects Branch in the Bioeffects Division. The research described in this report was completed prior to the reorganization, therefore prior project reports, contracts, and IACUC protocols are designated RHPB. This technical report was written as the Final Report for AFRL Work Unit ODAWP001. This project was partially funded by DARPA (in conjunction with UES contract FA8650-08-C-6832).

Research performed with Dr. Overall, University of Pennsylvania, under UES contract FA8650-08-C-6832. Henry M. Jackson Foundation employees were working under Cooperative Agreement FA8650-05-2-6518.

All studies involving animals were approved by the Wright-Patterson Institutional Animal Care and Use Committee, and were conducted in a facility accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care, International, in accordance with the *Guide for the Care and Use of Laboratory Animals*, National Research Council (1996). Studies were conducted under approved Air Force Research Laboratory Institutional Animal Care and Use Committee Protocol AFDR-2009-002A "*Genome-wide Association Mapping for Superior Intelligence in Military Working Dogs*" (Univ. of PA Protocol #802551).

ACKNOWLEDGEMENTS

Thanks to Dr. Walter Burghardt, Chief, Behavioral Medicine and Military Working Dog Studies Department of Defense Military Working Dog Veterinary Service, at 341 TRS, for his continued support of this work. A heartfelt thank you to LTC Rick Probst, DVM, DACLAM, US Army Veterinary Corps, who served as the Chief of the USAF Animal Use Programs, AFSGRC, during this research. The authors would also like to thank the numerous contractor groups, as well as multiple international governmental/military MWD programs, who had granted access to their dogs for behavioral testing and genotyping. Their support and confidence in this project and the research team are outstanding and essential for the accomplishments described in this report.

SUMMARY

In a collaborative effort between the Air Force Research Laboratory, Human Effectiveness Directorate, Applied Biotechnology Branch (now 711 HPW/RHDJ), and the University of Pennsylvania, this project aimed to genetically map superior intelligence in the military working dog (MWD) population. To achieve this goal, a total of 199 canine subjects were recruited from United States working dog contractors. Of the recruited subjects, 153 were tested for problem solving using a behavioral tests regimen, i.e. the Canine Intelligence Testing Protocol (CITP), developed by Dr. Karen Overall, a canine behavior expert. This testing regimen allowed quantitative assessment of intelligence in individual dogs using a scoring system based on the latency to response, success-in-effort time, attentiveness, interest in novelty exploration, response to signaling and showing, observational learning, problem solving/boldness and handedness. Blood samples were collected from all subjects in the cohort, and genomic DNA prepared from the whole blood was stored to maintain integrity prior to whole genome (WG) single nucleotide polymorphism (SNP) typing. One hundred and seventeen subjects, belonging to three breeds, German Shepherd Dog, Belgian Malinois and Labrador Retrievers, were down-selected for WG SNP typing by means of the Affymetrix Canine SNP Array v2, which contains a total of 127,132 SNPs, selected from the 2.5 million SNPs that were identified in the canine genome project. Due to premature termination of funding by DARPA, this project could not be completed as planned. For instance, behavioral testing of the subjects in the cohort was only partially completed, and the analysis of the available behavioral tests data could not be conducted. Despite these drawbacks, the principal investigators of this project were determined to complete the project as much as possible, especially for the WG SNP typing and advanced bioinformatics. As such, the second phase of this project mostly focused on the development of algorithms for unsupervised analysis of genome-wide association study (GWAS) data. As a proof-of-concept, a classification analysis of the WG SNP typing dataset of 117 phenotypically tested subjects in three breeds (German Shepherd Dog, Labrador Retrievers, and Belgian Malinois) was conducted. Using the algorithm that we have developed, the canine subjects were successfully clustered into the correct breeds with an accuracy ranging from 89 – 100%, solely based on the WG SNP profiles. Classification accuracy, however, was not significantly affected by data process methods, or by the quality of the annotations of the SNP. This result confirms that this algorithm is highly robust. The details of the development of this algorithm are described in the Technical Report AFRL-RH-WP-TR-2011-0081 entitled: "*Development of Advanced Classification Algorithm for Genome-Wide Single Nucleotide Polymorphism (SNP) Data Analysis*".

Keywords: military working dog, genome-wide association study, genetic marker, intelligence, Canine Intelligence Testing Protocol, classification technique, clustering analysis

Technical Report: September 2011

1. INTRODUCTION

"The capability they (Military Working Dogs) bring to the fight cannot be replicated by man or machine. By all measures of performance their yield outperforms any asset we have in our inventory. Our Army (and military) would be remiss if we failed to invest more in this incredibly valuable resource."

General David H. Petraeus, USA, 9 Feb, 2008

1.1 Intelligence and Genetics

The underlying molecular mechanism of intelligence (as well as its very definition) is complex and context-dependent (Gray *et al.* 2004). Although intelligence may have different meanings under different circumstances, it can be loosely defined as a general mental capability related to one's ability to learn, reason, plan, comprehend complex ideas, think abstractly, and solve problems by integrating the situational information with knowledge learnt from past experiences. Although it is widely accepted that there is a significant role of inheritance in the determination of intelligence levels, the exact genetic components and how they operate are far from understood. It is, however, certain that intelligence is not determined by a single gene, but by a complex interaction of a large number of genes, and that each of them may only have a very small effect size. Such genes of varying effect sizes that collectively contribute to a quantitative trait are called quantitative trait loci (QTL). Because QTLs contribute interchangeably and additively as probabilistic propensities, any particular QTL associated with a polygenic trait is neither necessary nor sufficient. This implies that the underlying molecular basis for two individuals with a similar level of intelligence may be different. Such genetic heterogeneity would significantly impact the power of genetic analysis of identifying intelligence-associated loci. Despite this complexity, multivariate genetic analyses suggest that overlapping gene sets may be involved in multiple cognitive abilities (Plomin *et al.* 1997).

Studies on family, twin and adoption data in humans demonstrated that there is a strong genetic influence on human intelligence. The intelligence quotient (IQ) scores of identical twins raised apart have been shown to be highly similar (nearly as similar as those of identical twins raised together), while those of fraternal twins are less similar (Daniel *et al.* 1963; Vandenberg 1968). Consistent with the notion that genetics contribute significantly to intelligence, the IQs of adopted children have only a small relationship to the IQs of the biological children of their adoptive parents, or to their adoptive parents. As the adopted children age, they become more similar to their biological parents and less similar to their adoptive parents. Model-fitting analysis and meta-analysis of these genetic data on IQ suggest that heritability may account for approximately 50% (i.e. 40-80% as suggested by different investigators) of the variance in IQ scores (Detterman, *et al.* 1990; Daniels, *et al.* 1997; Spady *et al.* 2008; Deary *et al.* 2006).

1.2 Genetics in Canine Behavior

Examination of a coding repeat microsatellite region in canines indicated that these segments contain fewer perfect repeat sets than those found in humans (Fondon *et al.* 2004). These findings indicate that the canine may have an innate ability to rapidly develop new alleles, thus a much shorter evolutionary time required for the development of new phenotypes (Fondon and Garner, 2007). Humans may have taken the advantage of this ease of genetic crossover for trait development to create the vast and varied breed-oriented canine behaviors such as herding, guarding, pointing, tracking, and retrieving (Coppinger and Scheider, 1995; Akey *et al.* 2010). As such, the dog displays the greatest behavioral diversity of all land mammals. Studies examining heritability of these traits indicate that, at least for these specific canine-oriented behaviors, the controlling gene set may actually be relatively small (Ruefenacht, *et al.* 2002).

It has recently been suggested that the canine exhibits more human-like behavior than any other animal, including primates (Udell *et al* 2008), making the dog an excellent animal model for cognitive research. In light of this, there have been recent attempts to understand canine aggression, PTSD, and other behaviors as correlated to equivalent functions/syndromes in human cognition (Markman, *et al.* 2004; Nippak *et al.* 2005; West *et al.* 2002). Using a candidate gene approach to identify contributing gene sets to canine behavior has met with little success, possibly due to small sample numbers, as well as poorly defined phenotype classifications of complex behavior (Masuda *et al*, 2004; Ogata *et al*. 2006; Våge *et al* 2010). However, with the completion of the canine genome project and identification of informative mapping SNPs, whole genome scans (genome-wide association studies or GWAS) can be conducted using high throughput microarray profiling techniques such as the Affymetrix GeneChip Technology Platform. With careful development of quantitative behavioral phenotype assessment, GWAS can be an invaluable method to examine high-resolution mapping of the entire genome for intelligence-related QTLs. However, extreme care must be taken in the development of the behavioral testing methodology to ensure that the testing is both quantifiable and repeatable and measures a very specific domain of intelligence and/or cognitive functions, i.e. endophenotype (Sabb *et al*, 2009; Amos, 2007). Additionally, canine breed differences in GWAS have been seen in linkage disequilibrium coverage, population structures, and SNP tagging, thus requiring a careful assessment of individual breeds prior to conducting such scans (Ke *et al*, 2010).

1.3 Increased Need for Military Working Dogs

Despite on-going research to develop new methods of improvised explosive device (IED) detection, the olfactory system of the military working dog still out performs equipment, with 80% versus 50% detection compared to sensor systems (Ackerman, 2010). With two theaters of military operation plus the needs of DoD, Transportation Security Administration (TSA), and Homeland Security in securing continental US locations, there has been a strain on the ability of the Air Force and US breeders/trainers to supply healthy, well trained MWDs. Additionally, the need for replacement animals due to injury and/or infection from deployment has also increased

the need for animals to new levels. This fact has been recognized by General David Petraeus (as quoted above) who has stated the strong need for more MWDs.

1.4 Military Working Dog Intelligence Genetics (MWDIG) Project

Developing genetic testing methods for use as a breeding tool will allow more consistent intelligence and behavior in MWD litters, decreasing the dropout rate and lowering training/selection costs. At this time, very few genetic approaches have been developed for use by the DoD to select for traits needed for outstanding performance in military-relation missions, although the use of genetic tests as a breeding tool has been used by the AKC and breeders since the mid 1990's. The use of such tests have become an industry standard for proactive prevention of diseased stock (http://www.caninehealthinfo.org/chicinfo.html). Because of this, genetic analysis is a logical approach to unlock the molecular mechanism of canine intelligence (and other desirable traits for military missions). Once genes contributing to intelligence are identified, canine genetic tests can be subsequently developed and used as a "pre-purchase" test requirement for acquisition and acceptance of dogs into the DoD MWD programs. They may also be developed as a breeding tool towards the creation of a superior intelligent Military Working Dog "Breed", containing desired attributes of several breeds such as the German Shepherd Dog and Belgian Malinois, yet displaying high levels of intelligence and independent decision-making not currently seen in any breeds. Such "super intelligent" canines may permit relatively autonomous missions in such a manner as currently used in UAV tactics, allowing for a single handler to monitor/direct multiple MWDs out of sight with sensor-activated vests (Miller 2010). However, even with advanced remote control vests, the rate limiting factors on the use of autonomous MWDs will not be device-oriented, but in the canine's trainability, response to environmental factors in theater, and independent decision-making capabilities.

The identification of intelligence-related genes has another significant implication that it would facilitate understanding how these genes interact with each other to contribute to overall intelligence and how they may be modulated for performance enhancement. Thus, gaining new knowledge in a complex polygenetic trait as intelligence will not only provide an invaluable quantitative tool for selection of MWD breeding stock, but also provide a better understanding of the additive gene effects on intelligence and cognitive functions, as well as defects in these functions (Sarasa, *et al.* 2009; Burghardt, *et al*. 2011). As there are interplays between genetic and environment components in intelligence/cognition, an understanding of how these genes interact with the environment could allow the modulation of environmental factors so that the genetic potentials of MWDs can be maximized. This might ultimately prove that the canine is an ideal model system for the investigation of human performance augmentation, an area of intense AF interest.

1.5 Canine Genome-Wide Single Nucleotide Polymorphism (SNP) Analysis

The completion of the canine genome sequence has resulted in many new genetic markers and thus provided unprecedented opportunities for the identification of genes involved in complex polygenic traits (Ostrander, 2000). The genome-wide scanning approach has many attractive aspects, such as the global assessment of linkage disequilibrium (LD) strength and high resolution mapping of the location of trait-associated loci (Amos 2007; Farrall *et al*. 2005; Pearson *et al.* 2008). Although there are multiple sources of genetic variations in mammalian genomes, single nucleotide polymorphisms (SNPs) have emerged as the marker of choice for whole genome linkage and association studies due to their high abundance, stability, and relative ease of scoring (Ding *et al*. 2009). These attributes make whole-genome SNP typing a powerful technique for conducting GWAS. Most of the SNPs used in GWAS are mapping markers, rather than functional mutations (i.e. they are not causative mutations or genetic variances). Despite this, a GWAS with an adequate genomic coverage will allow the identification of a subset of these SNPs that may be very close, in term of chromosomal distance, to a QTL. The discovery of a SNP associated with the QTL can thus result in an indirect association between the SNP and the trait itself (Sham *et al*. 2009; Almasy, *et al*. 2009). Therefore, association studies based on the underlying principle of LD are significantly facilitated by the whole-genome SNP profiling.

The initial Canine Genome Project produced a high-quality draft of the genomic sequence of a female boxer (Lindblad-Toh, *et al.* 2005). By comparing this genome sequence with that of other breeds, the project successfully compiled a comprehensive set of SNPs applicable to all dog breeds (Wayne, *et al.* 2007, Ostrander, *et al.* 2005). These selected SNP markers are spaced 25,000 to 30,000 base pairs (bp) apart (average distance). While the canine SNP marker set is not as dense as the human counterpart (averaging 3,000 bp in distance), it is, nonetheless, a useful tool for mapping the canine trait-associated loci of interest (Karlsson, *et al.* 2007). High-throughput analysis of genome-wide SNP markers in the canine genome can now be achieved using commercially available SNP microarrays (Butcher *et al.* 2008, Ostrander *et al.* 2005). Two versions of the canine SNP arrays exist. Although they both provide whole-genome coverage, they have significantly different resolution. Version 1 has ~27,000 high quality SNPs, while version 2 contains ~50,000 high-quality SNPs (among a total of 127,132 SNPs per chip). Because of the increased resolution, Version 2 was used in this study. This array is a 5-μm format, perfect match probes only (with 20 probes/SNP) Whole Genome Sampling Assay (WGSA) design. It contains probe sets for a total of ~127K SNPs. These SNPs were chosen from a total of over 2.5 million SNPs generated as part of the canine genome project and include the majority of the "gold" set of the Version 1 array (i.e. 26,625 SNPs derived from a panel of 10 diverse breeds). Similarly, a "platinum" set of 49,633 SNPs has been identified using a panel of 10 diverse breeds in the Version 2 array.

Two different library files can be used with the Version 2 arrays. While the library file ***DogSty06m520431*** will show the results for the full set of the SNPs on the chip (i.e. 127,132 SNPs), the library file ***DogSty06m520431P*** will mask out the SNPs that are not included in the

"platinum" set and thus only shows the results for the 49,633 SNPs that are considered as high-quality. Despite the concern of their annotation quality, some of the SNPs not included in the "platinum" set may in fact be associated with intelligence. Therefore, both library files were used in this study to generate two datasets that were analyzed independently.

One of the factors affecting the power of a genetic study is the information content that can be extracted from the samples. While the physical distance between the QTL and SNP markers is not the only factor that influences the strength of LD, it is still considered a major factor in most cases (Borecki *et al.* 2008, Gu *et al.* 1996). Some studies suggest that a highly dense map with about 500,000 SNP markers spanning the whole genome may be needed for a GWAS to be successful, while others have shown that strong LD can be extended up to 1 centiMorgan (*cM*) (Gu and Rao, 2003) and thus ~30,000 SNPs will probably be sufficient for a genome-wide scan. As the Version 2 of the canine SNP array can provide information content for 50-127K SNPs (depending on the library files used in data processing), high-resolution genome-wide coverage can thus be adequately achieved using the current canine array design.

1.6 Advanced Bioinformatics for Identification of Small-Effect-Size QLTs in GWAS

Since the contribution of each gene (or QTL) to a highly complex polygenic trait like intelligence could be extremely small (e.g. it might be as low as 0.4%), it is therefore necessary to develop a more robust computational method for the analysis of the genome-wide SNP datasets to be generated in this study. To achieve this goal, two different approaches, namely Biologically Guided Selection and Computational Based Feature Synthesis and Classification, were pursued in parallel. Techniques based on feature synthesis using genetic algorithm were explored. Initially, low dimensional feature vectors were synthesized from the original genotyping dataset that has high dimensional feature vectors using co-evolutionary genetic programming (CGP). The synthesized features were obtained by applying a series of operators (composite operator vectors) to the original features. These operators are binary trees with simple operators as the inner nodes and the original features as the leaf nodes. First, the internal nodes of the tree representing the composite operator were randomly determined in a recursive manner. After all the internal nodes are generated, the original features were randomly picked and attached to the leaf nodes. The genetic programming operations were then applied to the binary trees in the order of crossover, mutation and selection. In addition, an elitism replacement method was adopted to keep the best composite operator, in terms of classification accuracy, from generation to generation.

The classification accuracy of a Bayesian classifier in the synthesized, low-dimension feature space was used to assess the fitness of the synthesized features, as assessed by classification accuracy. The best-fitted synthesized features were generated using the CGP algorithm through the iteration of the mutation-selection process. To train the algorithm, CGP was used to run the training data and evolve through the mutation-selection process to select the best composite

operator based on the Bayesian classifier in the synthesized feature space. In the testing phase, the synthesized features were generated by applying the composite operator vector to the original features of the testing samples, and the Bayesian classifier used for the classification of the test samples.

As the first step of the development of this methodology, we analyzed the whole genome SNP profiles of 117 dogs from three breeds (German Shepherd Dog, Belgian Malinois, and Labrador Retriever) using this approach. We were able to classify these dogs into three groups, one for each breed, with 89 – 100% accuracy. The high degree of accuracy of this classification technique in clustering these canine subjects into their corresponding breeds in an unsupervised manner strongly suggests that this algorithm can be further developed and optimized for the analysis of complex traits such as intelligence. The details of the development of this algorithm are described in the Technical Report AFRL-RH-WP-TR-2011-0081 entitled: "*Development of Advanced Classification Algorithm for Genome-Wide Single Nucleotide Polymorphism (SNP) Data Analysis*".

2. MATERIALS AND METHODS

All studies involving animals were approved by the Wright-Patterson Institutional Animal Care and Use Committee, and were conducted in a facility accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care, International, in accordance with the Guide for the Care and Use of Laboratory Animals, National Research Council (1996). Studies were conducted under approved Air Force Research Laboratory Institutional Animal Care and Use Committee Protocol AFDR-2009-002A "Genome-wide Association Mapping for Superior Intelligence in Military Working Dogs" (University of PA Protocol #802551). All test equipment was carefully designed and prototyped to minimize risk of injury to the animals, and no injuries were reported during the course of this study.

2.1 Canine Cohort

In this pilot study, dogs already working or in advanced training were used. These dogs were mostly owned by three private US government contractor facilities or working dog breeders. All subjects tested could detect some sort of substance, and some of them could perform other tasks (e.g. patrolling) as well. Many of these dogs have completed the training and deployed in the theater of operations after the participation in this study.

Although permission to test DoD MWDs at the 341st Training Squadron, Lackland AFB, and the Army Special Operations Command (SOCOM) Ranger dogs has been received, these permissions were granted after the project was well under way. Therefore, no DoD MWDs were used in the study reported here. In fact, testing DoD MWDs was not the goal of this pilot study, which was clearly stated in the DARPA-approved proposal.

2.2 Behavioral Testing of Canine Subjects

2.2.1 Design and Construction of Test Equipment. To conduct the behavioral tests of the canine subjects, three devices as described below were designed by Dr. Overall and constructed:

a. Puzzle Box - for the assessment of problem solving ability and/or boldness;

b. Angled Fence around which dogs must detour to get the item they wish (or are supposed) to obtain - for the assessment of problem solving ability and/or boldness; and

c. Reward Box where dogs must push a lever to get the reward - for the assessment of observational learning and following command.

The design of devices requires careful consideration of many facets of animal safety and ease of transportation/shipment. In addition, these devices have to be able to withstand the abuse by claws/teeth of large powerful dogs. Consequently, expensive materials like "bullet-proof glass" (polycarbonate thermoplastic) were used to build these devices.

Prototypes were developed and completed for the 'Puzzle Box' and 'Angled Fence'. Behavioral tests using the 'Puzzle Box' have been conducted and subsequently validated. Due to premature termination of funding by DARPA, the 'Angle Fence' was prototyped and initial behavior tests were conducted, but its use was not validated. The lack of funds prevented prototyping of the 'Reward Box'.

2.2.2 Canine Intelligence Behavioral Tests Regimen. The CITP specifically developed for this study consists of 11 behavioral tests for attentiveness, novelty, interest, signaling/showing, observational learning/showing, problem solving/boldness and handedness. The tests are described below (a more in-depth description of the CITP regimen and the analysis of the behavioral tests data will be described in a separate report).

Attentiveness I, II

These tests examine a set of command responses given by either the Handler or a Tester (unknown person). Data is collected on latency to response, time needed to address the commands, attention, posture, and other behaviors of subject. For the Attentiveness II Test, the Handler or Tester moves a novel object. Data is collected on latency to response, actual response, attention, posture, and other behaviors of subject.

Novelty

This test examines the animal response to novel objects. The tester will collect data on latency to response, number of boxes checked, order of boxes checked, total time needed to check all five boxes, posture, and other behaviors of subject.

Interest I, II, and III

These tests examine subject's response to familiar objects. Data will be collected on latency to response, time needed to retrieve the objects, posture, and other behaviors of subject. Interest II Test is similar to Interest I test, except that it uses additional objects. Interest III Test is similar to Interest II Test, except that some objects are visually marked. Tester collects data on latency to response, time needed to retrieve the objects, number of objects checked, posture, and other behaviors of subject.

Signaling/Showing

In this test, the position of a hidden object is indicated to the dog by the Tester. Data is collected on latency to response, time needed to retrieve the object, number of mistakes (checking incorrect locations), posture, and other behaviors of subject.

Observational learning

This test requires the use of the 'Reward Box'. Object is placed in the box, which has a lever that can open one end of the box. Tester demonstrates correct retrieval method to the dog. Data is collected on latency to response, time needed to retrieve the object, posture, and other behaviors of subject.

Problem solving/Boldness I, II

The Problem solving/boldness I Test requires the use of the 'Puzzle Box'. Object is placed in the center of a clear box with several openings. Dog must move the object to a larger hole at one end of the box in order to successfully retrieve the object. Data is collected on latency to response, time needed to retrieve the ball, posture, and other behaviors of subject. The Problem solving/boldness II Test requires the use of the 'Angled Fence', a clear barrier with small holes every 3-6 inches so the dog can detect object odor through the holes. An object is placed on one side of the barrier, while the dog is located on the other side. Data is collected on latency to response, time needed to retrieve the treat, posture, and other behaviors of subject.

Handedness/Brain lateralization Test

The handedness of the dog is determined using the number of times a particular hand (paw) is manipulating an object. Data is collected on number of times the dog touches the object with the right paw verses the left paw.

All tests in the CITP regimen were videotaped for data analysis by a trained canine behavior expert not involved with the on-site testing (to eliminate operator bias/error). All test segments for each individual dog were compiled into a single video file (CITP video). The video file for each individual dog was converted from AVI to MPEG-2 format and recorded onto a DVD for long-term storage/archives.

2.3 Blood Sample Collection

A blood sample was collected by a licensed veterinarian from each dog after completion of the behavioral testing for conducting genome-wide SNP typing. Briefly, a total of 5-6 ml of blood was obtained from each tested subject via venipuncture of the cephalic vein and collected in EDTA-coated vaccutainer tubes. The blood samples were stored at 4 ^{o}C prior to shipment to AFRL/RHPB. Samples were maintained at 4 ^{o}C with ice packs during shipment.

2.4 Genomic DNA Isolation from Blood Samples

High-molecular-weight genomic DNA was extracted from blood leukocytes using the Qiagen QIAampR DNA Blood Midi Kit, as recommended by the manufacturer. Briefly, blood samples were added to the QIAGEN Protease in a 15-ml centrifuge tube. Lysis buffer was then added to each sample, followed by thorough mixing for at least 1 minute. The mixture was then incubated at 70 °C for 10 minutes. Ethanol (100%) was added to each sample, followed by thorough mixing. One half of the supernatant of each sample was then added onto a QIAamp Midi column (placed in a 15 ml centrifuge tube), and the samples centrifuged at 1,850 x g for 3 minutes. After the removal of the filtrate, the remaining half of the supernatant samples was loaded onto the QIAamp Midi column, and the centrifugation step was repeated. The bound DNA was washed using the washing buffers AW1 and AW2. High-molecular weight genomic DNA was subsequently recovered using the elution buffer AE. The purified DNA samples were stored in small aliquots at –20 ºC until being processed for target preparation.

2.5 Target Preparation, Chip Hybridization and Detection

The genomic DNA samples were first diluted to 50 ng/µL, using the reduced EDTA-TE buffer in a 96-well reaction plate. Restriction digestion of the DNA samples with Sty I was initiated by the addition of 14.75 µL Digestion Master Mix to each sample to produce a final volume of 20 µL containing 250 ng genomic DNA, 2 µg BSA and 1 unit Sty I in 1x restriction digestion buffer (NE Buffer #3: 50 mM Tris-HCl, 100 mM NaCl, 10 mM $MgCl_2$ and 1 mM dithiothreitol). The digestion mix was incubated at 37 ºC for 2 hours in a thermal cycler. Once the digestion was completed, the enzyme was inactivated by heating at 65 ºC for 20 minutes. Ligation was initiated by the addition of ligation mix containing DNA ligase and the Sty adaptors to the digested DNA samples. After incubating at 16 ºC for 3 hours, the reaction mix was heated to 70 ºC for 20 minutes to inactivate the DNA ligase. The ligation products were then diluted 4-fold in AccuGENE® water (Affymetrix) to yield a final volume of 100 µL.

A 10 µL aliquot of the ligation product from each sample was transferred to the corresponding well of a 96-well reaction plate, followed by the addition of the polymerase chain reaction (PCR) Master Mix (90 µL/sample) to produce a final volume of 100 µL containing 0.1 mmol GC-Melt, dNTPs (0.035 µmol each), 0.45 nmol PCR Primer #002 and 2 µL Titanium Taq DNA Polymerase (50x stock) in 1x Titanium Taq Buffer. PCR was carried out using the following setting:

a. 94 ºC for 3 minutes (1 cycle);

b. 94 ºC for 30 sec → 60 ºC for 45 sec → 68 ºC for 15 sec (30 cycles);

c. 68 ºC for 7 minutes (1 cycle); and

d. 4 ºC → HOLD

After the PCR was completed, the reaction plate was centrifuged at 2,000 rpm for 30 seconds to recover the condensates. The PCR products (3μL/sample) were analyzed using gel electrophoresis (2% agarose in TBE buffer). In general, this procedure produced PCR products of fragment size ranging from 250 – 1,100 bp.

The PCR products were purified using the Clontech Clean-Up Plate according to the procedure recommended by the manufacturer with three washes using AccuGENE® water, followed by the elution of the PCR products using RB Buffer. The concentration of the purified PCR products was determined by measuring its optical density (OD) at 260 nm (OD_{260}). Three dilutions for each PCR product were made and quantified independently. The average of the OD measurements for each sample was calculated and used as the final concentration. Once the concentrations of the samples were determined, they were diluted to 2 μg/μL in RB Buffer.

The purified, normalized PCR products were treated with Fragmentation Reagent at 37 ºC for 35 minutes, followed by heating at 95 ºC for 15 minutes. The size of the fragmented PCR products was determined using gel electrophoresis (4% agarose in TBE buffer). In general, the average fragment size of the PCR products was reduced to less than 180 bp after this step. The fragmented targets were labeled using the GeneChip® DNA Labeling Reagent (from Affymetrix) according to the Affymetrix Human Mapping 500K Array Technical Manual. Briefly, 19.5 μL of Labeling Master Mix was added to each sample, and the reaction mix was incubated at 37 ºC for 4 hours, followed by incubation at 95 ºC for 15 minutes. The labeled target for each sample was first mixed with 190 μL of hybridization master mix, and the resulting mix was denatured at 95 ºC for 10 minutes and kept at 49 ºC until use. The denatured target was then loaded onto a Canine SNP Array v2. The arrays (with hybridization cocktail loaded) were placed into a preheated hybridization oven and allowed to hybridize at 49 ºC for 18 hours.

After hybridization, the hybridization cocktail was removed from each chip and transferred to a tube. Array Holding Buffer was then added to each array. The washing, staining, and scanning of the hybridized arrays were performed using the Affymetrix Fluidics Station 450 and the GeneChip Scanner 3000 7G following the Affymetrix Human Mapping 500K Array Technical Manual.

2.6 Canine SNP Array Data Processing

Data processing was performed using the snp5 command line software downloaded from Affymetrix to make the genotype calls. Initially, a QC analysis was performed to assess the data quality. The information in the Intensity QC Table indicated the overall performance of the chip analysis. When all steps of the assay are working as expected, the QC call rate is typically ≥75% for the entire collection of 127K SNPs and ≥85% for the "platinum" set of SNPs. As described in section 1.1, both library files (***DogSty06m520431*** and ***DogSty06m520431P***) were used so that two datasets consisting of 127K SNPs or 50K "platinum" SNPs were generated for downstream data analysis. Initially, Dynamic Model algorithm was used to perform QC analysis on individual arrays. Once completed, genotype calls of the SNPs were determined using the Bayesian Robust Linear Model with Mahalanobis distance classifier (BRLMM) algorithm batch analysis tool (Miclaus *et al.* 2010, Hong *et al.* 2010, Hoggart *et al.* 2003).

In this study, a total of 117 canine subjects were genotyped using the Affymetrix canine SNP array version 2.0 in three batches. The SNP array datasets were processed using two different approaches:

i. **DP Method 1:** Each SNP array dataset was processed separately to generate the genotype calls, and the processed datasets were combined into a single large dataset.

ii. **DP Method 2:** The three SNP array datasets were combined into one large dataset, and the resultant dataset was processed to generate the genotype calls.

2.7 Unsupervised Breed Assignment Clustering Analysis

2.7.1 Clustering Analysis Steps. The clustering analysis pipeline consists of the following five steps:

a. Data cleanup;
b. Creation of a distance matrix;
c. Assign initial clusters based on the genotype call distance matrix;
d. Merge clusters with smallest genotype call distance; and
e. Construction of a hierarchical cluster containing all subjects.

2.7.2 Data Cleanup. To ensure data quality, a three-step filtering process was developed to filter out low-quality SNPs (and samples) prior to downstream data analysis (Lander *et al.* 1995). In the first filter, samples with an overall call rate of <75% will be excluded from the dataset. The filtered sample set was then subjected to the second data filter. Any SNP with <90% call rate across all the samples will be eliminated from subsequent data analysis. Following these two filtering steps, the final call rate of the remaining samples/SNPs will be examined, and samples with a call rate <95% will be excluded from the dataset. We reasoned that this data

cleanup procedure is especially important when the full set of 127K SNPs datasets are used since some SNPs in the full set are expected to be of suboptimal quality.

Before the implementation of this 3-step data cleanup procedure, two simple methods to handle missing data (no calls) were tested:

i. Removed all SNPs with any missing data points – this filter resulted in the removal of ~80% of the SNPs; and

ii. No data cleanup – the data was coded so that the metric for comparing how the two SNPs are related can account for the missing data.

It was decided that if this simple "all or none" approach failed to generate acceptable clustering results, the more sophisticated 3-step data cleanup procedure as described above will be implemented.

These datasets (with or without) data cleanup, were then used as input data for the development and validation of the advanced clustering techniques. The primary goal of the analysis was to develop a clustering technique that can separate dogs by breed, solely based on two pieces of information, the SNP profiles and the fact that there are three breeds in the population. Neither the information concerning the number of dogs in each breed, nor information on any breed-specific SNPs was used as input data. The secondary goal was to evaluate how data processing, data cleanup and SNP annotation quality may affect the final clustering result.

2.7.3 Creation of Genotype Call Distance Matrix. The distance matrix was generated using the following steps:

i. Compare the genotype of each SNP of all sample pairs and numerically code the distance of each pair-wise comparison:
   a. Distance = 0, if both alleles are the same
   b. Distance = 1, if only one allele is the same (for example, the genotype of a subject is AA or BB, while that of the other subject is AB)
   c. Distance = 2, if no allele is the same (for example, the genotype of a subject is AA, while that of the other subject is BB)
   d. Distance = N/A, if there is a no call (i.e. missing data) in one sample (or in both samples).

ii. Summarize the distance of all pair-wise comparison for all samples.

2.7.4 Development of Unsupervised Clustering Algorithm. The algorithm used for unsupervised breed assignment analysis was based on the hierarchical clustering technique of the Ward's algorithm for the calculation of the distance-based group assignment (Ward, *et al.* 1961). The analysis started with 117 clusters, each cluster containing only one canine subject. The

algorithm then identified the closest pair of clusters and merged them into one single cluster. The distances between the new cluster and all other clusters were then re-calculated, and the closest pair of clusters identified and merged. This process was reiterated until all the samples were merged in one single cluster. The distance from the root was selected to result in three separate clusters. The members in each of these clusters and the breed they belong to were identified.

3. RESULTS

3.1 Canine Cohort

In this study, a total of 199 canine subjects were recruited. Table 1 shows the entire list of all recruited subjects. Blood samples have been collected from all recruited subjects and shipped to AFRL Applied Biotechnology Branch for genome-wide SNP analysis.

Table 1: Compiled List of Subjects in the Cohort

Subject ID	Name	Gender	Breed	Behavioral Testing
U1	Slick	M	BOC	No
U2	Cody	MC	AUS	No
U3	Rocky	M	BOC	No
U4	Maddie	F	BOC	No
U5	Isidor	M	BDF	No
U6	Oya	FS	BDF	No
U7	Jessie Lynn	F	BOC	No
U8	Ricochet	F	BOC	No
U9	Thunder	M	GSD	No
U10	Hannah	F	BOC	No
U11	Dell	F	BOC	No
U12	Rhys	F	BOC	No
U13	Rivet	F	PRT	No
U14	Hillary	F	BOC	No
U15	Joyce	F	BOC	No
U16	Pepper	F	BOC	No
U17	Hawke	M	BOC	No
U18	Mac	MC	BOC	No
U19	Jan	FS	BOC	No
U20	Vegas	MC	AUS	No
U21	Sting	MC	AUS	No
U22	Opus	M	AUS	No
U23	Melica	F	AUS	No
U24	Kelly	FS	BOC	No

U25	Bouquet	F	AUS	No
U26	Cody	MC	AUS	No
U27	Breyer	MC	AUS	No
U28	Burdock	MC	AUS	No
U29	Orso	MC	AUS	No
U30	Colt	M	AUS	No
U31	Slinger	M	AUS	No
U32	Story	F	AUS	No
U33	Bounce	F	AUS	No
U34	Asa	M	AUS	No
U35	Riot	FS	AUS	No
U36	Chill/Chiel	M	AUS	No
U37	Numi	M	AUS	No
U38	Victoria	F	AUS	No
U39	Ivy	F	AUS	No
U40	Jackson	M	AUS	No
U41	Dolce	FS	AUS	No
U42	Oz	MC	AUS	No
U43	Baker	M	AUS	No
U44	Sydney	FS	MAL	No
U45	Hunter	MC	MAL	No
U46	Charlie	M	AUS	No
U47	Echo	M	LAB	Yes
U48	Balu	M	AUS/BOC	Yes
U49	King	M	LAB	Yes
U50	Karma	F	LAB	Yes
U51	Ben	M	LAB	Yes
U52	Johnny	MC	LAB	Yes
U53	Kira	F	MAL	Yes
U54	Mika	F	GSD	Yes
U55	Richa	F	MAL	Yes
U56	Elli	F	GSD	Yes
U57	Keno	M	GSD	Yes
U58	Brandy	F	LAB	Yes
U59	Tuky	M	GSD	Yes
U60	Chilli	M	MAL	Yes
U61	Hina	F	MAL	Yes
U62	Crogan	M	MAL	Yes
U63	Daryl	M	LAB	Yes
U64	Stevie	F	GR	Yes
U65	Cyna	F	MAL	Yes

U66	Sara	F	LAB	Yes
U67	Lady	F	LAB	Yes
U68	Hatos	M	GSD	Yes
U69	Bella	F	MAL	Yes
U70	Natalie	F	LAB	Yes
U71	Lobo	M	LAB	Yes
U72	Nova	F	LAB	Yes
U73	Rollo	F	GR	Yes
U74	Ringo	F	LAB	Yes
U75	Lucy	F	LAB	Yes
U76	Kaia	FI	LAB	Yes
U77	Woody	MI	LAB	Yes
U78	Casper	MI	LAB	Yes
U79	Szandi	FI	GSD	Yes
U80	Rony	MI	GSD	Yes
U81	Toni	MI	GSD	Yes
U82	Lola	FS	MAL	Yes
U83	Denny	MI	LAB	Yes
U84	Werci	MI	GSD	Yes
U85	Roppi	MI	GSD	Yes
U86	Amanda	FI	LAB	Yes
U87	Toti	MI	GSD	Yes
U88	Mickey (aka Rex)	MI	GSD	Yes
U89	Krisz	MI	GSD	Yes
U90	Lacey	FS	BEL	Yes
U91	Dark	MI	GSD	Yes
U92	Linda	FI	GSD	Yes
U93	Fritz	MC	LAB	Yes
U94	Lucky 6	FI	GR	Yes
U95	Santos I	MI	GSD	Yes
U96	Arco 13	MI	MAL	Yes
U97	Bieke I	FI	MAL	Yes
U98	Brenda II	FI	GSD	Yes
U99	Goliath	MC	PRT	Yes
U100	Bonsai	MI	GSD	Yes
U101	Flem	MI	MAL	Yes
U102	Hanna	FI	GSD	Yes
U103	Igan	MI	GSD	Yes
U104	Dasty	MI	GSD	Yes
U105	Lousie	FI	GSP	Yes
U106	Charon	MI	GSD	Yes

U107	Epos	MI	MAL	Yes
U108	Ado	MI	GSD	Yes
U109	Tank	MC	AST	Yes
U110	Nestor	MI	GSD	Yes
U111	Zorba	MI	LAB	Yes
U112	Bubi	MI	GSD	Yes
U113	Bax	MI	GSD	Yes
U114	Mali	MI	MAL	Yes
U115	Csoki	MI	GSD	Yes
U116	Gack	MI	GSD	Yes
U117	Roy	MI	GSD	Yes
U118	Tito	MI	GSD	Yes
U119	Nick	MI	GSD	Yes
U120	Bebop	F	AUS	Yes
U121	Story	F	AUS	Yes
U122	Sarah	F	AUS	Yes
U123	Louie	M	AUS	Yes
U124	Lola	F	AUS	Yes
U125	Spell	F	AUS	Yes
U126	Lock	M	AUS	Yes
U127	Lock & Bunny	M	AUS	Yes
U128	Nova	F	AUS	Yes
U129	Arson	M	AUS	Yes
U130	Roper	M	BOC	Yes
U131	Shine	F	AUS	Yes
U132	Sprite	F	AUS	Yes
U133	Ben	M	AUS	Yes
U134	Reba	F	AUS	Yes
U135	Flash	F	AUS	Yes
U136	Mo	M	AUS	Yes
U137	Pilot	M	AUS	Yes
U138	Dan	M	AUS	Yes
U139	Foxy	F	AUS	Yes
U140	Opal	F	AUS	Yes
U141	Peggs	F	AUS	Yes
U142	Taxi	F	AUS	Yes
U143	Riso	MI	MAL	Yes
U144	Szarik	MI	GSD	Yes
U145	Astor	MI	MAL	Yes
U146	Roy	MC	MAL	Yes
U147	Pluto	MI	MAL	Yes

U148	Houden	MI	MAL	Yes
U149	Aspi	MI	MAL	Yes
U150	Roy 2	MI	MAL	Yes
U151	Ana	FS	GSD	Yes
U152	Ben	MI	LAB	Yes
U153	Cora	FS	MAL	Yes
U154	Bona	FS	GSD	Yes
U155	Yana	FS	MAL	Yes
U156	Kim	FS	MAL	Yes
U157	Chester	MI	MAL	Yes
U158	Sjonnie	MI	GSD	Yes
U159	Kejsi	FS	MAL	Yes
U160	Lana	FS	MAL	Yes
U161	Tiger	MI	MAL	Yes
U162	Jara	FS	MAL	Yes
U163	Bajdy	MI	GSD	Yes
U164	Simba	FS	GSD	Yes
U165	Tiki	FS	AUS X	Yes
U166	Madison	FS	LAB X*	Yes
U167	Oliver	MC	LAB X*	Yes
U168	Shadow	MC	BOC X	Yes
U169	Dublin	FS	GSD	Yes
U170	Keegan	MC	BOC	Yes
U171	Rumble	MC	BOC	Yes
U172	Focus	MC	BOC	Yes
U173	Ben	MC	PWC	Yes
U174	Akiva	MI	GSD	Yes
U175	Roscoe	MC	LAB	Yes
U176	Zoomie	MC	BOC	Yes
U177	Stevie	MC	BOC	Yes
U178	Peyton	MC	CBR	Yes
U179	Tic Tac	MC	BOC	Yes
U180	Koda	MC	LAB X*	Yes
U181	Kelly	FS	MAL	Yes
U182	Lucy	FS	MAL	Yes
U183	Dany	MI	GSD	Yes
U184	Brit	MI	GSD	Yes
U185	Bouc	MI	MAL	Yes
U186	George	MI	LABX	Yes
U187	Jimmy	MI	LAB	Yes
U188	Palmito	MI	LAB	Yes

U189	Jake	MI	LAB	Yes
U190	Senta	F?	MAL	Yes
U191	Mimo	MC	SS	Yes
U192	Tosca	FS	MAL	Yes
U193	Robby	MI	MAL	Yes
U194	Willy	MI	GSDX	Yes
U195	Fero	MI	MAL	Yes
U196	Hannah	FS	LAB	Yes
U197	Egy	MI	GSD	Yes
U198	Bona II	FS	MAL	Yes
U199	Bonzo	MI	GSD	Yes

Legends:

a. Breed Abbreviations:

AST = American Staffordshire terrier
AUS = Australian shepherd
AUS X = Australian shepherd cross
BDF = Bouvier des Flandres
BEL = Belgian shepherd
BOC = Border collie
BOC X = Border collie mix
CBR = Chesapeake bay retriever
GR = Golden Retriever
GSD = German shepherd dog
GSDX = German shepherd dog cross
GSP = German shorthair pointer
LAB = Labrador retriever
LAB X* = Labradoodle (Labrador retriever x Poodle)
LABX = Labrador retriever cross
MAL = Malinois
PRT = Parson Russell Terrier
PWC = Pembroke Welsh corgi
SS = Springer Spaniel

b. Gender Abbreviations:

F or FI = female intact
FS = female spayed
M or MI = male intact
MC = male castrated
GTA = Global Training Academy, TX

3.2 Assessment of Canine Intelligence

To quantitatively and reliably evaluate the attentiveness, interest in novelty exploration, response to signaling and showing, observational learning, problem solving/boldness, and handedness of the canine subjects, we have developed the CITP, which consists of 11 behavioral tests (for details, see Materials and Methods).

Of the 199 dogs recruited in this study, a total of 153 dogs have been tested using the CITP. Due to premature termination of funding by DARPA, analysis of this behavioral testing dataset was not completed. However, the data of a subset of 108 dogs was partially analyzed. Subjects in this subpopulation are mostly from three breeds (see **Table 2**). Their age ranged from 1 to 10 years old, with the average age of 28 months (most were 2-5 years in age).

Table 2: Number of Canine Subjects with Behavioral Data Analyzed

Breed	Total Tested	Number Analyzed
German Shepherd (GSD)	47 (+ 1 GSD cross)	45
Belgian Malinois (MAL)	44	33
Labrador Retriever (LAB)	26 (+1 LAB cross)	22
Miscellaneous breeds	8	8
TOTAL	127	108

Empirical evaluation of the overall performance of these dogs allowed the identification of the overall top 25 and bottom 25 performers (**Table 3**). Pair-wise comparisons revealed that there is no statistically significant difference between the breeds with respect to the number of top or bottom performers. However, the result of statistical analysis did suggest that one of the kennels tested had significantly more top performers, whereas the other had significantly more bottom performers ($p<0.05$, G-test). The molecular basis for such observation is currently unclear. Should such difference be confirmed to be genetically related, the canine cohort described here could be proven to be an invaluable resource for the identification of gene loci contributing to canine intelligence.

Table 3: Numbers of Top and Bottom Performers in Each Breed

Breed	# Tested	# Top Performers	# Bottom Performers
German Shepherd dog	47	8 (17%)	13 (28%)
Belgian Malinois	44	11 (25%)	5 (11%)
Labrador Retriever	26	4 (15%)	6 (23%)

3.3 Genome-Wide Single Nucleotide Polymorphism Typing of Canine Subjects

Blood samples collected from the canine subjects that have been phenotypically tested were processed for genomic DNA extraction. A subpopulation of 117 dogs (see **Table 3**) with their behavioral tests data evaluated were selected for whole genome single nucleotide polymorphism (WG SNP) typing using the Affymetrix canine SNP Array v2. The ID and the breed of these canine subjects selected for this analysis are shown in **Table 4**.

Table 4: Subject ID and Breed of Canine Subjects Selected for WG SNP Typing

Subject ID	Breed
U47	Labrador Retriever
U49	Labrador Retriever
U50	Labrador Retriever
U51	Labrador Retriever
U52	Labrador Retriever
U53	Belgian Malinois
U54	German Shepherd
U55	Belgian Malinois
U56	German Shepherd
U57	German Shepherd
U58	Labrador Retriever
U59	German Shepherd
U60	Belgian Malinois
U61	Belgian Malinois
U62	Belgian Malinois
U63	Labrador Retriever
U65	Belgian Malinois
U66	Labrador Retriever
U67	Labrador Retriever
U68	German Shepherd
U69	Belgian Malinois
U70	Labrador Retriever
U71	Labrador Retriever
U72	Labrador Retriever
U74	Labrador Retriever
U75	Labrador Retriever
U76	Labrador Retriever
U77	Labrador Retriever

U78	Labrador Retriever
U79	German Shepherd
U80	German Shepherd
U81	German Shepherd
U82	Belgian Malinois
U83	Labrador Retriever
U84	German Shepherd
U85	German Shepherd
U86	Labrador Retriever
U87	German Shepherd
U88	German Shepherd
U89	German Shepherd
U91	German Shepherd
U92	German Shepherd
U93	Labrador Retriever
U95	German Shepherd
U96	Belgian Malinois
U97	Belgian Malinois
U98	German Shepherd
U100	German Shepherd
U101	Belgian Malinois
U102	German Shepherd
U103	German Shepherd
U104	German Shepherd
U106	German Shepherd
U107	Belgian Malinois
U108	German Shepherd
U110	German Shepherd
U111	Labrador Retriever
U112	German Shepherd
U113	German Shepherd
U114	Belgian Malinois
U115	German Shepherd
U116	German Shepherd
U117	German Shepherd
U118	German Shepherd
U119	German Shepherd
U143	Belgian Malinois
U144	German Shepherd
U145	Belgian Malinois
U146	Belgian Malinois

U147	Belgian Malinois
U148	Belgian Malinois
U149	Belgian Malinois
U150	Belgian Malinois
U151	German Shepherd
U152	Labrador Retriever
U153	Belgian Malinois
U154	German Shepherd
U155	Belgian Malinois
U156	Belgian Malinois
U157	Belgian Malinois
U158	German Shepherd
U159	Belgian Malinois
U160	Belgian Malinois
U161	Belgian Malinois
U162	Belgian Malinois
U163	German Shepherd
U164	German Shepherd
U181	Belgian Malinois
U182	Belgian Malinois
U183	German Shepherd
U184	German Shepherd
U185	Belgian Malinois
U187	Labrador Retriever
U188	Labrador Retriever
U189	Labrador Retriever
U190	Belgian Malinois
U192	Belgian Malinois
U193	Belgian Malinois
U195	Belgian Malinois
U196	Labrador Retriever
U197	German Shepherd
U198	Belgian Malinois
U199	German Shepherd
U200	German Shepherd
U201	German Shepherd
U202	German Shepherd
U203	Belgian Malinois
U204	Belgian Malinois
U205	Belgian Malinois
U206	Belgian Malinois

U207	Belgian Malinois
U208	German Shepherd
U209	Belgian Malinois
U210	Belgian Malinois
U211	German Shepherd
U212	German Shepherd
U213	Belgian Malinois

3.4 Characteristics of the SNP datasets

The SNP array datasets generated were processed using two different methods. In the first method, each SNP array dataset was processed separately to generate the genotype calls, and the processed datasets were combined into a single large dataset (i.e. Process-Merge Method). The resulting SNP datasets are designated as A+B+C_Full Set or A+B+C_Platinum Set (Table 5), dependent on the library files used. Due to the nature of this approach, it is anticipated that a significant portion of the batch effect generated during microarray analysis will remain. In the second method, the three SNP array datasets were combined into one large dataset, and the resultant dataset was processed to generate the genotype calls (i.e. Merge-Process Method). SNP datasets, generated using this method, are designated as ABC_Full Set or ABC_Platinum Set in Table 5, dependent on the library files used. Compared to the Process-Merge method described above, the Merge-Process method can effectively reduce the batch effect.

The resultant datasets, regardless the data processing methods used, thus contained the genotype calls of all interrogated SNPs (i.e. 127,132 SNPs, distributed across the entire canine genome) of 117 dogs belonging to three breeds. Additionally, datasets containing the genotype calls of a subset of these SNPs (a total of 49,663 SNPs) that represent the high-quality SNP set were also generated using the Platinum Set library file.

Table 5 shows the number (and percentage) of subjects, as well as SNPs with specific call rates in the four datasets generated using different data processing methods and library files. Comparing the two data processing methods, the Process-Merge Method appeared to produce a significantly better call rate in subjects, and a slightly better call rate in SNPs for the full set. However, a completely opposite result was observed when the platinum set library file was used: the Merge-Process Method produced a significantly better call rate in subjects and SNPs. Although the exact reason for this observation is not clear, this result thus suggested that the data processing method has differential influences on the call rate of the SNPs, which in turn depends on the quality of the SNPs.

Table 5: Number and Percentage of Subjects and SNPs with Specific Call Rates

Call Rate	A+B+C (Full Set)		A+B+C (Platinum Set)		ABC (Full Set)		ABC (Platinum Set)	
	Subject (%)	SNP (%)	Subject (%)	SNP (%)	Subject (%)	SNP (%)	Subject (%)	SNP (%)
100%	0 (0)	25123 (19.76)	0 (0)	12841 (25.86)	0 (0)	24234 (19.06)	0 (0)	14118 (28.43)
90% - 99.9%	0 (0)	47479 (37.35)	85 (72.65)	23775 (47.87)	0 (0)	46202 (36.34)	102 (87.18)	26786 (53.94)
85% - 89.9%	80 (68.38)	10314 (8.11)	24 (20.51)	3075 (6.19)	4 (3.42)	7856 (6.18)	6 (5.13)	2556 (5.15)
80% - 84.9%	35 (29.91)	8812 (6.93)	8 (6.84)	2336 (4.7)	68 (58.12)	6051 (4.76)	9 (7.69)	1530 (3.08)
70% - 79.9%	2 (1.71)	14184 (11.16)	0 (0)	3673 (7.4)	45 (38.46)	9473 (7.45)	0 (0)	1634 (3.29)
<70%	0 (0)	21220 (16.69)	0 (0)	3963 (7.98)	0 (0)	33316 (26,21)	0 (0)	3039 (6.12)
Total	117 (100)	127132 (100)	117 (100)	49663 (100)	117 (100)	127132 (100)	117 (100)	49663 (100)

3.5 Unsupervised Classification Algorithm for Breed Assignment

Due to lack of funding, behavioral testing of the subjects in the cohort was only partially completed. More importantly, the phenotype analysis of the behavioral tests data which was acquired could not be accomplished. Consequently, analysis of the genome-wide SNP typing datasets using traditional statistical methods was not possible. Under these circumstances it was decided that the aim of the study for the remaining time should focus on the development of advanced algorithms which would be robust enough for unsupervised analysis of genome-wide SNP typing datasets. Although this is a highly risky approach, success in such an attempt would have a far-reaching impact not only on the genetic analysis of canine intelligence, but also on data mining of genetic studies in general, and especially GWAS.

As a proof-of-concept, a classification analysis of the WG SNP typing dataset of a subpopulation of canine subjects (see Table 4) was conducted. The primary goal of the analysis is the separation of the dogs by breed analyzing the data in an unsupervised manner. Therefore, only two pieces of information were used: the genome-wide SNP profiles and the three subgroups (i.e. three canine breeds) in the population. Note that the number of dogs in each breed was *NOT* used as input data in the analysis nor was any information concerning potential breed-specific SNPs.

Initially the distance between all sample pairs based on the similarity/difference in the genotype calls was calculated for all SNPs. The result was then summarized as a distance matrix. The unsupervised breed assignment was achieved using a variant of hierarchical clustering algorithm for the calculation of the distance-based group assignment (Ward, *et al.* 1961). The analysis

starts with each dog in a separate cluster. The algorithm then identifies the closest pair of clusters and merges them into one single cluster. The distances between the new cluster and all other clusters are then re-calculated, and the closest pair of clusters identified and merged. This process is reiterated until all the samples are merged in one single hierarchical cluster. The distance from the root is selected to result in three separate clusters.

Of the three clusters generated, Cluster #1 closely resembled the breed of Belgian Malinois, while Clusters #2 and #3 resembled the breeds of Labrador Retriever and German Shepherd Dog, respectively. The algorithm developed can cluster the dogs of the Belgian Malinois breed (44 dogs) with an accuracy >90%. The result of Cluster #2 showed that all Labrador Retriever dogs were clustered into one group with 100% accuracy. As with the clustering results of Belgian Malinois and Labrador Retriever, this algorithm can cluster the German Shepherd Dog with an accuracy close to 90%. Interestingly, the data process method, the annotation quality of the SNP, and the data cleanup method seemed to have only a minor effect on the accuracy of the clustering results. The details of the algorithm and the classification results have been previously reported (Technical Report AFRL-RH-WP-TR-2011-0081 "*Development of Advanced Classification Algorithm for Genome-Wide Single Nucleotide Polymorphism (SNP) Data Analysis*").

4. SUMMARY AND CONCLUSIONS

This study was designed to genetically map superior intelligence in the military working dog population. Despite the challenges and drawbacks that have been encountered during the course of this research (for instance, less than half of the approved budget was received from DARPA), a number of significant milestones were achieved:

1. Recruitment of 199 canine subjects for this study and collection of blood samples from all recruited subjects.

2. Development and partial validation of the CITP for quantitative assessment of canine intelligence in attentiveness, interest in novelty exploration, response to signaling and showing, observational learning, problem solving/boldness, and handedness.

3. Phenotyping of 153 canine subjects using the CITP regimen and partial analysis of the test data of 108 dogs. Empirical evaluation of the performance of the canine subjects has also been conducted, resulting in the estimation of top 25 and bottom 25 candidates, with respect to their overall performance.

4. Completed genome-wide SNP typing of 117 dogs (German Shepherd Dog: 47; Belgian Malinois: 44; Labrador Retriever: 26).

5. Developed advanced classification algorithm and successfully achieved unsupervised breeds assignment, solely based on the SNP profiles of subjects.
6. Approval for access to testing of the MWDs at Lackland AFB was granted, as well as access to SOCOM 'Ranger' dogs, a unique first. While the testing reported here was not able to take advantage of the generous offers by both groups, nonetheless obtaining approvals indicated the high level of interest and support from both organizations. Offers for dog access from numerous MWD programs of NATO countries were also given.

Formal project milestones (as designated in the DARPA approved proposal) were completed either on time or early, up to the point of premature termination at 3 1/2 months into the project. Although the overall goal of this study was not achieved due to lack of funds, this work does lay a solid foundation by generating materials, datasets, and enabling tools for the mapping of genes contributing to canine intelligence. If funding is available in the future, this cutting-edge scientific endeavor can be readily revitalized and would provide a clear path towards the genetic mapping of canine intelligence. Gaining an understanding of the inherited factors of canine intelligence would institute a paradigm shift in the breeding and ultimate uses of the Military Working Dog.

5. REFERENCES

Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW. (2010) Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci U S A.* **107**:1160-5.

Almasy, L, Blangero, J. (2009) Human QTL linkage mapping. *Genetica* **136**:333-340.

Amos, CI. (2007) Successful design and conduct of genome-wide association studies. *Hum Mol Genet* **16**:220-225.

Bacanu SA, Devlin B, Roeder K. (2000) The power of genomic control. *Am J Hum Genet* **66**:1933-44.

Batt L, Batt M, Baguley J, McGreevy P. (2008) Stability of motor lateralisation in maturing dogs. *Laterality* **13**:468-79.

Borecki IB, Province MA. (2008) Linkage and association: basic concepts. *Adv Genet* **60**:51-74.

Branson, NJ, Rogers, LJ. (2006) *J Comp Psychol*. **120**:176-83.

Butcher LM, Davis OS, Craig IW, Plomin R. (2008) Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes Brain Behav* **7**:435-46.

Butcher LM, Meaburn E, Knight J, Sham PC, Schalkwyk LC, Craig IW, Plomin R. (2005) SNPs, microarrays and pooled DNA: identification of four loci associated with mild mental impairment in a sample of 6000 children. *Hum Mol Genet* **14**:1315-25.

Butcher, LM, Davis, OS, Craig, IW, Plomin, R. (2008) Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes Brain Behav* **7**:435-446.

Coppinger, R. and R. Schneider (1995). Evolution of working dogs. *In* J. Serpell (ed.), The Domestic Dog: Its Evolution, Behavior and Interactions with People. Cambridge: Cambridge University Press, pp. 21-47.

Daniel EE, Erlenmeyer-Kimling L, Jarvik LF. (1963) I. Q., Genetics, and Culture. *Science* **142**:1477-9.

Daniels, M., Devlin, B., Roeder, K. (1997) *In* B. Devlin, S. E. Fienberg., & K. Roeder (pp. 45-70). Intelligence, Genes, and Success: Scientists respond to The Bell Curve. New York: Springer.

Detterman DK, Thompson LA, Plomin R. (1990) Differences in heritability across groups differing in ability. *Behav Genet* **20**:369-84.

Devlin B, Bacanu SA, Roeder K. (2004) Genomic Control to the extreme. *Nat Genet* **36**:1129-30.

Devlin B, Roeder K. (1999) Genomic control for association studies. *Biometrics* **55**:997-1004.

Haverbeke A, Diederich C, Depiereux E, Giffroy JM. (2008) Cortisol and behavioral responses of working dogs to environmental challenges. *Physiol Behav* **93**:59-67.

Ding, C, Jin S. (2009) High-throughput methods for SNP genotyping. *Methods Mol Biol* **578**:245-254.

Eaves L, Meyer J. (1994) Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. *Behav Genet* **24**:443-55.

Evans, RI, Herbold, JR, Bradshow, BS, Moore, GE. (2007) Causes for discharge of military working dogs from service: 268 cases (2000-2004). *J Am Vet Med Asso*c **231**:1215-1220.

Farrall, M, Morris, AP. (2005) Gearing up for genome-wide gene-association studies. *Hum Mol Genet* **14**:157-162.

Fondon JW 3rd, Garner HR. (2004) Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci U S A*. **101**:18058-63.

Fondon JW 3rd, Garner HR. (2007) Detection of length-dependent effects of tandem repeat alleles by 3-D geometric decomposition of craniofacial variation. *Dev Genes Evol.* **217**:79-85.

Gu C, Todorov A, Rao DC. (1996) Combining extremely concordant sibpairs with extremely discordant sibpairs provides a cost effective way to linkage analysis of quantitative trait loci. *Genet Epidemiol* **13**:513-33.

Gu CC and Rao DC. (2003) Designing an optimum genetic association study using dense SNP markers and family-based sample. *Front Biosci* **8**:s68-80.

Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M. (2004) Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol* **27**:53-63.

Hochberg Y, Benjamini Y. (1990) More powerful procedures for multiple significance testing. *Stat Med* **9**:811-8.

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM. (2003) Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* **72**:1492-1504.

Hong H, Su Z, Ge W, Shi L, Perkins R, Fang H, Mendrick D, Tong W. (2010) Evaluating variations of genotype calling: a potential source of spurious associations in genome-wide association studies. *J Genet* **89**:55-64.

Karlsson, EK, Baranowska, I, Wade, CM et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet* **39**:1321-1328.

Ke X, Kennedy LJ, Short AD, Seppälä EH, Barnes A, Clements DN, Wood SH, Carter SD, Happ GM, Lohi H, Ollier WE. (2011) Assessment of the functionality of genome-wide canine SNP arrays and implications for canine disease association studies. *Anim Genet* **42**:181–90.

Lander E, Kruglyak L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**:241-247.

Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, de Andrade M, Henley JR, Rocca WA, Ahlskog JE, Maraganore DM.(2007) A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet* **3**:e98.

Lindblad-Toh, K, Wade, CM, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**:803-819.

Loehlin JC. (1989) Partitioning environmental and genetic contributions to behavioral development. *Am Psychol* **44**:1285-92.

Markman, EM, Abelev, M. Word learning in dogs? (2004) *Trends in Cognitive Sciences* **8**:479-481.

Masuda K, Hashizume C, Kikusui T, Takeuchi Y, Mori Y. (2004) Breed differences in genotype and allele frequency of catechol O-methyltransferase gene polymorphic regions in dogs. *J Vet Med Sci.* **66**:183-7.

Miclaus K, Wolfinger R, Vega S, Chierici M, Furlanello C, Lambert C, Hong H, Zhang L, Yin S, Goodsaid F. (2010) Batch effects in the BRLMM genotype calling algorithm influence GWAS results for the Affymetrix 500K array. *Pharmacogenomics J* **10**:336-346.

Nippak PM, Chan AD, Campbell Z, Muggenburg B, Head E, Ikeda-Douglas CJ, Murphey H, Cotman CW, Milgram NW. (2003) Response latency in Canis familiaris: mental ability or mental strategy? *Behav Neurosci* **117**:1066-75.

Nippak PMD, Milgram NW. (2005) An investigation of the relationship between response latency across several cognitive tasks in the beagle dog. *Progress in Neuro-Pshchopharmacology & Biological Psychiatry* **29**:371-377.

Ogata N, Hashizume C, Momozawa Y, Masuda K, Kikusui T, Takeuchi Y, Mori Y. (2006) Polymorphisms in the canine glutamate transporter-1 gene: identification and variation among five dog breeds. *J Vet Med Sci* **68**:157-9.

Olson, PN, Hall, MF, Peterson, JK, Johnson GS (2004) Using genetic technologies for promoting canine health and temperament. *Animal Reprod Sci* **82**:225-230.

Ostrander, EA, Galibert, F, Patterson, DF (2000) Canine genetics comes of age. *Trends Genet* **16**:117-124.

Ostrander, EA, Wayne RK. (2005) The canine genome. *Genome Res* **15**:1706-1716.

Overall, KL, Hamilton, SP. (2006) Working bitches and the neutering myth: sticking to the science. *J Vet Behavior* **1**:124-141.

Pearson, TA, Manolio TA. (2008) How to interpret a Genome-wide Association Study. JAMA 299:1335-1344.

Plomin R, Crabbe J. (2000) DNA. *Psychol Bulletin* **126**:806-28.

Province MA. (2000) A single, sequential, genome-wide test to identify simultaneously all promising areas in a linkage scan. *Genet Epidemiol* **19**:301-22.

Risch N, Zhang H. (1995) Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* **268**:1584-9.

Ruefenacht, S, Gebhardt-Henrich, S, Miyake T, Gaillard, C. (2002) A behavior test on German Shepherd dogs: heritability of seven different traits. *Applied Animal Behavior Science* **79**:113-132.

Sabb FW, Burggren AC, Higier RG, Fox J, He J, Parker DS, Poldrack RA, Chu W, Cannon TD, Freimer NB, Bilder RM. (2009) Challenges in phenotype definition in the whole-genome era: multivariate models of memory and intelligence. *Neuroscience* **164**:88-107.

Sham, PC, Cherny SS, Purcell S. (2009) Application of genome-wide SNP data for uncovering pairwise relationships and quantitative trait loci. *Genetica* **136**:237-243.

Skol AD, Scott LJ, Abecasis GR, Boehnke M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* **38**:209-13.

Spady, TC, Ostrander, EA. (2008) Canine behavioral genetics: pointing out the phenotypes and herding up the genes. *Am J Hum Gen* **82**:10-18.

Svartberg, K. (2002) Personality traits in the domestic dog (Canis familiaris). *Applied Animal Behaviour Science* **79**:157-174.

Svartberg, K. (2005) A comparison of behaviour in test and in everyday life. *Applied Animal Behaviour Science* **91**:103–128.

Todorov AA, Rao DC. (1997) Trade-off between false positives and false negatives in the linkage analysis of complex traits. *Genet Epidemiol* **14**:453-64.

Udell MA, Wynne CD. (2008) A review of domestic dogs' (*Canis familiaris*) human-like behaviors: or why behavior analysts should stop worrying and love their dogs. *J Exp Anal Behav* **89**:247-61.

Våge J, Wade C, Biagi T, Fatjó J, Amat M, Lindblad-Toh K, Lingaas F. (2010) Association of dopamine- and serotonin-related genes with canine aggression. *Genes Brain Behav* **9**:372-8.

van den Berg, L, Versteeg, SA, van Oost, BA (2003) Isolation and characterization of the canine serotonin receptor 1A gene (htr1A). *J Hered*. **94**:49-56.

Vandenberg SG. (1968) Primary mental abilities or general intelligence? Evidence from twin studies. *Eugen Soc Symp* **4**:146-60.

Ward, JH, Hook, ME. A Hierarchical Grouping Procedure Applied to a Problem of Grouping Profiles. Lackland Air Force Base, Texas: Personal Laboratory, Wright Air Development Division, March 1961.

Wayne, RK, Ostrander, EA. (2007) Lessons learned from the dog genome. *Trends in Genetics* **23**:557-567.

West, RE Young, RJ. (2002) Do domestic dogs show any evidence of being able to count? *Anim Cog*n **5**:183-186.

7. LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

CGP – co-evolutionary genetic programming
cM – centi Morgan
CITP – canine intelligence testing protocol
EDTA – ethylenediaminetetraacetic acid
GW – genome-wide
GWAS – genome-wide association study
LD – linkage disequilibrium
MWD – military working dog
OD – optical density
PCR – polymerase chain reaction
PM – perfect match
QC – quality control
QTL – quantitative trait loci
SNP – single nucleotide polymorphism
TE – Tris + EDTA
TBE – Tris + Boric Acid + EDTA
WG – whole genome
WGSA – whole genome sampling assay