

An Insertion in the *RSPO2* Gene Correlates with Improper Coat in the Portuguese Water Dog

HEIDI G. PARKER, KEVIN CHASE, EDOUARD CADIEU, KARL GORDON LARK, AND ELAINE A. OSTRANDER

From the Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892 (Parker, Cadieu, and Ostrander); and Department of Biology, University of Utah, Salt Lake City, UT 84112 (Chase and Lark).

Address correspondence to Dr E. A. Ostrander at the address above, or e-mail: eostrand@mail.nih.gov.

Abstract

We recently showed that genes at 3 loci account for the majority of variation in canine fur. Allelic variation at genes controlling length of fur, texture, and curl is responsible for the striking phenotypic variety observed among purebred dogs in the United States today. In this paper, we investigate the phenomenon of “improper coat” (IC) or a coat that is not typical of the breed. IC is occasionally observed among specific breeds, such as the Portuguese Water Dog (PWD), and is characterized by short hair on the head, face, and lower legs, rather than a thick and even coat covering the whole body. The IC is reminiscent of that observed on the curly or flat-coated retriever, thus making such dogs unable to compete effectively in conformation events. We have found that the presence of the wild-type allele, rather than the expected variant allele at the R-spondin 2 (*RSPO2*) gene, accounts for this phenotype. The development of a genetic test that distinguishes these 2 allelic types would allow breeders to easily avoid producing PWD with ICs.

Key words: fur, furnishings, genetics, morphology, mutation

Most of the phenotypic variation in the coat of domestic dogs can be traced to genetic alterations in just 3 genes. The first, fibroblast growth factor 5 (*FGF5*), is associated with long hair; keratin 71 (*KRT71*) is associated with the degree of curl; and a variant in the R-spondin 2 gene (*RSPO2*) is associated with both wiry texture and a growth pattern of the fur (Cadieu et al. 2009). The latter trait, known as “furnishings,” increases hair growth on the face and legs and is typified by the canine moustache and eyebrows such as those exemplified by the schnauzer breeds, among others.

The effects of these 3 mutations and combinations of the same were identified through a large genome-wide association study, in part because each has become fixed during the creation of multiple breeds displaying similar coat types (Cadieu et al. 2009). Occasionally, however, individuals within a single breed are born that do not adhere to the breed standard. In such cases, breeders often wish to have the underlying genetic variant identified so that a suitable genetic test may be developed. One such example is the improper coat (IC) of the Portuguese Water Dog (PWD), a breed which is expected to have “a profuse thickly planted coat of strong healthy hair, which covers the body evenly” including the face, top of the head, and legs (Figure 1A) (American Kennel Club (AKC) 1998).

A typical PWD, which meets the breed standard, is homozygous for a variant form of the *FGF5* gene (Cadieu et al. 2009), which encodes a Cys95Phe change in exon 1 that is associated with long rather than short hair. In addition, PWD may carry an Arg151Trp change in the coding region of the *KRT71* gene in exon 2 that is associated with curly hair. Because both curly and wavy coat types are allowed by the PWD standard, PWDs are as likely to be lacking that genetic variant as to have it. Finally, all PWD should have the 167 bp 3' UTR insertion in *RSPO2* that our data suggests affect expression levels of the gene and is associated with coat texture and the furnishings growth pattern (Cadieu et al. 2009). In the original study, however, we observed that 2 of 43 PWDs genotyped did not carry the expected *RSPO2* insertion. One of those 2 dogs was described in written material provided by the owner as having an “improper coat.” A PWD with IC will have short hair on the head and face, as well as on their lower legs (Figure 1B). In addition, their appearance is often described as that of a flat- or curly coated retriever.

We have shown that changes in the length of the hair and the regions on the body where hair grows long are controlled, respectively, by the *FGF5* and *RSPO2* genes (Cadieu et al. 2009). Based on this study, a normal coated PWD should carry 2 copies of the T allele in exon 1 of

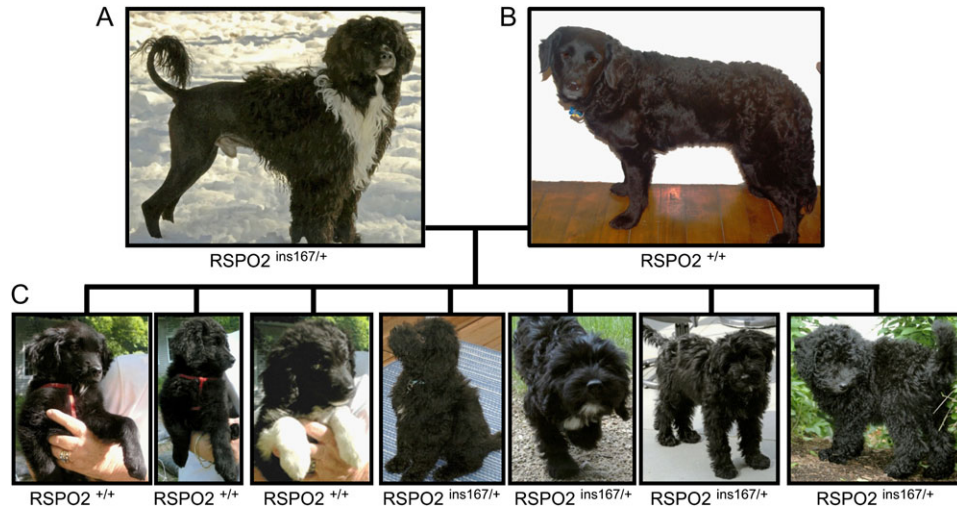


Figure 1. Inheritance of IC among PWD. (A) A male PWD showing a normal wavy coat and sporting the traditional lion cut, where the hindquarters and muzzle are shorn. (B) Female PWD with IC in its natural state, no trimming. Note the lack of facial hair, hair on the legs, and the shortened coat all over. (C) Offspring of the cross between A and B. Three of the 7 pups are homozygous for the wt allele and display the IC phenotype. The remaining 4 appear normal but are all carriers of the IC allele displaying perfect segregation with the lack of insertion in *RSPO2*.

FGF5, and at least one copy of the insertion in the 3' UTR of *RSPO2* (Cadieu et al. 2009). We therefore hypothesized that the IC of the PWD was due to variant alleles in either *RSPO2* and/or *FGF5*.

Methods

DNA Isolation and Sample Selection

DNA was isolated from whole blood drawn into acid citrate dextrose or ethylenediaminetetraacetic acid tubes using the Qiagen Genomic-tip 100/G kit and the manufacturer's specifications (Qiagen Corp., Valencia, CA). The dataset included 5 affected dogs, chosen because of their IC phenotype; 19 putative carriers selected based on the production of affected offspring; and 10 dogs that were unaffected and did not carry the variant genotype. The latter set of dogs were chosen because they produced no affected offspring when crossed to either known affected dogs (based on a minimum of 5 pups produced per dog) or when crossed with known carriers (minimum of 11 pups produced), thus assuring their genotypic status. An additional 253 PWDs were chosen from the parent Georgie study, which has a dataset of more than 1000 dogs, as a population representative sample (Chase et al. 2002).

Fragment Analysis

Polymerase chain reactions (PCRs) were carried out in a buffer of 10 mM Tris, 40 mM NaCl, 2 mM MgCl₂, 200 μM deoxynucleoside triphosphate, 200 μM spermidine, and 6 pM forward and reverse primer. Amplification was performed using the Applied Biosystems PCR block

GeneAmp PCR System 9700 as follows: 8 cycles of 94 °C for 20 s, 64 °C for 20 s, and 72 °C for 40 s, followed by 30 cycles of 94 °C for 20 s, 56 °C for 20 s, and 72 °C for 40 s, followed by a final extension at 72 °C for 10 min. Fragment separation was performed using an Applied Biosystems 3130xl Genetic Analyzer with a 36 cm capillary array and pop-7 polymer (Applied Biosystems, Santa Clara, CA). GeneMapper version 3.7 (Applied Biosystems) was used to call the alleles. Lab personnel were blinded to phenotype of dogs being genotyped. Each reaction plate contained a mixture of dogs of varying phenotypes and suspected genotypes.

Statistical Methods

We tested for association between a specific ancestor and the insertion using the correlation between consanguinity with the ancestor and allele count (0, 1, or 2) for all genotyped animals. Significance was established using Monte Carlo simulations in which a marker with a similar allele frequency was simulated in the population. Five thousand randomly simulated alleles were used to estimate the null distribution of the correlation coefficient (e.g. no association with a specific ancestor).

Modeling Gene Flow

Gene flow simulations were performed using EvolGenius5.1 (Kliman 2002) and Web PopGen (Sheehy 2004). These 2 programs simulate changes in allele frequencies over time through genetic drift given a population size, starting allele frequency, mutation rate, migration rate, and fitness. PopGen also allows modeling of a population bottleneck or temporary reduction in population size, in this

Table 1 Simple sequence repeat markers near FGF5. Positions given are on chromosome 32 based on the CanFam2 assembly. *P* values are for the single most significant allele from each marker

Name	Primer sequences	Position	P-value
Lfgf5b	F-CCAAGCTACCCTTGTCTCCA R-CCCTCTGCCTATGTCTCTGG	7503836–7503914	0.19
Lfgf5c	F-GAGCCTGCTTCCCCTTTTT R-CCAGGTTAGCTATTTAAGTAGGATGC	7420398–7420463	0.13
Lfgf5d	F-TGAGTTCCCCAAACTCCTTG R-TTCTCCCTCTGCCTGTGTCT	7592071–7592107	0.09
Lfgf5f	F-TCTCAATGATTATTCAAACCTTGC R-CCTGTCTGTGTCTCTGCCTCT	7750576–7750615	0.06
Lfgf5g	F-TGGAGAATAGATGAATATCCAAAGG R-TCCCCAAACAATAAGTGGAG	7218618–7218653	0.06

case representing the migration into the United States from a small subset of the European population.

In order to calculate an effective population size from which to model gene flow, we made the following assumptions about the PWD population in the United States, using AKC registration data and common breeding practices. First, we assumed that the current size of the population is about 20 000 dogs (Chase et al. 2006); second, only 10% of dogs from any given breed contribute to the next generation (Ostrander and Kruglyak 2000); and third, the use of popular sires reduces the number of males contributing to future generations compared with females. Starting with 12 dogs from the original US founders (Braund and Miller 1986; AKC 1998), and inferring a steady growth rate in the breeding population of 14% per

generation, with a modest estimation of 1:4 male to female breeding ratio, we can calculate an average long-term effective population size adjusted for sex contribution at $N_e = 25$ (Li 1997).

The scenarios modeled are not exhaustive but represent a subset of known methods for breed recreation. The maximum number of generations was set at 80 from the recreation of the breed in Portugal and a maximum of 40 generations since introduction into the United States. The “A” allele represents the normal PWD coat mutation, and the “a” represents the IC or wild-type (wt) allele. Initial allele frequency was set at 0.125, representing one homozygous wt dog in the 8 original European founders. The starting allele frequency in the United States was modeled as either the same as the European population

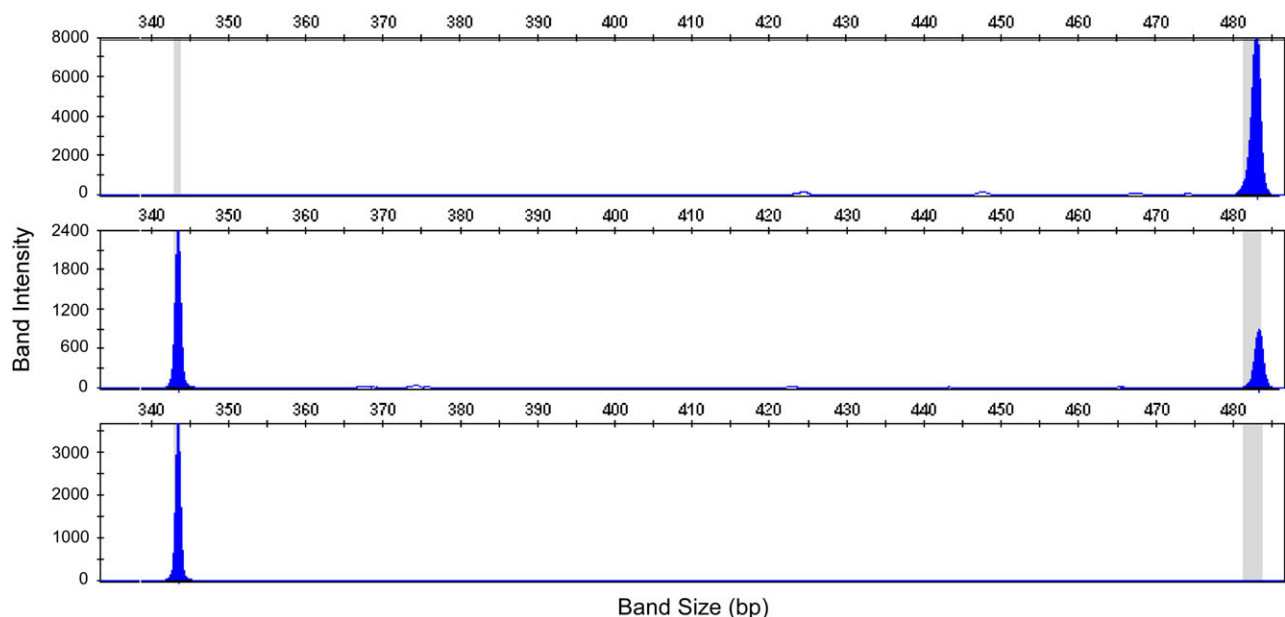
**Figure 2.** Size discrimination of amplicons containing the indel in the 3' UTR of *RSP02* at position 11,634,766. These 3 graphs define: (A) the variant (furnishings) allele—will not produce IC, (B) carrier—can produce IC when mated to another carrier or to an affected dog, and (C) affected (wt allele)—PWD with this genotype displays an IC. Fragment size is indicated on the x-axis and intensity on the y-axis. The sizes of the 2 fragments are 343 (wt) and 483 bp. Primers used to amplify the segment are forward: 5'-TGGCTAAAGAAAACCTCCACAA-3' and reverse: 5'-TGCATTGGCAAAACTACCTC-3'.

Table 2 Results of gene flow simulations modeled after the historical recreation of the PWD breed and subsequent importation into the United States

Original founders	Average generations to AA fixation	% Fixed AA	% Fixed aa	% Not fixed	Average "a" allele frequency ^a
80 generations, bottleneck at 40					
No selection	bb 70% ^b	83.00	6.00	11.00	0.55
0.9 fitness ^c	bb 70%	86.00	1.00	13.00	0.25
0.5 fitness	bb 94%	100.00	0.00	0.00	0.00
40 generations					
No selection	13.00	70.00	1.00	29.00	0.47
0.9 fitness	13.00	79.00	0.00	21.00	0.32
0.5 fitness	12.00	98.00	0.00	2.00	0.29
40 generations, low-level migration					
No selection	n.d.	23.00	0.00	77.00	0.16
0.9 fitness	n.d.	28.00	0.00	72.00	0.16
0.5 fitness	n.d.	46.00	0.00	54.00	0.06
40 generations, low-level migration, no "a" in founders					
No selection	n.d.	26.00	0.00	74.00	0.13
0.9 fitness	n.d.	40.00	0.00	60.00	0.14
0.5 fitness	n.d.	52.00	0.00	48.00	0.07
40 generations, one hybridization event in first generation					
No selection	13.00	73.00	1.00	26.00	0.28
0.9 fitness	10.00	90.00	0.00	10.00	0.26
0.5 fitness	10.00	95.00	0.00	5.00	0.13

^a Average calculated from only those simulations where neither allele was fixed. The average does not include all the simulations in which the frequency of "a" is 0 or 1.

^b Percentage of simulations in which the "A" allele was fixed in the population before the bottleneck (bb) occurred allowing no possibility of carrying the IC allele into the US population.

^c Fitness of the "aa" genotype relative to the "Aa" or "AA" genotype, where "a" is the allele responsible for IC.

(0.125) or 0. Low-level migration into the PWD population was simulated at 1 of every 20 dogs per generation, taken from a population with the same allele frequency as the assumed European founding population. The bottleneck was modeled as a reduction to 12 dogs in the 40th generation, lasting for 5 generations. In simulations where no migration was allowed, the number of generations to fixation of the "A" allele was calculated and averaged over all runs. If the "A" allele did not reach fixation in the allotted number of generations (40–80), the frequency of the "a" allele was noted and averaged. Reduction in fitness was only applied to the homozygous wt individuals (aa genotype) as heterozygous dogs would be indistinguishable based on phenotype alone. Simulations were repeated 100 times for all models.

Results

To determine the source of the abnormal coat in the PWD, we tested for the presence of variant alleles in both the

RSPO2 and *FGF5* genes using DNA isolated from 5 PWD with IC and 29 with normal coats. Of the latter, 19 were predicted to be carriers of the critical variant as they had produced IC offspring. The remaining 10 of 29 were expected to carry breed normal genotypes at both *RSPO2* and *FGF5*, based on the lack of IC dogs among their progeny.

We observed that all the PWDs, regardless of coat type, were homozygous for the T allele at *FGF5*, which is associated with long hair on the body. In addition, we found no association with the *FGF5* gene or surrounding region on canine chromosome 32 (CFA32) with IC (Table 1).

By comparison, none of the 5 dogs with IC carried the expected insertion in the 3' UTR of the *RSPO2* gene, which is associated with furnishings, whereas all the 29 PWD with standard coats carried the insertion. The 19 obligate carriers, as expected, each carried 1 copy of the wt allele and 1 copy of the variant (Figure 2).

We next analyzed a family of PWD that segregated IC and found that phenotype and genotype correlated perfectly (Figure 1C). All affected dogs being homozygous for the wt allele. This suggested that the lack of insertion in the *RSPO2* gene is the sole cause of IC in PWD.

To assess the prevalence of the wt allele in the PWD population, we genotyped an additional 253 dogs selected at random from the "Georgie Project" collection of PWDs (<http://www.georgieproject.com/>). As expected, the majority, 195, were homozygous for the insertion and an additional 58 were heterozygous. No additional dogs were identified that were homozygous for the wt allele. Based on the randomly chosen dogs, we calculated an allele frequency of ~12% for the wt allele at *RSPO2* in the PWD population. Assuming Hardy–Weinberg (HW) equilibrium, 1.4% of PWDs should display IC. Based on information collected through the Georgie Project, approximately 0.5% of PWD report the IC phenotype. Although the number of affected dogs reported is 3 times lower than the number expected, the distribution of genotypes in the dataset is not significantly different than expected based on HW calculations ($\chi^2 P = 0.19$). The true frequency of IC dogs is likely higher than what we report here, as owners with IC dogs are less likely to have participated in the parent study (Georgie Project) because their dogs did not meet the breed standard (<http://www.georgieproject.com/>). However, if we assume that the reported numbers are representative of the breed as a whole, then the reduced appearance of affected dogs could indicate that current selection against the allele only applies to the affected individuals, not to carriers of the IC allele.

Simulations based on historical number of founders and an average effective population size suggest that the presence of the wt allele at modern frequencies would require introduction of the allele into the population within the last 40 generations (Table 2). This event most likely occurred during breed formation in the US, as the majority of simulations end in fixation of the desired allele within 10–15 generations, once immigration is halted. This does not rule out the presence of a wt allele in the original founders; although without reintroduction and with

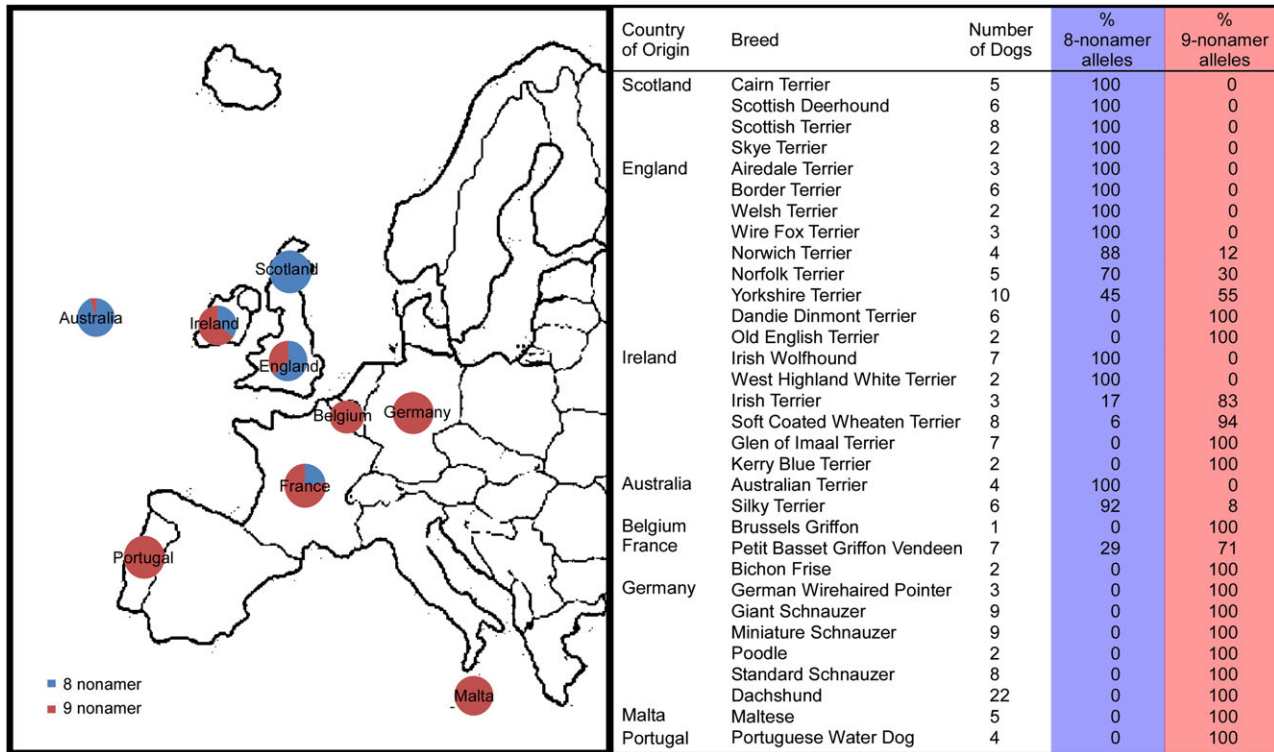


Figure 3. European distribution of the 2 alleles of the *RSPO2* insertion mutation in breeds with furnishings. On the left, pie charts on the map of Europe show the distribution of alleles in dogs originating from 8 different countries. The chart on the right gives the numbers and breeds of dog interrogated. Country of origin is assumed based on written histories (American Kennel Club 1998). All Australian breeds assessed were derived from European founders primarily of UK origin.

selection against the IC type, this allele would have quickly been lost.

Discussion

The modern PWD was recreated in the 1930s by a fancier of the ancient mariner dog (Braund and Miller 1986). The formal breeding program to re-establish the PWD breed is thought to be based on a single working dog that epitomized the breed standard. As a single founder, this dog would have been bred to several females that may have come from a variety of water-retrieving breeds, many of which do not have the furnishings phenotype. This offers a possible source for the wt allele in the current population. The phenotype associated with the wt allele is recessive therefore removing it from the population would be difficult especially because coat type does not affect the health of the dog.

Alternately, this allele may have been introduced throughout breeding in the United States, prior to breed recognition by the AKC and subsequent closing of the studbook. Analysis of data simulating genetic drift, assuming even moderate selection against the IC phenotype, would suggest that the IC allele was introduced into the PWD within recent generations (Table 2). Most scenarios that do not involve some form of migration into the US population

result in fixation of the proper coat allele in as few as 10 generations. These estimates are generous given that the simulations are based on average effective population size of the breed over 40–80 years. The effective size would have been much smaller in the first 10 generations when fixation would have taken place. Determination of the geographical origins of allele introduction could presumably be done by examination of foreign PWD populations whose ancestors were unilaterally exported to the US. If the allele was introduced in the US, we would not expect to see the wt allele in the foreign population. Alternately, if the allele was introduced from an imported PWD, the wt allele would be present elsewhere in the world at similar or higher frequencies.

In our original analysis of the *RSPO2* mutation, we observed 2 versions of the insertion, the first with 9 repeats (V9) of a 9 base-pair sequence (ATAATGAAC) and a second with only 8 (V8). The variant found among PWDs has 9 copies of the nonamer, as does that of their relatives the poodles and almost all the breeds developed in continental Europe. The V8 variation is found almost exclusively among breeds created in the British Isles and is fixed in the majority of terriers from the United Kingdom (Figure 3).

Based on the distribution of alleles, we hypothesize that the V8 variant likely arose as a deletion produced by replication slippage (Viguera et al. 2001) within the original

insertion, perhaps in 1 of the old English terriers such as the rough black and tan terrier, the presumed progenitor to most modern terrier breeds (AKC 1998). The pervasive contribution of the V8 allele of *RSPO2* in the breeds of the United Kingdom indicates a single common founder for all these breeds. In contrast, the widespread prevalence of the V9 variant throughout Europe suggests that it is the ancestral version of the insertion that first created the furnished coat found in many modern breeds (Figure 3).

Although IC in the PWD is not a life-threatening condition, it does pose a problem for owners intent on showing their dogs and selling their offspring, as these individuals do not meet the breed standard and are therefore ineligible to win championships and have little value as breeding stock. In addition, IC PWDs are reported to have an undercoat, and they shed, which can be an issue for owners expecting a nonshedding dog to avoid allergic reactions (<http://www.pwdca.org/health/conditions/allergy/>). Indeed, the fact that PWD have a single coat that does not shed, much like human hair, is one of their most desirable traits. Thus, there is strong motivation within the PWD community to prevent the spread of the wt allele within the breed.

The data presented here can serve as the basis for development of a genetic test for IC in the PWD, which will allow owners to identify carriers. By doing so, they can reduce the appearance of the undesirable phenotype in litters without removing individuals from the breeding pool. The latter issue is always critical when dealing with purebred dogs. Each dog contributes some amount of hybrid vigor to the breed and unnecessary removal of individuals from the breeding pool should be avoided. Careful selecting of mates through genetic testing is far more desirable (Ostrander and Wayne 2005).

Based on comparisons of normal-coated PWD to those with IC, we find that the *RSPO2* mutation not only controls furnishings but also contributes to overall hair length. It may also contribute to the nonshedding phenotype. These aspects of the mutation were not evident in the multibreed comparisons from the previous study because the genetic backgrounds of all breeds tested were very different and obscured minor changes in phenotype. Careful studies on the effect of a mutation in a single genetic background will allow us to gain additional insights regarding the complete phenotype created by each variant.

Funding

National Institutes of Health (GM063056) to K.C. and K.G.L.; Intramural program of the National Human Genome Research Institute.

Acknowledgments

We would like to thank the many dog owners and breeders who have so generously contributed samples and time to our studies.

References

- American Kennel Club. 1998. The complete dog book. New York: Howell Book House.
- Braund K, Miller DF. 1986. The complete portuguese water dog. New York: Howell Book House.
- Cadiou E, Neff M, Quignon P, Walsh K, Chase K, Parker HG, Vonholdt BM, Rhue A, Boyko A, Byers A, et al. 2009. Coat variation in the domestic dog is governed by variants in three genes. *Science*. 326:150–153.
- Chase K, Carrier DR, Adler FR, Jarvik T, Ostrander EA, Lorentzen TD, Lark KG. 2002. Genetic basis for systems of skeletal quantitative traits: principal component analysis of the canid skeleton. *Proc Natl Acad Sci U S A*. 99:9930–9935.
- Chase K, Sargan D, Miller K, Ostrander EA, Lark KG. 2006. Understanding the genetics of autoimmune disease: two loci that regulate late onset Addison's disease in Portuguese Water Dogs. *Int J Immunogenet*. 33:179–184.
- Kliman RM. 2002. A project based approach to teaching complex population genetics to undergraduates. *Bioscene*. 27:13–20.
- Li W-H. 1997. Molecular evolution. Sunderland (MA): Sinauer Associates, Inc.
- Ostrander EA, Kruglyak L. 2000. Unleashing the canine genome. *Genome Res*. 10:1271–1274.
- Ostrander EA, Wayne RK. 2005. The canine genome. *Genome Res*. 15:1706–1716.
- Sheehy B. 2004. Web PopGen. [cited 2009 Dec 04]. Available from: URL http://www.runet.edu/~rsheehy/Gen_flash/popgen/.
- Viguera E, Canceill D, Ehrlich SD. 2001. Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J*. 20:2587–2595.

Received December 4, 2009; Revised April 12, 2010;

Accepted May 12, 2010

Corresponding Editor: Robert Wayne