Whole-Genome Sequencing of African Dogs Provides Insights into Adaptations against Tropical Parasites
Yan-Hu Liu, Lu Wang, Tao Xu, Xiaomin Guo, Yang Li, Ting-Ting Yin, He-Chuan Yang, Yang Hu, Adeniyi Adeola, Oscar Sanke, et al.

To cite this version:
Yan-Hu Liu, Lu Wang, Tao Xu, Xiaomin Guo, Yang Li, et al.. Whole-Genome Sequencing of African Dogs Provides Insights into Adaptations against Tropical Parasites. Molecular Biology and Evolution, Oxford University Press (OUP), 2018, 35 (2), pp.287-298. 10.1093/molbev/msx258. hal-03127438

HAL Id: hal-03127438
https://hal.archives-ouvertes.fr/hal-03127438
Submitted on 16 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Whole-Genome Sequencing of African Dogs Provides Insights into Adaptations against Tropical Parasites

Yan-Hu Liu,†,1,2 Lu Wang,†,1 Tao Xu,†,3 Xiaomin Guo,3 Yang Li,3 Ting-Ting Yin,2,4 He-Chuan Yang,5 Yang Hu,6 Adeniyi C. Adeola,2,4 Oscar J. Sanke,7 Newton O. Otecko,2,4 Meng Wang,1 Yaping Ma,1 Olaogun S. Charles,8 Mikkel-Holger S. Sinding,9,10 Shyam Gopalakrishnan,9 José Alfredo Samaniego,9 Anders J. Hansen,9 Carlos Fernandes,11 Philippe Gaubert,12 Jane Budd,13 Philip M. Dawuda,14 Eli Knispel Rueness,15 Lubin Jiang,6 Weiwei Zhai,5 M. Thomas P. Gilbert,9,16,17 Min-Sheng Peng,2,4 Xiaopeng Qi,*3 Guo-Dong Wang,*,2,4 and Ya-Ping Zhang,*1,2,4

1State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, and School of Life Sciences, Yunnan University, Kunming, Yunnan, China
2State Key Laboratory of Genetic Resources and Evolution, and Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China
3Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences/Key Laboratory of Bioactive Peptides of Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China
4Sino-Africa Joint Research Center, Chinese Academy of Sciences, Kunming, Yunnan, China
5Human Genetics, Genome Institute of Singapore, A*STAR, Singapore
6CAS Key Laboratory of Molecular Virology and Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai, China
7Taraba State Ministry of Agriculture and Natural Resources, Jalingo, Taraba State, Nigeria
8Natural History Museum of Denmark, Copenhagen, Denmark
9Natural History Museum, University of Oslo, Blindern, Oslo, Norway
10CE3C—Centre for Ecology, Evolution and Environmental Changes, Departamento de Biología Animal, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
11Institut des Sciences de l’Évolution de Montpellier (ISEM)—UM-CNRS-IRD-EPHE, Université de Montpellier, Montpellier, France
12Breeding Centre for Endangered Arabian Wildlife, Sharjah, United Arab Emirates
13Department of Veterinary Surgery and Theriogenology, College of Veterinary Medicine, University of Agriculture Makurdi, Benue State, Nigeria
14Department of Veterinary Medicine and Biobank, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
15Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway
16Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA, Australia
17Norwegian University of Science and Technology, University Museum, Trondheim, Norway
†These authors contributed equally to this work.
*Corresponding authors: E-mails: qixiaopeng@mail.kiz.ac.cn; wanggd@mail.kiz.ac.cn; zhangyp@mail.kiz.ac.cn.
Associate editor: John Novembre

Abstract

Natural selection in domestic dogs is of great interest in evolutionary biology since dogs have migrated to every inhabited continent of the world alongside humans, and adapted to diverse environments. Here, we explored their demographic history and genetic basis of adaptation to the tropical African environment using whole genome analyses of 19 African indigenous dogs from Nigeria. Demographic analysis suggests that the ancestors of these dogs migrated into Africa from Eurasia 14,000 years ago and underwent a severe founder effect before population expansion. Admixture analysis further reveals that African dog genomes contain about 1.88–3.50% introgression from African golden wolves (Canis anthus). Population genetic analysis identifies 50 positively selected genes linked with immunity, angiogenesis, ultraviolet protection, as well as insulin secretion and sensitivity that may contribute to adaptation to tropical conditions. One of the positively selected genes, adhesion G protein-coupled receptor E1 (ADGRE1), has also been found to be association with severe malaria resistance in African human populations. Functional assessments showed that ADGRE1 provides protective host defense against Plasmodium infections. This result, together with the fact that the inflammatory response to canine babesiosis is similar to complicated falciparum malaria in humans, support the dogs as a model for the study of malaria control and treatment.

Key words: African dogs, adaptive evolution, antiparasite, demographic history.
The domestic dog (Canis lupus familiaris) dispersed across the planet alongside human migrations, and in doing so must have had to undergo adaptation to diverse environments. Natural selection played key roles in shaping the fitness of the domestic dog to these environments, with the underlying mechanisms being of great interest in evolutionary biology (Freedman et al. 2016). For example, previous studies on the adaptation of Tibetan dogs to the hypoxic Tibet Plateau consistently revealed EPAS1 and HBB as candidate genes for hypoxia adaptation (Gou et al. 2014; Li et al. 2014; Wang et al. 2014). Moreover, Wang et al. (2014) found that the hemoglobin levels were very similar between Tibetan dogs and human, suggesting a convergent evolution between dogs and humans living on the Tibetan Plateau.

The African continent is prone to a number of selective forces that have led to unique human genetic adaptations. One of the most pervasive is malaria—in most African regions malaria is endemic, and World Health Organization (WHO) estimates indicate that in 2015, 88% of global malaria cases and 90% of global malaria deaths occurred in Africa (WHO 2015). Previous studies have identified several genes associated with immune responses that protect some African populations against malaria (Kusi et al. 2008; Machado et al. 2010; Band et al. 2013; Kariuki et al. 2013; Apinjoh et al. 2014; Carstensen et al. 2014). In Africa, there also exists canine babesiosis, a malaria-like disease caused by the tick-borne parasite Babesia, which, like Plasmodium, is an apicomplexan hemoproteozoan parasite (Sasaki et al. 2007). It has been noted that infection rates of B. canis rossi and B. canis vogeli are lower in dogs from Nigeria than in those in South Africa or Sudan (Sasaki et al. 2007), suggesting that there might be differential local adaptation against the parasite across the African continent. A further well-documented series of genomic encoded human adaptations relate to protection against strong ultraviolet radiation (UVR) in the equatorial region (Lamason et al. 2005; Norton et al. 2007; Jablonski and Chaplin 2013), including maintaining a dark skin and kinky hair texture. The latter potentially confer protection to the brain against thermal stress (Jablonski and Chaplin 2014). The African cattle also revealed signatures of selection for environmental adaptive traits including coat color, heat tolerance, and tick resistance (Kim et al. 2017).

Domestic dogs are exotic to Africa, having first been domesticated from Eurasian gray wolves between 15,000 and 40,000 years ago (Vilà et al. 1997; Gemonpré et al. 2009). Little is currently known about the origin of African domestic dogs. Besides that, wolves have been considered to be absent in Africa (Mech 1981; Wang, Ma, et al. 2016). Although genome-wide analyses have estimated that a subset of ancestral dogs migrated toward the Middle East, Europe, and into Africa ~15 000 years ago (Wang, Zhai, et al. 2016), the oldest archaeological evidence for African dogs was found in Egypt, dating ca. 6300–5600 BC (Mitchell 2015). Therefore, considerable gaps exist in our understanding of the history of domestic African dogs. For example, a recent study based on 382 African dog mitochondrial D-loop sequences found that West African dogs share a subhaplogroup with southwest European dogs and inferred possible gene flow between West African dogs and southwest European dogs ~12,000 years ago (Adeola et al. 2017). However, mitochondrial DNA (mtDNA) studies are limited as they only account for a small proportion of the genetic material of an individual. Although analysis of whole-genome data could further contribute to our understanding, currently whole-genome information is only available for four African village dogs. Hence, additional nuclear genomes of African village dogs are needed to further investigate their origin and diversity. Moreover, dogs in Africa may possess adaptations related to protective mechanisms against selective forces like babesiosis and strong UVR. Such adaptations are largely unknown, but by identifying genomic regions under selection, genes possibly associated with adaptive responses could be revealed.

Whole-genome sequencing provides a powerful, holistic approach to understand the demographic history and natural selection of populations and species. Here, we sequenced the genomes of 15 indigenous dogs from Nigeria and 4 African golden wolves (Canis anthus) (Gaubert et al. 2012; Rueff et al. 2012; Koepfli et al. 2015) to study their demographic history and adaptive evolution, as well as possible gene flow between the two species.

Results

Sample Collection and Whole-Genome Sequencing

We sampled 15 indigenous dogs from six states in Nigeria (supplementary table S1, Supplementary Material online) and 4 African golden wolves from Algeria, Egyptian Sinai, Morocco, and Senegal. We performed whole-genome sequencing to an average depth of 13× for each sample after removing PCR redundancy (supplementary table S1, Supplementary Material online). Since tissue samples of three African golden wolves were collected from dead subjects, DNA damage was possible. We therefore analyzed C to T and G to A mutation enrichment near the end of the sequence reads (Binladen et al. 2006), which confirmed no DNA degradation of these samples (supplementary fig. S1, Supplementary Material online).

We also incorporated 36 published whole-genome sequencing data of dogs and gray wolves (fig. 1A: 4 Nigerian village dogs, 1 African ancient breed—Basenji, 7 gray wolves, 14 European breeds and Middle Eastern village dogs, and 10 indigenous dogs from Yingjiang in southern China) (Auton et al. 2013; Wang et al. 2013; Freedman et al. 2014; Gou et al. 2014; Wang, Zhai, et al. 2016). These samples cover all major Eurasian dog and wolf groups. After strict filtering, we identified ~24.4 autosomal million SNPs for further analysis (supplementary table S2, Supplementary Material online).

Population Structure and Phylogenetic Analysis

Principal component analysis (PCA) was performed to explore the relationships among the 55 individuals. In a 2D plot of the genotypes, there was a clear separation between dogs, Eurasian gray wolves (EGW), and African golden wolves (AGW) (fig. 1B). The dogs split into three groups along the
second dimension: 1) Chinese indigenous dogs from Yingjiang (CIDY), 2) Middle Eastern village dogs (MEVD) and European breeds (EB), and 3) the Basenji together with the Nigerian indigenous dogs (NID). We combined our data set with previous SNP array data for an expanded data set (vonHoldt et al. 2011) and observed a clustering of dog populations according to their geographic affinities (supplementary fig. S2, Supplementary Material online).

A maximum-likelihood phylogenetic tree agreed with the PCA results. At the base of the tree, dogs and Eurasian wolves were separated from African golden wolves (fig. 1C and supplementary fig. S3, Supplementary Material online). Dogs then diverged from Eurasian wolves and split into three clades. Chinese indigenous dogs from Yingjiang firstly split from other dogs, suggesting that the Yingjiang dogs may represent a very ancient divergence. The Middle Eastern village dogs and European breeds formed a distinct sister clade to African dogs, suggesting that the latter had an evolutionary history independent of the main process of dog domestication and diversification.

To explore the genetic relationships among individuals, we performed a structure analysis using a block relaxation algorithm to cluster individuals into different numbers of groups (supplementary fig. S4, Supplementary Material online). Five different clusters were inferred: 1) African golden wolves, 2) Eurasian wolves, 3) Yingjiang indigenous dogs, 4) Middle Eastern village dogs and European breeds, and 5) African dogs (fig. 1D). This is consistent with the result of PCA and phylogenetic analysis.

**Genetic Diversity**

In accordance with the above results, we grouped individuals into five populations (African golden wolves, Eurasian wolves, European breeds and Middle Eastern village dogs, Yingjiang indigenous dogs, and African dogs) and estimated nuclear diversity using the parameter $h_p$. As shown in figure 1E, there were significant differences ($P < 0.001$) in the levels of genetic diversity between populations. Both African golden wolves and Eurasian gray wolves exhibited high nuclear diversity (mean: $2.2 \times 10^{-3}$ and $1.8 \times 10^{-3}$, respectively). Yingjiang indigenous dogs had the highest diversity among the three dog populations, at $\sim 77.8\%$ of the diversity in Eurasian gray wolves. Although the genetic diversity of Middle Eastern village dogs and European breeds was lower than that of Yingjiang indigenous dogs, it was higher than in Nigerian indigenous dogs. These results suggest that the ancestors of the Nigerian indigenous dogs might have undergone a founder effect after their divergence from Eurasian dogs.
Admixture and Demographic History

To test for gene flow between populations, we performed D-statistics analysis (Patterson et al. 2012). The results indicate gene flow between African golden wolves and Nigerian indigenous dogs ($D = -0.013$ to $-0.019$, $Z = 8.66$), as well as between Yingjiang indigenous dogs and European breeds together with Middle Eastern village dogs ($D = -0.057$, $Z = 20.68$) (supplementary table S3, Supplementary Material online).

Because the D-statistics could not identify the direction of gene flow between African golden wolves and dogs, we used the joint site frequency spectrum (SFS) approach implemented in fastsimcoal2 to simulate demographic history (Excoffier et al. 2013). A total of 12 models of historical divergence and various gene flow scenarios were fitted to the allele-frequency spectrum of five populations (supplementary fig. S5, Supplementary Material online). A demographic model (fig. 2) in which mutual gene flow existed between African golden wolves and Nigerian dogs produced a better fit than the alternative models (supplementary table S4, Supplementary Material online). In the best fitting model, domestication of dogs occurred in the last 31,100 years, and the divergent time of Chinese dogs and other dogs is approximated at 16,600 years ago (supplementary table S5, Supplementary Material online). This analysis further indicated that about 14,800 years ago (95% CI 13,368–14,832), dogs migrated to Africa and introgressed with African golden wolves, followed by a 1.3-fold population expansion.

F4-ratio test were performed to calculate the proportion of gene flow between African golden wolves and African dogs (Patterson et al. 2012). As result, F4-ratio estimated about 5% ancestry from AGWs in NIDs (supplementary table S6, Supplementary Material online). PCAmix was used to identify segments of the African dogs’ genome belonging to AGWs or EBs (Brisbin et al. 2012). The ratio of the ancestral AGWs in African dogs range from 1.86% to 2.94% (average 2.45%), which is similar to the estimates from of fastsimcoal2 (1.43%, 95% CI [1.88–3.50%]), supplementary table S5, Supplementary Material online). The introgressed regions were about $\sim 12.62$ kb in size, and dispersed randomly across the genome. The lengths of regions range from 11.77 to 12.26 kb on an average (supplementary fig. S6, Supplementary Material online).

Adaptive Selection

We used Sweep Detector (SweeD) and the population branch statistic (PBS) (Yi et al. 2010; Pavlidis et al. 2013) to identify positive selection in African indigenous dogs. Considering the top 1% outliers in each method for each gene annotation, SweeD identified 399 genes and the PBS approach found 580 genes. Fifty genes are common in the two methods and as considered as potential candidate genes (supplementary table S7, Supplementary Material online).

Gene ontology (GO) identified significant overrepresentation of genes involved in local environmental adaptation (table 1). A number of relevant observations from other studies on the activity of these genes hint at their possible functions. For instance, enzyme activator activity (GO: 0008047) could be affected by infection. The expression profile of pig lung tissue postinoculation with Actinobacillus pleuropneumoniae showed a significant representation of genes belonging to this GO term (Zuo et al. 2012). Endoplasmic reticulum (GO: 0005783) plays key roles in important processes like protein transport and energy metabolism. The mRNA expression of GO: 0005783 is increased in mice after heat treatment (Yu et al. 2011). To detect the convergent evolution between African human and dogs, we compared our list of GO terms in dogs with that from African humans compiled by Barreiro et al. (2008). After performing GO enrichment analysis on the positively selected genes (PSGs) in African humans (Barreiro et al. 2008), we identified 34 enriched GO terms that showed statistical significance (supplementary table S7, Supplementary Material online). Interestingly, two terms: GTPase regulator activity (GO: 0030695) and nucleoside-triphosphatase regulator activity (GO: 0060589), both exist in the overlapping set between the two species (150-fold enrichment at a significance of $P = 6.42E-5$) (Barreiro et al. 2008).

Among the 50 PSGs identified in African dogs, three of them may play roles in innate or adaptive immune responses: adhesion G protein-coupled receptor E1 (ADGRE1), caspase recruitment domain family member 9 (CARD9), and vav guanine nucleotide exchange factor 1 (VAV1) (Mesecke et al. 2011; Hafalla et al. 2012; Kariuki et al. 2013). The human homologues of three genes are highly expressed in spleen and bone marrow (supplementary fig. S7, Supplementary Material online). The ADGRE1 homolog in mice (F4/80) is required for the differentiation of antigen-specific CD8+ regulatory T cells (Lin et al. 2005). A large multicenter case-control study revealed that ADGRE1 is associated with severe malaria (Malaria Genomic Epidemiology Network 2014). ADGRE1 has also been linked with complex malaria-associated seizures in African children (Kariuki et al. 2013). Moreover, ADGRE1 has been associated with hyperpyrexia, severe malaria anemia, and uncomplicated malaria in a...
Cameroonian population (Apinjoh et al. 2014), and its expression was induced in women with placental malaria infection (Muehlenbachs et al. 2007). The expression of ADGRE1 in mice is significantly increased by ~3.2-fold after 8 days following infection with Plasmodium chabaudi (Al-Quraishy et al. 2013).

Because positive selection can be difficult to distinguish from genetic drift produced by neutral processes related to a specific demographic history (Freedman et al. 2016; Wu et al. 2016). We performed simulations to generate 40-kb DNA sequences for 1,000,000 repetitions to calculate PBS, and to 1-Mb sequences for 10,000 times to calculate SweeD, based on demographic history as the posterior distributions (Ewing and Hermisson 2010). We then used these data to test whether the observed values of PBS and SweeD for ADGRE1 gene could be due to genetic drift instead of selection. As a result, the probability of the observed values of PBS and SweeD attributable to genetic drift was only 4E-4 and 3.17E-3, respectively (supplementary fig. S8, Supplementary Material online). For an in-depth examination of the role of dog ADGRE1 protein in the host defense against parasitic infection, we performed ectopic expression of dog ADGRE1 in RAW 264.7 macrophages (fig. 3A). Phagocytosis of pathogens by macrophages induces innate immune response, which in turn activates adaptive immunity. Although the murine homolog of ADGRE1 (F4/80) is predominantly expressed on eosinophils in mice, it is widely used as a cell surface marker of macrophage populations in mice. The precise role of F4/80 in innate and adaptive immunity remains elusive (Murray and Wynn 2011). To determine whether the dog membrane protein ADGRE1 is involved in the phagocytosis of Plasmodium parasites, RAW 264.7 macrophages transfected with GFP and ADGRE1 were infected with Plasmodium berghei ANKA, and the expression of 18S rRNA and Hsp70 specific to P. berghei ANKA was analyzed by real-time quantitative PCR. The expression levels of Plasmodium 18S rRNA and Hsp70 were significantly higher in ADGRE1 transfected cells compared with those transfected with GFP (fig. 3B). This result indicates that the phagocytosis of Plasmodium parasites is increased in the presence of dog ADGRE1.

Type I interferon plays important roles in various infectious diseases, including malaria. Intracellular DNA and RNA sensors such as cGAS, MDA5, and TLR7 are crucial activators for type I interferon production in response to Plasmodium infection (Gun et al. 2014; Yu et al. 2016). To determine whether increased intracellular parasites in ADGRE1 transfected cells results in a higher amount of type I interferon production, we analyzed Ifnb and Cxcl10 expression in GFP and ADGRE1 transfected cells and found that ADGRE1 indeed promoted Ifnb and Cxcl10 expression during Plasmodium infection (fig. 3C). Guanylate binding proteins (GBPs) are type I interferon inducible proteins associated with pathogen-containing vacuoles and are crucial for host defense against pathogens (Degrandi et al. 2008; Haldar et al. 2013). In line with the increased phagocytosis and type I interferon production in ADGRE1 transfected cells, the expression of Gbp1 and Gbp3 was markedly increased (fig. 3C). Type I interferon signaling plays a central role in neutrophil activation and malaria pathogenesis by mediating the production of proinflammatory cytokines such as IL-1 members and IL-6, and neutrophil migration (Roche et al. 2015). Accordingly, the expression of Il1a, Il1b and Il6 was increased in ADGRE1 transfected cells in response to Plasmodium infection (fig. 3C). In contrast, the expression of chemokine Cxcl1 was comparable between GFP and ADGRE1 transfected cells (fig. 3C). Moreover, the other critical innate immune responses during pathogenic infection including autophagy, lysosomal biogenesis, and apoptotic cell death pathways were not significantly affected by ectopic expression of ADGRE1 in response to Plasmodium infection (fig. 3D).

To identify the SNP which possibly enhances the function of ADGRE1, we used Fst values to identify highly differentiated sites between African dogs and European breeds, and took the top 1% of sites for annotation. We found a non synonymous G655A mutation within the ADGRE1 gene. This mutation had a very high allele frequency in African dogs (89.5%) compared with 25.0% in European breeds (fig. 4A). We subsequently explored whether this mutation contributes to elevated resistance to malaria in African dogs through the ectopic expression of mutated ADGRE1-G655A in RAW 264.7 macrophages and analysis of its effect on innate

### Table 1. GO Terms for Genes Identified by Both SweeD and PBS.

<table>
<thead>
<tr>
<th>Category</th>
<th>GO Term</th>
<th>No. of Genes</th>
<th>P Value</th>
<th>Citation on Environmental Adaptation and Immunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0030054--cell junction</td>
<td>5</td>
<td>0.035</td>
<td>Yu et al. (2011)</td>
</tr>
<tr>
<td>CC</td>
<td>0005783--endoplasmic reticulum</td>
<td>6</td>
<td>0.026</td>
<td>Zuo et al. (2012), Lv et al. (2014)</td>
</tr>
<tr>
<td>MF</td>
<td>0000847--enzyme activator activity</td>
<td>5</td>
<td>0.010</td>
<td>Barreiro et al. (2008), Lv et al. (2014)</td>
</tr>
<tr>
<td>MF</td>
<td>0030955--GTPase regulator activity</td>
<td>4</td>
<td>0.018</td>
<td>Lv et al. (2014)</td>
</tr>
<tr>
<td>MF</td>
<td>0030589--nucleosome-triphosphatase regulator activity</td>
<td>5</td>
<td>0.020</td>
<td>Barreiro et al. (2008), Lv et al. (2014)</td>
</tr>
<tr>
<td>MF</td>
<td>0070011--peptidase activity, acting on L-amino acid peptides</td>
<td>5</td>
<td>0.049</td>
<td></td>
</tr>
</tbody>
</table>
immune response during parasite infection. Interestingly, the expression of \textit{Plasmodium} \textit{Hsp70}, host \textit{Ifnb} and its inducible genes \textit{Gbp1}, \textit{Gbp2}, \textit{Gbp3}, and \textit{Gbp5}, and proinflammatory cytokine gene \textit{Il6} were higher for \textit{ADGRE1} than for \textit{ADGRE1} (fig. 4B). Collectively, these results suggest that dog \textit{ADGRE1} protein mediates \textit{Plasmodium} internalization and type I interferon production, the latter being a master regulator of the immune response against malaria in dogs.

Besides the three genes involved in immune responses, we also detected 13 PSGs from 50 candidate genes that are functionally involved in angiogenesis, ultraviolet protection, and the secretion sensitivity to insulin. Collagen and calcium binding EGF domains 1 (\textit{CCBE1}), plexin domain containing 2 (\textit{PLXDC2}), and Ras and Rab interactor 2 (\textit{RIN2}) play crucial roles in angiogenesis (Hogan et al. 2009; Sandri et al. 2012; Cheng et al. 2014). Angiogenesis is extremely important for thermoregulation of warm-blooded animals (Arens and Zhang 2006). We also identified seven genes associated with DNA repair and melanin accumulation that we hypothesize may provide protection from the negative effects of exposure to ultraviolet light. Those genes were the ubiquitin carboxyl-terminal esterase L3 (\textit{UCHL3}), LON peptidase N-terminal domain and ring finger 1 (\textit{LONRF1}), SNF2-related CREBBP activator protein (\textit{SRCAP}), ankyrin repeat domain 32 (\textit{ANKRD32}), dystonin (\textit{DST}), membrane-bound transcription factor peptidase site 1 (\textit{MBTPS1}), and phospholipid phosphatase-related protein type 5 (\textit{LPPR5}) (Lalonde et al. 2005; Sano et al. 2006; Matsuoka et al. 2007; Brandl et al. 2009; Weger et al. 2011; Park et al. 2013; Räschle et al. 2015). Several PSGs such as potassium voltage-gated channel interacting protein 1 (\textit{KCNIP1}), nucleobindin-1 (\textit{NUCB1}), and an enhancer of mRNA decapping 3 (\textit{EDC3}) were associated with insulin secretion and sensitivity (Lee et al. 2014a; Ramesh et al. 2015; Kim et al. 2016). In humans, it has been argued that these genes relate to specific traditional African diets and are
linked to the current health challenges among African American populations as explained by the thrifty gene hypothesis (Marshall 2005).

**Discussion**

**African Dogs Are a Distinct Population**

In this study, we applied whole-genome sequencing and performed population genomic analysis of 15 African indigenous dogs and 4 African golden wolves. We observed a split into three major dog groups: 1) southern Chinese indigenous dogs, 2) a cluster containing Middle Eastern village dogs and European breeds, and 3) African dogs (fig. 1B–D). The ancestors of the African dogs studied here were estimated to have entered the African continent about 14,000 years ago (fig. 2). This is consistent with the findings of earlier studies on dog population structure and history based on mtDNA (Adeola et al. 2017), and full nuclear genomes (Wang, Zhai et al. 2016). Interestingly, human Y chromosome haplotypes provide evidence of human migration to the Mediterranean coast of North Africa at the end of the Pleistocene (Underhill et al. 2001), and genome-wide SNPs analysis suggested a “back-to-Africa” migration >12,000 years ago (Henn et al. 2012). Therefore, it is plausible that the entry of dogs into Africa was in the company of the migrating humans. We detected gene flow from both African golden wolves and European breeds/Middle Eastern village dogs into the African dogs. A previous mtDNA study showed that West African dogs share a subhaplogroup with southwestern European dogs, possibly due to gene flow events (Adeola et al. 2017).

**African Dogs Have the Lowest Genetic Diversity**

In our study, the genetic diversity and effective population size of African indigenous dogs are the lowest among the major dog groups (figs. 1E and 2). Genetic diversity was higher in the Middle Eastern village dogs/European breeds, and highest in the Chinese dogs. The lowest genetic diversity of African dogs may be due to in the availability of samples only from Nigeria out of the vast African continent. But the dogs from Yingjiang, only a small county in China, have the highest diversity. This genetic diversity gradient is derived from studies of mtDNA (Savolainen et al. 2002; Pang et al. 2009), Y chromosome (Ding et al. 2012), and whole-genome (Wang, Zhai, et al. 2016). Moreover, the effective population sizes of dogs also show a similar gradual upward trend from African dogs, to European breeds/Middle Eastern village dogs, and then southern Chinese dogs (fig. 2). One study reported similar mtDNA haplotype diversity in African and East Asian village dogs (Boyko et al. 2009). But another study of dog mtDNA diversity found that 318 African village dogs...
was revealed (Kariuki et al. 2013). Our study suggests that malaria-associated seizures in African children and malaria in African humans (Kariuki et al. 2013). Dogs, has been proved to be associated with immune responses to malaria in African dogs (Storb and Thomas 1985), natural selection may work on a similar evolutionary direction among the two species. In this study, genes associated with insulin secretion and sensitivity, immunity, angiogenesis, and ultraviolet protection showed adaptive selection (supplementary table S8, Supplementary Material online). These genes were significantly clustered in seven GO terms (table 1). Two of these GO terms also existed in the GO terms of African human selection genes. Both terms modulate the rate of guanosine triphosphate (GTP) hydrolysis. GTP is involved in processes like signal transduction, genetic translation, energy transfer within the cell. Moreover, the two terms were associated with climate-mediated selection in sheep (Lv et al. 2014). This suggests possible convergent evolution between African dogs and humans under the same environmental exposures.

Similar evolutionary adaptations have been reported in some African and African-derived human populations. For example, African American children have higher levels of insulin and insulin resistance than those of various non-Africans populations (Svec et al. 1992; Arslanian and Suprasongsin 1996; Arslanian et al. 1997; Gower et al. 1999). In addition, many African populations are generally characterized by darker integumentary phenotypes as a protective adaptation against strong UVR (Lamason et al. 2005; Horton et al. 2007; Jablonski and Chaplin 2013). Other studies have demonstrated evidence of selection in genes associated with immune responses that protect African human populations from malaria and other diseases (Kusi et al. 2008; Machado et al. 2010; Band et al. 2013; Carstensen et al. 2014). In particular, the ADGRE1 gene, one of the strongly PSGs in African dogs, has been proved to be associated with immune response to malaria in African humans (Kariuki et al. 2013). In a previous case–control study, a relationship between ADGRE1 and malaria-associated seizures in African children was revealed (Kariuki et al. 2013). Our study suggests that ADGRE1 contributes to defense against Plasmodium infection in African dogs which could be a suitable model for biomedical research for malaria.

In summary, based on whole-genome analyses, this study expands our understanding of the genetic diversity, evolutionary history, and tropical adaptation of African dogs. The inflammatory-response against canine babesiosis is similar to that of complicated falciparum malaria in humans (Reyers et al. 1998), and dogs generally share ~360 diseases with humans (Shearin and Ostrander 2010). There are several advantages of Canis familiaris as a model for genetic susceptibility to disease, like greater homology to human, susceptibility to many diseases with humans variants (Shearin and Ostrander 2010; Boyko 2011). The Lupa Project aimed to enhance the use of the dog as an effective model to study common complex diseases in human and has got some achievements so far (Shearin and Ostrander 2010; Boyko 2011; Lequarré et al. 2011). Dogs are thus not only a useful species for mapping disease loci but also an excellent model for biomedical research of human diseases.

Materials and Methods

Sample Collection and Sequencing

We sampled 15 indigenous dogs from six different states in Nigeria (Oyo, Ondo, Akwa Ibom, Cross River, Taraba, and Ekiti) and 4 wolves from different countries in Africa (Algeria, Egypt, Morocco, and Senegal). Blood samples were collected from each dog and African golden wolf from Egypt. Tissue samples were collected from other African golden wolves accidentally knocked down on roads. Total genomic DNA was extracted from blood samples using the phenol–chloroform method, and 1–3 μg of DNA from each individual was sheared into fragments using 200–800 bp with the Covaris system. Tissue samples were extracted using the DNaseasy Blood and Tissue kit (Qiagen, Hilden, Germany), and DNA was fragmented into 300–400 bp fragments using a Bioruptor NGS sonicator (Diagenode, Denville, NJ). The DNA fragments were then sequenced using the Illumina HiSeq 2000 or 2500 platforms. Sequencing data from this study have been submitted to the Genome Sequence Archive (GSA, http://gsa.big.ac.cn/) under project number PRJCA000335.

Sequence Data Preprocessing and Variant Calling

Raw sequence reads were mapped to the dog reference genome (Canfam3) using the BWA-MEM version 0.7.10-r789 (Li 2013). Reads with identical start/end points were filtered using PICARD (version 1.87). Sequences were then locally realigned and base-recalibrated using the Genome Analysis Tool Kit (GATK, version 2.5-2-gf57256b) (Depristo et al. 2011). Specifically, after genome alignment and removing PCR duplicates, the distribution of misincorporation near the ends of the reads were carried out by mapDamage2.0 (Jónsson et al. 2013). Then, variant calling of sequence data were handled using the UnifiedGenotypeCaller in GATK. During base and variant recalibration, a list of known SNPs/indels was downloaded from the Ensembl database to serve as a training set.

Genetic Diversity and Population Structure

Genetic diversity was calculated from a nonoverlapping 40-kb windows across the genome using VCFtools v0.1.12b (Danecek et al. 2011). Principal component analysis was carried out using the smartPCA program from the EIGENSOFT package v5.0.1 (Patterson et al. 2006). Maximum-likelihood phylogenetic tree was built by SNPhylo (Lee et al. 2014b) and dhole (Cuon alpinus) was used as the outgroup. After thinning to a single SNP per 50-kb window, population structure analysis was performed...
using the block relaxation algorithm implemented in the ADMIXTURE software (Alexander et al. 2009).

Evolutionary History
To mitigate the effect of linkage disequilibrium, we took one SNP every 10kb, and then SNPs located 10 kb away from genes, were used to convert SFS by easySFS (https://github.com/isaacovercast/easySFS#easysfs). Demographic history was simulation by fastsimcoal2 (Excoffier et al. 2013). Mutation ratio was set to 6.6E-9 per sites per generation and generation time as 3 years (Kumar and Subramanian 2002; Wang et al. 2013; Wang, Zhai, et al. 2016). We used a recombination of 9.7E-9 (Wong et al. 2010). 12 models of various historical events were fitted to the joint SFS of dogs, grey wolves, and African golden wolves (Qiang et al. 2015). We ran the fastsimcoal2 50 times for each model with varying starting points to ensure convergence, and retained the fitting one with the highest likelihood. Demographic estimates were obtained from 100,000 simulations (~100,000) per parameter file, 40 Expectation/Conditional Maximization cycle (-L40). The best model was selected through the maximum log-likelihood of the SweeD likelihood for subsequent separate gene annotations. For gene annotations, we extracted the genes which present in both gene sets were considered significant candidates. The terms with $P < 0.05$ were considered significantly enriched. The terms with $P < 0.05$ were considered significantly enriched. To detect convergent evolution between African human and dogs, selection regions of African humans were obtained from a previous publication (Barreiro et al. 2008). After gene annotation, GO enrichment analysis was done as dogs.

Simulation
We performed simulations for PBS to generate 40-kb DNA sequences for 1,000,000 times based on the demographic history as the posterior distributions by the software msms (Ewing and Hermisson 2010). For each run, we calculated PBS values by 40-kb windows from the simulated data sets. For SweeD, we simulated 1MB sequences for 10,000 times. African golden wolves were used as outgroup to distinguish folded or unfolded of African dogs’ SNPs, and then SweeD was performed using unfolded site in African dogs for each 1MB sequence. $P$ values were got by calculating the proportion sets in simulation data larger than observed value.

Plasmid Construction and Retroviral Infection
The full-length cDNA of the dog ADGRE1 gene was amplified from cDNA of dog liver tissue using the primers ADGRE1 forward (5’-AAATAGATCTATGTGGAGCTTCAACTTGCTC CTC-3’), and ADGRE1 reverse (5’-AAATGGCAGCCGCTTA ATCCGTCCTAGAGTGGAGGGG-3’).

The PCR product was digested with BglII and NotI (restriction sites are underlined in the primer sequences) and cloned into the retroviral expression vector MSCV2.2 to obtain MSCV-ADGRE1. The cloned construct was verified by DNA sequencing. For generation of the ADGRE1-G65SA, the 1072 bp of the 5’ portion of the ADGRE1 gene was synthesized and cloned into MSCV-ADGRE1 by replacing the analogous fragment. Retrovirus production in 293 T cells and infection of RAW 264.7 macrophages was performed as previously described (Qi et al. 2013). GFP positive cells were isolated by fluorescence activated cell sorting (FACS) and expanded for further analysis.

Plasmodium Culture and Infection of Cells
Plasmodium berghei ANKA was cultured as described previously (Jiang et al. 2013). All animal experiments were conducted in accordance with the guidelines of animal care from the Institut Pasteur of Shanghai, Chinese Academy of Sciences and were approved by the Animal Care and Use Committee of the Kunming Institute of Zoology, Chinese Academy of Sciences. For parasite infection, infected red blood cells from mice were suspended in phosphate-buffered saline and subsequently administered to the RAW 264.7 macrophages for stimulation as described previously (Yao et al. 2016).

Real-Time Quantitative PCR
The total RNA was isolated from RAW 264.7 macrophages using TRIzol reagent (Invitrogen), and cDNA was reverse transcribed using Superscript III reverse transcriptase (Invitrogen). Real time quantitative PCR was performed using the CFX96 Real-Time System (BIO-RAD). Primer sequences are listed in supplementary table S9, Supplementary Material online.
Immunoblot Analysis and Antibodies

Samples were separated by 12% SDS–PAGE and then electrophoretically transferred onto polyvinylidene fluoride membranes. Membranes were blocked with 5% nonfat milk and then incubated overnight in primary antibody at 4 °C. The following primary antibodies were used: anti-LC3B (NB600-1384; Novus Biologicals), anticaspase 3 (9661 S, 9491 S, and 8592 S; Cell Signaling Technology), anti-GFP antibody (SC-9996, Santa Cruz Biotechnology), anti-TFEB (A303-673 A; Bethyl Laboratories, Inc.), and anti-GAPDH (5174 S; Cell Signaling Technology). The secondary antibodies used were HRP-labeled antirabbit antibodies (Jackson ImmunoResearch Laboratories, Inc.).

Statistical Analysis

Data were given as means ± SD. Statistical analyses were performed using two-tailed Student’s t tests. P < 0.05 were considered significant.

Supplementary Material

Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgments

This work was supported by grants from the NSFC (91531303), and the Breakthrough Project of Strategic Priority Program of the Chinese Academy of Sciences (XDB13000000), the Chinese Academy of Sciences (CXJ-17-M141) and (Y602381081), Sino-Africa Joint Research Center, Chinese Academy of Sciences (SAJC201611). ERC Consolidator Grant (681396) ‘Extinction Genomics’, Marie Skłodowska-Curie Actions Fellowship (655732) ‘Wherewolf’, and Carlsbergfondet fellowship (CF14-0995). G.-D.W. and M.-S.P. are supported by the Youth Innovation Promotion Association, Chinese Academy of Sciences. N.O.O. is supported by the CAS-TWAS President’s Fellowship Program for Doctoral Candidates. C.F. acknowledges support from Fundação para a Ciência e Tecnologia (FCT, MCTES, Portugal) and Faculdade de Ciências da Universidade de Lisboa through, respectively, the Ciência 2007 contract C2007-UL-342-CBA1 and the contract as Invited Assistant Professor.

References


