

## RESEARCH ARTICLE

# True Colors: Commercially-acquired morphological genotypes reveal hidden allele variation among dog breeds, informing both trait ancestry and breed potential

Dayna L. Dreger<sup>1</sup>✉‡, Blair N. Hooser<sup>1</sup>✉‡, Angela M. Hughes<sup>2</sup>, Balasubramanian Ganesan<sup>2</sup>, Jonas Donner<sup>3</sup>, Heidi Anderson<sup>3</sup>, Lauren Holtvoigt<sup>2</sup>, Kari J. Ekenstedt<sup>1\*</sup>

**1** Department of Basic Medical Sciences, College of Veterinary Medicine, Purdue University, West Lafayette, IN, United States of America, **2** Wisdom Health, Vancouver, WA, United States of America, **3** Wisdom Health, Helsinki, Finland

✉ These authors contributed equally to this work.

‡ These authors are joint first authors on this work.

\* [kje0003@purdue.edu](mailto:kje0003@purdue.edu)



## OPEN ACCESS

**Citation:** Dreger DL, Hooser BN, Hughes AM, Ganesan B, Donner J, Anderson H, et al. (2019) True Colors: Commercially-acquired morphological genotypes reveal hidden allele variation among dog breeds, informing both trait ancestry and breed potential. PLoS ONE 14(10): e0223995. <https://doi.org/10.1371/journal.pone.0223995>

**Editor:** Carlos E. Ambrósio, Faculty of Animal Sciences and Food Engineering, University of São Paulo, BRAZIL

**Received:** June 5, 2019

**Accepted:** October 2, 2019

**Published:** October 28, 2019

**Copyright:** © 2019 Dreger et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The full genotype dataset is available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.ghx3ffbj4>.

**Funding:** Partial support for KJE was provided by the Office of the Director, National Institutes of Health (NIH) under award number K01-OD027051. This funder had no role in study design, data collection and analysis, decision to publish, or

## Abstract

Direct-to-consumer canine genetic testing is becoming increasingly popular among dog owners. The data collected therein provides intriguing insight into the current status of morphological variation present within purebred populations. Mars WISDOM PANEL™ data from 11,790 anonymized dogs, representing 212 breeds and 4 wild canine species, were evaluated at genes associated with 7 coat color traits and 5 physical characteristics. Frequencies for all tested alleles at these 12 genes were determined by breed and by phylogenetic grouping. A sub-set of the data, consisting of 30 breeds, was divided into separate same-breed populations based on country of collection, body size, coat variation, or lineages selected for working or conformation traits. Significantly different ( $p \leq 0.00167$ ) allele frequencies were observed between populations for at least one of the tested genes in 26 of the 30 breeds. Next, standard breed descriptions from major American and international registries were used to determine colors and tail lengths (e.g. genetic bobtail) accepted within each breed. Alleles capable of producing traits incongruous with breed descriptions were observed in 143 breeds, such that random mating within breeds has probabilities of between  $4.9e^{-7}$  and 0.25 of creating undesirable phenotypes. Finally, the presence of rare alleles within breeds, such as those for the recessive black coloration and natural bobtail, was combined with previously published identity-by-descent haplotype sharing levels to propose pathways by which the alleles may have spread throughout dog breeds. Taken together, this work demonstrates that: 1) the occurrence of low frequency alleles within breeds can reveal the influence of regional or functional selection practices; 2) it is possible to visualize the potential historic connections between breeds that share rare alleles; and 3) the necessity of addressing conflicting ideals in breed descriptions relative to actual genetic potential is crucial.

preparation of the manuscript. This work was additionally made possible through the financial support of a commercially-affiliated funder, Wisdom Health, Vancouver, WA, USA, as follows: several co-authors are employed by (AMH, BG, JD, HA, and LH) or have previously consulted (KJE) for this company. Wisdom Health additionally funded a Veterinary Summer Scholars position (BNH) in summer 2018. Wisdom Health and the employee co-authors had a significant role in providing raw data (data collection) and assisting with the final editing of the manuscript, but only a very minor role in study design, data analysis, and the decision to publish. The study was managed and executed primarily by researchers at Purdue University (DLD, BNH, and KJE), and these authors are responsible for final scientific content. The specific roles of all authors are articulated in the 'author contributions' section.

**Competing interests:** The following authors have a current (employees: AMH, BG, JD, HA, LH) or former (consultancy: KJE) commercial affiliation with Wisdom Health. Wisdom Health markets an array of genetic tests to consumers, including the Wisdom Panel, from which data was used in the present work. Our funding sources do not alter our adherence to PLOS ONE policies on sharing data and materials, excepting those portions of the Wisdom Panel which are proprietary intellectual property.

## Introduction

Guided by human selection, the domestic dog (*Canis lupus familiaris*) has become one of the most physically diverse species, with hundreds of recognized breeds, differentiated from each other by specific morphological and behavioral characteristics [1]. One of the most readily accessible and easily visualized defining characteristics of dog breeds is in their presentation of pigmentation and color patterns. The nomenclature used currently to refer to dog pigmentation genetics was outlined by Little in 1957 [2]. Scientific advancement since that time has allowed for the validation of many of his postulated loci and the identification of causal genetic variants for many of the common coat colors in the species. These coat color and trait variants are increasingly included in commercially-available genetic test panels, the emergence of which has contributed greatly to the acceptance and implementation of genetic screening of purebred dogs by owners, breeders, and veterinarians. In addition to providing valuable information regarding the genetic status of potential breeding dogs in terms of disease state and morphological traits, the combined genotype databases collected by these commercial entities provide a valuable resource for monitoring the diversity of selected alleles, such as those driving coat color, within breeds.

The entirety of coat color patterns in mammals consist of spatial and temporal production of pheomelanin, yellow- to red-based pigment, and eumelanin, black- to brown-based pigment, controlled by the interaction of multiple genes. In dogs, variation at the *Agouti Signaling Protein (ASIP)* gene determines the distribution of eumelanin and pheomelanin across the body of the dog and along the individual hair shafts, as dictated by a four-allele dominance hierarchy:  $a^y$  (fawn)  $>$   $a^w$  (wolf sable)  $>$   $a^t$  (tan points)  $>$   $a$  (recessive black) [3–5]. However, regardless of the accompanying *ASIP* genotype, the *Melanocortin 1 Receptor (MC1R)* gene determines the ability of a melanocyte to produce eumelanin at all, presenting its own dominance hierarchy of four alleles:  $E^M$  (melanistic mask)  $>$   $E^G$  (grizzle/domino)  $>$   $E$  (wild type)  $>$   $e$  (recessive red) [6–8]. While the  $E^M$ ,  $E^G$ , and  $E$  alleles all allow for the production of both eumelanin and pheomelanin, dependent on the pattern dictated by *ASIP*, homozygous inheritance of  $e$  prevents eumelanin production, resulting in a completely pheomelanin color [6–8]. Further, the dominant derived  $K^B$  allele of *Canine Beta-Defensin 103 (CBD103)* prevents the production of patterning by *ASIP*, resulting in a solid eumelanin phenotype when combined with a dominant *MC1R* genotype, or a solid pheomelanin phenotype when combined with a *MC1R* genotype of  $e/e$  [9]. In this way, inheritance of homozygous  $e$  at *MC1R* or dominant  $K^B$  at *CBD103* are epistatic to any of the *ASIP* phenotypes [8,9]. In addition to the dominant  $K^B$  and recessive  $k^y$  alleles of *CBD103*, an intermediate allele,  $k^{br}$ , produces the brindle coat color pattern with alternating stripes of eumelanin and pheomelanin [10]. However, due to the complex structural nature of the  $k^{br}$  allele, which has never been reliably defined and remains unpublished, it was not possible to distinguish between it and  $K^B$  for the purposes of this paper.

The base pheomelanin and eumelanin pigments can be altered by a number of modifier genes, of which *Tyrosinase Related Protein 1 (TYRP1)*, *Proteasome Subunit Beta 7 (PSMB7)*, *Microphthalmia-associated Transcription Factor (MITF)*, and *RALY heterogeneous nuclear ribonucleoprotein (RALY)* are explored further in the scope of this paper. *TYRP1* alters all eumelanin in hair and skin evenly from black to brown, and presents as a compound heterozygote with at least four alternate recessive alleles:  $b^s$ ,  $b^c$ ,  $b^d$ , and a variant that has thus far only been described in the Australian Shepherd [11–13]. *PSMB7* expression produces the harlequin pattern of white, dilute, and deeply pigmented patches only when inherited along with the merle phenotype of *Premelanosome 17 (PMEL17)*, not tested in the present study [14–16]. The harlequin variant is homozygous lethal, so the pigmentation phenotype is only expressed when

inherited heterozygously [14]. *MITF* expression produces white spotting on top of a regularly pigmented background and, depending on the breed background involved, is inherited as a co-dominant or recessive phenotype [17,18]. Expression of *RALY*, in combination with other yet-unknown genetic modifiers, alters a tan point background pattern, produced by the  $a^t$  allele of *ASIP*, such that the tan points, normally restricted to the paws, muzzle, eyebrows, and chest, extend up the extremities, forming a eumelanin “saddle-shaped” pattern on the dog’s dorsal surface [19].

Expanding beyond pigmentation patterns, other phenotypic traits that are controlled by a relatively small number of genes and can be easily visualized include hair length and texture, tail length, muzzle length, and ear shape. Three genes, *Fibroblast Growth Factor 5* (*FGF5*), *Respondin 2* (*RSPO2*, not tested in the present study), and *Keratin 71* (*KRT71*), define much of the coat type variation between dog breeds [20–24]. Specifically, expression of recessive *FGF5* variants produce a long coat, a dominant *RSPO2* variant produces longer hair specifically on the muzzle and eyebrows, and at least two variants in *KRT71*, one of which was tested in this study, produce a curly coat [20–24]. The tailless trait, whereby a dog is born with a truncated or absent tail, is caused in some breeds by a variant in the *T brachyury transcription factor* (*T*) gene [25,26]. Like harlequin, taillessness is homozygous lethal, so will be expressed viably only in heterozygotes [25,26]. Ear set and muzzle length are traits that are very likely impacted by numerous genetic variants, however a marker on canine chromosome 10 is known to segregate in some cases for erect versus drop ears [27,28], and the *Bone Morphogenic Protein 3* (*BMP3*) gene is associated with foreshortening of the face [29]. A variant in *SMOC2* has also been shown to contribute to muzzle length, though it is not included in the present study [30].

Epistatic effects are prevalent within coat color and morphological variation in the dog. We have previously mentioned epistasis between a homozygous non-functional genotype at *MC1R* or a dominant variation at *CBD103* with the ability to express *ASIP*-driven phenotypes. However, canine coat color also presents scenarios whereby a specific allelic background is required for expression of a modifier. This is exemplified in the requirement of the merle phenotype of *PMEL17* in order to express the harlequin pattern of *PSMB7*, the necessity for an  $a^t$  tan point base pattern at *ASIP* to display the *MC1R* grizzle or *RALY* saddle tan phenotypes, and a moderate to long hair length to produce a *KRT71* curly coated phenotype. On an incompatible background, presence of some gene variants may remain unexpressed for generations or, as many breeds are fixed for only a small number of phenotype options, may remain completely unobserved. As national and international breed organizations are charged with defining their breed’s characteristics, which are then regulated by registering bodies such as the American Kennel Club (AKC), United Kennel Club (UKC), The Kennel Club (KC) in the United Kingdom, or Fédération Cynologique Internationale (FCI), failure of breed standards to account for rare—though naturally occurring—variation can lead to frustration or confusion when unexpected traits are expressed. That same existence of rare variants within breeds—which, due to epistasis and genetic background, may never express the associated phenotypes—can provide intriguing information regarding the development of, or relationships between, breeds throughout history.

We utilized custom genotyping array data from Mars Wisdom Health for 11,790 canids, representing 212 pure breeds and 4 wild canine populations (S1 Table), genotyped for seven coat color and five physical characteristic genes, as curated by OMIA [31] (Table 1). For each of these genes, we determined the frequency of each allele within each breed. Note that for one gene (*TYRP1*) not all alleles could be tested, these are indicated in Table 1. Our primary objective for this study was to evaluate: 1) the breed-type distribution of morphologic variants; 2) the implications of founder effects and/or selection preferences between geographically or behaviorally independent populations of the same breed; 3) the breed-specific carrier status of

**Table 1. Genes and markers genotyped for breed analysis.**

Gene	OMIA Reference <sup>a</sup>	Allele	Variant Details	Reference
<i>Agouti signaling protein</i>	000201–9615	<i>a<sup>y</sup></i>	p.A82S and p.R83H	[4]
		<i>a<sup>w</sup></i>	(wild-type)	
		<i>a<sup>t</sup></i>	g.23365298ins239	[5]
		<i>a</i>	p.R96C	[3]
<i>Melanocortin receptor 1</i>	001199–9615	<i>E<sup>M</sup></i>	p.V264M	[6]
		<i>E<sup>G</sup></i>	p.G78V	[7]
		<i>E</i>	(wild-type)	
		<i>e</i>	p.R306ter	[8]
<i>Canine beta-defensin 103</i>	001416–9615	<i>K<sup>B</sup>/k<sup>br</sup></i>	g.58965449_58965451ins>del	[9,10]
		<i>k<sup>y</sup></i>	(wild-type)	
<i>Tyrosinase related protein 1</i>	001249–9615	<i>B</i>	(wild-type)	
		<i>b<sup>c</sup></i>	p.C41S	[11]
		<i>b<sup>s</sup></i>	p.Q331ter	[11]
		<i>b<sup>d</sup></i>	(not tested)	[11]
		<i>b<sup>a</sup></i>	(not tested)	[12,13]
<i>Microphthalmia associated transcription factor</i>	000214–9615	<i>S</i>	(wild-type)	
		<i>s<sup>p</sup></i>	g.21836232_21836427ins>del	[17]
<i>Proteasome subunit beta 7</i>	001454–9615	<i>H</i>	(wild-type)	
		<i>h</i>	p.V6G	[14]
<i>RALY heterogeneous nuclear ribonucleoprotein</i>	001806–9615	<i>saddle</i>	g.23252754_23252770dup CCCCAGGTCAGAGTTT	[19]
		<i>wild-type</i>	(wild-type)	
<i>Fibroblast growth factor 5</i>	000439–9615	<i>short</i>	(wild-type)	
		<i>long</i>	p.C95F	[20]
<i>Keratin 71, type II</i>	000245–9615	<i>curl</i>	p.R151W	[20]
		<i>no curl</i>	(wild-type)	
<i>T brachyury transcription factor</i>	000975–9615	<i>tailless</i>	p.I63M	[25]
		<i>wild-type</i>	(wild-type)	
<i>Bone morphogenetic protein 3</i>	001551–9615	<i>wild-type</i>	(wild-type)	
		<i>short muzzle</i>	p.F448L	[29]
<i>Genome-wide association region</i>	-	<i>drop ear</i>	Chr10:11072007 <sup>b</sup>	[27]
		<i>erect ear</i>	(wild-type)	

<sup>a</sup>OMIA is the Online Mendelian Inheritance in Animals database, which catalogs broadly resourced information on the genetics of traits and disorders in animals [31].

<sup>b</sup> CanFam2.0

<https://doi.org/10.1371/journal.pone.0223995.t001>

variants disallowed within breed standard descriptions; and 4) the ancestral connections between breeds that share rare trait-causing variants.

## Results

### Unexpectedly broad distribution of trait-causing alleles across breeds

Marker genotypes were combined to interpret the actual biallelic genotypes for each of the queried genes for every dog in the dataset. Individuals of the same breed were combined to calculate breed allele frequencies and are reported in S2 Table. Breeds were assigned to phylogenetic clades as previously reported [32,33] (S1 Table). Breeds not included in the earlier phylogenetic studies were assigned to defined clades based on known breed history, phenotypic commonalities, and geographic region of origin (indicated by parentheses in S1 Table).

Over all combined breeds, the ancestral allele (Table 1) predominates at all genes, except for *ASIP* and *MC1R*, where derived alleles account for 82% and 57% of alleles, respectively. The derived alleles at lowest frequency across all 11,790 canids are  $E^G$  (grizzle/domino) of *MC1R* and *tailless* of *T*, each representing ~1% of all alleles detected within their respective genes.  $E^G$ , which produces the grizzle or domino pattern when combined with  $a^t$  tan points at *ASIP*, is present in 28 breeds, with highest frequencies recorded in Borzois (56%), Polish Greyhounds (43%), and Salukis (31%). Cladistic representation of the  $E^G$  allele is heavily weighted to those breeds in the Mediterranean and UK Rural clades (Fig 1), with 60% and 31% of all  $E^G$  alleles in our dataset arising from those clades, respectively.

Assignment of biallelic genotypes for the *ASIP* gene based on genotype tests for  $a^y$ ,  $a^t$ , and  $a$ , and a genotype of  $a^w$  by exclusion of the other three alleles, revealed the presence of unusual triple-allele combinations in a few dogs. In these cases, dogs would appear to genotype as  $a^y/a^t$ , plus an additional third allele. This genotyping outcome reflects a scenario where the  $a^y$  point mutations and the  $a^t$  SINE insertion occur on the same chromosome. The resultant combinatorial allele, termed  $a^{yt}$ , was identified in 14 domestic breeds across 10 clades, and the Dingo (S2 Table). Phenotype descriptions of the evaluated dogs were not available, so the effect of the  $a^{yt}$  allele on phenotype is presently unknown. The breed with the highest frequency of  $a^{yt}$  is the Dogo Argentino (33%), a predominantly white mastiff breed.

The *tailless* allele was detected in 48 breeds, with greatest frequencies in the Tenterfield Terrier (30%), Swedish Vallhund (18%), Spanish Water Dog (14%) and Australian Shepherd (13%). The remaining 44 breeds carry the *tailless* allele at <10%.

The recessive  $a$  allele of *ASIP* has previously been reported in six breeds (German Shepherd Dog, Belgian Shepherd, Schipperke, Australian Shepherd, Shetland Sheepdog, and Eurasier) [3–5]. We identified the allele in 89 breeds, documenting for the first time the presence of the allele in an additional 83 breeds (S2 Table). The *RALY* duplication required for the *saddle* modification of the tan point phenotype was detected in 203 breeds (S2 Table). The  $E^M$  “melanistic mask” allele of *MC1R* has previously been identified in only 11 breeds [3,7]. We identified  $E^M$  as present in 164 breeds, 153 not previously documented (S2 Table). In many breeds, the mask phenotype is not observable due to epistatic interactions with non-compatible genotypes, such as solid eumelanin or solid white.

Four populations of wild canines were included in the analyses, although they are not counted as “breeds” in the analyses presented here. The wild canine populations had high levels of wild-type fixation across all genes evaluated (S2 Table). Eastern coyotes ( $n = 29$ ) show a 2% frequency of the *MC1R*  $E^G$  allele and a 2% frequency of the *TYRP1*  $b^s$  allele. The Western coyotes ( $n = 19$ ) have a 21% frequency of the *FGF5* *long* allele. The Dingo samples ( $n = 12$ ) show an 8% frequency of the *TYRP1*  $b^s$  allele. The marker for ear shape, and the causal variants in *RALY*, *MITF*, and *ASIP* are moderately variable across all four wild populations.

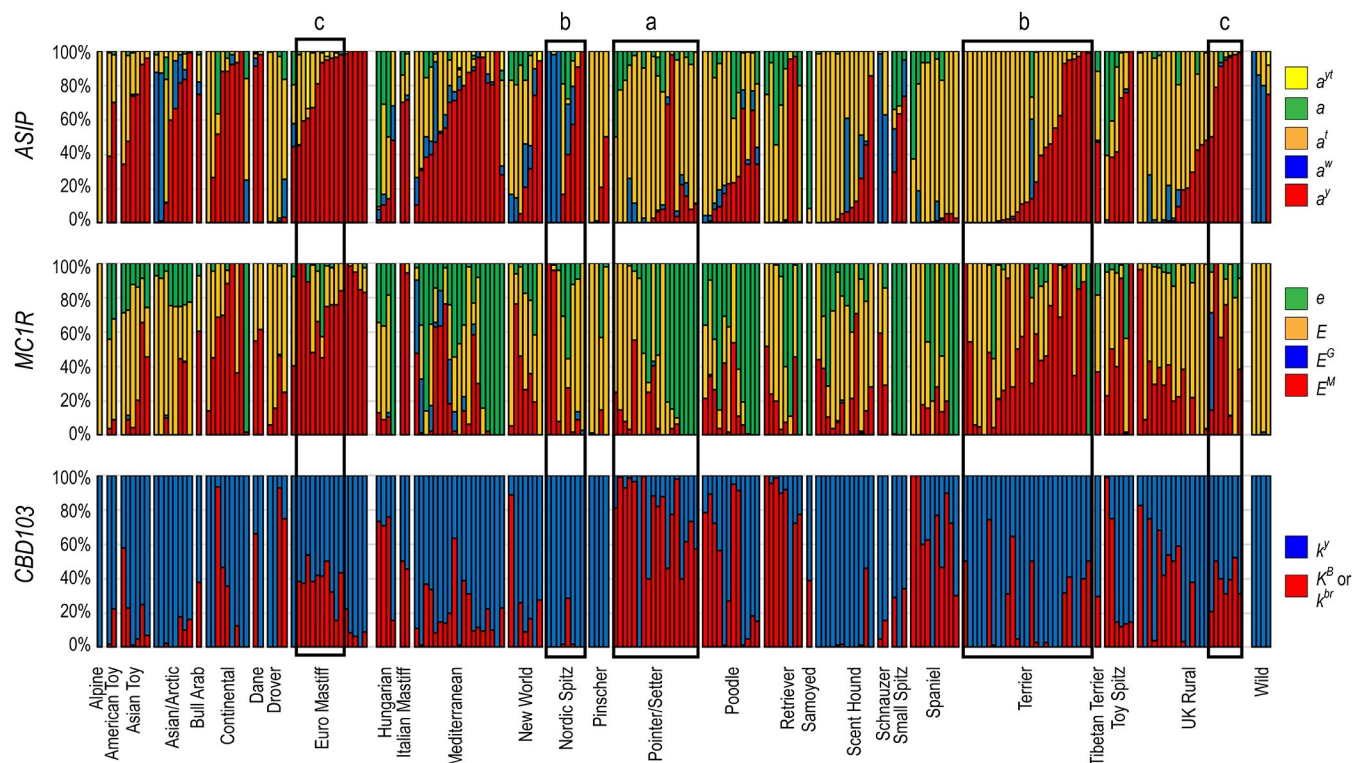
### Cladistic patterns of allele distribution

Allele frequencies of *ASIP*, *MC1R*, and *CBD103* are graphically represented in Fig 1, relative to the phylogenetic groupings of the breeds. Association of allele frequency with cladistic assignment highlights multiple key tendencies in regards to color preference within breed-type groups. The cladistic distribution of *CBD103* alleles reflects a tendency toward fixation of a single allele within related breeds. The Hungarian, Pointer/Setter, Poodle, and Retriever clades contain breeds nearing fixation for the  $K^B/k^{br}$  variant, presumably primarily  $K^B$  due to the predominantly solid-colored phenotypes of breeds in these clades. Conversely, the Asian/Arctic, Mediterranean, New World, Scent Hound, and Terrier clades consist of breeds nearing fixation for  $k^y$ . Since the  $K^B$  allele produces a solid colored animal and is epistatic to alleles of *ASIP*

and all *MC1R* alleles except for *e*, variation of *MC1R* and *ASIP* alleles is relatively uncontrolled in breeds with a high frequency of  $K^B$ . Comparatively, clades with high frequency of the  $k^y$  allele show greater degrees of preference for the patterning alleles of *ASIP*, specifically  $a^y$ ,  $a^w$ , and  $a^t$ . Clades with strong heterogeneity of *CBD103*, represented as multiple breeds within a clade with  $K^B$  frequency of 40–50% (e.g. Euro Mastiff, portions of UK Rural), are also those that consist of breeds presenting the brindle coloration. This is an accurate reflection of the present inability to distinguish between the brindle  $k^{br}$  and dominant black  $K^B$  alleles due to the complex genetic structure of the variants. Comparatively, clades with multiple breeds that show a 40–50% frequency of  $K^B$  likewise have a high frequency of  $a^y$ , the *ASIP* background that is necessary to produce brindling across the entire body of the dog.

The wild-type  $a^w$  allele of *ASIP*, while predominant in three of the four wild canine populations, is at greatest frequency in the spitz clades (Asian/Arctic, Nordic Spitz, Schnauzer). For *MC1R*, the *e* allele was identified in 24 of the 28 clades represented in this data, but it is seen at highest frequencies in hunting breeds (Pointer/Setter, Poodle, Retriever, Spaniel clades), continental sighthounds and flock guardians (Hungarian, Mediterranean clades), and white spitz breeds (Samoyed, Small Spitz clades). The melanistic mask allele,  $E^M$ , was likewise widespread, observed here in 164 breeds across 24 clades. Finally, the  $E^G$  allele is most prevalent in the Mediterranean clade, represented only sporadically in 11 other clades.

Allele frequencies for the nine remaining genes tested, sorted relative to phylogenetic breed relationships, are presented in S1 Fig. With the exception of the harlequin mutation, detected



**Fig 1. Allele frequencies for *ASIP*, *MC1R*, and *CBD103* by phylogenetic breed relationship.** Breeds are grouped by phylogenetic clade, as described, and sorted within clade by frequency of *e* and then  $a^y$  to demonstrate patterns of phenotype expression across interacting genes. Thick black boxes highlight examples of color preference influencing interacting genes: a) Pointer/Setter breeds are commonly seen in solid colors, caused by the  $K^B$  allele of *CBD103* or the *e/e* genotype of *MC1R*. Since the solid-color genotypes are epistatic to the dominant alleles of *MC1R* and all alleles of *ASIP*, variation at those genes does not follow a trend for color preference. b) Related breeds with high frequency of the  $k^y$  allele of *CBD103* have a more structured pattern of *ASIP* allele preference. c) Breeds with a preference for the brindle pattern show heterogeneity for  $K^B/k^{br}$  (reflective of a  $k^{br}$  phenotype) and high frequency of  $a^y$ , required for the expression of brindle across the whole body.

<https://doi.org/10.1371/journal.pone.0223995.g001>

only in Great Dane and Yorkshire Terrier, all remaining gene variants were present broadly across all clades. There is no apparent excess of specific *TYRP1* brown alleles in any given clade or breed, despite, or perhaps due to, the identical phenotype produced by both alleles. Breeds fixed for the *MITF* SINE insertion, associated with piebald white spotting, are most abundant in the Pointer/Setter and Terrier clades, but are still widely distributed. Similarly, the *saddle* variant of *RALY* is seen in numerous clades, but is most abundant in the Euro Mastiff, Mediterranean, Scent Hound, and Terrier clades. Only eight breeds (AIRT, BEDT, IWSP, KOMO, LAKE, PULI, WELT, WFOX), in the Hungarian, Retriever, and Terrier clades are entirely fixed for the hair *curl* variant of *KRT71*. The *long* haired variant of *FGF5* is at the highest levels in the American Toy, Asian Toy, Continental, Hungarian, Poodle, Retriever, Small Spitz, and Spaniel clades. Conversely, the highest representation of the *short* haired wild-type genotype is in the Euro Mastiff clade. Toy breeds, Mastiffs, and Terriers represent the highest frequencies of the *BMP3 short muzzle* genotype. Finally, the only clades with breeds that are fixed for the *drop* ear marker are the Pointer/Setter, Retriever, Scent Hound, and Spaniel clades.

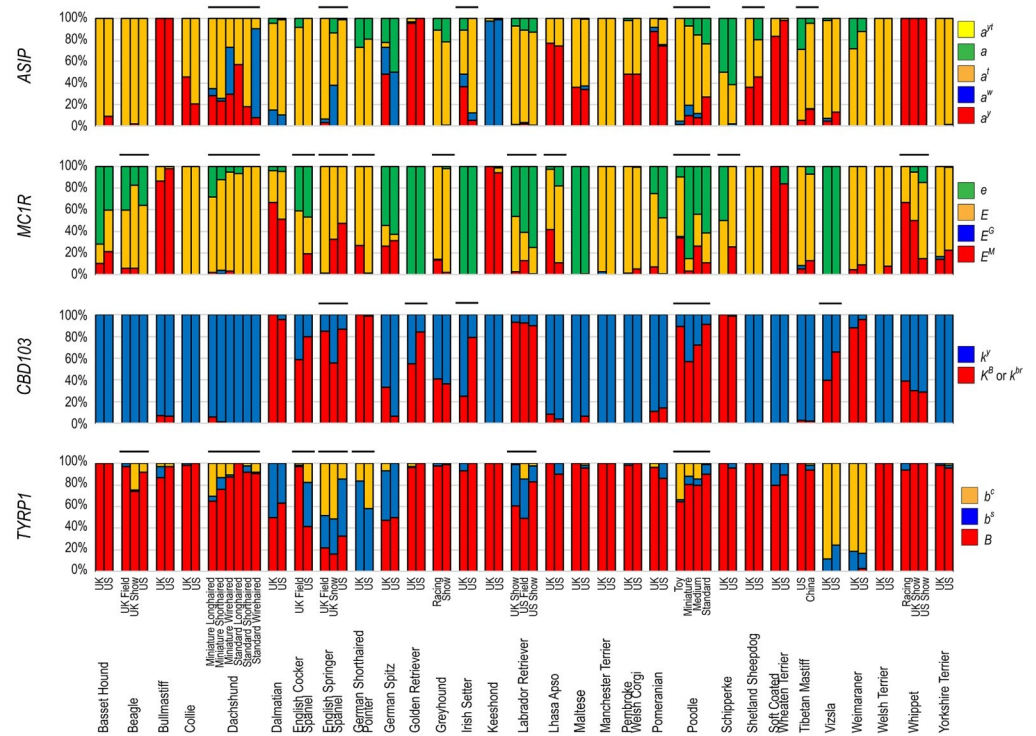
### Allele frequency is influenced by within-breed selection and geographic separation

Thirty breeds were divided into two or more distinct populations based on geographic region (27 breeds), body size (Dachshund and Poodle), coat type (Dachshund), and lineage or application, such as those selected for working applications (i.e. field, racing) or conformational traits (i.e. show) (6 breeds). All such populations were genetically distinguishable via principle components analysis (PCA). For the present study, allele distributions of the 12 tested genes between same-breed populations were evaluated for significance using either Pearson  $\chi^2$  or Fisher's Exact test (the latter was used when any allele grouping presented at <5 individuals). Distributions were deemed statistically significantly different at  $p \leq 0.00167$ , after correction for multiple testing. Nine breeds (BASS, BULM, DALM, KEES, MANT, POM, SCWT, WEIM, WELT) showed no significant difference between populations in allele distribution at any evaluated gene. The remaining 21 breeds had significant allele distribution differences between populations for at least one gene (Fig 2, S2 Fig). The gene with the greatest number of breeds displaying differences in allele distribution by population is *MC1R*, which was statistically different in 11 breeds (Fig 2). The *T* gene, responsible for the *tailless* phenotype, is not statistically different between populations of any breed (S2 Fig).

The effect of same-breed population divergence is readily observable in the system of coat color gene interactions and epistasis. Five breeds (ESSP, GOLD, ISET, POOD, VIZS) have significantly different *CBD103* allele distributions between same-breed sample populations (Fig 2). Of these five breeds, three are uniformly fixed for an *MC1R* genotype of *e/e* (GOLD, ISET, VIZS), effectively preventing the expression of the phenotype variability that the different alleles of *CBD103* would otherwise produce. Conversely, regional phenotype preference can be seen in breeds such as the Labrador Retriever (Fig 2). The breed shows significantly different allele distributions by population for *MC1R* and *TYRP1*, indicating that US conformation show lines prefer yellow (*MC1R e/e*) dogs with black noses (*TYRP1 B/\_*), UK conformation show lines prefer black (*MC1R E/\_*, *TYRP1 B/\_*) dogs, and US field lines consist mainly of brown (*MC1R E/\_*, *TYRP1 b/b*) or yellow dogs (*MC1R e/e*) ( $p \leq 0.003$ ).

### The natural probability of disallowed traits in pure breeds

The approved phenotypes for each breed were determined based on the written breed standards of the AKC, FCI, UKC, and KC (Britain). Of the 212 breeds evaluated, 143 were



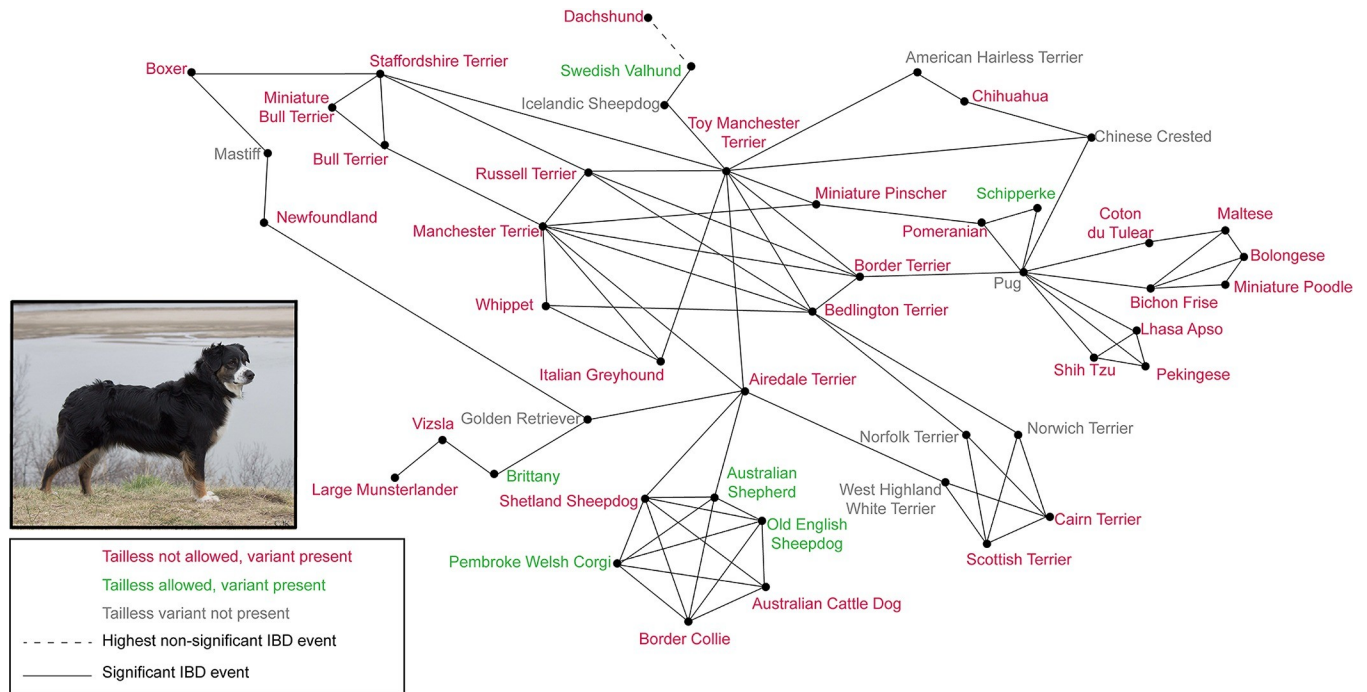
**Fig 2. Allele frequencies for *MCI1R*, *ASIP*, *TYRP1*, and *CBD103* across same-breed populations.** Horizontal black bars indicate within-breed allele distributions that are significantly different ( $p \leq 0.00167$ ).

<https://doi.org/10.1371/journal.pone.0223995.g002>

observed to carry at least one allele that would result in an unfavorable phenotype, termed “fault”, relative to at least one of the four queried breed registries (S3 Table). The breeds with the highest number of fault-causing alleles are the Treeing Walker Coonhound and Great Dane with 6 fault alleles present in each, and the Yorkshire Terrier and Rottweiler with 5 fault alleles present in each. Seventy-eight percent of the total number of fault allele occurrences produce phenotypes disallowed by all queried registries that recognize the breed. For 8.9% of fault allele observations, one or more of the queried breed registries allow the phenotype produced without bias, with the remaining registries disallowing it entirely or imparting a preference bias. In 15.1% of the fault allele observations, at least one of the four registries describes the allowed phenotypes indistinctly. This consists of situations where the trait is preferred to a lesser extent than alternatives, is not described at all, or is worded in such a way as to lead to ambiguous interpretation. The most frequently observed fault-causing alleles are the recessive brown alleles of *TYRP1*, representing 29.8% of all fault allele instances. Notably, 79.6% of all fault alleles are recessively inherited.

The probability of producing the disallowed phenotype associated with each fault allele was calculated assuming a random breeding same-breed population, and taking into consideration the inheritance patterns and gene interactions required for the phenotype expression. The fault-producing probabilities range from  $4.9e^{-7}$  (any non-solid color in the Black Russian Terrier) to 0.25 (red in the UK population of Schipperkes). Overall, the fault alleles were detected at low frequencies and, due to complex inheritance hierarchies and epistasis, have a  $<0.01$  probability ( $<1\%$  chance) of producing the fault phenotype in 58.9% of instances. Only 4.2% of fault alleles have a  $>0.10$  probability ( $>10\%$  chance) of producing the fault phenotype.





**Fig 3. Identity-by-decent (IBD) haplotype sharing breed relationships connect breeds carrying the *T* allele for *tailless*.** Solid black lines represent instances of significant haplotype sharing levels between breeds. The color of the breed names reflects the proposed carrier status of the tailless variant in the sampled members of those breeds, indicating not present (grey), present and permitted within the breed standard (green), and present and not permitted within the breed standard (red). The Dachshund breed shows no significant haplotype sharing with any other breed, however, its highest non-significant haplotype sharing value is with the Swedish Valhund (dashed line). Inset, the Australian Shepherd breed permits natural taillessness.

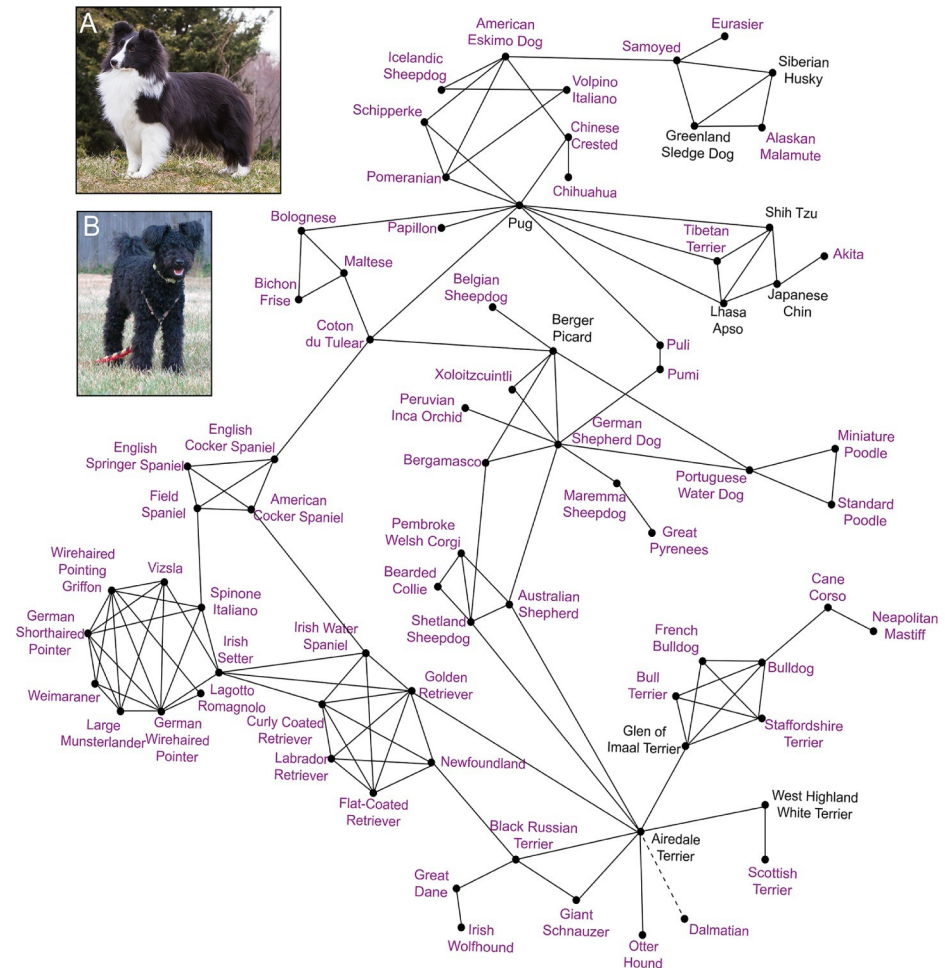
<https://doi.org/10.1371/journal.pone.0223995.g003>

### Ancestral routes of allele transmission

Previous research has demonstrated the ability of identity-by-decent (IBD) haplotype sharing to reveal shared ancestry events between modern dog breeds, accurate to approximately 150 years ago [32]. We used these previously identified significant breed relationships to successfully connect breeds genotyped to possess the rare phenotype alleles: *T tailless* (Fig 3) and *ASIP a* (Fig 4). Forty-eight breeds were identified as carrying the *tailless* allele of *T*, in many cases at very low frequencies, in the present study. Thirty-eight of these breeds were represented in previous research [32,33], allowing identification of IBD haplotype sharing relationships. By using these pre-determined breed relationships, each of the 38 breeds (Fig 3, represented by red or green text) can be connected into a single relationship matrix, providing a potential route of transmission of the *tailless* allele. Only nine non-carrier breeds (Fig 3, represented by grey text) were required to serve as potential transmitters of the *tailless* allele, reflecting populations that have successfully eliminated the allele from the breed, or have decreased the frequency enough so as to not be detected in our sampling. Likewise, 89 breeds were identified as carrying the *ASIP a* allele in the present study, 68 of which were also analyzed for IBD haplotype sharing previously [32,33]. Of these 68, all but three breeds (Tibetan Mastiff, Kuvasz, and Anatolian Shepherd) could be connected via significant IBD haplotype sharing events (Fig 4).

### Discussion

Most publications to date that assess allele frequency within dog breeds have been conducted on a relatively small scale, limited to select sets of breeds and/or variants of known relevance to specific breeds [34–40]. A recent publication focused on a very large set of purebred and



**Fig 4. Identity-by-descent (IBD) haplotype sharing breed relationships connect breeds carrying the *a* allele of *ASIP*.** Solid black lines indicate instances of significant haplotype sharing between breeds. Breed names in purple indicate observed frequency of the *a* allele within the breed in the present study. Breed names in black indicate that the *a* allele was not detected within the sampled individuals of that breed. The dashed line connecting the Dalmatian to the Airedale Terrier indicates that no significant haplotype sharing was detected with the Airedale Terrier, however the highest non-significant level of haplotype sharing was measured with the Airedale Terrier. Inset images show examples of dogs with the recessive black phenotype, A) Shetland Sheepdog, B) Pumi.

<https://doi.org/10.1371/journal.pone.0223995.g004>

mixed-breed dogs and a panel of 152 disease markers [41], but did not evaluate morphological or pigmentation gene variants. Here, we have leveraged an expansive number of DNA samples, both in terms of dog numbers and breed representation for which 12 genes impacting coat color and morphological variation have been genotyped on a commercial genotyping platform (Table 1). In doing so, we have revealed patterns of allele frequency and distribution that inform our understanding of breed development and relationships, regional phenotypic preferences, and the impact of canine registering bodies on the prevalence of rare alleles.

### Unexpected breed distribution of low frequency alleles

The majority of the alleles evaluated in the present study are broadly recognized across many modern dog breeds. However, a small number of the variants have only previously been recognized in select breeds. For example, the  $E^G$  allele of *MC1R* was initially detected in Salukis and Afghan Hounds [7]. We have here described 26 additional breeds in which the  $E^G$  allele has

not previously been reported (S2 Table). Seven breeds (ANAT, CAAN, CASD, CAUC, KKLG, POLG, TAIG) carrying  $E^G$  are phylogenetically related to the Saluki and Afghan Hound, suggesting shared traits due to common ancestry, but  $E^G$  was also found in the Asian/Arctic, Hungarian, Nordic Spitz, Pointer/Setter, Poodle, and Scent Hound clades. The  $E^G$  allele requires an *ASIP* tan point phenotype in order to express as an observable pattern [7], the lack of which would allow the allele to persist within a breed undetected. This may be the case in breeds such as the Anatolian Shepherd, Black Russian Terrier, Maltese, Norwegian Buhund, Puli, and Toy Poodle, where the tan point phenotype is rare or absent.

The *tailless* allele of the *T* gene has previously been investigated in breeds known to have the natural tailless phenotype, confirming the cause of the trait in 18 breeds [25,26]. The frequency of the variant has not, however, been evaluated in breeds not necessarily expected to harbor the trait. We report the occurrence of the *tailless* allele in 38 new dog breeds (S2 Table, Fig 3). The tailless phenotype is variably expressed, ranging from complete anury to a truncated or kinked tail [25,26]. While absence of a tail in a breed expected to have a full length tail would be immediately noticeable, a mildly shortened or bent tail may be dismissed as inconsequential, thus allowing a low frequency of the allele to persist within some breeds. Likewise, surgical tail docking is still a common practice among certain breeds in some countries, rendering the natural length of the tail unknown beyond a neonatal age. Among the 38 newly reported carrier breeds of *tailless*: two allow a natural tailless phenotype (MCNB, OES); seven follow tail docking procedures in the US (AIRT, BOX, MPIN, MPOO, TENT, VIZS, WELT); eight have breed standards that describe a shorter than average ideal tail length (BORT, BULT, CAIR, MANT, MBLT, RWST, SCOT, TMNT); and nine have naturally curled tails (BICH, BOLO, COTO, GSPZ, LHAS, MALT, PEKE, POM, SHIH), which would mask small variations in length or structure. Indeed, anecdotal reports of abnormally short and/or kinked tails have been reported in at least eight of the newly reported *tailless* carrier breeds (BEDT, BULT, DACH, LHAS, SSHP, STAF, SCOT, VIZS) [42–46] (personal communications with breeders). It is important to note that many of the breeds in which we have now detected the *tailless* allele show the allele as present at very low frequencies. This allows for the possibility that some of these calls may be false positive errors that have escaped our quality control efforts.

The identification of an allele with the combined  $a^y$  and  $a^t$  variants in the *ASIP* gene poses a number of intriguing implications. The variant associated with the  $a^t$  allele is a SINE insertion located on CFA chr24:23,365,298–23,365,537, in the upstream regulatory region of *ASIP*, and has been attributed to dorsoventral and banded hair patterning [5]. The  $a^y$  variant consists of two adjacent amino acid substitutions in exon four of the *ASIP* gene (p.A82S R83H), located at CFA chr24:23,393,510–23,393,515 [4]. At approximately 28 kb apart, a canine recombination rate of 1.2 cM/Mb [47] would predict a 0.042% rate of crossover between the two variants. We observed the  $a^y$  and  $a^t$  variants on the same chromosome, briefly termed the  $a^{yt}$  allele, a total of 48 times within the 11,790 genotyped dogs, and within 14 seemingly unrelated breeds, and the Dingo. The phenotypes of individual genotyped dogs were not available, but relying on the accepted colorations with each breed, the remaining 11 breeds with  $a^{yt}$  all allow for the standard  $a^y$  fawn phenotype. Frequencies of  $a^{yt}$  varied within the fawn breeds from 1% to 8%, suggesting that any phenotypic difference caused by the allele combination would not be immediately recognized as outside the acceptable  $a^y$  fawn coloration. Indeed, with the added  $a^t$  SINE insertion, preventing the production of banded hair patterning,  $a^{yt}$  may result in a fully phaeomelanin color, lacking the eumelanin tips on the hairs that are commonly present in  $a^y$  fawn dogs. The current method of determining the biallelic genotype of *ASIP* relies on assaying the  $a^y$ ,  $a^t$ , and  $a$  variants, and assigning the presence of  $a^w$  in the absence of the other detectable markers. In this way, a genotype of  $a^{yt}/a^w$  and  $a^y/a^t$  would appear to be the same. Seven breeds (ANAT, DANE, GPYR, LAGO, MARM, TIBM, TIBS) in which the  $a^{yt}$  allele was

detected also showed observable levels of the  $a^w$  allele, presenting situations where the true frequencies of  $a^w$  may be higher than reported here. An additional 63 breeds have the  $a^y$ ,  $a^w$ , and  $a^t$  alleles, reflecting the opportunity for the  $a^{yt}/a^w$  genotype to be incorrectly recorded as  $a^y/a^t$ , which would also result in underestimate of  $a^w$  frequency. There is, of course, the possibility that either the  $a^y$  or  $a^t$  variants previously described are not causal variants, but rather very accurate markers. However, both variants have been successfully used to predict coat color genotypes in multiple studies without raising concern [7,19,48,49]. Further pursuit of the phenotypic impact of the  $a^{yt}$  allele, along with determining its inheritance pattern, is ongoing.

The *RALY* duplication associated with the saddle tan modification of the tan point phenotype was detected in 203 breeds (S2 Table), thirty-eight of which are either fixed or variable for the saddle tan phenotype, and 119 of which have epistatic variation at additional coat color genes that prevent the expression of *saddle tan*. However, both the saddle tan and tan point mutations were detected in 14 breeds for which only a tan point phenotype is permitted or expected. For instance, Bernese Mountain Dogs have a fixed phenotype of black, tan points, and white spotting; despite the 100%  $a^t$  allele frequency and the addition of an 18% *saddle tan* allele frequency, no readily identifiable dogs exist with the saddle tan phenotype. These instances reiterate the findings of the initial canine *RALY* research, suggesting at least one additional modifier gene that is required for the production of the saddle tan phenotype [19].

The *Harlequin* allele was identified, as expected, in the Great Dane population [14], but was also identified—unexpectedly—in the Yorkshire Terrier population, where it has not been previously described. Yorkshire Terriers are not known to possess the variant at *PMEL17* that produces the merle phenotype and is required for the expression of harlequin when combined with the *PSMB7* variant. Therefore, even with the *Harlequin* mutation, it would presently be impossible to produce a harlequin patterned Yorkshire Terrier. Despite this, there is no recent haplotype sharing between the Yorkshire Terrier and Great Dane breeds, which could have resulted in transfer of the allele [32]. The detected presence of the allele in our data set could be explained by a *de novo* mutation in the Yorkshire Terrier breed, inaccurate marker selection, or genotype quality in the SNP array. The latter is unlikely due to the accuracy of this test on the positive control population (Great Danes) and the negative control population (all other dog breeds). The *de novo* mutation theory is supported by the fact that *Harlequin* was only identified in the UK Yorkshire Terrier population (2 dogs with the *Harlequin* allele out of 24 UK Yorkshire Terriers) and not at all in the US Yorkshire Terrier population ( $n = 107$  dogs); these two populations interbreed only rarely, suggesting the *de novo* mutation arose privately in the UK. Further investigation is required to draw definitive conclusions.

Given the expectation of ancestral alleles in wild canine species, the level of derived alleles detected in the wild canine populations was somewhat surprising. While 99% of *MC1R* alleles in the grey wolf, Eastern coyote, Western coyote, and Dingo were found to be wild-type *E*, only 72% of *ASIP* alleles from wild canines were wild-type  $a^w$ , with only the Western coyote fixed for  $a^w$ . The genotyped Dingos were predominantly (75%)  $a^y$  fawn, consistent with their usual phenotype. Somewhat unexpectedly, however, the  $a^t$  tan point allele of *ASIP* was detected at levels of 20% in Eastern coyotes and 14% in grey wolves. Likewise, derived coat color alleles at *TYRP1* and *MITF*, and the morphologic variants associated with long hair, and drop ears, were detected at within-species levels between 2–46% ( $b^s$  in Eastern coyotes and drop ear in Grey wolves, respectively). The derived alleles at *ASIP*, *MC1R*, *CBD103* have previously been detected at low frequency in various populations of coyotes and wolves, most of which have been postulated to occur due to introgression with domestic dogs [49–52]. Such introgression between grey wolves and domestic dogs has been documented to occur at low levels across Eurasia, but detailed studies of domestic dog introgression with wolves and other species (coyotes, dingoes) in other geographic areas are sparse [53,54]. This is the first report of the tan

point (*ASIP*), and long hair (*FGF5*) variants in wild canine populations. A recent paper verified the *CFA10* ear shape locus, and demonstrated that, while definitively a contributing locus, it does not perfectly account for ear phenotype [28]. Thus, the higher frequency of ear shape and hair length variant alleles in wild populations is perhaps not surprising, as these phenotypes are known to be driven by more than one locus [20,27,52].

### Intended working application influences phenotypic variation

Phylogenetic analysis of dog breeds is conducted in the absence of phenotypic data and provides intriguing insight into the history of breed formation. Reversing that scenario by assessing phenotypic association relative to phylogenetic relationships begins to unravel the individual characteristics that define cladistic groupings. The comparison of allele frequencies of the interacting coat color genes *ASIP*, *MC1R*, and *CBD103* (Fig 1) reveal preferred clade-specific colorations. *ASIP*-driven patterns are widely preferred in the Scent Hound, Spitz, and Terrier clades; conversely, single-pigment patterns, characterized by high frequencies of *CBD103*  $K^B$  or *MC1R*  $e$  alleles, are prevalent in the clades with a history of hunting applications, namely those of the Pointer/Setter, Poodle, Retriever, and Spaniel clades. Similarly, the hunting-related clades also generally display a higher incidence of brown pigment production, though they do not show preference for a specific recessive *TYRP1* allele (S1 Fig).

Historically, white has been selected for in breeds to improve their visibility. A human's ability to visually locate their dog, aided by bright white patterning, is important for hunting and terrier breeds that traditionally work in dense vegetation to locate, subdue, or retrieve quarry [55–59]. Conversely, when a dog is meant to remain camouflaged and not startle game, as with retriever breeds, a bland or solid color that will blend into the surrounding foliage is desired [55,58]. White spotting, resultant of variation at the *MITF* gene [17], is broadly present across all clades. However, reflecting the applicable purpose of the color pattern, the presence of breeds fixed for white markings appears to be greatest in the Pointer/Setter, Spaniel, and Terrier clades (S1 Fig).

### Geography and lineage influence allele distribution among same-breed populations

Dog breeds, while standardized and often interchangeable around the world, have previously been shown to differ in their genetic composition based on geographic locale [32,60–63]. We detected significant ( $p \leq 0.00167$ ) allele distribution differences based on geographic site of sample collection in 17 of the 26 breeds for which genomic sub-populations could be identified based on geography. Similarly, six breeds (BEAG, ESSP, ECKR, GREY, LAB, WHIP) encompassed multiple lineages differentiated by the application of the dogs to hunting, racing, or conformation competitions. In each case, at least one of the genes queried showed significant allele distribution differences between lineages. These disparate allele frequencies can represent either regional differences in preference or influence of prominent ancestor bias.

*TYRP1* is a representative example. The various recessive alleles of *TYRP1* all ultimately produce the same pigmentation shade (brown instead of black), regardless of the specific alleles present. Therefore, significant variation between *TYRP1* recessive allele frequencies relative to population may indicate a founder genotype or the effect of an influential ancestor. The frequency of the  $b^c$  allele is significantly higher in US Field Beagles ( $p \leq 0.003$ ) and UK English Springer Spaniels ( $p \leq 0.003$ ), compared to conformation lineage Beagles and US English Springer Spaniels which each have higher frequencies of  $b^s$  (S2 Fig). As the specific *TYRP1* allele cannot be purposefully selected for through phenotype observation alone, these differences between populations may reflect divergent selection of the lineages. There is no

apparent pattern to the distribution of the  $b^s$  and  $b^c$  alleles relative to clade distinctions (S1 Fig). This would imply that these two  $b$  alleles are substantially old enough so as to have propagated throughout dog types prior to the development of closed-breed populations.

Traditionally, many dog breeds have had their tails surgically docked to a shorter length, predominantly for the purposes of preventing injury to the tail while the dog carries out its defined function. With the changing use of dogs over time, the necessity for tail docking has recently become a source of discussion among canine enthusiasts and the veterinary community. Between 1987 and 2018, 35 countries have banned or restricted tail docking or the ability to exhibit dogs with docked tails in regulated kennel club events [64–68]. There are currently no restrictions placed on docking in the US, while the UK has banned tail docking since 2006 [64]. Dog breeders in countries with docking bans wanting to uphold the traditional appearance of a dock-tailed breed could therefore select for the *tailless* variant of the  $T$  gene to produce naturally shortened tails in their dogs. However, none of the breeds from the US and the UK that were genotyped in the present study show significant differences in the *tailless* variant frequencies. Eleven breeds collected in both locations traditionally have docked tails, while only four of these breeds (PEMB, SKIP, VIZS, WELT) had any measureable level of the *tailless* variant present. While the level of *tailless* is numerically higher in the UK populations of the Pembroke Welsh Corgi, Schipperke, and Vizsla compared to US populations, the differences were not significantly different ( $p \leq 0.00167$ ). This lack of disparity likely reflects the age of the samples used in this dataset, which were collected between 2005 and 2016. It is probable that insufficient time has passed to accurately reflect the efforts of selection for natural taillessness resulting from the procedural ban, and the  $T$  allele frequencies are expected to increase in select breeds over time in those countries banning surgical docking.

It is worth noting that, while overall numbers for most breeds were quite high, once split into geographic or lineage groups, the cohort size in some cases became smaller. These small numbers may ultimately influence the accuracy of allele frequency estimates in these subpopulations, and larger sample sizes in future studies would solidify true differences between such populations.

### The natural probability of disallowed traits in pure breeds

A breed standard is a written description of a given breed, as determined by a committee of educated breeders, that details how a typical dog of that breed should look and behave. Historically, the standards functioned as a framework for breeders to aim at when producing dogs of that breed. Breed standards outline traits that are disallowed, so that breeders can opt to breed away from or minimize their occurrence. Each canine registering organization, usually specific to country or region, employs its own set of breed standards and, for the most part, these standards are concordant breed-to-breed across registries. However, there are instances where the wording between standards is not precisely mirrored, such that variations in color, size, or morphology are more or less tolerated from registry to registry.

Our analyses detected genetic variants that would cause disallowed phenotypes in 67.5% of breeds tested (S3 Table). The majority of these, 58.9%, have a <1% probability of producing that undesirable phenotype given random breeding. These values represent multi-generational efforts to eliminate unfavorable phenotypes within breeds, however they also highlight that conformity to breed standard is certainly not yet universal or complete. They also illustrate the difficulty in selecting against alleles that are masked not just by dominance, but also epistasis, even in a highly-visible trait such as coat color. In general, these findings exemplify three separate scenarios: 1) traits broadly disallowed but carried at low frequencies, 2) traits allowed under some registries but not others, and 3) single traits that persist due to breed-specific



**Fig 5. Purebred dogs exhibiting color traits deemed inappropriate by one or more breed registries.** A) Bull Terrier with a brown nose and brown patch above its eye. B) Shetland Sheepdog with piebald white spotting. C) Great Dane with the harlequin pattern on a fawn base color.

<https://doi.org/10.1371/journal.pone.0223995.g005>

allowances for particular trait combinations (Fig 5). For example, regardless of breed registry (AKC, UKC, KC, and FCI), the Bull Terrier is described as always having a black colored nose, and never showing brown hair pigmentation. However, we detected the  $b^c$  and  $b^s$  alleles of *TYRP1* at a frequency of 3% each in our population of pure bred Bull Terriers. This results in a probability of 0.0036 in producing a brown colored purebred Bull Terrier, present as brown pigmented skin and/or coat, assuming random breeding (Fig 5; S3 Table). Conversely, Shetland Sheepdog breed standards differ in that the AKC disallows piebald spotting in the breed, the FCI and KC tolerate, but do not prefer, excessive white spotting, and the UKC allows any variation from no white to fully piebald. We detected the *MITF* variant for white spotting, known to cause the piebald phenotype in this particular breed [18], at a frequency of 16% in the UK population of the breed and 6% in the US population (Fig 5; S2 and S3 Tables). Finally, the Great Dane breed standard is relatively cohesive across breed registries, but defines acceptable coat colors in terms of pattern combinations; for example, white spotting and the harlequin pattern are only allowed on a black base color. However, because black base color is dictated by *CBD103* on chromosome 16 with a breed frequency of 66%, white spotting is controlled by *MITF* on chromosome 20 with a breed frequency of 6%, and harlequin is caused by *PSMB7* on chromosome 9 with a breed frequency of 21%, the realistic potential to produce a fawn-based harlequin or a fawn and white spotted Great Dane is unavoidable (Fig 5, S2 Table). These values result in a 1.33% probability of producing a fawn and white Great Dane through random breeding of purebred dogs, or a 3.80% probability of producing a fawn Great Dane with one copy of the harlequin mutation. Since the frequency levels of the merle variant—required for production of the harlequin phenotype—are not known, the latter value does not necessarily reflect the number of fawn-based harlequin dogs produced.

The Schipperke breed, collected from populations in the US and the UK, presents a clear example of how regional acceptance of certain characteristics can drive the frequency of a variant within a population. The Schipperke breed standards for AKC, UKC, and FCI all state that the dog must be entirely black in color. However, the KC in the UK states that “any solid color” is permissible. As such, the allele frequencies for the *MC1R* recessive  $e$  allele, which produces a solid red color when homozygous, was observed at 50% among the dogs sampled in the UK ( $n = 6$ ), and 0% among the dogs sampled in the US ( $n = 44$ ) (Fig 2).

Much recent emphasis has been placed on the importance of genetic diversity within breeds [69–76]. With the conservation of diversity in mind, breeders and breed organizations must

weigh the relative value of breed standard conformity with preservation of genetic diversity. The existence of unfavorable, though arguably benign pigmentation or morphological variations, has here been quantified and can be addressed by applied genetic screening to reduce the carrier frequency of breeding stock, or by reassessing breed standards to broaden the acceptance of preexisting variation. Likewise, though our analyses have indicated that production of disallowed phenotypes is generally quite low, the occurrence of an undesirable pigmentation trait should not necessarily exclude a dog from purebred status if that variant has been detected in the appropriate population. As a recent example, effective 1 January 2019, the Great Dane Club of America revised their breed standard to allow merle coloring on a black base. Canine genetic research has clarified that the presence of the merle allele is required for the Harlequin phenotype [14]; since this relationship was previously unclear, the breed had not allowed merle (without the Harlequin modifier) until this change. These revisions demonstrate the purebred dog community recognizing and willingly implementing the findings from canine genetic research. The present work will guide similar decision-making by breed clubs regarding definition of acceptable breed colors.

### Diverse breed representation of rare alleles informed by ancestral haplotype sharing

The unique structure of the dog genome dictates that, even across breeds, large regions of linkage disequilibrium can accompany trait-causing variants [77], as measured by IBD haplotype sharing. As such, we can assume that the presence of a given allele in two breeds may indicate shared ancestry between those breeds. For example, it has long been assumed that the recessive black allele of *ASIP* (*a*) is predominantly found in herding breeds [3–5]. However, among the 212 breeds analyzed here, the *a* allele was identified in 89 breeds, 83 of which are not previously reported as carrying the allele, representing 23 of the possible 28 clades assigned using IBD haplotype sharing in previous work [32,33]. While herding breeds, present in the Continental Shepherd, Hungarian, New World, Nordic Spitz, and UK Rural clades, comprise 14 of the *a* allele-possessing breeds—an expected result—the greatest breed representation was among the Pointer/Setter clade with 13 breeds possessing the *a* allele. The only clades without measured frequency of the *a* allele are the Alpine, American Terrier, Asian Toy, Pinscher, and Standard/Miniature Schnauzer clades.

IBD haplotype sharing can reflect a shared ancestry between populations, and the amount of haplotype sharing between breeds correlates significantly with the time point at which those breeds shared a common ancestry [32]. Using instances of significant levels of haplotype sharing between breeds, with reliably dating introgression events to as early as the late 1800's [32], 65 of the 89 breeds displaying carrier frequencies >0% for the *a* allele can be connected (Fig 4). While not necessarily reflecting the exact mode of allele sharing between breeds, the measured haplotype sharing instances successfully demonstrate a recent ancestral history between the breeds. There are 21 breeds with positive *a* allele frequencies that cannot be connected via haplotype sharing due to not being included in the previous haplotype sharing analyses [32], but are predicted to be in phylogenetic clades already represented among *a*-carrying breeds. Three breeds, the Anatolian Shepherd, Kuvasz, and Tibetan Mastiff, carry the *a* allele and are included in the haplotype sharing analyses, but do not show significant ancestry with other breeds. Therefore, while the extensive 65-breed relationship matrix supports the potential for the *a* allele to spread between breeds via recent introgression events, the presence of the allele in the Anatolian Shepherd, Kuvasz, and Tibetan Mastiff, for which no recent introgression events have been detected, suggests that the allele itself arose early in the history of the domestic dog, establishing a broad distribution well before the development of modern breeds in the late 1800's.



The *tailless* allele of the *T* gene was identified in 48 breeds, representing 14 clades. While 10 of these breeds have been previously identified as carriers of the *tailless* allele [25,26], 38 are reported here for the first time. Thirty-eight of the 48 *T*-carrying breeds are represented in the identity-by-descent dataset [32], all of which can be connected into a single relationship matrix (Fig 3). In some instances, such as with the Brittany and Newfoundland, which have *tailless* carrier frequencies of 0.04 and 0.01 respectively, there is no direct identity-by-descent relationship with a potential source breed of taillessness. However, both show ancestral relationships with the Golden Retriever, who then further shows ancestry with the Airedale Terrier, a carrier of *tailless* at a frequency of 0.02. It is possible that the Golden Retriever either carries *tailless* at a frequency not detected in our screening, or the trait previously existed within the breed and has since been selected against and eliminated. Other possible historic sources of *tailless* that have decreased or eliminated their current carrier frequency include the Icelandic Sheepdog, Keeshond, Mastiff, and Pug.

While the allele frequencies reported herein are intended to provide general information regarding the existence of particular alleles within breeds, they are not likely to be perfectly accurate estimates of actual breed-wide allele frequencies. For instance, while the German Shepherd Dog, Lagotto Romagnolo, and Australian Shepherd are represented by 162, 139, and 137 dogs, respectively, our dataset only includes eight Small Munsterlanders, and nine each of Redbone Coonhounds and Bergamascos. Despite these limitations, no other published genotype dataset matches the presently reported size and diversity of dog samples across a dozen pigmentation and morphologically relevant genes.

## Conclusions

The broad adoption of commercial genotyping services yields an immense amount of genetic information. We have demonstrated how this data can be utilized to detail the current phenotypic diversity of 212 dog breeds, and the impact of population divergence due to geographic separation or selection practices. While most dog breeds have existed as closed breeding populations since the late 19<sup>th</sup> century, we have shown that rare trait-causing variants continue to persist within most breeds. Conflicting breed standard descriptions of multiple registering organizations may have facilitated the persistence of these traits within certain populations. In addition, epistatic masking effects have contributed to the continuance of various trait-causing alleles, due to the complicated multi-gene pathways whereby recessive alleles can remain unexpressed over multiple generations, and making selection for or against them challenging without the use of genotyping services.

The modern existence of domestic dogs is such that coat color is primarily a matter of aesthetics. Consideration should be given by breed associations to unifying breed standards across registering bodies, either with the intent of increasing selective pressure against truly undesirable characteristics, or expanding the standards to permit phenotypes for which the causal variants exist ancestrally within the breed. Simultaneously, conservation of genetic diversity within breeds must be weighed. The present study documents for the first time the frequencies of alleles at 12 coat color and morphological genes, across 11,790 dogs, representing 212 breeds, and demonstrates not only the anticipated genotypic variations within breeds, but also rare and unexpected alleles not previously reported.

## Materials and methods

### Sample collections

Genotype data from 11,790 canids, representing 212 pure dog breeds and 4 wild canine populations, was compiled during the development and implementation of the Mars WISDOM

PANEL platform (Wisdom Health, Vancouver, WA, USA). DNA collections herein represent a subset of those initially reported in Donner, *et al.* [41]. Dog DNA samples were obtained by Wisdom Health (formerly Mars Veterinary) and Genoscooper Laboratories (Helsinki, Finland), between January 2005 and October 2016, as owner-submitted, non-invasive cheek swab collections. Dog owners provided consent for use of their dog's DNA in research. The dogs sampled predominantly originated from the US and UK, though samples were also obtained in smaller numbers from several other countries. Dogs were considered to be purebred if registered with a relevant all-breed registry, the predominant ones being: Fédération Cynologique Internationale, American Kennel Club, United Kennel Club, and UK Kennel Club, or an applicable single-breed registry for rare breeds. Breed and species classification was further verified through principal component analysis (PCA) and genotyping on the WISDOM PANEL platform (Wisdom Health), particularly for the minority newer/rarer breeds not currently (at the time, or even now) recognized by a national registry. PCA further revealed that 30 breeds formed subpopulations based on geography, body size, coat type, or function (e.g., show vs. field). Archived wild canine samples were used to represent: 1) the grey wolf (*Canis lupus lupus*, sampled primarily from Eastern North America,  $n = 12$ ); 2) coyote (*Canis latrans*) collected from Eastern North America (Eastern coyote,  $n = 29$ ) or British Columbia and Southwestern US (Western coyote,  $n = 19$ ), populations segregated as determined by PCA cluster analysis; and 3) dingo (*Canis lupus dingo*) ( $n = 12$ ) populations. All dog breeds were represented by  $\geq 10$  individuals, with the exception of the Small Munsterlander ( $n = 8$ ), Bergamasco ( $n = 9$ ), and Redbone Coonhound ( $n = 9$ ).

## Genotyping

Genotyping of seven coat color and five morphological trait variants (Table 1) was conducted on a custom-designed Illumina Infinium HD bead chip using manufacturer-recommended protocols ([78]; Illumina, San Diego, CA, USA). The validation and genotyping quality control measures for this platform were previously described in detail [41,78]. Trait variant assays specifically were validated through extensive correlation of genotypes with established breed phenotypes and owner-submitted pictures of individual dogs.

## Allele frequencies and statistical analysis

Allele frequencies for each variant were determined for each breed and, when appropriate, breed subpopulations, and converted to binary genotypes for genes with multiple alleles. The total number of dog samples per breed are reported in S1 Table. Due to sporadic failure, the number of genotypes obtained per breed for each gene may vary from the total number of dogs per breed (S2 Table). For breeds with subpopulations ( $n = 30$  breeds, delineated primarily by geography, but also body size, breeding line of conformation versus working, etc.), within-breed statistical significance of the differences in genotype distribution between the subpopulations was evaluated by Pearson's chi-square contingency tables, or Fisher's exact tests (when allele counts in any single cell fell below 5), for each gene. Calculations were conducted with the MASS package in R [79,80]. A p-value of  $< 0.00167$  was chosen to indicate a significantly different allele distribution between the same-breed subpopulations; this is the Bonferroni correction for multiple testing of 30 breeds for each gene (i.e.,  $0.05/30$ ). In the case of *FGF5*, no genotypes were available for the US population of German Spitz, thus significance was corrected instead for 29 breeds, to a p-value of  $\leq 0.00142$  for this gene. Further correction (e.g., for testing of 12 different genes) was not applied because a balanced approach was desired and the initial Bonferroni correction was deemed suitable. Further analysis was conducted for the six breeds sampled from  $> 2$  subpopulations. When these breeds differed significantly at any

gene using the initial analysis described above, they were subsequently evaluated for pairwise significance using Pearson's chi-square or Fisher's exact tests. Significance for each breed subpopulation was determined by  $p\text{-value} = 0.05/n$ , where  $n$  = maximum number of subpopulations remaining in analysis. This number therefore varied for each gene, depending on remaining significant breeds. For example, if Poodles—divided into four subpopulations—remained, the correction was  $0.05/4$ , resulting in a  $p$ -value cut-off of 0.0125; however if Dachshunds—divided into six subpopulations—remained, the correction was  $0.05/6$ , resulting in a  $p$ -value cut-off of 0.0083. Specific statistical applications and  $n$  values are indicated in the appropriate figure, table legends or footnotes.

### Probability of phenotype expression

When alleles that would produce an undesirable or disallowed phenotype were observed in a given breed, relative to AKC, UKC, KC (Britain), or FCI breed standard descriptions, the probability of producing that phenotype was calculated with the following equation:

$$\text{Probability of phenotype } X = \left( p^2 + \sum_{q=a}^c 2pq \right) * \left( r^2 + \sum_{s=d}^f 2rs \right) * \left( t^2 + \sum_{u=g}^i 2tu \right)$$

Such that:

$p$  = the frequency of the fault-producing allele,  $P$ , of the gene causal for  $X$

$q$  = the frequency of any same-gene allele recessive to  $P$ , of which there can be up to 3 ( $a$  to  $c$ )

$r$  = the frequency of the most dominant allele,  $R$ , at an interacting gene that is required for production of the fault phenotype

$s$  = the frequency of any same-gene allele recessive to  $R$ , of which there can be up to 3 ( $d$  to  $f$ )

$t$  = the frequency of the most dominant allele,  $T$ , at an interacting gene that is required for production of the fault phenotype

$u$  = the frequency of any same-gene allele recessive to  $T$ , of which there can be up to 3 ( $g$  to  $i$ )

The traits for which genotypes were obtained present scenarios in which up to two known interacting genes can influence the expression of a phenotype. Namely, expression of an *ASIP* phenotype relies on corresponding *CBD103* and *MC1R* genotypes, the *MC1R* phenotypes caused by  $E^M$  and  $E^G$  require specific genotypes at *ASIP* as well as a homozygous wild-type *CBD103* genotype, and a recessive genotype at *CBD103* can result in multiple possible *ASIP* phenotypes, only some of which may be undesirable. Conversely, recessive homozygosity at *TYRP1* will result in brown pigment of the hair and keratinized skin, which may present as a coat fault if coupled with eumelanin-producing *MC1R* and *ASIP* genotypes, or a nose pigment fault regardless of hair pigmentation. Taillessness is only expressed in the heterozygous state, as it is embryonic lethal when homozygous, therefore, probability values were corrected to reflect outcomes among live births.

### Supporting information

**S1 Fig. Allele distribution.** Distribution of alleles for a) *TYRP1*, b) *MITF*, c) *PSMB7*, d) *RALY*, e) *KRT71*, f) *FGF5*, g) *T*, h) *BMP3*, i) chr10 ear set marker. Breeds are grouped by phylogenetic relationship.

(DOCX)

**S2 Fig. Allele frequencies in breeds with multiple populations.** Allele frequencies for a) *MC1R*, b) *ASIP*, c) *TYRP1*, d) *CBD103*, e) *MITF*, f) *PSMB7*, g) *RALY*, h) *KRT71*, i) *FGF5*, j) *T*, k) *BMP3*, l) chr10 ear marker, for all breeds with multiple populations. Initial  $X^2$  significance ( $p < 0.0167$ ) is indicated by horizontal black bars. Pairwise significance was conducted for all significant breeds with greater than two populations, with significance level indicated by \*\*. Letters under horizontal black bar denote significant groupings.

(DOCX)

**S1 Table. Sample information.** List of breeds, their abbreviations used throughout the paper, and the number of samples genotyped. Each breed was assigned to a phylogenetic clade based on previously published results [32,33]. Clade names in parenthesis indicate breeds not included previously [32,33], but for which a clade was assigned based on known breed history and phenotypes.

(DOCX)

**S2 Table. Allele frequencies for all breeds and genes tested.** Allele frequencies for (a) coat color genes *ASIP* and *MC1R*, (b) the brown (*TYRP1*) and dominant black (*CBD103*) genes, (c) the white spotting (*MITF*), harlequin (*PSMB71*), and saddle tan (*RALY*) genes, (d) hair length (*FGF5*), hair curl (*KRT71*), and ear set, and (e) skull shape (*BMP3*) and natural taillessness (*T*). Breeds fixed for a single allele at any gene are indicated with bold text.

(DOCX)

**S3 Table. Unfavorable or “fault” phenotypes possible by breed and breed registry.** Breeds genotyped to have alleles that would produce phenotypes considered as a “fault” by either the American Kennel Club (AKC), Fédération Cynologique Internationale (FCI), United Kennel Club (UKC), or The Kennel Club of the UK (KC). The level of tolerance within each breed registry is designated as either not allowed (N), not preferred (n.p.), allowed (Y), or ambiguously worded (amb.). A breed not recognized by a given organization is indicated with a dash (-). Inheritance of the fault-causing allele is designated as dominant (D), recessive (R), or compound heterozygote (CH). Breed name abbreviations are as listed in S1 Table. Probabilities for producing the non-standard phenotype were calculated assuming random mating within the breed, and account for multi-gene inheritance, expression, and epistatic effects.

(DOCX)

## Acknowledgments

The authors gratefully thank all the dog owners, breeders, and veterinarians who contributed to this study by testing their dogs and through their interest in advancing canine genetics research. They would also like to thank Dr. Hsin-Yie Weng for statistical assistance and Stephen Davison for classification of sub-populations.

## Author Contributions

**Conceptualization:** Dayna L. Dreger, Angela M. Hughes, Kari J. Ekenstedt.

**Data curation:** Blair N. Hooser, Balasubramanian Ganesan, Jonas Donner.

**Formal analysis:** Dayna L. Dreger, Blair N. Hooser, Balasubramanian Ganesan.

**Funding acquisition:** Angela M. Hughes, Lauren Holtvoigt, Kari J. Ekenstedt.

**Methodology:** Dayna L. Dreger, Angela M. Hughes, Jonas Donner, Heidi Anderson, Kari J. Ekenstedt.

**Resources:** Angela M. Hughes, Lauren Holtvoigt, Kari J. Ekenstedt.

**Supervision:** Angela M. Hughes, Kari J. Ekenstedt.

**Visualization:** Dayna L. Dreger.

**Writing – original draft:** Dayna L. Dreger, Blair N. Hooser, Kari J. Ekenstedt.

**Writing – review & editing:** Dayna L. Dreger, Angela M. Hughes, Balasubramanian Ganesan, Jonas Donner, Heidi Anderson, Kari J. Ekenstedt.

## References

1. The American Kennel Club. *The Complete Dog Book*. New York: Ballantine Books; 2006.
2. Little CC. *The inheritance of coat color in dogs*. Ithaca, New York: Comstock Publishing Associates; 1957.
3. Kerns JA, Newton J, Berryere TG, Rubin EM, Cheng JF, Schmutz SM, et al. Characterization of the dog Agouti gene and a nonagouti mutation in German Shepherd Dogs. *Mamm Genome*. 2004; 15(10):798–808. <https://doi.org/10.1007/s00335-004-2377-1> PMID: 15520882
4. Berryere TG, Kerns JA, Barsh GS, Schmutz SM. Association of an Agouti allele with fawn or sable coat color in domestic dogs. *Mamm Genome*. 2005; 16(4):262–72. <https://doi.org/10.1007/s00335-004-2445-6> PMID: 15965787
5. Dreger DL, Schmutz SM. A SINE insertion causes the black-and-tan and saddle tan phenotypes in domestic dogs. *J Hered*. 2011; 102 Suppl.
6. Schmutz SM, Berryere TG, Ellinwood NM, Kerns JA, Barsh GS. MC1R Studies in Dogs with Melanistic Mask or Brindle Patterns. *J Hered*. 2003; 94(1):69–73. <https://doi.org/10.1093/jhered/esg014> PMID: 12692165
7. Dreger DL, Schmutz SM. A new mutation in MC1R explains a coat color phenotype in 2 “old” breeds: Saluki and Afghan Hound. *J Hered*. 2010; 101(5):644–9. <https://doi.org/10.1093/jhered/esq061> PMID: 20525767
8. Newton JM, Wilkie AL, He L, Jordan SA, Metallinos DL, Holmes NG, et al. Melanocortin 1 receptor variation in the domestic dog. *Mamm Genome*. 2000; 11(1):24–30. <https://doi.org/10.1007/s003350010005> PMID: 10602988
9. Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, et al. A beta-defensin mutation causes black coat color in domestic dogs. *Science*. 2007; 318(5855):1418–23. <https://doi.org/10.1126/science.1147880> PMID: 17947548
10. Kerns JA, Cargill EJ, Clark LA, Candille SI, Berryere TG, Olivier M, et al. Linkage and segregation analysis of black and brindle coat color in domestic dogs. *Genetics*. 2007; 176(3):1679–89. <https://doi.org/10.1534/genetics.107.074237> PMID: 17483404
11. Schmutz SM, Berryere TG, Goldfinch AD. TYRP1 and MC1R genotypes and their effects on coat color in dogs. *Mamm Genome*. 2002; 13(7):380–7. <https://doi.org/10.1007/s00335-001-2147-2> PMID: 12140685
12. Hrckova Turnova E, Majchrakova Z, Bielikova M, Soltys K, Turna J, Dudas A. A novel mutation in the TYRP1 gene associated with brown coat colour in the Australian Shepherd Dog Breed. *Anim Genet*. 2017; 48(5):626.
13. Jancuskova T, Langevin M, Pekova S. TYRP1: c.555T>G is a recurrent mutation found in Australian Shepherd and Miniature American Shepherd dogs. *Anim Genet*. 2018;500–1. <https://doi.org/10.1111/age.12709> PMID: 30109695
14. Clark LA, Tsai KL, Starr AN, Nowend KL, Murphy KE. A missense mutation in the 20S proteasome  $\beta$ 2 subunit of Great Danes having harlequin coat patterning. *Genomics*. 2011; 97(4):244–8. <https://doi.org/10.1016/j.ygeno.2011.01.003> PMID: 21256207
15. Clark LA, Wahl JM, Rees CA, Murphy KE. Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. *Proc Natl Acad Sci*. 2006; 103(5):1376–81. <https://doi.org/10.1073/pnas.0506940103> PMID: 16407134
16. Murphy SC, Evans JM, Tsai KL, Clark LA. Length variations within the Merle retrotransposon of canine PMEL: Correlating genotype with phenotype. *Mob DNA. Mobile DNA*; 2018; 9(1):1–11.
17. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NHC, Zody MC, Anderson N, et al. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet*. 2007; 39(11):1321–8. <https://doi.org/10.1038/ng.2007.10> PMID: 17906626

18. Schmutz SM, Berryere TG, Dreger DL. MITF and white spotting in dogs: A population study. *J Hered.* 2009; 100(Suppl 1):S66–74.
19. Dreger DL, Parker HG, Ostrander EA, Schmutz SM. Identification of a mutation that is associated with the saddle tan and black-and-tan phenotypes in Basset Hounds and Pembroke Welsh Corgis. *J Hered.* 2013; 104(3):399–406. <https://doi.org/10.1093/jhered/est012> PMID: 23519866
20. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, et al. Coat variation in the domestic dog is governed by variants in three genes. *Science.* 2009; 326(5949):150–3. <https://doi.org/10.1126/science.1177808> PMID: 19713490
21. Parker HG, Chase K, Cadieu E, Lark KG, Ostrander EA. An insertion in the RSPO2 gene correlates with improper coat in the Portuguese water dog. *J Hered.* 2010; 101(5):612–7. <https://doi.org/10.1093/jhered/esq068> PMID: 20562213
22. Bauer A, Hadji Rasouliha S, Brunner MT, Jagannathan V, Bucher I, Bannoehr J, et al. A second KRT71 allele in curly coated dogs. *Anim Genet.* 2018;2016–9.
23. Salmela E, Niskanen J, Arumilli M, Donner J, Lohi H, Hytönen MK. A novel KRT71 variant in curly-coated dogs. *Anim Genet.* 2018; 50:101–4. <https://doi.org/10.1111/age.12746> PMID: 30456859
24. Dierks C, Mönke S, Philipp U, Distl O. Allelic heterogeneity of FGF5 mutations causes the long-hair phenotype in dogs. *Anim Genet.* 2013; 44(4):425–31. <https://doi.org/10.1111/age.12010> PMID: 23384345
25. Haworth K, Putt W, Cattanach B, Breen M, Binns M, Lingaas F, et al. Canine homolog of the T-box transcription factor T; failure of the protein to bind to its DNA target leads to a short-tail phenotype. *Mamm Genome.* 2001; 12(3):212–8. <https://doi.org/10.1007/s003350010253> PMID: 11252170
26. Hytönen MK, Grall A, Hédan B, Dréano S, Seguin SJ, Delattre D, et al. Ancestral T-box mutation is present in many, but not all, short-tailed dog breeds. *J Hered.* 2009; 100(2):236–40. <https://doi.org/10.1093/jhered/esn085> PMID: 18854372
27. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Pielberg GR, Sigurdsson S, et al. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet.* 2011 Oct 13; 7(10):e1002316. <https://doi.org/10.1371/journal.pgen.1002316> PMID: 22022279
28. Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat Commun.* 2019; 10(1):1489. <https://doi.org/10.1038/s41467-019-09373-w> PMID: 30940804
29. Schoenebeck JJ, Hutchinson SA, Byers A, Beale HC, Carrington B, Faden DL, et al. Variation of BMP3 Contributes to Dog Breed Skull Diversity. *PLoS Genet.* 2012; 8(8):1–11.
30. Marchant TW, Johnson EJ, McTeir L, Johnson CI, Gow A, Liuti T, et al. Canine Brachycephaly Is Associated with a Retrotransposon-Mediated Missplicing of SMOC2. *Curr Biol.* 2017; 27(11):1573–1584.e6. <https://doi.org/10.1016/j.cub.2017.04.057> PMID: 28552356
31. Faculty of Veterinary Science University of Sydney. Online Mendelian Inheritance in Animals, OMIA [Internet]. [cited 2019 Jan 14]. Available from: <http://omia.angis.org.au/>
32. Parker HG, Dreger DL, Rimbault M, Davis BW, Mullen AB, Carpintero-Ramirez G, et al. Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Rep.* 2017; 19(4):697–708. <https://doi.org/10.1016/j.celrep.2017.03.079> PMID: 28445722
33. Talenti A, Dreger DL, Frattini S, Polli M, Marelli S, Harris AC, et al. Studies of modern Italian dog populations reveal multiple patterns for domestic breed evolution. *Ecol Evol.* 2017; 8:2911–25.
34. Eckardt J, Kluth S, Dierks C, Philipp U, Distl O. Population screening for the mutation associated with osteogenesis imperfecta in dachshunds. *Vet Rec.* 2013; 172(14):364. <https://doi.org/10.1136/vr.101122> PMID: 23315765
35. Holder AL, Price JA, Adams JP, Volk HA, Catchpole B. A retrospective study of the prevalence of the canine degenerative myelopathy associated superoxide dismutase 1 mutation (SOD1:c.118G > A) in a referral population of German Shepherd dogs from the UK. *Canine Genet Epidemiol.* 2014; 1(1):10.
36. Monobe M, Bulla C, da Silva R, Lunsford K, Araujo JP Jr. Frequency of the MDR1 mutant allele associated with multidrug sensitivity in dogs from Brazil. *Vet Med Res Reports.* 2015;(January):111.
37. Mizukami K, Yabuki A, Kohyama M, Kushida K, Rahman MM, Uddin MM, et al. Molecular prevalence of multiple genetic disorders in Border collies in Japan and recommendations for genetic counselling. *Vet J.* 2016; 214:21–3. <https://doi.org/10.1016/j.tvjl.2016.05.004> PMID: 27387721
38. Takanosu M. Different allelic frequency of progressive rod-cone degeneration in two populations of Labrador Retrievers in Japan. *J Vet Med Sci.* 2017; 79(10):1746–8. <https://doi.org/10.1292/jvms.17-0243> PMID: 28855430

39. Crespi JA, Barrientos LS, Giovambattista G. von Willebrand disease type 1 in Doberman Pinscher dogs: genotyping and prevalence of the mutation in the Buenos Aires region, Argentina. *J Vet Diagnostic Investig.* 2018; 30(2):310–4.
40. Ahonen S, Seath I, Rusbridge C, Holt S, Key G, Wang T, et al. Nationwide genetic testing towards eliminating Lafora disease from Miniature Wirehaired Dachshunds in the United Kingdom. *Canine Genet Epidemiol. Canine Genetics and Epidemiology*; 2018; 5(1):2.
41. Donner J, Anderson H, Davison S, Hughes AM, Bouirman J, Lindqvist J, et al. Frequency and distribution of 152 genetic disease variants in over 100,000 mixed breed and purebred dogs. *PLOS Genet.* 2018; 14(4):1–20.
42. Pearson D. The Bedlington Terrier [Internet]. Ashcroft Bedlington Terriers. 2015 [cited 2019 Aug 8]. Available from: <http://www.ashcroftterriers.com/about-bedlington-terriers>
43. Lhasa Apso [Internet]. Dog Breed Info Center. 2018 [cited 2019 Apr 30]. Available from: <https://www.dogbreedinfo.com/lhasaapso.htm>
44. Think my new sheltie is docked [Internet]. Sheltie Nation. 2011 [cited 2019 Apr 30]. Available from: <https://sheltieforums.com/threads/think-my-new-sheltie-is-docked.10163/>
45. Short stumpy tail [Internet]. Staffordshire bull terrier. 2011 [cited 2019 Apr 30]. Available from: <http://staffy-bull-terrier.niceboard.com/t10059-short-stumpy-tail>
46. Opinions on having tail docked [Internet]. r/vizsla. 2017 [cited 2019 Apr 30]. Available from: [https://www.reddit.com/r/vizsla/comments/6ksvt6/opinions\\_on\\_having\\_tail\\_docked/](https://www.reddit.com/r/vizsla/comments/6ksvt6/opinions_on_having_tail_docked/)
47. Dumont BL, Payseur BA. Evolution of the genomic rate of recombination in mammals. *Evolution (N Y)*. 2008; 62(2):276–94.
48. Oguro-Okano M, Honda M, Yamazaki K, Okano K. Mutations in the Melanocortin 1 Receptor,  $\beta$ -Defensin103 and Agouti Signaling Protein Genes, and Their Association with Coat Color Phenotypes in Akita-Inu Dogs. *J Vet Med Sci.* 2011; 73(7):853–8. <https://doi.org/10.1292/jvms.10-0439> PMID: 21321476
49. Schmutz SM, Berryere TG, Barta JL, Reddick KD, Schmutz JK. Agouti sequence polymorphisms in coyotes, wolves and dogs suggest hybridization. *J Hered.* 2007; 98(4):351–5. <https://doi.org/10.1093/jhered/esm036> PMID: 17630272
50. Brockerville RM, McGrath MJ, Pilgrim BL, Marshall HD. Sequence analysis of three pigmentation genes in the Newfoundland population of *Canis latrans* links the Golden Retriever *Mc1r* variant to white coat color in coyotes. *Mamm Genome.* 2013; 24(3–4):134–41. <https://doi.org/10.1007/s00335-012-9443-x> PMID: 23297074
51. Schweizer RM, Durvasula A, Smith J, Vohr SH, Stahler DR, Galaverni M, et al. Natural selection and origin of a melanistic allele in North American Gray Wolves. *Mol Biol Evol.* 2018; 35(5):1190–209. <https://doi.org/10.1093/molbev/msy031> PMID: 29688543
52. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* 2010; 8(8):49–50.
53. Pilot M, Greco C, VonHoldt BM, Jędrzejewski W, Sidorovich VE, Konopinski MK, et al. Widespread, long-term admixture between grey wolves and domestic dogs across Eurasia and its implications for the conservation status of hybrids. *Evol Appl.* 2018;(July 2017):662–80. <https://doi.org/10.1111/eva.12595> PMID: 29875809
54. Dufresnes C, Remollino N, Stoffel C, Manz R, Weber J, Fumagalli L. Two decades of non-invasive genetic monitoring of the grey wolves recolonizing the Alps support very limited dog introgression. 2019;(November 2018):1–9.
55. Croxton-Smith A, editor. *Hounds and Dogs, their care, training & working for hunting, shooting, coursing, hawking, police purposes.* London: Lonsdale Library; 1932. <https://doi.org/10.3390/ani9070390>
56. Lee RB. *A history and description of the modern dogs of Great Britain and Ireland: Sporting Division.* London: Horace Cox; 1897.
57. Graham JA. *The Sporting Dog.* New York: The MacMillan Company; 1904.
58. Smith S, editor. *The encyclopedia of North American sporting dogs: Written by Sportsmen for Sportmen.* 1st ed. Minocqua, WI: Willow Creek Press; 2002.
59. Campbell KL, Campbell JR. *Companion Animals: Their biology, care, health and management.* 2nd ed. Upper Saddle River, NJ: Pearson Prentice Hall; 2009.
60. Quignon P, Herbin L, Cadieu E, Kirkness EF, Hédan B, Mosher DS, et al. Canine Population Structure: Assessment and Impact of Intra-Breed Stratification on SNP-Based Association Studies. *PLoS One.* 2007 Dec 19; 2(12):e1324-. <https://doi.org/10.1371/journal.pone.0001324> PMID: 18091995
61. Shearin AL, Hédan B, Cadieu E, Erich SA, Schmidt EV., Faden DL, et al. The MTAP-CDKN2A locus confers susceptibility to a naturally occurring canine cancer. *Cancer Epidemiol Biomarkers Prev.* 2012; 21(7):1019–27. <https://doi.org/10.1158/1055-9965.EPI-12-0190-T> PMID: 22623710

62. Edwards SM, Woolliams JA, Hickey JM, Blott SC, Clements DN, Sánchez-Molano E, et al. Joint genomic prediction of canine hip dysplasia in UK and US labrador retrievers. *Front Genet.* 2018; 9(March):1–12.
63. Pedersen NC, Liu H, McLaughlin B, Sacks BN. Genetic characterization of healthy and sebaceous adenitis affected Standard Poodles from the United States and United Kingdom. *Tissue Antigens.* 2012; 80:46–57. <https://doi.org/10.1111/j.1399-0039.2012.01876.x> PMID: 22512808
64. Sinmez CC, Yigit A, Aslim G. Tail docking and ear cropping in dogs: a short review of laws and welfare aspects in the Europe and Turkey. *Ital J Anim Sci.* 2017; 16(3):431–7.
65. South African Veterinary Council. The South African Veterinary Council policy on tail docking in dogs [Internet]. 2008 [cited 2019 Jan 14]. Available from: <https://www.savc.org.za/tail-docking>
66. New Zealand Veterinary Association. Canine Tail Docking [Internet]. 2018 [cited 2019 Jan 14]. Available from: <https://www.nzva.org.nz/page/policytaildock>
67. Australian Veterinary Association. AVA welcomes end to tail docking in Western Australia [Internet]. 2010. Available from: <https://www.ava.com.au/mediarelease/ava-welcomes-end-tail-docking-western-australia>
68. Animal Welfare Veterinary Division. Information on dog tail docking provided for the Animal Welfare Division [Internet]. 2002. Available from: <https://web.archive.org/web/20070626075746/http://www.defra.gov.uk/animalh/welfare/domestic/awbillconsulttaildocking.pdf>
69. Wijnrocx K, Francois L, Stinckens A, Janssens S, Buys N. Half of 23 Belgian dog breeds has a compromised genetic diversity, as revealed by genealogical and molecular data analysis. *J Anim Breed Genet.* 2016; 133:375–83. <https://doi.org/10.1111/jbg.12203> PMID: 26927793
70. Pedersen NC, Pooch AS, Liu H. A genetic assessment of the English bulldog. *Canine Genet Epidemiol. Canine Genetics and Epidemiology;* 2016; 3(1):6.
71. Kettunen A, Daverdin M, Helfjord T, Berg P. Cross-breeding is inevitable to conserve the highly inbred population of puffin hunter: The Norwegian Lundehund. *PLoS One.* 2017; 12(1):1–16.
72. Wiener P, Sánchez-Molano E, Clements DN, Woolliams JA, Haskell MJ, Blott SC. Genomic data illuminates demography, genetic structure and selection of a popular dog breed. *BMC Genomics.* 2017; 18(1):609. <https://doi.org/10.1186/s12864-017-3933-x> PMID: 28806925
73. Mastrangelo S, Filippo B, Auzino B, Marco R, Spaterna A, Ciampolini R. Genome-wide diversity and runs of homozygosity in the “Braque Français, type Pyrénées” dog breed. *BMC Res Notes.* 2018; 11(1):11–6. <https://doi.org/10.1186/s13104-017-3119-2>
74. Pohjoismäki JLO, Lampi S, Donner J, Anderson H. Origins and wanderings of the finnish hunting spitzes. *PLoS One.* 2018; 13(6):1–17.
75. Jansson M, Laikre L. Pedigree data indicate rapid inbreeding and loss of genetic diversity within populations of native, traditional dog breeds of conservation concern. *PLoS One.* 2018; 13(9):1–17.
76. Dreger DL, Rimbault M, Davis BW, Bhatnagar A, Parker HG, Ostrander EA. Whole-genome sequence, SNP chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. *Dis Model Mech.* 2016 Dec 1; 9(12):1445–60. <https://doi.org/10.1242/dmm.027037> PMID: 27874836
77. Ostrander EA, Kruglyak L. Unleashing the Canine Genome Unleashing the Canine Genome. *Genome Res.* 2000; 10:1271–4. <https://doi.org/10.1101/gr.155900> PMID: 10984444
78. Donner J, Kaukonen M, Anderson H, Möller F, Kyöstiä K, Sankari S, et al. Genetic panel screening of nearly 100 mutations reveals new insights into the breed distribution of risk variants for canine hereditary disorders. *PLoS One.* 2016; 11(8):1–18.
79. Venables WN, Ripley BD. *Modern Applied Statistics with S.* 4th ed. New York: Springer; 2002.
80. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available from: <https://www.r-project.org>.