Oct 4, 2016
Lauren H Henson
Smithsonian Conservation Biology Institute
Center for Conservation and Evolutionary Genetics
3001 Connecticut Ave., NW
Washington, DC 20008
Phone 250-891-7883
hensonlh@gmail.com

CHARACTERIZATION OF GENETIC VARIATION AND BASIS OF INFLAMMATORY BOWEL DISEASE IN THE TOLL-LIKE RECEPTOR 5 GENE OF THE RED WOLF AND THE MANED WOLF

*HENSON, LAUREN H, Smithsonian Conservation Biology Institute, Center for Conservation and Evolutionary Genetics, 3001 Connecticut Ave., NW, Washington, DC 20008: Environmental Science and Policy, George Mason University, 4400 University Dr., Virginia 22030

SONGSASEN, NUCHARIN, Smithsonian Conservation Biology Institute, Center for Species Survival, 1500 Remount Rd., Front Royal, VA 22630

WADDELL, WILL, Point Defiance Zoo and Aquarium, 5400 N Pearl St, Tacoma, WA 98407

WOLF, KAREN N, Point Defiance Zoo and Aquarium, 5400 N Pearl St, Tacoma, WA 98407

EMMONS, LOUISE, Smithsonian National Museum of Natural History, 10th St & Constitution Ave., NW. Washington, DC 20560

GONZALEZ, SUSANA, Instituto deInvestigaciones Biológicas Clemente Estable. Ministerio de Educación y Cultura. , Av. Italia 3318- 11600, Montevideo

FREEMAN, ELIZABETH, School of Integrative Studies, George Mason University, 4400 University Dr., Virginia 22030

MALDONADO, JESUS, Smithsonian Conservation Biology Institute, Center for Conservation Genomics, 3001 Connecticut Ave., NW, Washington, DC 20008

ABSTRACT Characterizing Toll-like receptors across taxa can lead to an increasingly accurate documentation of the evolutionary processes acting within this receptor class, as well as a greater understanding of the diseases associated with these receptors. This study examines two sequenced portions of the Toll-like receptor 5 protein coding gene in two imperiled canid species: the near threatened maned wolf (*Chrysocyon brachyurus*) and the critically endangered red wolf (*Canis rufus*), to characterize genetic variation and investigate the presence of single nucleotide polymorphisms (SNPs) previously associated with canine inflammatory bowel disease (IBD). Both maned and red wolves suffer from inflammatory bowel disease, threatening the sustainability of their crucial *ex situ* populations. Here we report novel polymorphic positions found in maned and red wolf TLR5, differences in variation with regard to nucleotide polymorphisms and resulting amino acid variation between maned wolves, red wolves, gray wolves and domestic dogs. Domestic dog SNPs associated with IBD were not found to be polymorphic in maned wolves and red wolves. Samples of both focal species and gray wolves lack the protective alleles present in many dog breeds, suggesting a potential genetic predisposition for IBD in these two wild canid species and a possible development of these protective alleles post domestication. This potential predisposition informs *ex situ* management practices and treatment for IBD.

KEY WORDS maned wolf; red wolf; Toll-like receptor 5; Inflammatory Bowel Disease

With the perils of habitat fragmentation, hunting and disease threatening *in situ* populations, the viability of *ex situ* populations is becoming increasingly relevant to the survival of many species. This study focuses on two canid species: the critically endangered North American red wolf (*Canis rufus*) and the near threatened South American maned wolf (*Chrysocyon brachyurus*) (IUCN 2016). The red wolf, a canid that was once endemic to the southeastern United States, was considered extinct in the wild in 1980 due to habitat loss, hunting and the introgression of coyote genes (Fredrickson & Hedrick 2006). Due to this introgression, the taxonomic status of red wolves is highly controversial with many studies indicating that the red wolf is a hybrid species between gray wolves and coyotes while others insist it is a distinct species (Gese et al. 2015). The *in situ* population of red wolves was recovered with an intensive reintroduction program in North Carolina beginning in 1987, that has resulted in a population of approximately 75 individuals residing within the current 6,000 km^2 reintroduction area (Harrison 2014). The maned wolf faces many of its own challenges (Ratter et al. 1997, Deem & Emmons 2005, Maia & Gouveia 2002) and as of 2008 the highly fragmented *in situ* maned wolf population was estimated at only 17,000 mature individuals and is predicted to decline by at least 10% in the next ten years (Paula et al. 2008).

Though the importance of viable *ex situ* populations for both maned and red wolves is becoming increasingly apparent, both species suffer health (Phillips & Scheck 1991, Gilioli & Silva 2000) and reproductive difficulties (Rabon 2011, Ginsberg 1994, Rodden et al. 1996, Johnson et al. 2014). Gastrointestinal disease is a major factor in mortalities in both red and maned wolves and has a high prevalence in both *ex situ* populations (Acton et al. 2000, Maia & Gouveia 2002, Stirling et al. 2008, Seeley et al. 2016).

87 Inflammatory bowel disease (IBD), a common diagnosis in both species, is characterized
88 by inflammation of the gastrointestinal tract (Craven et al. 2004). IBD is a multifaceted disorder
89 (German et al. 2003) that has both microbial (Simpson & Jergens 2011, Inness et al. 2007,
90 Xenoulis et al. 2008) and genetic bases (Cario & Podolsky 2000, Himmel et al. 2008) in other
91 species. In the domestic dog (*Canis familiaris*), the disease is linked to single nucleotide
92 polymorphisms (SNPs) in the Toll-like receptor 4 (TLR4) and Toll-like receptor 5 (TLR5) genes
93 (Kathrani et al. 2010), with two protective alleles against IBD identified in TLR5 across 38
94 different breeds (Kathrani et al. 2011).
95 Toll-like receptors (TLRs) are pattern recognition receptors essential to the functioning of
96 the innate immune system that belong to a large family of interleukin 1 receptors (Akira 2003).
97 These transmembrane receptors consist of a cytoplasmic Toll-interleukin 1 receptor (TIR)
98 domain, responsible for downstream signal transduction, transmembrane domains and leucine
99 rich repeat (LRR) extracellular domains (Kawai & Akira 2010). These extracellular LRR motifs
100 form a ligand-binding horseshoe shaped solenoid-like structure that recognizes pathogen
101 associated molecular patterns (PAMP) (Leulier & Lemaitre 2008, Bell et al. 2003) on the surface
102 of bacterial cells and helps to initiate an appropriate immune response through the production of
103 cytokines (Akira et al. 2001). The pathogen recognition capability of TLRs make them an
104 important component of the innate immune system and indicates a greater specificity for this
105 system (Kawai & Akira 2010). Six TLR families have been identified in vertebrates, with each
106 family recognizing a specific class of PAMP (Roach et al. 2005). Receptors within the TLR5
107 family are responsible for detecting bacterial flagellin and mammalian TLR5 recognizes flagellin
108 from both Gram-negative and Gram-positive bacteria (Hayashi et al. 2001).

Phylogenetic analysis places the origin of TLRs at 700 mya (Leulier & Lemaitre 2008). Studies attempting to identify the applicable model of evolution for these immune genes have oscillated between claiming that vertebrate TLRs are highly conserved because of the functional constraint of PAMPs (Roach et al. 2005) or that they are experiencing positive selection as a result of their constant interaction with rapidly evolving pathogens (Areal et al. 2011). Signatures of positive selection have been found in all mammalian TLRs across carnivores, lagomorphs, rodents, primates and artiodactyla, with non-viral TLRs having higher rates of positive selection than viral TLRs (Areal et al. 2011). In all species where adaptive selection has been identified, selective pressure is focused on the LRR extracellular domain because of its interaction with PAMPs (Areal et al. 2011) while a large portion of the TLR domain remains relatively conserved (Akira 2003). In TLR5, evidence of adaptive evolution has been found in the LRR pattern recognition extracellular domain (Wlasiuk et al. 2009, Smith et al. 2012) and a specific signature of adaptively evolving codons within this domain identified in domesticated mammals (Smith et al. 2012).

Toll-like receptors, because of their important role in the innate immune system, have been associated with many maladies. Mutations in TLRs, or their associated signaling pathways, have been linked to pneumococcal disease, systemic lupus erythematosus, chagas cardiomyopathy, malaria and tuberculosis in humans (Netea et al. 2012). Of particular interest to the present study, TLRs have also been implicated in the pathogenesis of gastrointestinal disorders (Netea et al. 2012). A healthy gut is characterized by its ability to regulate its immune response to food antigens and commensal bacteria while maintaining the ability to respond to pathogens. When this balance is disrupted, it can lead to inflammation and IBD. TLRs play an important role in maintaining this balance (Fukata & Abreu 2008). In humans, polymorphisms in

the TLR2 gene as well as the TLR4 gene are more likely to be present in patients with colorectal cancer, and a TLR9 polymorphism has been associated with Crohn's disease. SNPs in TLR1, 2 and 6 associate with both ulcerative colitis and Crohn's disease (Fukata & Abreu 2008), and genomic methods implicate TLR7 and 8 in celiac susceptibility (Netea et al. 2012). In mice, individuals that lack TLR2, 4, 5 or 9 are more likely to develop colitis and have a higher mortality risk (Maloy & Powrie 2011, Vijay-Kumar et al. 2007). Polymorphisms in the TLR5 gene are significantly associated with IBD in domestic dogs (Kathrani et al. 2010), with two protective SNPs identified across many different dog breeds (Kathrani et al. 2011).

Our aims in the present study were: 1) to characterize polymorphisms within two selected regions of the TLR5 gene in maned and red wolves; and 2) to determine the role of previously identified genetic markers for IBD in these two disease-prone species. We report novel polymorphisms in these previously unsampled threatened species and differing variability in TLR5 among four canid species: the maned wolf, red wolf, gray wolf and domestic dog. We also demonstrate a potential role for IBD SNPs in the pathogenesis of IBD within maned wolves and red wolves.

METHODS

Thirty one maned wolves (24 *ex situ* and 7 *in situ)* and fifteen red wolves were sampled for this study. Due to the opportunistic collection of samples an IACUC was not required by either the Smithsonian Conservation Biology Institute's IACUC committee or George Mason University's IACUC committee. *Ex situ* maned wolf samples were collected from individuals housed at the Smithsonian Conservation Biology Institute in Front Royal, VA and at four other Association of Zoos and Aquariums (AZA) accredited institutions. Maned wolf *in situ* samples represent populations situated in Bolivia (N=5), Argentina (N=1) and Brazil (N=1). Red wolf *ex*

155 *situ* samples were collected from individuals at the Point Defiance Zoo and Aquarium in Tacoma,
156 WA. For extant *ex situ* individuals of both species, blood was collected opportunistically during
157 routine veterinary examinations. For deceased *ex situ* individuals, necropsy samples of liver were
158 collected. *In situ* maned wolf samples were obtained from DNA extracted for previous studies
159 investigating maned wolf genetic variability throughout their range (Gonzalez et al. 2015).
160 Samples from Argentina and Brazil are from samples stored at the Conservation Genetics
161 Laboratory at Departamento de Biodiversidad y Genética-IIBCE-Uruguay. Bolivian samples
162 were obtained from populations in Noel Kempff Mercado National Park (Emmons et al. 2012).
163 DNA was extracted from whole blood and tissue using a Qiagen DNeasy blood and tissue
164 kit (QIAGEN). For extraction from whole blood Qiagen tissue lysis buffer was substituted for
165 phosphate buffered saline (PBS) and DNA eluted in 100 µl buffer AE with no repeat elution. For
166 tissue samples a 1g tissue sliver was used for extraction, incubated at 56 ºC overnight for
167 thorough lysis and eluted in 100 µl buffer AE with no repeat elution. DNA concentration and
168 quality was measured using a Nanodrop 1000 Spectrophotometer (Thermo Scientific).
169 Two fragments surrounding the three IBD associated SNPS in domestic dogs (Kathrani et
170 al. 2010) were selected for amplification (Fig. 1). Both fragments are approximately 350bp
171 (trimmed) with one fragment (Frag1) containing SNPs G22A (G727A in this study) and C100T
172 (C805T); and the second fragment (Frag2) containing SNP T1844C (C2549T). Primers (Supp
173 Fig. 1) were designed to amplify these fragments using the Primer3 software (Untergasser et al.
174 2012) against domestic dog TLR5 (Genbank accession NW_0119176 and Ensembl accession
175 ENSCAFT00000018059). AmpliTaq Gold Taq and buffer (Applied Biosystems) were used for
176 all polymerase chain reactions (PCR) but cycling conditions varied between fragments (Supp Fig.
177 2). All reactions were run on a Biorad DNA engine Peltier thermal cycler tetrad (Bio-Rad). To

178 inspect products for specific binding, and for the quality and quantity of amplified DNA, PCR
179 products were run on a 1.5% agarose gel using GelRed dye (Biotium), a BioRad PowerPac Basic
180 gel box and Tris-Acetate (TAE) buffer. Gels were visualized using a MultiDoc-it Digital
181 Imaging System (UVP).

182 Effective purification methods varied based on fragment and species due to the variation
183 in size of nonspecific bands appearing in PCR products. For maned wolf samples nonspecific
184 bands were typically less than 100 bp, and fragments were purified using 2 µl of EXOsapIT
185 (Affymetrix) per 7 µl of PCR product and incubated at 37 ºC for 25 min followed by 80 ºC for
186 15 min. In red wolves, where contaminating products were typically larger than 100 bp, products
187 were purified with solid phase reversible immobilization using carboxyl coated magnetic beads
188 (SPRI beads). Samples were incubated for 10 min at room temp, then 5 min on a magnetic
189 separation plate, subsequently washed with 100% ethanol and eluted with Qiagen Elution Buffer
190 (QIAGEN) and 20% Tween (EBT) (CSH protocol).

191 Purified products were sequenced using Big Dye Terminator v3.1 (Applied Biosystems).
192 Samples were heated to 96º C for 2 min, followed by 24 cycles of 96 ºC for 10 sec, 50 ºC for 10
193 sec and 60 ºC for 4 min. Sequenced fragments were cleaned using a Sephadex G50 (GE
194 Healthcare) column. After the application of water to dry Sephadex powder and the subsequent
195 solidification of the powder, sequencing products were applied to the column, and centrifuged at
196 2500 RPM for 5 min in an Allegra X-15R plate centrifuge (VWR). Ten microliters of Hi-Di
197 Formamide (Life Technologies) was added to each well of sample and the plate was sequenced
198 on an ABIPRISM3100 genetic analyzer (Life Technologies). All fragments were sequenced on
199 both the forward and reverse strands to confirm polymorphic positions.

Sequenced fragments were aligned using the software program Sequencher 5.3 (Gene Codes) and inspected manually for the presence of polymorphic positions. Subsequent contigs were aligned with available published domestic dog sequences for TLR5 (Genbank accession NW_0119176 and Ensembl accession ENSCAFT00000018059). SNP position was reported in reference to ENSCAFT00000018059. The number of SNPs in the two amplified regions of red wolf and maned wolf TLR5 were counted and compared to the number of SNPs in the same two regions in domestic dogs (Cusco et al. 2014). Heterozygous positions were identified in Sequencher and corroborated by manual inspection. For heterozygous loci, the gametic phase was determined using the software PHASE (Stephens et al. 2001). Mean heterozygosity was calculated and compared between *ex situ* and *in situ* maned wolf samples using a Mann-Whitney U test and between maned wolves, red wolves and previously published values for domestic dog and gray wolf (Cusco et al. 2014) using a one way ANOVA with a Bonferroni post hoc in SPSS. Nucleotide diversity (Θ) was calculated using a Tajima's test of neutrality in MEGA 5.22 (Tamura et al. 2011). To investigate patterns of selection rates of non-synonymous (dN) and synonymous (dS) substitutions were calculated using both the codon based HyPhy selection model and the Nei and Gojobori (1986) method, with a Jukes- Cantor correction using MEGA software version 5.2.2 (Tamura et al. 2011).

Translation of fragment sequences into amino acids was performed in Sequencher (Genecodes). Amino acid change ratio was calculated by dividing the length of the resulting translation for each fragment by the number of amino acid changes created by non-synonymous SNPs. The amino acid change ratio for each fragment was compared to published data for domestic dog TLR5 (Cusco et al. 2014). Protein domain predictions were made in SMART (Letunic et al. 2014) and used to identify domains encompassed by Frag1 and Frag2 as well as

the domain type for identified SNPs. PROVEAN (Choi & Chan 2015) was used to predict the functional impact of SNPs resulting in non-synonymous mutations by taking into consideration the amino acid sequence surrounding the residue of interest and classifying the mutation as either deleterious or neutral.

RESULTS

We detected two polymorphic positions in maned wolves, both in Frag1, and six polymorphic positions in red wolves with four in Frag1 and two in Frag2 (Table 1 & Fig. 1). In contrast, inspection of previously published data (Cusco et al. 2014) revealed that domestic dogs and gray wolves have more SNPs within these two TLR5 regions. Domestic dogs have seven SNPs in Frag1 and three SNPs within Frag2 and gray wolves have 5 SNPs in Frag1 and 6 SNPs in Frag2 (Table 2). No polymorphic positions were shared between maned wolves and red wolves. Neither of the SNPs identified in maned wolves were found to be polymorphic in domestic dog and only one red wolf SNP was common to domestic dogs (A729G). A729G was also seen to be polymorphic within the published gray wolf SNP data set in addition to G2274A (Table 1).

In contrast with the finding of more SNPs in TLR5, dogs and gray wolves did not significantly differ from maned and red wolves in mean heterozygosity at these SNPs ($P \geq 0.1$) (Table 2). Tajima's D nucleotide diversity measures for maned and red wolves found a greater average variability in maned wolves ($\Theta = 0.002599$) than in red wolves ($\Theta = 0.0013765$) echoing the trend seen in heterozygosity with more variability in maned wolves than red wolves. Within maned wolves there was no significant difference in mean heterozygosity between *ex situ* and *in situ* samples ($P \geq 0.1$).

A Z-test of selection for each species by fragment revealed no evidence of non-neutral selection for Frag1 in maned wolves and Frag2 in red wolves ($P \geq 0.05$). The ratio of dN/dS could not be calculated for these fragments because of the lack of synonymous mutations in each. In red wolf Frag1, although the codon based Z test of selection showed only neutral selection ($P \geq 0.05$), HyPhy calculated dN/dS at 0.146, indicating a slight evidence of purifying selection (Table 3). Tests of selection were not performed for maned wolf Frag2 due to the lack of synonymous or non-synonymous mutations.

Tests of selection between species found evidence for both purifying and positive selection. The dN/dS ratio for Frag1 between maned and red wolves was 0.3362 indicating purifying selection. The Nei Gojobori method with a Jukes Cantor correction for purifying selection also found purifying selection between these species in Frag1 ($P \leq 0.05$) with the probability at 0.05. For maned wolf and red wolf Frag2, strong evidence was found for positive selection ($P \leq 0.05$) with a codon based Z test of selection using a Nei Gojobori model with Jukes Cantor correction yielding an overall probability of 0.03 (Table 3).

We used an amino acid change ratio to compare the effect of these described polymorphic sites on resulting proteins. Domestic dogs and gray wolves had a higher amino acid change ratio than either maned or red wolves, which may be due to a population bottleneck in red wolves or to a higher level of evolutionary conservation of the TLR5 locus in both threatened species (Table 4).

We predicted protein domain structure for the selected fragments. Frag1 for both maned wolves and red wolves consisted of three unknown domains and two low complexity regions. Frag2 consisted of three leucine rich repeat (LRR) regions, one leucine rich repeat C-terminal (LRR-CT) region and one unknown region. All SNPs in Frag1 in both species were in areas with

268 unknown SMART predictions while both SNPs in red wolf Frag2 were in the LRR region (Table
269 1).

270 Both maned wolf SNPs were non-synonymous compared with two of six red wolf SNPs,
271 five of eleven gray wolf polymorphisms and four of ten domestic dog polymorphisms. All non-
272 synonymous maned wolf and red wolf SNPs were in Frag1 while domestic dog and gray wolf
273 non-synonymous SNPs were more evenly distributed between the two fragments.

274 The functional impact of these non-synonymous SNPs was tested using PROVEAN and
275 all identified red wolf and maned wolf non-synonymous SNPs were shown to have a neutral
276 effect on protein function (Table 1). Comparatively, three non-synonymous domestic dog SNPs
277 and four gray wolf non-synonymous SNPs present within Frag1 and Frag2 were reported to have
278 a probably or possibly damaging impact on protein function (Cusco et al. 2014). One of these
279 SNPs is T1844C, a SNP previously associated with domestic dog IBD (Kathrani et al. 2010),
280 which was shown to be deleterious (Cusco et al. 2014). All identified domestic dog and gray
281 wolf SNPs (Cusco et al. 2014) with a potential functional impact are not present as polymorphic
282 positions in maned or red wolves.

283 The SNPs identified as associated with domestic dog IBD (G727A, C805T and C2549T)
284 were not polymorphic in maned or red wolves. However, both red wolves and maned wolves
285 lacked the protective T allele in C805T and C2549T and the risk allele A in G727A (Fig. 2).
286 Gray wolves also lack these protective alleles (Cusco et al. 2014) indicating that the non-
287 protective C is potentially ancestral. Provean predictions show that the deleterious impact of the
288 leucine to serine amino acid change in the C2549T SNP is retained in maned wolves and red
289 wolves (Table 5).

290 **DISCUSSION**

291 Toll-like receptors are increasingly becoming a target of research due to their crucial
292 roles as sentinels of the innate immune system and their associations with many common and
293 debilitating diseases in both humans and in animal models (Netea et al. 2012). Our
294 characterization of the TLR5 locus is the first description of polymorphism in this locus in two
295 threatened canid species in need of careful captive management. A greater understanding of the
296 genetic diversity of these immune genes should contribute to maintaining healthy *ex situ*
297 populations.
298 The larger number of SNPs in domestic dog and gray wolf and the lack of significant
299 difference between mean heterozygosity in all four species implies that these regions may be
300 more variable in gray wolf and domestic dog but that heterozygosity has been maintained over
301 evolutionary time. This would suggest a role for balancing selection in this system, which has
302 been implicated in the evolution of innate immunity in humans (Ferrer-Admetlla et al. 2008).
303 The higher amino acid change ratio in domestic dogs and gray wolves indicates that the observed
304 genetic variation results in changes in amino acid composition within the two TLR5 regions.
305 Future studies should screen for variation across a larger number of canid species and individuals
306 to confirm levels of variation across this family.
307 Consistent with reports that identify the leucine-rich repeat region of TLR5 as a site
308 under adaptive selection due to its direct interaction with evolving pathogens (Areal et al. 2011),
309 signatures of adaptive selection were detected within the LRR here between maned and red
310 wolves, indicating that this ligand binding pocket is potentially adapting to compete with
311 evolving microbes, as in other mammalian species. However this hypothesis needs to be further
312 tested by conducting comparative microbiome analyses in maned and red wolves. The higher
313 number of SNPs in both gray wolf and domestic dog and their higher propensity to be non-

synonymous and damaging, is suggestive of deleterious allele accumulation in this region of TLR5. Deleterious allele accumulation is seen in domestic dogs and gray wolves and could be a result of a previously documented bottleneck in domestic dogs and in the European population of wolves referenced in this study (Cusco et al. 2014, Cruz et al. 2008, Pilot et al. 2014). Some of the non-damaging mutations found in maned and red wolves could be products of adaptation to different microbial and dietary environments that require a species-specific function of TLR5 (Bergman et al. 2010). The complete lack of overlap in variable sites between all four canid species and the conservation of just one polymorphic position between red wolves, gray wolves and domestic dogs further suggests a potential species-specific function for TLR5 in maned wolves and red wolves, as is seen in other species (Werling et al. 2009).

The finding of less variation in maned and red wolf TLR5 fragments than in domestic dog and gray wolf is supported by: (1) the low SNP number in maned wolves and red wolves compared to domestic dogs and gray wolves, (2) the neutral functional impact of observed non-synonymous SNPs, (3) the low amino acid change ratio and the low number of SNPs in both Frag1 and the putatively variable LRR region in Frag2. This potential within-species conservation supports previous studies that identify TLRs as a conserved class of proteins (Roach et al. 2005). The finding of distinct variation between species, especially in the ligand binding site, points to a potential specificity of function of TLR5 in maned wolves and red wolves, likely influenced by differences in their microbial environments (Takeda et al. 2003). The implications of these findings for studies of the adaptive nature of TLR5 are limited by the lack of a robust demographic analysis that would incorporate the distinct recent evolutionary histories of these four species. That said, we chose samples that capture the full range of

336 diversity in the current red wolf and maned wolf populations, especially through the inclusion of
337 *in situ* maned wolf samples.

338 Within these two now characterized fragments lie the three SNPs previously associated
339 with domestic dog IBD (G727A, C805T and C2549; Kathrani et al. 2010). The non-polymorphic
340 nature of these SNPs in maned wolves and red wolves makes them unsuitable as diagnostic
341 markers for inflammatory bowel disease. However, all sampled *ex situ* and *in situ* maned wolves
342 and red wolves lacked the protective thymine present in the alleles validated across all dog
343 breeds (C805T and C2549T; Kathrani et al. 2011), suggesting that both taxa may carry a genetic
344 predisposition to IBD. The high prevalence of IBD in captive populations of both species, and
345 the retention of the deleterious effect of the C2549T SNP, are further evidence for this
346 predisposition. A large population of gray wolves also lacks the protective allele T in both
347 C805T and C2549T, indicating that the non-protective allele C is ancestral and that the T allele
348 emerged in domestic dogs (Cusco et al. 2014). The lack of the domestic dog protective alleles in
349 these threatened canids is not enough to prove a genetic predisposition to IBD. Future work is
350 needed to correlate standardized molecular and pathological markers of disease status with these
351 genetic findings.

352 Recent work on dog domestication has identified a host of genes containing a signature of
353 domestication, typically in mutations that allow dogs to better adapt to association with humans
354 (Wang et al. 2013). Adaptation to a starch-based diet plays an essential role in domestication and
355 provides a relevant link to IBD (Axelsson et al. 2013). Since TLR5 recognizes bacterial flagellin,
356 a shift in gut microbiome composition as a result of diet change (Middelbos et al. 2010) can
357 result in an inappropriate hypo- or hyper- activation of the TLR pathway and lead to

358 inflammation (Cario 2010). It is possible that the T allele in domestic dog C805T and C2549T
359 developed as a protection against this type of inflammation.
360 *Ex situ* maned and red wolves in the United States are primarily fed artifical starch based
361 diets (Songsasen 2014, Harrison 2014) in contrast to their *in situ* omnivorous and carnivorous
362 diets respectively (Aragona & Setz 2001, Paradiso & Nowak 1972). Differences in diet between
363 captive and wild individuals can result in changes to the intestinal microbial community
364 (Turnbaugh et al. 2009). Feeding *ex situ* wild canids diets developed for domestic dogs may
365 promote the development of microbial communities more similar to those of domestic animals
366 than to those of their *in situ* conspecifics (De Jesús-Laboy et al. 2011). Since Toll-like receptor 5
367 specifically recognizes bacterial flagellin, a dog-like microbiome interacting with a maned or red
368 wolf Toll-like receptor evolutionarily adapted to the native microbiome of these species could
369 result in a hypo- or hyper-activation of this portion of the innate immune system. While dogs
370 have had about 10,000 years (Wang et al. 2013, Freedman et al. 2014) since the domestication of
371 grains to adapt to a starch based diet, maned and red wolves have been managed by species
372 survival plans in captivity for only 30 and 35 years respectively. Adaptation to starch based diets
373 should not be a goal of captive breeding programs and should be avoided at all costs for any *ex*
374 *situ* programs that may eventually result in reintroduction. With further investigation, this
375 relationship between putative genetic predisposition, and inappropriate diet in combination with
376 the resulting foreign microbial community could explain the high prevalence of IBD in *ex situ*
377 maned and red wolves.

378 Future studies should focus on documenting and correlating the gastrointestinal
379 microbiome compositions of *ex situ* and *in situ* maned wolves and red wolves with clinical,
380 histopathological and serum markers of IBD. Further characterizations of the full sequence of

TLR5 for these sampled populations of maned and red wolves, in addition to other canid species, could inform the evolutionary nature of toll like receptors within the Canis genus. Additionally, with more research on the accurate diagnosis of maned and red wolf IBD, future work can focus on correlating the TLR5 SNPs identified in domestic dogs with a definitive IBD diagnosis in maned and red wolves.

ACKNOWLEDGMENTS

We would like to thank the Center for Conservation and Evolutionary Genetics for their incredible support throughout this project, specifically Nancy McInerney for her invaluable guidance, Lilly Parker for her advice, Nandanevi Cortes for training in PHASE, and all graduate students, postdocs, contractors and fellows that participate in writing group for their assistance with editing. Priscilla Joyner was instrumental in assisting with sample collection and advice regarding IBD pathogenesis. Additionally we would like to thank Olga Francino for providing information regarding SNP allele frequencies for the gray wolf population and Aarti Kathrani for answering questions regarding amplification protocols. We would also like to acknowledge the red wolf and maned wolf species survival plan groups for their assistance with samples and openness to implement suggested management changes. Louis Emmons' maned wolf research in Bolivia was in collaboration with the Museo de Historia Natural Noel Kempff Mercado, Universidad Gabriel René Moreno, Santa Cruz, Bolivia, under permits from Dirección General de Biodiversidad and the Servicio Nacional de Áreas Protegidas of the Estado Plurinacional de Bolivia, and it was supported by the Smithsonian Institution, the National Geographic Society, and the Wildlife Conservation Society. Susana González' research was supported by PEDECIBA, CSIC-UdelaR from Uruguay.

LITERATURE CITED

Acton AE, Munson L, Waddell WT (2000) Survey of necropsy results in captive red wolves (Canis rufus), 1992-1996. J Zoo Wildl Med 31:2–8

Akira S (2003) Toll-like receptor signaling. J Biol Chem 278:38105–38108

Akira S, Takeda K, Kaisho T (2001) Toll-like receptors: critical proteins linking innate and acquired immunity. Nat Immunol 2:675–680

Aragona M, Setz EZF (2001) Diet of the maned wolf, Chrysocyon brachyurus (Mammalia: Canidae), during wet and dry seasons at Ibitipoca State Park, Brazil. J Zool 254:131–136

Areal H, Abrantes J, Esteves PJ (2011) Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. BMC Evol Biol 11:368

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar Å, Lindblad-Toh K (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495:360–364

Bell JK, Mullen GED, Leifer CA, Mazzoni A, Davies DR, Segal DM (2003) Leucine-rich repeats and pathogen recognition in Toll-like receptors. Trends Immunol 24:528–533

Bergman IM, Rosengren JK, Edman K, Edfors I (2010) European wild boars and domestic pigs display different polymorphic patterns in the Toll-like receptor (TLR) 1, TLR2, and TLR6 genes. Immunogenetics 62:49–58

Cario E (2010) Toll-like receptors in inflammatory bowel diseases: A decade later. Inflamm Bowel Dis 16:1583–1597

Cario E, Podolsky DK (2000) Differential alteration in intestinal epithelial cell expression of Toll-like receptor 3 (TLR3) and TLR4 in Inflammatory Bowel Disease. Infect Immun 68:7010–7017

Choi Y, Chan AP (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics

Craven M, Simpson JW, Ridyard AE, Chandler ML (2004) Canine inflammatory bowel disease: retrospective analysis of diagnosis and outcome in 80 cases (1995–2002). J Small Anim Pract 45:336–342

Cruz F, Vilà C, Webster MT (2008) The legacy of domestication: accumulation of deleterious mutations in the dog genome. Mol Biol Evol 25:2331–2336

Cusco A, Sanchez A, Altet L, Ferrer L, Francino O (2014) Non-synonymous genetic variation in exonic regions of canine Toll-like receptors. Canine Genet Epidemiol 1:1–12

439 Deem SL, Emmons LH (2005) Exposure of free-ranging Maned Wolves (Chrysocyon
440 brachyurus) to infectious and parasitic disease agents in the Noel Kempff Mercado
441 National Park, Bolivia. J Zoo Wildl Med 36:192–197

442 De Jesús-Laboy KM, Godoy-Vitorino F, Piceno YM, Tom LM, Pantoja-Feliciano IG, Rivera-
443 Rivera MJ, Andersen GL, Domínguez-Bello MG (2011) Comparison of the fecal
444 microbiota in feral and domestic goats. Genes 3:1–18

445 Ferrer-Admetlla A, Bosch E, Sikora M, Marquès-Bonet T, Ramírez-Soriano A, Muntasell A,
446 Navarro A, Lazarus R, Calafell F, Bertranpetit J, Casals F (2008) Balancing selection is
447 the main force shaping the evolution of innate immunity genes. J Immunol 181:1315–
448 1322

449 Fredrickson RJ, Hedrick PW (2006) Dynamics of hybridization and introgression in red wolves
450 and coyotes. Conserv Biol 20:1272–1283

451 Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, Silva PM, Galaverni M,
452 Fan Z, Marx P, Lorente-Galdos B, Beale H, Ramirez O, Hormozdiari F, Alkan C, Vilà C,
453 Squire K, Geffen E, Kusak J, Boyko AR, Parker HG, Lee C, Tadigotla V, Siepel A,
454 Bustamante CD, Harkins TT, Nelson SF, Ostrander EA, Marques-Bonet T, Wayne RK,
455 Novembre J (2014) Genome sequencing highlights the dynamic early history of dogs.
456 PLoS Genet 10:e1004016

457 Fukata M, Abreu MT (2008) Role of Toll-like receptors in gastrointestinal malignancies.
458 Oncogene 27:234–243

459 German A, Hall E, Day M (2003) Chronic intestinal inflammation and intestinal disease in dogs.
460 J Vet Intern Med 17:8–20

461 Gese EM, Knowlton FF, Adams JR, Beck K, Fuller TK, Murray DL, Steury TD, Stoskopf MK,
462 Waddell WT, Waits LP (2015) Managing hybridization of a recovering endangered
463 species: The red wolf Canis rufus as a case study. Curr Zool 61:191–205

464 Gilioli R, Silva FA (2000) Frequency of parasites and Salmonella infection in captive maned-
465 wolf, Chrysocyon brachyurus, kept in Zoos at the State of São Paulo, Brazil. Arq Bras
466 Med Vet e Zootec 52:337–341

467 Ginsberg JR (1994) Captive breeding, reintroduction and the conservation of canids. In: Olney
468 PJS, Mace GM, Feistner ATC (eds) Creat Conserv. Springer Netherlands, p 365–383

469 Harrison R (2014) Red Wolf Species Survival Plan Meeting.

470 Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, Goodlett DR, Eng JK, Akira S, Underhill
471 DM, Aderem A (2001) The innate immune response to bacterial flagellin is mediated by
472 Toll-like receptor 5. Nature 410:1099–1103

473 Himmel ME, Hardenberg G, Piccirillo CA, Steiner TS, Levings MK (2008) The role of T-
474 regulatory cells and Toll-like receptors in the pathogenesis of human inflammatory bowel
475 disease. Immunology 125:145–153

476 Inness VL, McCartney AL, Khoo C, Gross KL, Gibson GR (2007) Molecular characterisation of
477 the gut microflora of healthy and inflammatory bowel disease cats using fluorescence in
478 situ hybridisation with special reference to Desulfovibrio spp. J Anim Physiol Anim Nutr
479 91:48–53

480 International Union for the Conservation of Nature (2016) The IUCN Red List of Threatened
481 Species.

482 Johnson AEM, Freeman EW, Wildt DE, Songsasen N (2014) Spermatozoa from the maned wolf
483 (Chrysocyon brachyurus) display typical canid hyper-sensitivity to osmotic and freezing-
484 induced injury, but respond favorably to dimethyl sulfoxide. Cryobiology 68:361–370

485 Kathrani A, House A, Catchpole B, Murphy A, German A, Werling D, Allenspach K (2010)
486 Polymorphisms in the Tlr4 and Tlr5 Gene Are Significantly Associated with
487 Inflammatory Bowel Disease in German Shepherd Dogs. PLoS ONE 5:e15740

488 Kathrani A, House A, Catchpole B, Murphy A, Werling D, Allenspach K (2011) Breed-
489 independent toll-like receptor 5 polymorphisms show association with canine
490 inflammatory bowel disease. Tissue Antigens 78:94–101

491 Kawai T, Akira S (2010) The role of pattern-recognition receptors in innate immunity: update on
492 Toll-like receptors. Nat Immunol 11:373–384

493 Letunic I, Doerks T, Bork P (2014) SMART: recent updates, new developments and status in
494 2015.

495 Leulier F, Lemaitre B (2008) Toll-like receptors — taking an evolutionary approach. Nat Rev
496 Genet 9:165–178

497 Maia OB, Gouveia AMG (2002) Birth and mortality of maned wolves Chrysocyon brachyurus
498 (Illiger, 1811) in captivity. Brazil J Biol 62:25–32

499 Maloy KJ, Powrie F (2011) Intestinal homeostasis and its breakdown in inflammatory bowel
500 disease. Nature 474:298–306

501 Middelbos IS, Vester Boler BM, Qu A, White BA, Swanson KS, Fahey GC Jr (2010)
502 Phylogenetic characterization of fecal microbial communities of dogs fed diets with or
503 without supplemental dietary fiber using 454 pyrosequencing. PLoS ONE 5:e9768

504 Netea MG, Wijmenga C, O'Neill LAJ (2012) Genetic variation in Toll-like receptors and disease
505 susceptibility. Nat Immunol 13:535–542

506 Paradiso JL, Nowak RM (1972) Canis rufus. Mammalian Species:1

507 Paula RC, Medici P, Morato RG (2008) Maned wolf action plan—population and habitat
508 viability assessment.

509 Phillips MK, Scheck J (1991) Parasitism in Captive and Reintroduced Red Wolves. J of Wildl
510 Dis 27:498–501

511 Pilot M, Greco C, vonHoldt BM, Jędrzejewska B, Randi E, Jędrzejewski W, Sidorovich VE,
512 Ostrander EA, Wayne RK (2014) Genome-wide signatures of population bottlenecks and
513 diversifying selection in European wolves. Heredity (Edinb) 112:428–442

514 Rabon DRJ (2011) Factors affecting reproduction in the red wolf (Canis rufus). ProQuest, UMI
515 Dissertation Publishing

516 Ratter JA, Ribeiro JF, Bridgewater S (1997) The Brazilian Cerrado vegetation and threats to its
517 biodiversity. Ann Bot 80:223–230

518 Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A (2005)
519 The evolution of vertebrate Toll-like receptors. PNAS 102:9577–9582

520 Rodden MD, Sorenson LG, Sherr A, Kleiman DG (1996) Use of behavioral measures to assess
521 reproductive status in maned wolves (Chrysocyon brachyurus). Zoo Biol 15:565–585

522 Seeley KE, Garner MM, Waddell WT, Wolf KN (in press) A survey of diseases in captive red
523 wolves (Canis rufus), 1997-2012. J Zoo Wildl Med

524 Simpson KW, Jergens AE (2011) Pitfalls and Progress in the Diagnosis and Management of
525 Canine Inflammatory Bowel Disease. Vet Clin N Am: Small 41:381–398

526 Smith SA, Jann OC, Haig D, Russell GC, Werling D, Glass EJ, Emes RD (2012) Adaptive
527 evolution of Toll-like receptor 5 in domesticated mammals. BMC Evol Biol 12:122

528 Songsasen N (2014) Maned Wolf Species Survival Plan.

529 Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction
530 from population data. Am J Hum Genet 68:978–989

531 Stirling J, Griffith M, Blair I, Cormican M, Dooley JSG, Goldsmith CE, Glover SG, Loughrey A,
532 Lowery CJ, Matsuda M, McClurg R, McCorry K, McDowell D, McMahon A, Cherie
533 Millar B, Nagano Y, Rao JR, Rooney PJ, Smyth M, Snelling WJ, Xu J, Moore JE (2008)
534 Prevalence of gastrointestinal bacterial pathogens in a population of zoo animals.
535 Zoonoses Public Hlth 55:166–172

536 Takeda K, Kaisho T, Akira S (2003) Toll-Like Receptors. Annu Rev Immunol 21:335–376

537 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular
538 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
539 maximum parsimony methods. Mol Biol Evol 28:2731–2739

Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI (2009) The Effect of Diet on the Human Gut Microbiome: A metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med 1(6):14–16

Untergasser A, Cutcutache I, Koressar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer 3- new capabilites and interfaces. Nucleic Acids Res 40(15):e115

Vijay-Kumar M, Sanders CJ, Taylor RT, Kumar A, Aitken JD, Sitaraman SV, Neish AS, Uematsu S, Akira S, Williams IR, Gewirtz AT (2007) Deletion of TLR5 results in spontaneous colitis in mice. J Clin Invest 117:3909–3921

Wang G, Zhai W, Yang H, Fan R, Cao X, Zhong L, Wang L, Liu F, Wu H, Cheng L, Poyarkov AD, Poyarkov Jr NA, Tang S, Zhao W, Gao Y, Lv X, Irwin DM, Savolainen P, Wu C-I, Zhang Y (2013) The genomics of selection in dogs and the parallel evolution between dogs and humans. Nat Commun 4:1860

Werling D, Jann OC, Offord V, Glass EJ, Coffey TJ (2009) Variation matters: TLR structure and species-specific pathogen recognition. Trends Immunol 30:124–130

Wlasiuk G, Khan S, Switzer WM, Nachman MW (2009) A history of recurrent positive selection at the Toll-like receptor 5 in primates. Mol Biol Evol 26:937–949

Xenoulis PG, Palculict B, Allenspach K, Steiner JM, Van House AM, Suchodolski JS (2008) Molecular-phylogenetic characterization of microbial communities imbalances in the small intestine of dogs with inflammatory bowel disease. FEMS Microbiol Ecol 66:579–589

Figures and Tables

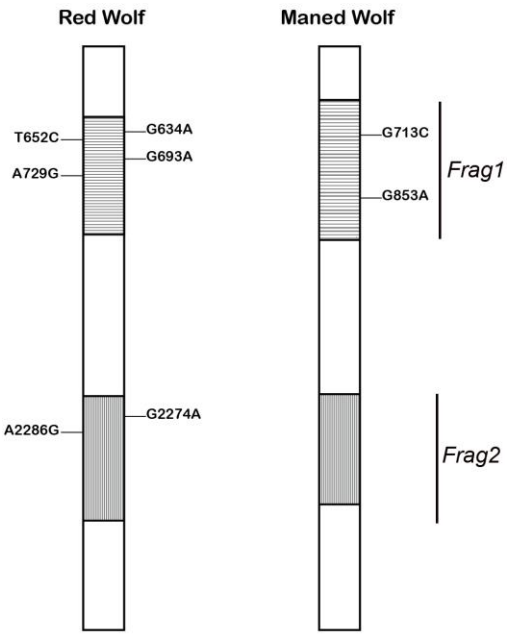

Figure 1: Single nucleotide polymorphisms identified in TLR5 in 29 maned wolves and 15 red wolves within the indicated Fragment 1 and 2 regions.

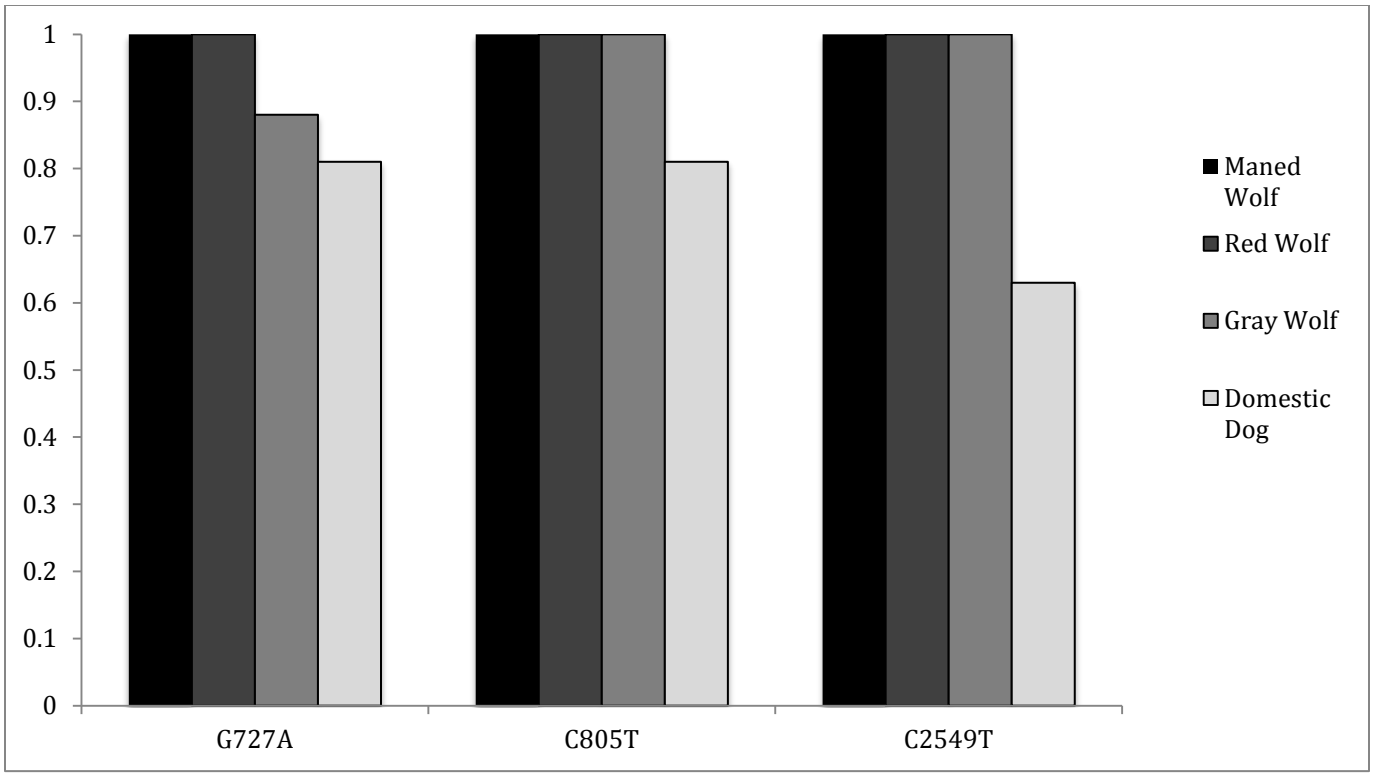

Figure 2: Observed allele frequency for IBD associated SNPs (A in G727A, C in C805T and C2549T) in maned wolves and red wolves. Allele frequencies for gray wolf and domestic dog IBD SNPs provided by Cusco et al. 2014.

Table 1: Polymorphic sites in TLR5 in 29 maned wolves and 15 red wolves

Position	SNP ID	Codon		Aa Subst[a]	Protein Domain[b]	Provean Output[c]	Allele Freq			
		Allele 1	Allele 2				Maned Wolf	Red Wolf	Gray Wolf[d]	Domestic Dog[d]
634	G634A	**G**TC	**A**TC	val/ile*	ncp[1]	neutral	G(1)	G(.86)		G(1)
652	T652C	**T**GG	**C**GG	trp/arg*	ncp	neutral	T(0)	T(.73)		T(1)
693	G693A	CC**G**	CC**A**	pro/pro	ncp		G(1)	G(.91)		G(1)
713	G713C	C**G**C	C**C**C	arg/pro*	ncp	neutral	G(.75)	G(1)		G(1)
729	A729G	GC**A**	GC**G**	ala/ala	ncp		A(0)	A(.82)	A(.4)	A(.39)
853	G853A	**G**TC	**A**TC	ile/val*	ncp	neutral	G(.54)	G(0)		G(0)
2274	G2274A	CG**G**	CG**A**	arg/arg	LRR[2]		G(1)	G(.92)	G(.76)	G(1)
2286	A2286G	GC**A**	GC**G**	ala/ala	LRR		A(0)	A(.92)		A(0)

SNP position is in reference to ENSCAFT00000018059 and SNP ID includes the most frequent allele first, followed by the position and the least frequent allele.

[a] Amino acid substitution * non-synonymous aa change

[b] Protein domain predicted by SMART [1] ncp, no confident prediction [2] LRR, leucine rich repeat region

[c] Provean function prediction of non-synonymous SNPs only

[d] Gray wolf and red wolf allele frequencies provided by Cusco et al. 2014

Table 2: Comparison of SNP number and SNP heterozygosity in maned wolves, red wolves, gray wolves and domestic dogs

Species[a]	Number of SNPs			Mean Heterozygosity[d]
	Syn[b]	Non-Syn[c]	Σ	
Maned Wolf	0	2	2	0.44 ± 0.09*
Red Wolf	4	2	6	0.23 ± 0.10*
Gray Wolf	6	5	11	0.27 ± 0.17*
Domestic Dog	6	4	10	0.26 ± 0.15*

Differences in mean heterozygosity between species were tested for significance by a one way ANOVA with a Bonferroni post hoc
* $P \geq 0.10$
[a] Mean heterozygosity for gray wolf and domestic dog extrapolated from data published in Cusco et al. 2014
[b] Synonymous
[c] Non-synonymous
[d] Mean ± SD

Table 3: Tests for Neutral, Purifying and Positive Selection and dN/dS for maned wolves (MW) and red wolves (RW)

Species/Frag[a]	dN/dS[b]	Selection Test[c]					
		Neutral Selection		Purifying Selection		Positive Selection	
		Probability[d]	Test Statistic[e]	Probability[f]	Test Statistic	Probability[g]	Test Statistic
MW[1] Frag1		0.17	1.39	1.00	-1.35	0.08	1.43
RW[2] Frag1	0.146	0.27	1.11	0.13	1.15	1.00	-1.14
RW Frag2		0.14	1.48	1.00	-1.40	0.07	1.45
MW and RW Frag1	0.3362	0.09	1.73	**0.05**	-1.67	1.00	-1.69
MW and RW Frag2		0.07	-1.81	1.00	-1.76	**0.03**	1.88

MW Frag2 not included due to lack of variable sites, significant values in bold
[a]Species abbreviations [1] Maned Wolf [2] Red Wolf
[b] dN/dS for fragments with both non-synonymous and synonymous changes only
[c] Codon-based Z test of Selection, Nei-Gojobori method with Jukes-Cantor correction
[d] Probability of rejecting the null hypothesis of dN=dS
[e] Test statistic= dN-dS
[f] Probability of rejecting the null hypothesis of dN=dS for dN<dS
[g] Probability of rejecting the null hypothesis of dN=dS for dN>dS

Table 4: Comparison of amino acid change ratio in maned wolf, red wolf, gray wolf and domestic dog TLR5

Species[a]	Protein length (aa)	AA change ratio[b]
Maned Wolf	288	1/144
Red Wolf	279	1/139.5
Gray Wolf	288	1/57.6
Domestic Dog	288	1/72

[a] Amino acid change ratio for gray wolf and domestic dog extrapolated from data published in Cusco et al. 2014
[b] Amino acid ratio: amino acid changes caused by nsSNPs divided by the protein length

Table 5: Polymorphic sites associated with Inflammatory Bowel Disease in maned wolves, red wolves, gray wolves and domestic dogs

Position	SNP ID	Codon		AA Subst[a]	Protein Domain[b]	Provean Output[c]	Allele Freq			
		Allele 1	Allele 2				Maned Wolf	Red Wolf	Gray Wolf[d]	Domestic Dog[d]
727	G727A	**G**CG	**A**CG	ala/thr*	ncp[1]	neutral	G(1)	G(1)	G(.88)	G(.81)
805	C805T	**C**GC	**T**GC	arg/cys*	ncp	neutral	C(1)	C(1)	C(.99)	C(.81)
2549	C2549T	T**C**G	T**T**G	ser/leu*	LRR CT[2]	deleterious	C(1)	C(1)	C(.98)	C(.63)

SNP position is in reference to ENSCAFT00000018059 and SNP ID includes the most frequent allele first, followed by the position and the least frequent allele.

[a] Amino acid substitution * non-synonymous aa change

[b] Protein domain predicted by SMART [1] ncp, no confident prediction [2] LRR CT, leucine rich repeat C-terminal region

[c] Provean function prediction of non-synonymous SNPs only

[d] Gray wolf and red wolf allele frequencies provided by Cusco et al. 2014

Supplement

Supp. Fig. 1

Primer	Manufacturer	Direction	Primer sequence
TLR5 Fragment 1	Eurofins MWG (Huntsville, USA)	Forward	5'-GTT TCT CAA GGA CCC AGC AC-3'
		Reverse	5'-TCC TGA AGG CTT CTC TGT CG-3'
TLR5 Fragment 2	Eurogins MWG (Huntsville, USA)	Forward	5'-GCT GCA CCT GAA CCA CAA C-3'
		Reverse	5'-TGA AGA GGG AGA ACG TGA GG-3'

Supp. Fig. 2

Frag1 cycling conditions		
Cycle Number	Settings:	
1	95°C	10 minutes
35	95°C	1 minute
	57°C	1 minute
	72°C	2 minutes
1	72°C	7 minutes
Frag2 cycling conditions		
Cycle Number	Settings:	
1	95°C	8 minutes
2	95°C	30 seconds
	64°C	30 seconds
	72°C	1 minute
2	95°C	30 seconds
	62°C	30 seconds
	72°C	1 minute
2	95°C	30 seconds
	60°C	30 seconds
	72°C	1 minute
2	95°C	30 seconds
	58°C	30 seconds
	72°C	1 minute
2	95°C	30 seconds
	56°C	30 seconds
	72°C	1 minute
1	72°C	15 minutes

Supp. Fig.3

```
DD   ATGGGAGGAGAGCGTGCGTGCAGAGGAAGCAGCACGTGCCAAGTCCCAGC 50
MW
RW

DD   CCTGGAAGTGCGTGGAGAGGCGGCCGGAGCCCCCGTGTTGCACCCCCCGC 100
MW
RW

DD   CTCCCCGCTGGGCCTCCTTCCTTTGCATCCCTGGGGCCCCTGGATGTCAT 150
MW
RW

DD   CGAGTCATCGGGGGCCTTCCTGGCCACTCTGGCCGCTGCCGCCTGCCCCA 200
MW
RW

DD   GGACGTGCGCCCCAGCTGGCCCGTGTGTCCAGCCCCTGCTGCACCCTGTG 250
MW
RW

DD   CACAGAGCGGACGTGTGGCACTTGTCCAGATGACGGGCGCCCTGAGCCGC 300
MW
RW

DD   GCCGCGCTCGCCCCCACAGGCCTGGGCAGGGGGTGGGAGGGGGATGCACT 350
MW
RW

DD   GACCCGTCAGGCTGGGCGCTTCGCGGATGGTGGCCCGAAGGACCAGCGTG 400
MW
RW

DD   TGCGTGCTGACCCGGGCCGTGTGTGGAGCGCGCAGGGGGCGGAGGGGCGG 450
MW
RW

DD   GCCCCCGGCACTGGGCGGGGGCGGCACCAGTATCCTCCGCCTGCCATTTT 500
MW
RW

DD   CCCCGAAGCCCTGCACGCATCCCGATTGAGTGACGGCAAACAGACTCTCC 550
```

```
MW
RW

DD   TCAAGGTAAGTGTTTCTCAAGGACCCAGCACGGCGCTGAGTGCGCGTCCC 600
MW                                  -----------------
RW

DD   GCCGGGCGCACGTGTGGGGGGAGGCAGGTGCCCGTCCAGGGGCCCCCGTC 650
MW   ---A---C-----------------------------------------A--
RW             ----------------------R----------------


DD   CTGGGCCCCGGGGTGGCGTTGGGCTTGCACGGCTGTGTTTCCGTCCCGCA 700
MW   -C-------------------------------------------------
RW   -Y-----------------------------------------R-------

DD   GGATCATGGGCCGCCAGCTGGGCCGCACGCTGGGGCTGCTGCTTGTGGCC 750
MW   ------------S----------T--G----------------------
RW   --------------------------G-R--------------------

DD   GGCGCCGTGGCCGCAGCATCCTGCTGCGTGGCTGACGGCCGGAGGGCCCT 800
MW   --------------G-----------------C--T--------------
RW   --------------G-----------------------------------

DD   GTACCGCTCCTGCAACCTCAGCCAGGTGCCCCCGGTCCCCAGCACCACCG 850
MW   ---------------------------------------------T-------
RW   -------------------------------------------------

DD   AGATCCTCCTGCTGAGCTTCAACTACATCCGGGCCGTCACCCGCGCCTCG 900
MW   --R----------------------------------------------
RW   -------------------------------------------------

DD   TTCCCCCTCCTGGAGCGGCTGCAGCTGCTGGAGCTGGGGACGCAGCAGAC 950
MW   --------------------
RW   ------------------------------------------

DD   GCCCTTCAGCGTCGACAGAGAAGCCTTCAGGAACCTGCCCAACCTGCGCA 1000
MW
RW

DD   CCCTGGACCTGGGCAACAGCCGGGTGGATTTCCTGCATCCCGACGCCTTC 1050
MW
RW
```

```
DD   CAGGGGCTGCCCCACCTGCAGGAACTCCGGCTGTTCGCCTGTGGCCTCTC 1100
MW
RW

DD   CGACGTCGTGTTGACAGACGGTTATTTCAGAAACCTGGGGGCTTTGTTGC 1150
MW
RW

DD   GCCTGGACCTGTCCAAAAATCAGATTGGGAGCCTCGAGCTTCACGCCTCC 1200
MW
RW

DD   TTCCGGGAGCTGGGTTCCCTGAGGTCCGTGGACTTTTCCCTCAACCGGAT 1250
MW
RW

DD   CCCGGCTGCGTGTGAGCAGGGGCTCAGGCCCCTGCAGGGCAAGGCGCTCT 1300
MW
RW

DD   CCCTTCTGAACCTCGCGGCCAATGGCCTGTACAGCCGGGCCCCCGTGGAC 1350
MW
RW

DD   TGGGGGCGGTGCGGGAACCCGTTCAGGAATGTGGTCCTGGAGACCCTGGA 1400
MW
RW

DD   CGTGTCTAACAACGGCTGGACCGCAGACGTCACGGGCAACGTCACCAGGG 1450
MW
RW

DD   CCATCGGTGGGAGCCAGATCTCCTCCTTGGTGCTCGCCCACCACATCATG 1500
MW
RW

DD   GGGCAGGGGTTTGGCTTCCGGAACATCCGGGACCCTGACCGGAGCACGTT 1550
MW
RW

DD   CGCGGGGCTGGCCGGGAGCTCGGTGCTGCGGCTGGATCTGTCGCACGGCT 1600
MW
RW
```

```
DD   TCGTCTTCTCCCTGAACGCCCGACTGTTCGAGGTGCTCGGGGACCTGAAG 1650
MW
RW

DD   CTCCTGGACCTCGCCCACAACAAGATCAACAGGATCGCGGGAGAAGCGTT 1700
MW
RW

DD   TCACGGCCTCGGCAGCGTCCAGGTTCTCAACCTGTCGCACAATCTCCTGG 1750
MW
RW

DD   GCGAGCTCTATGACTCTGACTTCTCGGGGCTCGCGGAGGTCGCCTACATT 1800
MW
RW

DD   GACCTGCAGCACAATCACATCGGGATCATCCAGGACCAGACGTTCAGATT 1850
MW
RW

DD   CCTGGGGGCGCTTCGGACCCTGGATCTCCGCGACAACGCCCTCAAAACCG 1900
MW
RW

DD   TTTCCTTCGTGCCCAGCATAGACACCATCTTCCTGGGCAACAACAAGCTG 1950
MW
RW

DD   GAGACCGTGTCCCACATGGACCTCACAGCCAGCTTCCTGGAGCTGTCGGA 2000
MW
RW

DD   CAACAGGCTGGAGGACCTGGGCGACCTCTACTCGCTCCTCCGGGTCCCTG 2050
MW
RW

DD   CCCTGCAGGTCCTCATCCTCAACCGCAACCGCCTGTCCGCGTGCCGTGGC 2100
MW
RW

DD   GGACACGGCCCCACGGGCAGCGTCGGCCCAGAGAGGCTCTTCCTCGGGAG 2150
MW
RW
```

```
DD   CAACATGCTGCAGCTGGCCTGGGAGACCGGGCGGTGCTGGGACGTGTTCC 2200
MW
RW

DD   GGGGGCTGCCCCGGCTCCGGGTGCTGCACCTGAACCACAACTACCTGGCC 2250
MW
RW

DD   GCCCTCCCGCCGGGGCTGCTGCGGGACCTCACGGCGCTGAGGGGCCTCGA 2300
MW                     ------------------------------
RW                     ---R-----------R--------------

DD   CCTGAGCGCCAACAGGCTGAGCACGCTGTCCCGGGGCGACCTGCCTGCTG 2350
MW   ------T---------------------------------------------
RW   ---------------------------------------------------

DD   CCTTGGAGGTGCTGGATGTGTCCAGGAACCAGCTCCTGTCCCTGGACCCC 2400
MW   ---------------------------------------------------
RW   ---------------------------------------------------

DD   GGGCTGCTCGCCCCGCTCAGAGCCGTGGACCTAACGCACAACAAGTTCAT 2450
MW   ---------------------------------------------------
RW   ---------------------------------------------------

DD   CTGCGGCTGCGAGCTCCGTCCCTTGGTGAGGTGGCTCAACCGGACCAACG 2500
MW   ---------------------------------------------------
RW   ---------------------------------------------------

DD   TCACTGTGTTCGGGTCCCGCGCAGACGTGCGCTGCGCCTACCCCAGCTTG 2550
MW   ----C---------------A---------------------------C-
RW   ----C-------------------------------------------C-

DD   CTTGCGGGGACGCCCCTGTCCTCTGTCTCCATGGAGGGCTGTGACGACGA 2600
MW   ----------------------
RW   ----------------------

DD   GGAGGCCCTGCGGACCCTCACGTTCTCCCTCTTCATCTTCTCCACCGTCG 2650
MW
RW

DD   GGGTCACGCTGTTCCTCCTGGCCGTCCTCGTGGCCGCCAAGCTCCGGGGC 2700
MW
RW
```

```
DD   CTTTGCTTCCTCTGTTACAAGGCGGCCCGGCGCCTCCTGCCTGCGGGGCC 2750
MW
RW

DD   CGCCGAGGACGGAGCGCCCGACGCGTACCAGTACGACGCCTACCTGTGCT 2800
MW
RW

DD   TCAGCGGCAGAGACTTCGAGTGGGTGCAGCGCGCGCTGCTCAGGCACCTG 2850
MW
RW

DD   GACGCTCAGTACAGCTCCCGAAACAGGCTGAACCTGTGCTTCGAGGAGAG 2900
MW
RW

DD   GGACTTCGTCCCGGGGCGGGAGCACATCGCCAACATCCAGGACGCCGTGT 2950
MW
RW

DD   GGAGCAGCCGCAAGGTGGTCTGTCTGGTGAGCAGGCACTTCCTCCGCGAC 3000
MW
RW

DD   GGGTGGTGCCTGGAGGCCTTCGCGGCCGCGCGGAGCCGCTGCGCGTCCCA 3050
MW
RW

DD   CCTGGACGGCGCCCTCGTCCTGGTGGTCGTGGGCTCCCTGTCGCAGTACC 3100
MW
RW

DD   AGCTGAGGAGGCACCCGGCCATCGGGGGCTTCGTGCGGCAGCGCCGGTAC 3150
MW
RW

DD   TTGAGGTGGCCCGAGGATCTGCAGGACGTGGGCTGGTTCCTGGACACGCT 3200
MW
RW

DD   CTCCCGACACATCCTGCAGGAGCAGAGGGGCGCGCGCGGGGATGGCGGCA 3250
MW
RW
```

```
DD   TCCCGCTGCGCACCGTGGCGGCCGGGCGCCGACCTCACTGCACCAGGGTC 3300
MW
RW

DD   GGGAGGCGCCGACCTCACTGCACCAGGGTCCGGGGGCGCCGACCTCACTG 3350
MW
RW

DD   CACCGGGGCCCCGGGGGGCGCCGACCTCACTGCACCGGGGTCCGGGGGGC 3400
MW
RW

DD   GCCGACCTCACTGCACCGGGGTCCGGGGGACGCCGACCTCACTGCACCGG 3450
MW
RW

DD   GGCCCCGGGGGGCGCCGACCTCACTGCACCGGGGCCCCGGGGGGCGCCGA 3500
MW
RW

DD   CCTCACTGCACCAGGGTCGGGAGGCGCCGACCTCAGTGCACCAGGGACCG 3550
MW
RW

DD   GGGGCGCCGACCTCACTGCACCGGGGCCCCGGGGGCGCTGCTCCTCCGCG 3600
MW
RW

DD   GCGGGCGCCCGGGTCCGACAAAGGGCGAGGGCGCGAGCGGTGCGGCGCAG 3650
MW
RW

DD   GAGCTCAGGGTCCGCGCGGAGGACCCGGGAGCACACACGGCCCGAGGAGC 3700
MW
RW

DD   CGCCCCGCGCCCGCCCCGCCCTCGGTCGCGCCCCCGCCGCAGCCCCAGCC 3750
MW
RW

DD   CCCCGCGCCCCCGCCCGGCCCCGCCGCCGCCCCCCGCGGCTGCCCCGACC 3800
MW
RW
```

```
DD   CTGCCCGCCCTCCGCCCGCGCGGGCTGCCCGGCCCCGCGTCCTTACCCGG 3850
MW
RW

DD   TCTCCCGGCCCCGCGGCGCGGGGGGCGGCGGCGGCGGCGGCGGCGGC 3900
MW
RW

DD   GGGGCCGGGACGCGTCCACGCAGGAGACAGGCGCCCCCGACGCGCCGGCC 3950
MW
RW


DD   CCGATGCGGACCCGGAGCCGGCTTCCGCCTCCCGCCGGAAGGCGTCCCGA 4000
MW
RW

DD   GCAGGACCGGAAGTCCCGCCCCGGCGGCTGAGGGGGGCGGCCGGGGGCGG 4050
MW
RW

DD   TCGTGTCAGCAGTTCGGCGGGGCGGGGGGTCACACCGACGTCCGTGGGCT 4100
MW
RW

DD   GCCCCCGGCCTCGGGCGGGCCGCGTGTTTCTCCGTCCGCCCGCTCCTCGC 4150
MW
RW

DD   CAGACCGCCGGGTCGCGGCGGGGGAGGGGGCGGGGCGGGGCGGGACGCAG 4200
MW
RW

DD   GAGGGGCGGGGCGTGGGAGAGGCGGGGGCGGGGCGCGAGAGGGGCGGGGC 4250
MW
RW

DD   GGGGCGGGGCGTGGGA 4266
MW
RW
```