Title

Heritability and complex segregation analysis of diabetes mellitus in American Eskimo Dogs

Journal

Journal of Veterinary Internal Medicine, 33(5)

ISSN

0891-6640

Authors

Cai, Stephen V

Famula, Thomas R

Oberbauer, Anita M

et al.

Publication Date

2019-09-01

DOI

10.1111/jvim.15570

Received: 17 January 2019 | Accepted: 8 July 2019

DOI: 10.1111/jvim.15570

STANDARD ARTICLE

Journal of Veterinary Internal Medicine Open Access

ACVIM American College of Veterinary Internal Medicine

Heritability and complex segregation analysis of diabetes mellitus in American Eskimo Dogs

Stephen V. Cai[1] | Thomas R. Famula[2] | Anita M. Oberbauer[2] | Rebecka S. Hess[1]

[1]Department of Clinical Sciences and Advanced Medicine, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

[2]Department of Animal Science, University of California, Davis, California

Correspondence
Rebecka S. Hess, Department of Clinical Sciences and Advanced Medicine, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA 19104.
Email: rhess@vet.upenn.edu

Abstract

Background: Heritability and mode of inheritance of spontaneous diabetes mellitus (DM) in American Eskimo Dogs (AED) are unknown.

Objective: Investigate the heritability and mode of inheritance of DM in AED.

Animals: An extended family of AED including 71 AED without DM, 47 AED with an unknown phenotype, and 38 AED with spontaneous DM.

Methods: Retrospective evaluation of inheritance. A logistic regression model was formulated to evaluate the heritability of DM, including effects of sex and neuter status. Subsequently, complex segregation analysis was employed to investigate the inheritance pattern of DM in AED. Six plausible models were considered, and the Akaike Information Criterion was used to determine the best of the biologically feasible models of inheritance of DM in AED.

Results: Heritability of DM in AED is estimated at 0.62 (95% posterior interval 0.01-0.99). Predicted DM probabilities for neutered females (NF), intact females (IF), neutered males (NM), and intact males (IM) were 0.76, 0.11, 0.63, and 0.12, respectively. There was no overlap between the 95% posterior intervals of disease probabilities in NF and IF or in NF and IM. Complex segregation analysis suggested that the mode of inheritance of DM in AED is polygenic, with no evidence for a single gene of large effect.

Conclusions and Clinical Importance: The estimated heritability of DM in AED is high but has low precision. Diabetes mellitus transmission in AED appears to follow a polygenic inheritance. Breeders could successfully implement a breeding program to decrease the incidence of DM in AED.

KEYWORDS

canine, genetic risk, non-Mendelian, polygenic

1 | INTRODUCTION

Heritability and mode of inheritance of middle-aged onset diabetes mellitus (DM) have not been reported in dogs with spontaneous disease. However, breed predisposition to DM suggests a genetic component, which can vary with geographic location. In the United States,

Abbreviations: AED, American Eskimo Dog; AIC, Akaike Information Criterion; AKC, American Kennel Club; DM, diabetes mellitus; IF, intact females; IM, intact male; NF, neutered females; NM, neutered males; ROC curve, receiver operating characteristic curve.

the Samoyed, Miniature Schnauzer, Miniature Poodle, Pug, Toy Poodle, and Australian Terrier breeds are at increased risk for DM.[1,2] In Sweden, the Australian Terrier, Samoyed, Swedish Elkhound, and Swedish Lapphund breeds have increased incidence of DM, whereas in the United Kingdom, the Samoyed, Tibetan Terrier, Cairn Terrier, Dachshund, Doberman Pinscher, Miniature Schnauzer, Siberian Husky, Scottish Terrier, West Highland White Terrier, Miniature Poodle, and Border Collie breeds are at highest risk for DM.[3,4] Familial DM occurs in Samoyeds and Keeshonds although the age of onset of DM in this inbred family of colony Keeshonds was unusually young (2-6 months of age) compared to the mean age of onset of DM in the general dog population, which is about 9 years of age.[5-7] The mode of inheritance of this unusually young age of onset DM in Keeshonds was suggested to be autosomal recessive.[6] The American Eskimo Dogs (AED) have not been reported to be at an increased risk for DM, although AED breeders have recognized the problem and subsequently charged an Ad Hoc Diabetes Committee (https://www.aedca.org/AdHocDiabetes/diabetes-main.html).

There are associations between DM and single nucleotide polymorphisms in different genes in various breeds of dogs[4,8-11] but not in AED. There is an association between the risk of DM and single nucleotide polymorphisms in the insulin gene region in Jack Russell Terriers and Cocker Spaniels, in the cytotoxic T-lymphocyte-associated protein 4 in mixed breed dogs, Samoyed, Miniature Schnauzer, West Highland White Terrier, Border Terrier, and the Labrador Retriever breeds, in inflammatory mediators, protein tyrosine phosphatase nonreceptor type 22, mitochondrial transfer RNA for protein translation, paired box 4, and hepatocyte nuclear factor 4a in several breeds, in the zinc finger protein 57 of Bichon Frise and Samoyed breeds, and in the dog leukocyte antigen of dogs not stratified by breed.[4,8-11]

Although ample evidence exists that there is a genetic risk for DM in dogs, the heritability and mode of inheritance of DM in dogs are not known. Heritability and mode of inheritance must be investigated in breed-specific studies because prior research suggests that the genetic architecture of DM differs by breed.[1-6,9-11] The goals of this study were therefore to investigate the heritability and mode of inheritance of DM in AED. It was hypothesized that the heritability of DM in AED is high and that the mode of inheritance is polygenic with several large effect loci.

2 | MATERIALS AND METHODS

2.1 | Study population

An online questionnaire (http://www.vet.upenn.edu/diabetes) designed to investigate the prevalence of spontaneous DM in dogs across the United States was launched in April 2017 and widely distributed (Appendix S1). Owners of all dogs, regardless of breed and whether the dog had DM, were encouraged to complete the survey. The survey was promoted electronically among veterinary students and faculty at 21 academic institutions and directly to owners of dogs who were examined at 10 private referral practices and 8 academic institutions by placing self-standing flyers advertising the survey in waiting areas (Appendix S1). The survey was also featured on the American Kennel Club (AKC) website and numerous breed-specific kennel clubs including the AED National Club, which also promoted the survey to their members electronically. Additionally, social media was used to promote the survey (Appendix S1). Data collected for all dogs included age at the time of survey completion, breed, sex, neuter status, age at the time of neuter, presence or absence of DM, AKC or United Kennel Club number if available, names and contact information of immediate relatives if known, and owner contact information. Additional data collected for dogs with DM included neuter status before diagnosis of DM, age at the time of DM diagnosis, clinical signs recorded at the time of DM diagnosis, and the dose, frequency, and name of the insulin product used for treatment of the dog. For the purposes of this study, only data regarding AED were analyzed. American Eskimo Dogs were included in the study if they were entered into the survey by the end of April 2018. In addition to the survey, owners and breeders of AED within the United States were contacted directly and identical data were collected from them in this manner. Purebred status and ancestry of AED were confirmed with review of AKC or United Kennel Club pedigrees. American Eskimo Dogs with DM were included only if the owner had an AKC or United Kennel Club registered pedigree and if the owner could be contacted to confirm the diabetic phenotype. Furthermore, dogs with DM were included only if at least 1 owner of 1 ancestor could be contacted to verify the presence or absence of DM in this ancestor. American Eskimo Dogs without DM or with an unknown phenotype were included only if they were directly related (eg, sibling, offspring, parent, grandparent) to a dog with DM. American Eskimo Dogs from countries other than the United States and Canada were excluded.

2.2 | Definition of phenotype

Dogs were defined as cases with DM if their owner asserted that a veterinarian had diagnosed the dog with DM. Dogs were defined as controls that do not have DM if their owner reported that the dog had no clinical signs consistent with DM (polyuria, polydipsia, polyphagia, weight loss) and was not being treated with insulin. Dogs were defined as having an unknown phenotype if their owner could not be contacted. Sex was classified as neutered females (NF), intact females (IF), neutered males (NM), and intact males (IM).

2.3 | Estimation of heritability

Logistic regression (logit) was chosen to model this binary disease outcome (dogs with or without DM). Disease risk was modeled as a function of sex and neuter class, and a presumed quantitative genetic contribution. The probability of disease was defined as p_{ij} for the ith sex and neuter class (i = NF, IF, NM, IM), and the jth dog and the logit of this probability was defined as $\theta_{ij} = \log[p_{ij}/(1 - p_{ij})]$. The logit was modeled as a function of sex and neuter class and a quantitative genotype, and the linear model was as follows: $\theta_{ij} = \mu + \text{sex}_i + a_j + e_j$, where μ is an unknown constant common to all dogs, sex_i is the additive contribution of the ith sex and neuter class to the risk of disease, a_j is the

additive genetic contribution to the risk of disease for the jth (j = 1, 2, 3, …) dog, and e_j is an unknown random residual contribution to the risk of disease particular to the jth dog. This unknown residual risk of disease (e_j) represents a function of unknown environmental factors, which could include diet, exercise, climate, and veterinary care. It is important to recognize that the dogs in this set of data are related to one another such that the covariance among relatives is accommodated by the inclusion of the additive genetic contribution (a_j).

Likelihood-based software to evaluate a binary trait among relatives is not readily available. Among publicly available software, there is a Bayesian-oriented package called MCMCglmm, available in the public domain language R. However, in data sets of this size and structure, the built-in priors for the variances used in MCMCglmm (ie, the inverse-Wishart) often lead to poorly converged posterior densities. Such was the case with the present set of data. Accordingly, for the analyses presented here, the Bayesian statistical language Stan, executed with the public domain language R, was used (Appendix S2).[12,13] The Stan language provides the user with more flexibility in the definition of prior densities, especially for the unknown variances. In this way, it was possible to build a hierarchical Bayesian model with weakly informative prior distributions for the unknown effects, a model that can better stabilize the estimation of the posterior density, especially in data sets with such a limited number of samples accompanied by a complex pedigree.[14]

The outline for the Bayesian animal model used in this study is the one established by Damgaard.[15] In this approach, which accommodates the contributions of relatives to the risk of disease, the user must provide Mendelian sampling terms for gametic effects, which are values that incorporate known parentage and parental inbreeding.[16] Once computed, the Mendelian sampling terms can then be included as input to the analysis so that the genetic covariance structure dictated by the pedigree is built properly.[15]

An important component of any Bayesian analysis is the definition of the prior distributions of the unknown parameters. Although a thorough description of Bayesian analysis is beyond the scope of this report, readers can nonetheless be aware that a Bayesian analysis is built around the recognition that the "posterior density is proportional to the likelihood times the prior density." It is the posterior density ("posterior" to the collection of data) that serves as the mechanism for evaluating the patterns (if any) observed in the risk of disease in this pedigree. The likelihood for disease was defined above, that of a simple binary variable (formally called a Bernoulli variable). Next, we must outline our assumptions about the prior distributions of our unknown parameters (where "prior" is meant to convey our thoughts about these unknown values before data collection). The adjective "weakly," which was used above in the discussion of prior distributions, is intended to convey that we harbor no strongly held opinions about the values of the unknown parameters before data collection, preferring to let the data inform us of the patterns reflected in the data. Accordingly, if the prior beliefs are "weakly" held, then the posterior largely reflects the likelihood (ie, "posterior density is proportional to the likelihood times the prior"), which is a result that is not too distant from the basis of classical statistics where likelihoods form the basis for all parameter estimates, hypothesis tests, and detection of patterns in the data.

Defining the prior distributions of these unknown parameters, it was assumed that the intercept (μ) and sex contribution (sex_i) were each drawn from a normal distribution, 1 with mean zero and a variance of 16 (ie, N[0, 16]). This choice for priors is intended to reflect, as best as can be done, the classical application of a mixed linear model; a model that contains fixed effects (ie, sex) and random effects (ie, animal contributions). Moreover, this binary trait is evaluated on the logit scale, and a variance of 16 on this log scale is sufficiently large so as to serve as a weakly informative prior for these unknown intercept and sex effects.[17,18] The additive genetic effects (a_j) in this model were assumed to be drawn from the multivariate normal distribution, with a null mean and a variance-covariance matrix of $A\sigma_a^2$. Naturally, A is the known numerator relationship matrix (ie, a table of values that holds the relationships among all pairs of animals in the study, as well as every dog's inbreeding coefficient) and σ_a^2 is defined as the unknown additive genetic variance of disease risk. In this formulation, it was assumed that the prior distribution for the unknown genetic standard deviation (σ_a) was drawn from the positive values of a Cauchy (0, 2.5), as advocated for weakly informative prior distributions for unknown variances in logistic models.[17,19] Finally, the residual term (e_j) was assumed to be drawn from a standard normal density [i.e., N(0, 1)] as required for this parameterization of a binary trait analysis.[20] Heritability (h^2), a measure of the inheritance of disease risk, is the proportion of all the observed variation in disease risk that can be assigned to variation in additive genetic effects. Accordingly, the heritability of disease risk was estimated as $h^2 = \sigma_a^2/\left(\sigma_a^2 + 1\right)$.[20]

Application of this Bayesian framework to the analysis of the data is reliant upon the efficient simulation of random numbers in a process broadly titled as Markov Chain Monte Carlo.[14] The objective is to draw random numbers that behave as if they were drawn from the posterior distribution, remembering that the goal is to learn about the posterior distribution. That is, the Markov Chain Monte Carlo is capable of determining which random draws make sense in the context of the assembled data and priors and which random draws are less likely to lead to a meaningful posterior. As a result, much of the emphasis in evaluating the qualities of a Bayesian analysis focuses on this series of randomly drawn numbers. One of the additional advantages of the Stan language, beyond its flexibility, is the implementation of a Hamiltonian Monte Carlo simulation, which is a directed-search process borrowed from physics (specifically, Hamiltonian mechanics) and made additionally more efficient by the application of a No U-Turn sampler (the so-called NUTS sampler). The objective of this simulation process is to provide representative samples of the unknown parameters of the model (eg, the effects of sex, the unknown genetic variance), thereby producing values which can be interpreted as representative of the possible effects of sex and genetic variation. The simulation process was conducted from 4 independent sampling sets, called chains, where each chain was built on a draw of 40 000 total samples. However, the early samples are usually not informative because of the Markov Chain nature of this process. For the purpose of this analysis, the first 15 000 samples of the total 40 000 were discarded. Moreover, because consecutive samples in a Markov Chain

are likely to be correlated to one another (and a truly random sample with which to work is required), only every 25th generated sample was kept and the preceding 24 sampled values were discarded. In this way, each chain generated 1000 parameter estimates (ie, [40000 – 15 000]/25 = 1000), and thus 4000 samples were generated across all 4 chains. Ensuring the convergence of this process to a consistent set of random samples was visualized through trace and other diagnostic plots of all the unknown parameters and computation of the Gelman-Rubin statistic for convergence being below 1.05.[21,22]

Finally, it is important to determine how well this logistic model fits the data that have been collected. The goodness-of-fit for logistic models is most easily visualized through a receiver operating characteristic (ROC) curve.[23] This plot is taken from the sensitivity and specificity (actually, 1 – specificity) of the model to accurately predict animals with and without disease. The most common way to summarize the ability of the model to discriminate between cases and controls is through the area under the curve, where values exceeding 0.90 are usually interpreted as evidence of a model with an excellent fit. One can also use an ROC curve to estimate the threshold probability of declaring a dog as affected. The predictions of a logistic regression are probabilities on a scale from 0 to 1, whereas the actual observations of disease are binary (eg, yes or no). The ROC curve can be used as a guide in evaluating at what predicted probability value there is the greatest accuracy in assigning a dog to affected versus unaffected. Because the Stan language also permits the generation of model predictions, the predicted probability of disease values was used to create an ROC curve, evaluate the area under the curve, and estimate the threshold in probability for dogs to be scored as affected using the R-language package pROC. [21,24]

2.4 | Complex segregation analysis

An additional objective of this work was to evaluate whether a single gene of large effect might impact the risk of DM in AED. A complex segregation analysis was performed using the publicly available package SEGREG, 1 of several programs available in the S.A.G.E. (v6.4) library.[25,26] Implementation of this analysis, however, required the elimination of "loops in the pedigree," a well-known challenge in the application of the Elston-Stewart algorithm.[27,28] Charting the path of a putative single locus through a pedigree requires an unambiguous means of ascertaining the probability that an offspring has received a given allele from both parents. This is the foundation of the Elston-Stewart algorithm, which computes the genotype transmission probabilities from parents to offspring. "Loops in the pedigree," or more precisely the presence of consanguineous matings, confound the ability to track the origin of alleles and genotypes. In such scenarios, the calculation of genotype probabilities, genotypic transmission probabilities, and most importantly, the likelihood of a given phenotype being impacted by all possible putative genotypes, becomes impossible. Accordingly, before the complex segregation analysis, a loop-breaking algorithm was implemented.[29] After the implementation of the loop-breaking algorithm, the reconfigured pedigree included 200 dogs in 1 large family, where dogs were duplicated to remove loops generated by inbreeding. For example, if 1 dog was related to 2 different families, this dog was duplicated and 1 duplicate remained in 1 family whereas the other duplicate was introduced to the other family. Although likely to decrease the power to detect the linkage of a major locus, the strategy employed was intended to minimize the impact of this pedigree simplification.

The subsequent analysis applied a model intended to mimic the logistic regression model outlined above, including a term for the sex of each dog, and a parameter to accommodate shared polygenic terms of family members, as well as the putative major locus effects.[28] As outlined, various models with and without Mendelian transmission were evaluated to establish or exclude the potential presence of a single gene of large effect.[25] Six models were considered. The simplest was a sporadic model, which considers no putative major locus effect, but does consider a term for sex and an accommodation of a polygenic contribution to disease.[28] This was followed by an evaluation of 3 simple mixed major locus models, considering a dominant, recessive, or codominant major locus, all of which follow the expected transmission of alleles outlined by Mendel. That is, for putative major genotypes AA, AB, and BB, the transmission probability for the A allele of these genotypes is set at 1.0, 0.5 and 0.0, respectively. Next, a model that considers environmental transmission of disease, where the polygenic term is removed and the transmission probabilities are set to being identically equal to the estimated allele frequency for all 3 putative major genotypes, was examined. Finally, a general model was considered where the transmission probabilities of the A allele were estimated from the data set. The Akaike Information Criterion (AIC), in which the goodness-of-fit is evaluated with a penalty for the number of parameters estimated, was used to compare models. The AIC is intended for model comparisons, where the model with the smallest AIC is typically considered the best balance between maximizing the likelihood with the smallest number of meaningful parameters. Such is the case here, with 1 subtle distinction. That is, the model considered to be best should also be biologically feasible. Of the biologically feasible models, the 1 with the smallest AIC was considered the most appropriate for the data.

The age at DM diagnosis was normally distributed based on visual inspection and the Skewness/Kurtosis test. Analysis of variance was therefore used to compare the age at DM diagnosis in dogs of different sex and neuter status. However, age of death or age at the time of data entry was not normally distributed among dogs without DM. Therefore, a two-sample Wilcoxon rank-sum (Mann-Whitney) test was used to compare the median age of dogs without DM to the median age of dogs with DM. A P-value <.05 was considered significant. These statistical analyses were performed using a statistical software package (Stata, version 14.0 for Mac; Stata Corp, College Station, Texas).

3 | RESULTS

3.1 | Study population

The study population included 156 AED, all of which were members of an extended family of inter-matings. Of these 156 AED, 109 had a known phenotype: 71 dogs were unaffected by DM (controls), and

38 dogs had DM (cases). Forty-seven dogs had an unknown phenotype. Of the 38 AED with DM, 16 were enrolled from the survey and 22 were enrolled from other outreach. Ten AED with DM resided in Canada but were sired by a dog residing in the United States and born from the United States AED breeding pool. All other study dogs resided in the United States.

Among the 38 AED with DM, 13 were NF, 10 were IM, 10 were IF, and 5 were NM at the time of DM diagnosis. Mean (±SD) age of the 38 AED at the time of DM diagnosis was 6.3 ± 2.4 years. There was no significant difference between the mean age of NF (5.7 ± 2.6 years), IF (6.7 ± 1.9 years), NM (6.7 ± 2.7 years), and IM (6.6 ± 2.6 years) at the time of DM diagnosis (P = .72). Median age of 66 of 71 dogs without DM, for which age at death or age at the time of data entry was known, was 13 years (range, 1-16.5 years) and was significantly higher than the median age of dogs with DM at the time of DM diagnosis (6.5 years, range 1-11 years, P < .0001). Age of neuter was known in 25 dogs with DM and 5 dogs without DM. Median age of neuter in 17 females with DM, 8 males with DM, 3 females without DM, and 2 males without DM was 4.0 years (range 0.3-10 years), 2.9 years (range 0.5-8 years), 6.5 years (range 5.6-14.6 years), and 1.5 years (range 0.5-2.5 years), respectively.

Thirty-seven of the DM cases were confirmed to have not received any steroid medication within 1 month before the diagnosis of DM; 1 intact female AED diagnosed with DM received a steroid medication within 1 month of DM diagnosis. This dog received twice daily insulin injections for 2.5 years after the steroid medication was discontinued and before its death of lymphosarcoma. Thirty-four of the 38 AED with DM were treated with twice daily SC insulin injections on a regular basis. Four AED with DM were not insulin treated. Three of these 4 dogs were euthanized at the time of DM diagnosis and the fourth dog survived untreated for 1 month before euthanasia.

Owners of 596 AED completed the survey. Five hundred sixty-seven of these entries were excluded because they described AED without DM that were not directly related to dogs with DM or whose owners could not be contacted to verify the phenotype. However, 13 dogs with DM who were entered into the survey were also excluded from the study because they were not AKC or United Kennel Club registered, or because the owner could not be contacted to

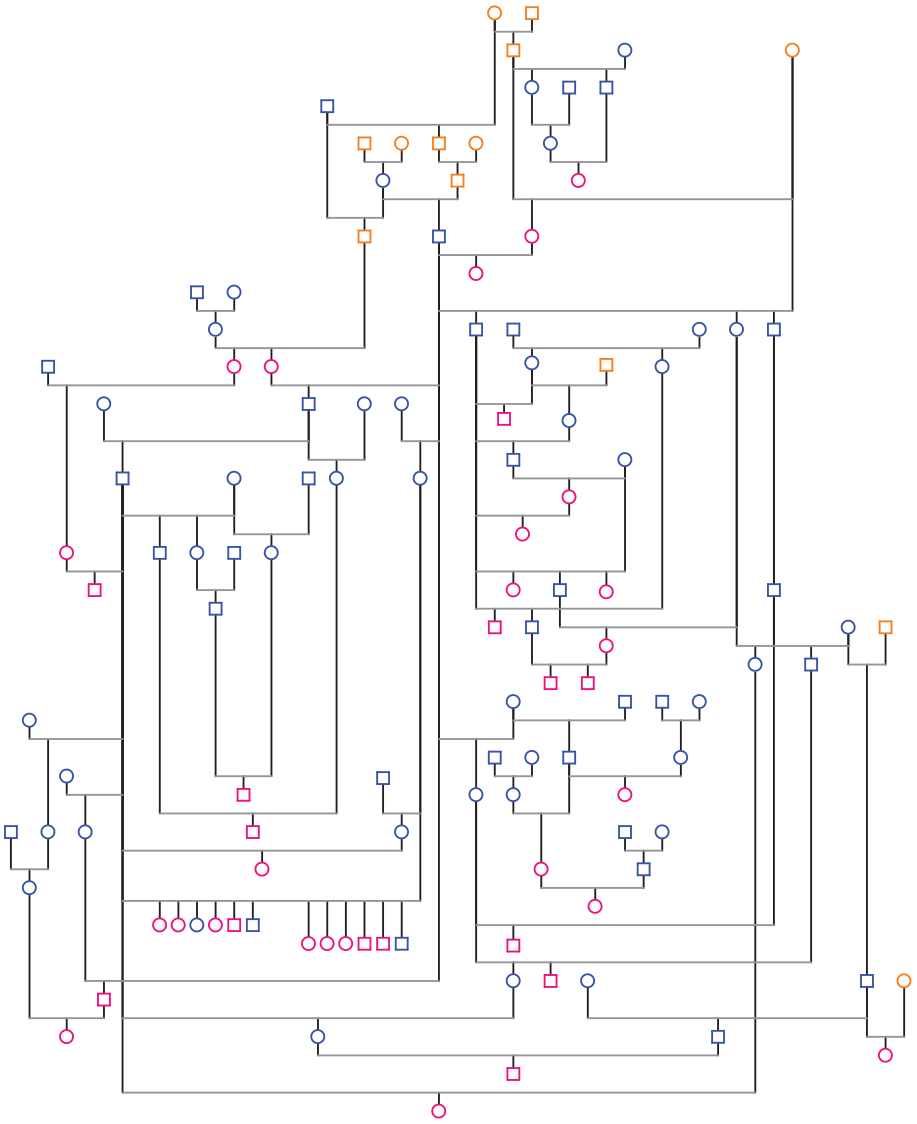

FIGURE 1 An illustration of a pedigree of 122 American Eskimo Dogs. Males are portrayed as squares and females are portrayed as circles. The illustration includes all 38 case dogs with diabetes mellitus (DM) (designated in red), all 71 control dogs without DM (designated in blue), and 13 of 47 dogs with unknown phenotype (designated in orange). Only 13 dogs with unknown phenotype are portrayed for graphic logistical reasons

TABLE 1 Heritability and predicted probabilities of diabetes mellitus (DM) by sex, as estimated by a mixed logistic regression model in American Eskimo Dogs

Parameter	Estimate	Lower 95% posterior interval limit	Upper 95% posterior interval limit
Predicted probability [DM|IF]	0.11	0.02	0.27
Predicted probability [DM|NF]	0.76	0.36	0.96
Predicted probability [DM|IM]	0.12	0.01	0.29
Predicted probability [DM|NM]	0.63	0.14	0.96
h^2	0.62	0.01	0.99

Abbreviations: DM, diabetes mellitus; IF intact female; IM intact male; NF neutered female; NM neutered male.

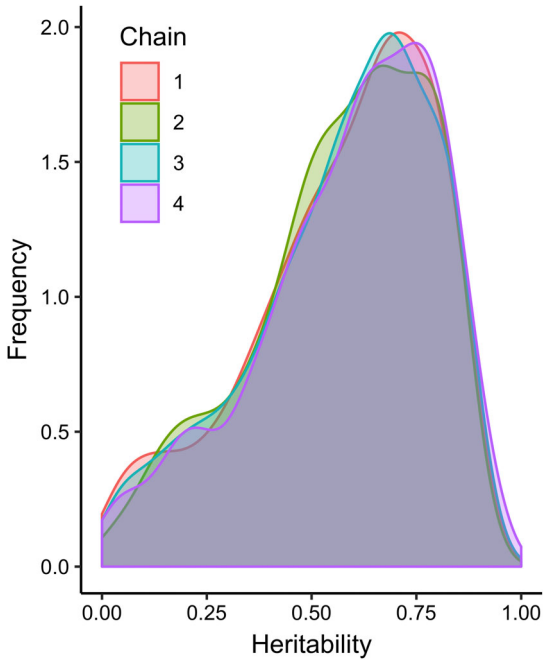

FIGURE 2 Posterior density plot of heritability for each of the 4 Markov Chain Monte Carlo simulations

verify the phenotype. The AKC states that the number of AED puppies from AKC litters between 2008 and 2017 was 5673 (personnel communication, AKC Canine Health Foundation Raleigh, NC 27675).

3.2 | Pedigree analysis

An illustration of the pedigree of 122 AED is provided in Figure 1. Results of the logistic regression analysis are summarized in Table 1. Most notable in this table is the estimate of heritability, where the mean value of the Markov Chain Monte Carlo samples is 0.62, accompanied by a posterior interval that spans all possible values of heritability. Figure 2 demonstrates this visually, being a plot of the frequency of each heritability value for each of the 4 chains that were simulated. Although unimodal, one can readily see that in a sample of this size where all the animals in the data are relatives of one another, the peak of this distribution still encompasses a broad set of plausible estimates.

The half-Cauchy distribution (half because only positive values were considered), which was used for the prior of the additive genetic SD, does not have a defined mean. With a simple random number generator available in R,[2] the mean and median of a half-Cauchy (0, 2.5) were readily computed as 21.4 and 2.5, respectively. Translating these half-Cauchy variables to represent heritability, the expected mean and median of the prior for heritability would be 0.71 and 0.86, respectively. In a sample of this size, the prior density is strongly, although not completely, reflected in the posterior density.

Table 2 provides a summary of the results of the complex segregation analysis, which are reported on a logistic scale. For brevity, results in Table 2 are reported for intact males, although all sexes provided similar values. The major locus models (dominant, recessive, and codominant) did not provide a significant improvement over the sporadic model, as judged by the larger AIC values for the major locus models. Although the general model provided the smallest AIC, it would be misleading to characterize this as providing the best explanation for the observed patterns of DM in this pedigree. The general model is not biologically feasible because the major locus transmission probabilities, which were estimated to be 0, 1.0, and 0.93 for the putative major locus genotypes of AA, AB, and BB, respectively (Table 2), are far from those dictated by the laws of Mendelian segregation. The Mendelian values, should a major locus be segregating in this pedigree, would approximate the expected transmission probabilities of 1.0, 0.5, and 0.0, for AA, AB, and BB, respectively. The sporadic model, which included only a polygenic term, provided a better biological explanation for the underlying genetic mechanisms of DM than did the general model permitting non-Mendelian segregation of a putative major locus.

Figure 3 presents the ROC curve taken from the predicted probabilities of disease as they were estimated in our Bayesian logistic model and serves as a means to quantify how well this model fits the data used to estimate the unknown sex and additive genetic parameters. The area under the curve, computed with the R-package pROC using the average values of the predicted probability of disease, was found to be 0.96, which suggests a very strong fit between observed and predicted disease states in this relatively small sample of dogs. [24] Moreover, the estimated threshold for predicting dogs as affected would be when the logistic model predictions of disease exceed 0.23, a value that would result in a specificity of 0.90 and a sensitivity of 1.0.

TABLE 2 Genetic models tested and their results from the complex segregation analysis of diabetes in intact male American Eskimo Dogs

Model	q	μ_{AA}	μ_{AB}	μ_{BB}	τ_{AA}	τ_{AB}	τ_{BB}	AIC
Sporadic[a]	…	−1.14	…	…	…	…	…	148.6
Dominant[b]	0.76	−1.78	−1.78	2.82	1.0	0.5	0.0	148.9
Recessive[b]	0.25	1.86	−1.77	−1.77	1.0	0.5	0.0	148.8
Codominant[b]	0.73	−1.54	−2.64	2.86	1.0	0.5	0.0	150.6
Environmental[c]	0.34	−2.01	−2.01	0.38	0.34[d]	0.34[d]	0.34[d]	152.1
General[e]	1.00	−3.31	−0.08	−2.59	0.0	1.0	0.93	123.9

Abbreviations: q = frequency of the A allele; μ_{AA}, μ_{AB}, μ_{BB} are logistic model parameter estimates for the putative major locus genotypes for intact males; τ_{AA}, τ_{AB}, τ_{BB} are the transmission probabilities for the putative A allele; AIC is the Akaike Information Criterion.

[a]The sporadic model considers no putative major locus effect but does consider a term for sex and an accommodation of a polygenic contribution to disease.

[b]The dominant, recessive, or codominant models are simple mixed major locus models, which follow the expected transmission of alleles outlined by Mendel.

[c]In the environmental model the polygenic term is removed and the transmission probabilities are set to being identically equal to the estimated allele frequency for all 3 putative major genotypes.

[d]0.34 = q in the environmental model.

[e]In the general model the transmission probabilities of the A allele were estimated from the data set.

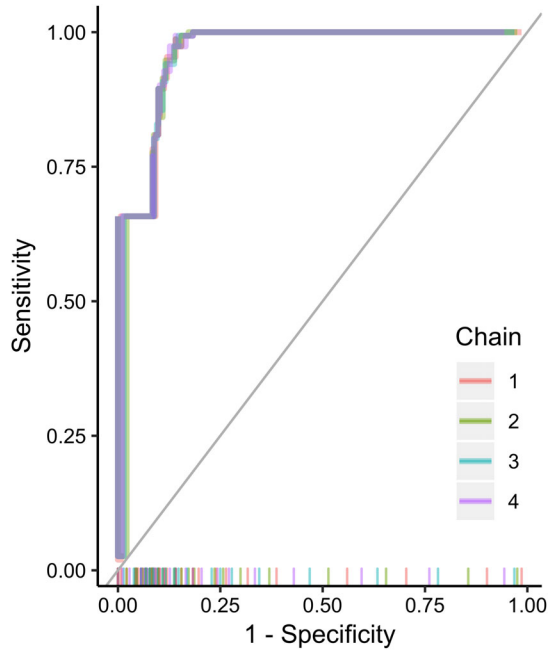

FIGURE 3 Receiver operating characteristic (ROC) curve for the predicted probability of diabetes mellitus in the logistic model for each of the 4 Markov Chain Monte Carlo simulations

4 | DISCUSSION

The large heritability point estimate of 0.62 indicates that there could be 1 or more loci with a profound impact on DM in AED. However, the wide 95% posterior interval, which spans the entire region of possible heritability estimates from 0 to 1, indicates that the magnitude of inheritance of DM should be interpreted with caution. It is important to be mindful of the fundamental of Bayesian analysis, which assumes that the "posterior is proportional to the likelihood times the prior." That is, in a data set of this size, the likelihood will not exert as large an influence on the posterior as we might prefer. Although the prior for the additive genetic SD is intended to be only weakly informative, the prior can exert more impact on the posterior density than is intended if the likelihood is also weakly informative. The estimated point heritability of 0.62 is slightly smaller than that expected from the half-Cauchy variables, indicating that the likelihood is providing some, although not a substantive amount, pressure to shrink heritability from our prior values. Nevertheless, these results are still encouraging, suggesting that breeders of AED could be successful in mounting a breeding program directed at reducing the incidence of DM in this breed. If the mean of the posterior of heritability is accepted as the most reasonable estimate for the data available, there remains the suggestion that there could be 1 or more loci with a significant effect on DM in AED, prompting the implementation of a complex segregation analysis.

The sporadic model, which included only a polygenic term, provided a better explanation for the underlying genetic mechanisms of DM than the major locus models. The general model is used to affirm a suspicion that 1 of the major locus models can explain the mode of inheritance if a major locus model has a smaller AIC than the sporadic model. However, in this study, the sporadic model had a smaller AIC than any of the major locus models, and, therefore, the general model was not used to affirm the superiority of any of the major locus models. The small AIC of the general model reflects a statistical rather than a biological finding. In any other statistical setting, that model with the lowest AIC is thought to provide the best explanation of the phenomenon under study. In this instance, however, that would not be the case, for the general model presented in Table 2 would lead to the conclusion that the major locus is inherited in a non-Mendelian fashion, which is a biological impossibility. As a result, the patterns of disease inheritance witnessed in this pedigree do not seem to be the result of a segregating locus of large effect. The likelihoods can also be used to contrast models with the likelihood ratio test, providing a possible test of significance. However, the small differences in AIC are a clear enough indication that no significant improvement of the model can be found in the incorporation of a putative major locus.

The complex segregation analyses suggest that the best model to describe the mode of inheritance of DM in AED is the polygenic sporadic model. Therefore, it may be concluded that a mixed model of inheritance (one with a locus of large effect modified by many additional loci of relatively smaller impact on disease risk) is not a plausible explanation for the pattern of inheritance of DM in this complex AED pedigree. It can be further concluded that DM transmission in AED follows a polygenic inheritance pattern with several large effect loci. This conclusion is in agreement with prior studies that have demonstrated associations between DM and single nucleotide polymorphisms in different genes in various breeds of dogs.[4,8-11] Breed-matched genome-wide association studies could help identify single nucleotide polymorphisms associated with DM in AED.

The challenge of estimating heritability in a relatively small data set is clear. However, over 600 AED were screened for possible inclusion into this study. Given the small population size of AED, with fewer than 6000 AKC registered AED born in the past 10 years, a reasonably large proportion of the AED population was actually screened for inclusion in this study. However, it is possible that sampling was not random and that the results are therefore not reflective of the heritability of DM in the breed at large. For example, it is possible that owners of dogs with DM were overrepresented because they were interested in contributing information to the study with the hope of improving the understanding of DM heritability in their dogs. Alternatively, it is possible that owners of dogs with DM were underrepresented because they were hesitant to share the disease status of their dogs.

An unexpected observation is reflected in the predictions of disease risk across sex classifications reported in Table 1. The 95% posterior interval for the predicted probability of DM in NF AED does not overlap with the 95% posterior interval for the predicted probability of DM in IF or IM AED. The conclusion from these data is that in AED, NF are at significantly higher risk for DM than IF and IM. This finding contradicts the current understanding that IF dogs are at increased risk for DM because higher progesterone and growth hormone concentrations contribute to insulin resistance, and because neutering can result in remission of DM, but is consistent with other inherited endocrine conditions that have a lower prevalence in intact dogs when compared to neutered dogs.[30-33] Importantly, the age of dogs with DM in the different sex categories was not different, suggesting that another unaccounted for variable associated with neuter status could be influencing this finding. Recall bias in which owners erroneously reported the neuter status of their dog in relation to the timing of DM diagnosis could contribute to this finding. Selection bias, in which owners of IF dogs with DM were hesitant to contribute data to the study because the dog was still breeding or had been bred, is also possible. Given the many biases that could have contributed to this finding and the physiologic rationale for prevention of DM with neutering, it is still recommended to neuter dogs as a measure of decreasing the risk of DM. Additional breed-wide prospective studies are warranted to confirm the relationship between neuter status and DM in AED.

American Eskimo Dogs from countries other than the United States and Canada were excluded by design because geographic location can influence disease risk. [34] The Canadian dogs included in this study were directly related to dogs residing in the United States. Future studies of DM in AED in other geographic regions are needed to determine if the heritability patterns of DM in AED reported here are unique to North America or observed worldwide. Future studies focusing on the heritability patterns of DM in the Nordic Spitz clade, which includes the AED, Icelandic Sheepdog, Keeshonds, Norwegian Elkhound, and Swedish Vallhund breeds, would also be interesting.[34]

This study has several limitations, including small sample size and possible recall and selection biases as described above. Furthermore, some dogs that are currently healthy could develop DM in the future, and this might contribute to phenotype misclassification. However, dogs in the control group were significantly older than dogs with DM at the time of DM onset, minimizing the risk that control dogs will develop DM after study enrollment. Finally, not all dog owners could be contacted, and as a result, some dogs had an unknown phenotype.

In conclusion, the heritability estimate of DM in this AED population is high with a mode of inheritance that is consistent with a polygenic, non-Mendelian pattern of disease transmission. Judicial breeding practices should be successful at decreasing the incidence of DM in AED; breeders will need to be particularly diligent as DM onset in AED occurs at about 6 years of age, so it is likely for dogs to be bred before their DM phenotype is known. This study demonstrates that the genetic nature of DM in AED underscores the importance of tracking lineage of breeding dogs. For instance, breeders could select sires and dams, who are not related to dogs with DM or closely related to one another, while being mindful of avoiding genetic bottlenecks in small populations. Future genome-wide association studies could help identify single nucleotide polymorphisms in genes that increase the risk for DM in AED. Studies of the heritability and mode of inheritance of DM in other breeds are needed to determine if the findings of this study are unique to AED or shared by other breeds.

ACKNOWLEDGMENTS

The study was made possible by a gift from Ms. Catharine Adler. The authors thank the American Kennel Club and American Eskimo Dog Club of America (AEDCA) for promoting access to the online survey (Appendix S1).

CONFLICT OF INTEREST DECLARATION

Authors declare no conflict of interest.

OFF-LABEL ANTIMICROBIAL DECLARATION

Authors declare no off-label use of antimicrobials.

INSTITUTIONAL ANIMAL CARE AND USE COMMITTEE (IACUC) OR OTHER APPROVAL DECLARATION

Authors declare no IACUC or other approval was needed.

HUMAN ETHICS APPROVAL DECLARATION

Authors declare human ethics approval was not needed for this study.

ORCID

Rebecka S. Hess https://orcid.org/0000-0002-3134-1348

REFERENCES

1. Hess RS, Kass PH, Ward CR. Breed distribution of dogs with diabetes mellitus admitted to a tertiary care facility. *J Am Vet Med Assoc*. 2000; 216:1414-1417. https://doi.org/10.2460/javma.2000.216.1414.
2. Guptill L, Glickman L, Glickman N. Time trends and risk factors for diabetes mellitus in dogs: analysis of veterinary medical data base records (1970-1999). *Vet J*. 2003;165:240-247. https://doi.org/10.1016/S1090-0233(02)00242-3.
3. Fall T, Hamlin HH, Hedhammar Å, Kämpe O, Egenvall A. Diabetes mellitus in a population of 180,000 insured dogs: incidence, survival, and breed distribution. *J Vet Intern Med*. 2007;21:1209-1216. https://doi.org/10.1111/j.1939-1676.2007.tb01940.x.
4. Catchpole B, Adams J, Holder A, Short A, Ollier W, et al. Genetics of canine diabetes mellitus: are the diabetes susceptibility genes identified in humans involved in breed susceptibility to diabetes mellitus in dogs? *Vet J*. 2013;195(2):139-147. https://doi.org/10.1016/j.tvjl.2012.11.013.
5. Kimmel SE, Ward CR, Henthorn PS, Hess RS. Familial insulin-dependent diabetes mellitus in Samoyed dogs. *J Am Anim Hosp Assoc*. 2002;38:235-238. https://doi.org/10.5326/0380235.
6. Kramer JW, Klaassen JK, Baskin DG, Prieur DJ, Rantanen NW, et al. Inheritance of diabetes mellitus in keeshond dogs. *Am J Vet Res*. 1988;49:428-431.
7. Hess RS, Saunders M, Van Winkle TJ, Ward CR. Concurrent disorders in dogs with diabetes mellitus: 221 cases (1993-1998). *J Am Vet Med Assoc*. 2000;217:1166-1173.
8. Kennedy L, Davison L, Barnes A, Short A, Fretwell N, et al. Identification of susceptibility and protective major histocompatibility complex haplotypes in canine diabetes mellitus. *Tissue Antigens*. 2006;68:467-476. https://doi.org/10.1111/j.1399-0039.2006.00716.x.
9. Short A, Saleh N, Catchpole B, Kennedy L, Barnes A, et al. CTLA4 promoter polymorphisms are associated with canine diabetes mellitus. *Tissue Antigens*. 2010;75:242-252. https://doi.org/10.1111/j.1399-0039.2009.01434.x.
10. Short A, Holder A, Rothwell S, Massey J, Scholey R, et al. Searching for "monogenic diabetes" in dogs using a candidate gene approach. *Canine Genet Epidemiol*. 2014;1:8. https://doi.org/10.1186/2052-6687-1-8.
11. Short AD, Catchpole B, Kennedy LJ, et al. Analysis of candidate susceptibility genes in canine diabetes. *J Hered*. 2007;98:518-525.
12. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, et al. Stan: a probabilistic programming language. *J Stat Softw*. 2017;76:1-33. https://doi.org/10.18637/jss.v076.i01.
13. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018 https://www.R-project.org/.
14. Gelman A, Carlin JB, Stern HS, Dunson D, Vehtari A, Rubin D. *Bayesian data analysis*. 3rd ed. Boca Raton: CRC Press; 2013.
15. Damgaard LH. Technical note: how to use Winbugs to draw inferences in animal models. *J Anim Sci*. 2007;85:1363-1368. https://doi.org/10.2527/jas.2006-543.
16. Quaas RL. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics*. 1976;32:949-953. https://doi.org/10.2307/2529279.
17. Gelman A, Jakulin A, Pittau MG, Su Y-S. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat*. 2008;2:1360-1383. https://doi.org/10.1214/08-AOAS191.
18. Gelman A, Simpson D, Betancourt M. The prior can often only be understood in the context of the likelihood. *Entropy*. 2017;19:555. https://doi.org/10.3390/e19100555.
19. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal*. 2006;1:515-534. https://doi.org/10.1214/06-BA117A.
20. Gianola D, Foulley J. Sire evaluation for ordered categorical data with a threshold model. *Genet Sel Evol*. 1983;15:201-224. https://doi.org/10.1186/1297-9686-15-2-201.
21. Fernández-i-Marín X. Ggmcmc: analysis of MCMC samples and Bayesian inference. *J Stat Softw*. 2016;70(9):1-20. https://doi.org/10.18637/jss.v070.i09.
22. Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7:457-472.
23. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med*. 2000;45:23-41. https://doi.org/10.1016/S0167-5877(00)00115-X.
24. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77. https://doi.org/10.1186/1471-2105-12-77.
25. Lalouel J, Rao D, Morton N, Elston RC. A unified model for complex segregation analysis. *Am J Hum Genet*. 1983;35:816.
26. S.A.G.E. Statistical Analysis for Genetic Epidemiology (v6.4) http://darwin.cwru.edu. 2016.
27. Elston R, Stewart J. A general model for the genetic analysis of pedigree data. *Hum Hered*. 1971;21:523-542.
28. Karunaratne PM, Elston RC. A multivariate logistic model (MLM) for analyzing binary family data. *Am J Med Genet*. 1998;76:428-437.
29. Axenovich TI, Zorkoltseva IV, Liu F, Kirichenko AV, Aulchenko YS. Breaking loops in large complex pedigrees. *Hum Hered*. 2008;65:57-65. https://doi.org/10.1159/000108937.
30. Fall T, Kreuger J, Juberget A, Bergstrom A, Hedhammar A. Gestational diabetes mellitus in 13 dogs. *J Vet Intern Med*. 2008;22:1296-1300. https://doi.org/10.1111/j.1939-1676.2008.0199.x.
31. Fall T, Hedhammar Å, Wallberg A, Fall N, Ahlgren K, et al. Diabetes mellitus in elkhounds is associated with Diestrus and pregnancy. *J Vet Intern Med*. 2010;24:1322-1328. https://doi.org/10.1111/j.1939-1676.2010.0630.x.
32. Pöppl AG, Mottin TS, González FH. Diabetes mellitus remission after resolution of inflammatory and progesterone-related conditions in bitches. *Res Vet Sci*. 2013;94:471-473. https://doi.org/10.1016/j.rvsc.2012.10.008.
33. Sundburg CR, Belanger JM, Bannasch DL, Famula TR, Oberbauer AM. Gonadectomy effects on the risk of immune disorders in the dog: a retrospective study. *BMC Vet Res*. 2016;12:278. https://doi.org/10.1186/s12917-016-0911-5.
34. Parker H, Dreger D, Rimbault M, et al. Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. *Cell Rep*. 2017;19:697-708. https://doi.org/10.1016/j.celrep.2017.03.079.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Cai SV, Famula TR, Oberbauer AM, Hess RS. Heritability and complex segregation analysis of diabetes mellitus in American Eskimo Dogs. *J Vet Intern Med*. 2019;33:1926–1934. https://doi.org/10.1111/jvim.15570