Journal of Veterinary Behavior 9 (2014) 281–289

Contents lists available at ScienceDirect

Journal of Veterinary Behavior

journal homepage: www.journalvetbehavior.com

Research

The Swedish Armed Forces temperament test gives information on genetic differences among dogs

Per Arvelius*, Erling Strandberg, W. Freddy Fikse

Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

ARTICLE INFO

Article history:
Received 10 March 2014
Received in revised form
16 June 2014
Accepted 18 June 2014
Available online 26 June 2014

Keywords:
temperament test
behavior
German shepherd dog
breeding
heritability
genetic correlation

ABSTRACT

With the dual purpose of selecting both breeding animals and dogs for training, all German shepherd dogs in the Swedish Armed Forces (SAF) breeding program are subjected to a temperament test. During the test, the dog's behavioral responses are rated with 2 different methods. In a previous study, using Principal Components Analysis (PCA) on the test items, 5 and 3 underlying behavioral dimensions from each rating method were defined. Three of the dimensions were reported to correlate significantly to training success. Using the test results from 873 dogs, we estimated heritabilities of, and genetic correlations among, the 38 test items and the 8 underlying behavioral dimensions. Parameters were estimated using a mixed linear animal model including fixed effects of sex, training level, test age and test year-location combination, and random effects of litter, genetic effect of the individual, and residual. Heritabilities ranged from 0.00 to 0.28 (standard error [SE] = 0.05-0.10), which is similar to what has been reported in previous studies of traits defined and measured in a comparable way. Genetic correlations were high (r_g = 0.92-0.98, SE = 0.08-0.12) between dimensions derived from each rating method and defined as either confidence, engagement, or aggressiveness, but relatively weak among these dimensions within rating method (r_g = 0.00-0.45, SE = 0.29-0.41). Our results imply that the test measures 3 separate behavioral dimensions and that the SAF temperament test as a whole is possible to use for selection of dogs for breeding, but also that some test items should be measured differently to be meaningful for genetic selection purposes. Furthermore, aggregating variables based on a PCA performed on phenotypic data might be suboptimal when defining dimensions for breeding purposes; taking genetic parameters into consideration resulted in generally higher heritabilities for the dimensions.

Introduction

In many parts of the world, police, customs, and military authorities as well as guide dog schools report difficulties in finding dogs suitable for service (e.g., Goddard and Beilharz, 1982; MacIsaac et al., 2005; Slabbert, 2008; Tjänstehundsutredningen, 2005; Vanderloo, 2005). This has led several of these organizations to initiate their own breeding programs. The programs often have a limited production and to be able to recruit good animals for breeding, they depend on the private breeding outside the program (e.g., MacIsaac et al., 2005). As a contrast, in a few working dog breeding programs, the breeding population, and thus breeding goal and selection, is under strong control (e.g., Cole et al., 2004). This enables more independence (Leighton, 2009) and faster genetic progress but typically at a higher cost per dog produced, at least unless the production is very large (Tjänstehundsutredningen, 2005).

The Swedish governmental inquiry Tjänstehundsutredningen (2005) described how the governmental Swedish Dog Training Centre (SDTC) during several decades of the 1900s ran breeding programs for German shepherd dogs (GSD) and Labrador retrievers. In the 1990s, the Swedish Government considered SDTC to be too expensive and it was therefore privatized. In 2002, the private successor went bankrupt. Simultaneously with this process, authorities using dogs, such as police, military, and customs, reported increasing problems of finding dogs with appropriate temperament. In 2005, the Swedish Armed Forces (SAF) started to build its own breeding program, with the aim to improve behavioral traits of importance to substance detection and personnel protection dogs (Wilsson and Sinn, 2012). They decided on a solution with control

* Address for reprint requests and correspondence: Per Arvelius, MSc, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, PO Box 7023, S-750 07 Uppsala, Sweden. Tel: +46-18-67-1970; Fax: +46-18-67-2848.

E-mail addresses: per.arvelius@slu.se, per.arvelius@hgen.slu.se (P. Arvelius).

1558-7878/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jveb.2014.06.008

over breeding goal and selection; a more or less closed population of GSD of approximately 80 bitches and 15 males, all owned by SAF, producing around 200-300 puppies per year. Both working dogs and replacement breeding animals are recruited from the dogs produced.

When a dog born within the SAF breeding program is around 16 months old, it is subjected to a standardized temperament test developed by the SAF. The test is in the form of a test battery containing 12 subtests. Based on test results and health data, it is decided whether the dog should be rejected or if it should be used for breeding or put into training. However, no genetic evaluation is carried out, and no one has verified if there is genetic variation for the traits measured in the SAF temperament test. The dog's behavior during the test is simultaneously rated using 2 separate rating protocols; in the first protocol, the rating method is termed "behavioral ratings" (BR), in the second "subjective ratings" (SR). Wilsson and Sinn (2012) defined BR as "ratings that are based on observed behaviors in a particular test situation, and attempt to rate behavior as objectively as possible", whereas SR "attempt to evaluate global behavior of the subject, and are based on a human scorer's overall perception of the dog's aggregate behavioral disposition, sometimes (but not always) irrespective of test situation". In their study, they analyzed SAF temperament test results from approximately 400 dogs to see how well each method characterized dogs in terms of their ability to successfully complete the training to become service dogs. They hypothesized that on one hand, BR might be easier to define and therefore more objective, with higher reliability of the measurements as a result. On the other hand, SR can be regarded as an average of several observations. Thus, SR reliability might benefit from a smaller measurement error. The SR may also better capture behaviors relevant to training outcome, with the potential to yield higher predictive validity.

Wilsson and Sinn (2012) performed principal component analyses (PCAs) and defined 5 underlying behavioral dimensions based on the BR and 3 dimensions based on the SR. Next, they used mixed model methodology to predict training success based on aggregated behavior scores (ABS) for the underlying dimensions, age, sex, test location, and litter identity. The proportion of dogs correctly classified (will or will not succeed in training) was slightly higher when using BR dimensions (72.0-78.3%) compared with when SR dimensions were used (70.3-71.7%), but the difference was considered small. One of the 5 BR dimensions and 2 of the 3 SR dimensions had a significant effect in predicting training success.

To predict the future success of a given dog, it makes sense to use a PCA based on the phenotypic correlations among ratings. However, the genetic correlation between 2 traits can differ both in magnitude and in sign compared with the corresponding phenotypic correlation (Falconer and Mackay, 1996). Consequently, it is not self-evident that a PCA based on phenotypic records is optimal when constructing composite traits ("underlying behavioral dimensions"), if these are to be used for selection of breeding animals.

The study by Wilsson and Sinn (2012) focused on the validity of the test, more exactly how well test results correlated to training success, and whether it was different between rating methods (BR vs. SR). This is important knowledge when using test results to select dogs for training; if the test has poor validity, dogs will more often be wrongly categorized as suitable or not suitable for training. For similar reasons, validity is important also for breeding purposes. Moreover, when using a test to select breeding animals, it is essential that the superiority of the selected parents is inherited by the offspring. This is often expressed as the heritability (h^2), which is defined as the proportion of the phenotypic variance that has additive genetic background. Thus, the heritability varies from 0 to 1, and the higher the heritability, the easier it becomes to accurately select breeding animals and to achieve genetic progress. To our knowledge, no study has analyzed and compared heritabilities of BR and SR based on the same temperament test in dogs.

Genetic parameters for behavioral characteristics in dogs, measured using a test battery similar to the SAF temperament test, have been published in only a handful of studies, and never for the SAF test. Heritabilities for traits defined and rated similarly as the SR items in the SAF test have been estimated to 0.10-0.30 (Wilsson and Sundgren, 1997; Ruefenacht et al., 2002; van der Waaij et al., 2008; Meyer et al., 2012). For behavior reactions rated in a way comparable to the SAF BR items, heritabilities generally are lower, 0.05-0.20 (Saetre et al., 2006; Arvelius et al., 2010). Only 1 of these previously studied tests, the SDTC temperament test (Wilsson and Sundgren, 1997; van der Waaij et al. 2008), resembles the SAF temperament test in the way that it was used in a professional breeding program for military or police working dogs, with supposedly limited variation for several environmental factors that can be expected to influence the ratings (only 1 judge assessing all dogs, few test locations, *et cetera*).

Our aim was to analyze to what extent the SAF temperament test measures genetic differences between dogs. We also wanted to study if the heritabilities of the underlying behavioral dimensions could be improved by taking genetic parameters into consideration when defining the dimensions. Finally, we wanted to explore whether heritabilities were higher for behavioral dimensions based on SRs than for dimensions based on BRs.

Materials and methods

Data

The analyses were based on results from a standardized temperament test given to GSD bred within the SAF breeding program and pedigree data for these dogs.

All puppies in the SAF breeding program are born at the same breeding station. The bitches normally live as ordinary companion dogs but are brought to the station 2 weeks before parturition. The pups stay with their mother until 8 weeks of age and are then placed in ordinary families, who have volunteered for raising the dogs. A dog normally stays with the family until it is 16 months old (mean = 16.0, standard deviation [SD] = 1.0, min = 10.7, max = 23.1 months), when it is subjected to a temperament test developed by the SAF.

The SAF temperament test data contained records from 873 dogs tested from October 2006 to December 2012, with a close to even sex ratio. In addition to the test ratings, the SAF data held information on sex, training level, birth and test date, litter identity and size, and test location (Table 1). Training level is an estimate of how much an individual dog has been trained by its puppy raiser, made by the judge and based on the number of training sessions arranged by SAF that the dog had participated in before the test. In total, 4 sessions were arranged.

Pedigree data on GSD born from June 1957 to August 2012 were received from the Swedish Kennel Club. After initial editing, removing obvious errors, and excluding all dogs except those tested and their ancestors, it contained 5012 individuals. The pedigree depth and completeness is described in Table 2.

The full test took about 40 minutes per dog to complete. All dogs were subjected to all 12 subtests in the same order, with the exception of 14% of the dogs that showed extreme and remaining fearfulness and for which the test was discontinued. All dogs were assessed by the same person, and most of the dogs were handled during the test by their puppy raisers. During each subtest, 1-4 BR were given; and after finishing the full test battery, 13 SR were given (Table 3). The 25 BR were given using predefined scales containing typical behaviors characterizing each step of the scale for each item.